ICTCA 2025

ISBN : 979-11-996391-0-2

16th International Conference on Theoretical and Computational Acoustics

August 18[Mon] - 22[Fri], 2025
BEXCO, Busan Korea

Proceedings

SAEVIT™
TECHNOLOGIES

Professional company in infrasound industry

SAEVIT Technologies

The first WNRS installed in Korea
(Wind Noise Reduction System)

We have received technology transfer of WNRS, which has been internationally certified by CTBTO(Comprehensive Nuclear-Test-Ban Treaty Organization) IMS(International Measurement System) from Enviroearth Co. in France.
Currently installed in a national institution (33 stations) to prove its technological capabilities.

As the exclusive distributor of Seismo wave Co. in South Korea and a specialized company with unrivaled technology and expertise in infrasound.

Providing comprehensive solutions for infrasound

- Equipment sales
- Data collection
- Analytics S/W development
- Maintenance
- Derivative

HEADQUARTERS

042-936-8686 36, Techno 5-ro, Yuseong-gu, Daejeon, Republic of Korea

sales@saevit.co.kr , www.saevit.co.kr

CONTENTS

1. Welcome Message

2. Organizing Committee

3. Venue

4. Contact Information

5. Social Program

6. Lunch Voucher

7. Sponsors

8. Program Overview

9. Plenary Lectures

10. Daily Program

11. Poster Session

12. Notes

WELCOME MESSAGE

On behalf of the organizing committee, we are delighted to extend an invitation to you for the International Conference of Theoretical and Computational Acoustics (ICTCA 2025), which will take place at the Busan Exhibition and Convention Center (BEXCO) in Busan, Korea from August 18 to 22, 2025.

ICTCA 2025 aims to serve as a premier platform for scholars, researchers, and practitioners worldwide to discuss the latest advancements and research in the fields of theoretical and computational acoustics. The conference will feature a comprehensive program, including plenary lectures, technical sessions, and poster presentations, on-line presentations, all designed to facilitate the exchange of ideas and the dissemination of groundbreaking research findings.

Beyond the academic program, we have arranged a variety of social events to enable participants to experience the vibrant Korean culture and the renowned hospitality of Busan. Known for its beautiful beaches, dynamic culture, and historical landmarks, Busan offers the perfect setting for both scholarly inspiration and relaxation.

As Korea's leading port city and a cultural hub, Busan provides the ideal backdrop for ICTCA 2025, promising not only a stimulating academic exchange but also an unforgettable cultural experience. We are eager to welcome you to what promises to be an engaging and fruitful conference in Busan.

We look forward to your participation in ICTCA 2025 and to welcoming you to Busan, Korea.
Sincerely,

Jea Soo Kim

Prof. Jeasoo Kim
Chair of the Organizing Committee
Department of Ocean Engineering
Korea Maritime & Ocean University (KMOU)

ORGANIZING COMMITTEE

- **Conference Chair**

  Jeasoo KIM (Korea Maritime and Ocean University)

- **Secretary General**

  Gihoon BYUN (Korea Maritime and Ocean University)

  Yonghwa CHOI (Korea Maritime and Ocean University)

- **Technical Program**

  Seongil KIM (Agency for Defense Development)

  Wan-Ho CHO (Korea Research Institute of Standards and Science)

- **Finance**

  Gihoon BYUN (Korea Maritime and Ocean University)

  Imseon HAN (Korea Maritime and Ocean University)

- **Publication**

  Sung-Hoon BYUN (Korea Research Institute of Ships & Ocean engineering)

  Jungpyo Hong (Changwon National University)

- **Student Program**

  Kyusik CHANG (VRSound)

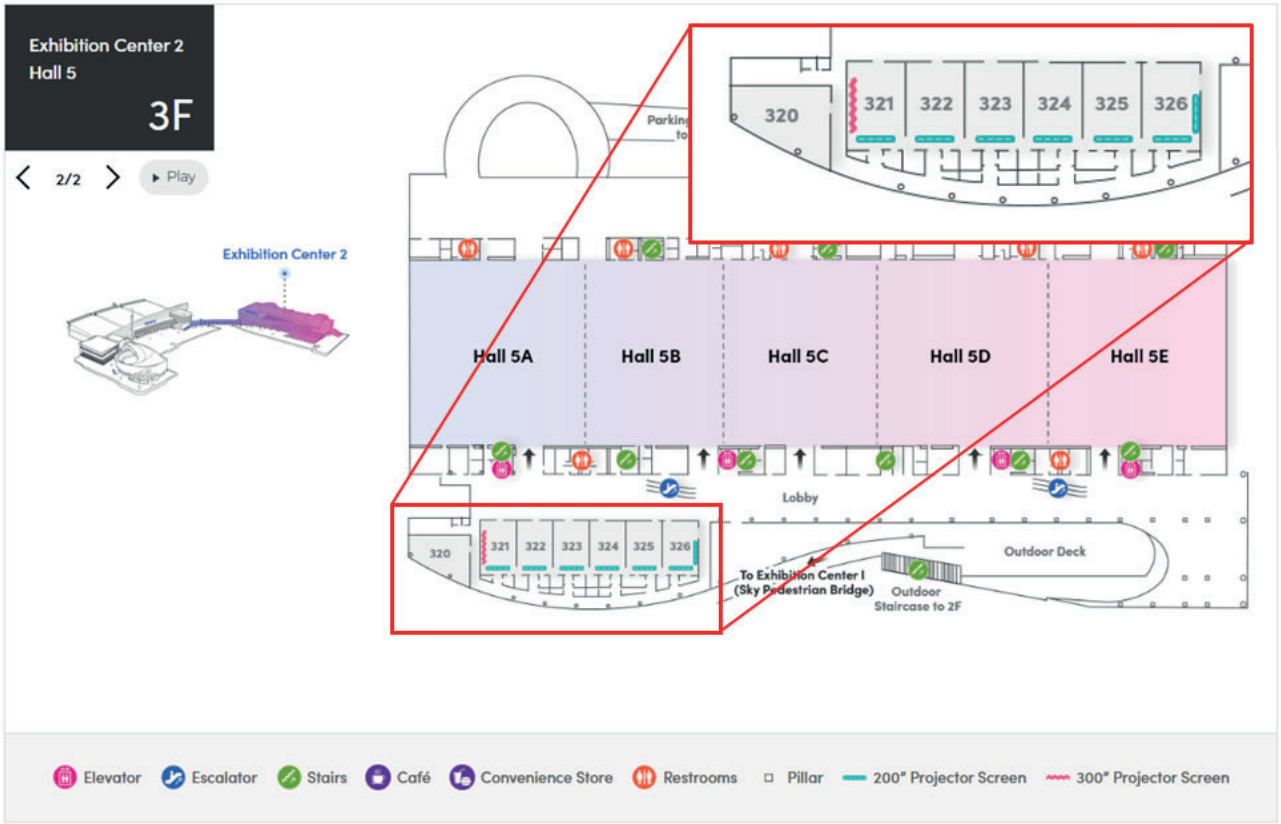
- **Conference Chair**

  Il SUNG (Agency for Defense Development)

VENUE

BEXCO, Exhibition Center II

* Room 320: Preparation room equipped with a laptop and printer.
* Rooms 321 and 322: Plenary Lectures
* Rooms 323 to 325: Oral Presentations
* Room 326: Poster Presentations



CONTACT INFORMATION

Conference Help Desk (On-site)

- Location : Registration Desk, 3F Lobby, BEXCO Exhibition Center II
- Email : ictca2025@gmail.com

Emergency Services (Korea)

- Police / Fire : ☎ 112 / 119
- Nearest Hospital :
  Pusan National University Hospital
  179 Gudeok-ro, Seo-gu, Busan
  Phone: +82-51-240-7000

ICTCA2025 Emergency Coordinator

- Name : Yonghwa Choi
- Mobile : +82-10-7232-1477
- Language : Korean / English available

SOCIAL PROGRAM

a. Monday 18 August 2025 – Welcome Reception

- The reception will offer light snacks and drinks.
- Place: Room 320 (VIP Room), Exhibition Center II, BEXCO
- Time: 16:20 – 17:40

b. Tuesday 19 August 2025 – Student Reception

- The Student Reception is open to all registered participants.
  Non-student attendees may join with a donation of 10,000 KRW.
  Light snacks and drinks will be served.
- Place: Room 320 (VIP Room), Exhibition Center II, BEXCO
- Time: 16:00 – 17:40

c. Wednesday 20 August 2025 – City Tour

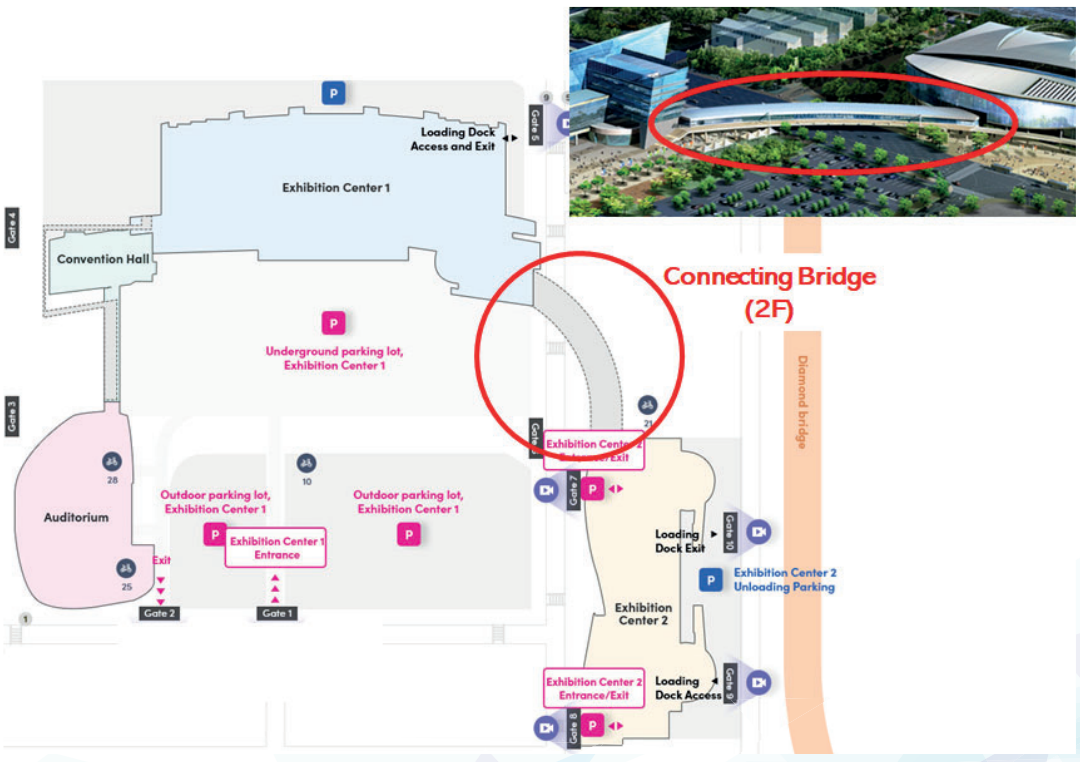
- Assembly Point: Gate 8 at the Main Entrance of Exhibition Center II, BEXCO

- Time: 13:30 – 17:30
- Tour Itinerary
  Stop 1: Hae-dong Yonggungsa Temple
  Stop 2: Gijang Balhyo (Traditional Makgeolli-making experience)

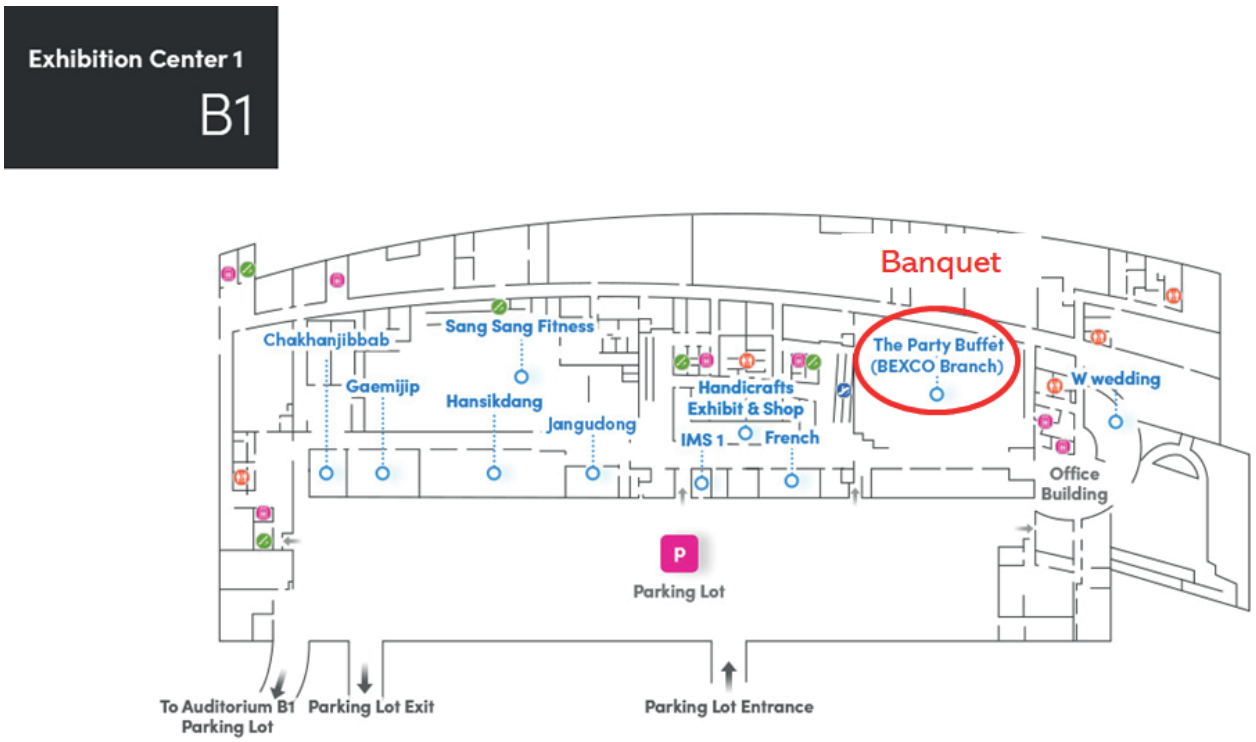
※ Participants may register for the city tour at the registration desk until the conclusion of Tuesday's sessions. (First come, first served.)

d. Wednesday 20 August 2025 – Banquet

- Place: The Party Buffet, Exhibition Center I (Basement Level 1), BEXCO
- Time: 18:00 – 20:00 (following the social tour)



SOCIAL PROGRAM



e. Thursday 21 August 2025 – Farewell party & Student award ceremony

- Place: Room 320 (VIP Room), Exhibition Center II, BEXCO
- Time: 16:00 – 17:40

LUNCH VOUCHER

ICTCA 2025
16th International Conference on Theoretical and Computational Acoustics

Lunch Voucher
₩12,000

Lunch Voucher Usage Guide

* Valid only for lunch from August 18 to August 21, 2025
* Location: Restaurants on B1 floor, Exhibition Hall 1, BEXCO (Korean Restaurant, Gaemijip, Chakan Jipbap, French, Jang Udon)

※ Any amount exceeding the voucher value must be paid by the individual.

Lunch Voucher Information

- Each registered participant will receive two lunch vouchers.
- Each voucher is valued at **12,000 KRW**, based on the Korean Won.
- Vouchers can be used during the conference period at participating restaurants located on the **basement level of Exhibition Center I, BEXCO.**
- A total of five restaurants accept the lunch vouchers. The list is provided below.

Restaurants

Jangudong

Location Exhibition Center 1, B1F
Tel +82-51-740-7475
Industry Restaurant (Korean)
Time 09:00 ~ 18:00

French

Location Exhibition Center 1, 1F
Tel +82-51-740-7897
Industry Restaurant (Pork Cutlet, Pasta, etc.)
Time 10:00 ~ 18:00

Chakhanjibbab

Location Exhibition Center 1, B1F
Tel +82-51-740-8008
Industry Restaurant (Korean)
Time 09:00 ~ 18:00

Gaemijip

Location Exhibition Center 1, B1F
Tel +82-51-740-7666
Industry Restaurant (Stir-fried Octopus)
Time 09:00 ~ 18:00

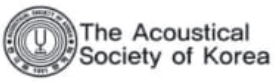
Hansikdang

Location Exhibition Center 1, B1F
Tel +82-51-740-7776
Industry Restaurant (Korean)
Time 09:00 ~ 18:00

SPONSORS

ASA
ACOUSTICAL SOCIETY
OF AMERICA

EAA


The Acoustical
Society of Korea

TQV

SAEVIT
TECHNOLOGIES

ORUN

HUENS

stx Engine

BOSTECH

KOREA
TOURISM
ORGANIZATION

BUSAN

SonarTech

Hanwha Systems

LIG Nex1

PROGRAM OVERVIEW

Monday, August 18, 2025					
ROOM	**321+322**	**323**	**324**	**325**	**326**
10:40-13:20	**Registration (3F)**				
13:20-14:00	Plenary Lecture 1 (Chi-Fang Chen, Chair : Jeasoo Kim)				
14:00-14:20	Coffee Break				
14:20-14:40		Structural Acoustics and Vibration	Acoustic Scattering and Geoacoustic Inversion in Marine Sediments	Model-Based and Data-Driven Approaches in Underwater Acoustics	
14:40-15:00					
15:00-15:20					
15:20-15:40					
15:40-16:00					
16:00-16:20					
16:20-17:40	**Welcome Reception**				
Tuesday, August 19, 2025					
ROOM	**321+322**	**323**	**324**	**325**	**326**
09:00-09:20		Beamforming and DOA Estimation in Underwater Acoustics	Underwater Acoustic Source Localization and Tracking	Inverse Design of Acoustic Metamaterials	
09:20-09:40					
09:40-10:00					
10:00-10:20					
10:20-10:40					
10:40-11:00	Coffee Break				
11:00-11:40	Plenary Lecture 2 (Haiqiang Niu, Chair : Gihoon Byun)				
11:40-13:20	Lunch				
13:20-14:00	Plenary Lecture 3 (Wonju Jeon, Chair : Wanho Cho)				
14:00-14:20	Coffee Break				
14:20-16:00					Poster
16:00-17:40	**Student Reception**				

PROGRAM OVERVIEW

Wednesday, August 20, 2025					
ROOM	**321+322**	**323**	**324**	**325**	**326**
09:00-09:20		Low-Frequency Acoustic Propagation and Sonar Signal Processing	Environmental Effects on Deep Ocean Acoustic Propagation	Acoustic Metamaterials and Metasurfaces for Noise Reduction and Wave Control	
09:20-09:40					
09:40-10:00					
10:00-10:20					
10:20-10:40					
10:40-11:00	Coffee Break				
11:00-11:40	Plenary Lecture 4 (Andrés Prieto Aneiros, Chair : Steffen Marburg)				
11:40-13:20	Lunch				
13:30-17:30	**City Tour**				
18:00-20:00	**Banquet**				
Thursday, August 21, 2025					
ROOM	**321+322**	**323**	**324**	**325**	**326**
09:00-09:20		Advanced Tracking Algorithms and Acoustic Estimation for Underwater Targets	Receiver Design and Robust Signal Processing for Underwater Acoustic Communications	Ultrasound and Acoustic Materials for Performance Enhancement and Structural Monitoring	
09:20-09:40					
09:40-10:00					
10:00-10:20					
10:20-10:40					
10:40-11:00	Coffee Break				
11:00-11:40	Plenary Lecture 5 (Marcus Maeder, Chair : Jeasoo Kim)				
11:40-13:20	Lunch				
13:20-14:00	Plenary Lecture 6 (Sean F. Wu, Chair : Chi-Fang Chen)				
14:00-14:20	Coffee Break				
14:20-14:40		AI and Acoustic Sensing Applications in Industry, Ecology, and Medicine	Modeling and Detection of Acoustic Signatures from Complex Underwater Structures	Computational and Intelligent Approaches in Speech and Acoustics	
14:40-15:00					
15:00-15:20					
15:20-15:40					
15:40-16:00					
16:00-17:40	**Farewell Party & Student Award Ceremony**				

PROGRAM OVERVIEW

Friday, August 22, 2025					
ROOM	321+322	323	324	325	326
09:00-11:40	**Committee Meeting**				

PLENARY LECTURES

Plenary Lecture 1
13:20–14:00, Aug 18 (Mon)

LONG-TERM MARINE SOUNDSCAPE MONITORING IN TAIWAN' S OFFSHORE WINDFARM AREAS: ECOLOGICAL INSIGHTS AND CONSERVATION INPLICATIONS

Chi-Fang Chen
National Taiwan University, Taiwan

Taiwan has set an ambitious target of achieving carbon neutrality by 2050, with offshore wind power anticipated to generate 40–55 GW. The rapid development of offshore wind farms along the Eastern Taiwan Strait (ETS) introduces significant underwater noise from construction activities such as pile driving, increased vessel traffic during surveys and construction, and continuous operational noise over a projected lifespan of 20–30 years. This escalating anthropogenic noise imposes considerable stress on the marine soundscape and its ecological constituents. Our research focuses on long-term acoustic monitoring at wind farm sites (2014–2024), investigating the changes in sound levels and their ecological consequences, particularly the acoustic behaviors of fish and the critically endangered Taiwanese humpback dolphin (Sousa chinensis taiwanensis). For the first time, our findings reveal clear impacts of pile-driving and operational noise on fish vocalization patterns and demonstrate how vessel traffic and windfarm-related noise influence the acoustic behavior of Taiwanese humpback dolphins. Additionally, we have developed automated detection algorithms to effectively identify vocalizations of both fish and dolphins from extensive acoustic datasets. As offshore wind energy expands, these long-term acoustical datasets serve as essential baseline data for monitoring evolving marine soundscapes and understanding the acoustic responses of marine species throughout the lifecycle of offshore windfarms. These insights are critical for informing conservation strategies and sustainable management practices in Taiwan' s transition to renewable energy. This research is funded by Taiwan National Science and Technology Council, Unitech Inc., Ørsted Taiwan, Copenhagen Infrastructure Partners (CIP), Taiwan Power Company.

Plenary Lecture 2
11:00–11:40, Aug 19 (Tue)

ADVANCES AND APPLICATIONS OF MACHINE LEARNING IN UNDERWATER ACOUSTIC SOURCE LOCALIZATION AND PROPAGATION MODELING

Haiqiang Niu
Institute of Acoustics, Chinese Academy of Sciences, People's Republic of China

In this talk, I will present our recent advances in applying machine learning to underwater acoustic source localization and propagation modeling. Our studies on source localization cover a range of environments, from shallow waters to deep sea scenarios, utilizing both real and synthetic data sets for training. The results demonstrate the superiority of machine learning methods over traditional approaches in handling the environmental uncertainties. We will also discuss some conclusions and the challenging problems encountered in source localization applications. The second focus of this presentation is on acoustic propagation modeling using neural operators. Unlike physics-informed neural network (PINN) methods, neural operators derive underlying relationships primarily from extensive, well-prepared data sets. Instead of learning mappings between finite-dimensional Euclidean spaces, these data-driven neural operators learn mappings between infinite-dimensional function spaces, which is particularly attractive for sound propagation modeling tasks. We will examine the generalization capabilities of neural operators when applied to sound propagation modeling in range-independent shallow water environments.

PLENARY LECTURES

Plenary Lecture 3
13:20–14:00, Aug 19 (Tue)

ACOUSTIC BLACK HOLES AND META-SURFACES: NEW SOLUTIONS FOR OLD PROBLEMS

Wonju Jeon
Korea Advanced Institute of Science and Technology, Republic of Korea

In this talk, recent advances in acoustic and elastic meta-structures to control noise and vibration are presented aiming at practical applications to real-world problems in our daily life and various industries. The first example is an ultra-light (20 times lighter than existing materials) soundproofing meta-panel to insulate broadband noise generated from electric vehicles, with the aid of negative mass density in low-frequency range (road noise) and negative bulk modulus in high-frequency range (motor noise). The second one is an acoustic meta-liner to insulate noise in a duct while allowing flow, by designing ultra-thin acoustic meta-surface with the consideration of visco-thermal losses in deep-subwavelength-scale Helmholtz resonators. The third one is a meta-damper to suppress vibration in beams or plates using waveguide absorbers based on Archimedean spiral acoustic black holes. For the three meta-structures, we present how to design them theoretically and validate their performance experimentally with a couple of audiovisual demonstrations.

Plenary Lecture 4
11:00–11:40, Aug 20 (Wed)

PERFECTLY MATCHED LAYERS: ACHIEVEMENTS AND FUTURE CHALLENGES

Andrés Prieto Aneiros
Galician Center of Mathematical Research and Technology, University of A Coruña, Spain

The Perfectly Matched Layer (PML) technique has become a reliable and efficient method for computing free-field numerical approximations in time-domain and frequency-domain models over the last three decades. Its combination with the widely used finite difference and finite element methods has spread its popularity and generalized its use to broad different areas of applications such as outdoor acoustics, aeroacoustics, structural mechanics, underwater acoustics, electromagnetism, optics, or geophysics, all of them with a common denominator: the wave propagation phenomena settled initially in an unbounded physical domain of interest. Currently, the PML technique is a key component in many well-known, established commercial and open-source software packages in computational acoustics. The study of its robustness, usability, and accuracy has gained attention from various research communities, facilitating the numerical analysis of its properties, extending the PML techniques to a number of different models, and enhancing its computational performance with optimal settings. This talk reviews the theoretical properties and computational capabilities of different state-of-the-art PML methods, and discusses the open questions and challenges that will be addressed in the near future of PML technique development.

PLENARY LECTURES

Plenary Lecture 5
11:00–11:40, Aug 21 (Thu)

IN THE AGE OF BIG DATA AND AI – RISKS AND OPPORTUNITIES FOR VIBROACOUSTICS

Marcus Maeder
Technical University of Munich, Germany

By looking back at the technology sector, common phrases such as the Internet of Things, Industry 4.0, Big Data, and Artificial Intelligence dominated the discussions through various fields of research, industry, and economy. Due to the rapid development of parallel processing hardware technology, algorithms, and the corresponding software solutions since 2015, this trend experienced unprecedented acceleration, with the release of the first ChatGPT version in late 2022, highlighting the age of AI. Despite the ongoing development of Artificial Intelligence, the excessive expectations in the technology followed by depression and disappointment as one part of a typical hype cycle, this technology offers a wide range of possibilities for researchers and engineers in the field of theoretical, experimental, and computational vibroacoustics if applied correctly and with care. This presentation examines current developments in artificial intelligence together with risks and opportunities when utilizing the technology to solve problems in the field of vibroacoustic. The importance of high-quality data and its associated nature are stressed, leading to a knowledge-and-experience-enhanced artificial intelligence (keeAI) incorporating problem-specific domain expertise of the developer and engineer, respectively. The presented examples serve as a possible template for future developments and help to facilitate the application of Big Data approaches and Artificial Intelligence beyond the disappointment of excessive expectations.

Plenary Lecture 6
13:20–14:00, Aug 21 (Thu)

ADVANCES IN UNDERWATER ACOUSTICS, STRUCTURAL ACOUSTICS, COMPUTATIONAL ACOUSTICS, AND AEROACOUSTICS

Sean F. Wu
Department of Mechanical Engineering, Wayne State University, U.S.A.

This paper presents a comprehensive review of key developments in four major branches of acoustics—Underwater Acoustics, Structural Acoustics, Computational Acoustics, and Aeroacoustics—as reflected in contributions published in the Journal of Theoretical and Computational Acoustics over the past thirty years. For each of these areas, we systematically classify the literature into three major categories:

Review Articles that synthesize foundational knowledge and provide historical context, New Methodologies that introduce cutting-edge theoretical frameworks and advanced computational techniques, and Emerging Applications that showcase the practical deployment of recent innovations, often validated through experimental investigations.

This structured review highlights major trends, breakthroughs, and evolving research directions in both theory and application. In Underwater Acoustics, advancements in signal propagation modeling and sonar system design are emphasized. Structural Acoustics features innovations in wave-structure interaction and vibroacoustic coupling. Computational Acoustics showcases high-performance numerical solvers, boundary element and finite element methods, and hybrid algorithms. In Aeroacoustics, recent progress includes improved modeling of turbulent flows, noise prediction for aircraft and wind turbines, and experimental validation of simulation results. By curating and contextualizing this body of work, we aim to provide researchers, educators, and practitioners with a valuable reference point and roadmap for future research. We hope this review serves as a gateway to deeper exploration, enabling readers to trace developments through the referenced literature and engage with ongoing innovations in these dynamic fields of acoustics.

DAILY PROGRAM

AUGUST 18 (MON)

14:20 – 16:00	Structural Acoustics and Vibration	323 (3F)
Chair	Seong-Hyun Lee (Korea Institute of Machinery and Materials)	
14:20	Continuation methods applied to topology optimization with discrete variables	
	Sang Won Kang, Gil Ho Yoon, Young Hun Choi	
14:40	Reduction of low-frequency and broadband vibrations in plates using vortex-shaped embedded acoustic black holes	
	Taehwan Son, Seongmin Park, Wonju Jeon	
15:00	Temperature-dependent sound absorption characteristics of microlattice materials backed with air cavity	
	Xiaozhen Li, Xiaobing Cai	
15:20	Sound absorption performance of bead-filled honeycomb composite structure	
	Long Xu, Xiaobing Cai	
15:40	Meta-fence for Rayleigh wave isolation	
	Hongjun Fan, Yongquan Liu	

Continuation methods applied to topology optimization with discrete variables

Gil Ho Yoon[1]; Young Hun Choi[1]; Sang Won Kang[1]
[1] Department of Mechanical Engineering, Hanyang University, Seoul, South Korea

ABSTRACT

This study proposes a numerical method that improves the conventional topology optimization of binary structures (TOBS) by allowing gradual control over the number of discrete values that design variables can take. The conventional TOBS method has the advantage of generating the design without gray elements by restricting design variables to either 0 or 1. However, it typically relies on linearization and integer programming, which can lead to instability when addressing highly nonlinear objective functions. This is primarily because abrupt changes in element density between 0 and 1 make linear approximations inaccurate. To overcome this limitation, the proposed method introduces a continuation strategy in which the number of allowable discrete values is initially large and gradually reduced throughout the optimization process. This enables smoother transitions in the early stages and ensures convergence to a binary solution in the final stage, thereby enhancing stability and convergence for highly nonlinear problems. The effectiveness of the proposed method is demonstrated through its application to a topology optimization problem for minimum dynamic compliance. Comparative analyses with a SIMP method and the conventional TOBS method show that the proposed method demonstrates superior performance and confirms its potential applicability in vibration and acoustic engineering.

Keywords: Topology optimization, Integer programming

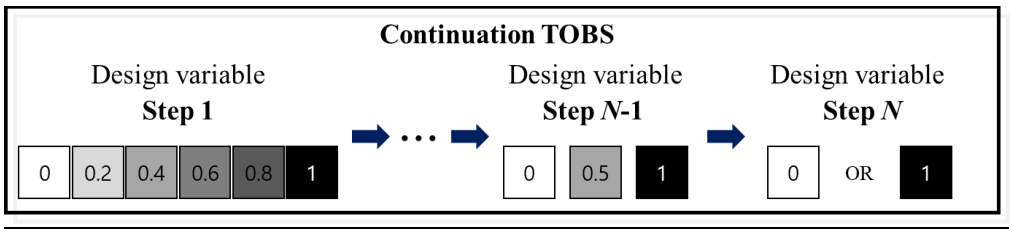


Figure 1 – Description of continuation TOBS method

ACKNOWLEDGEMENTS

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (RS-2024-00351611).

REFERENCES

1. Sivapuram, Raghavendra, Renato Picelli, and Yi Min Xie. "Topology optimization of binary microstructures involving various non-volume constraints." *Computational Materials Science* 154 : 405-425, 2018.
2. Sivapuram, Raghavendra, and Renato Picelli. "Topology optimization of binary structures using integer linear programming." *Finite Elements in Analysis and Design* 139 :49-61, 2018.
3. Yoon, Gil Ho. "Structural topology optimization for frequency response problem using model reduction schemes." *Computer Methods in Applied Mechanics and Engineering* 199.25-28 (2010): 1744-1763.

Reduction of low-frequency and broadband vibrations in plates using vortex-shaped embedded acoustic black holes

Taehwan SON[1]; Seongmin PARK[2]; Wonju JEON[3]
[1, 2, 3] Korea Advanced Institute of Science and Technology, Republic of Korea

ABSTRACT

This study proposes a novel compact design for two-dimensional embedded acoustic black holes (ABHs) aimed at damping vibrations across low and broadband frequencies. The proposed embedded ABH comprises multiple vortex-shaped ABHs arranged rotationally along the edge of a circular region, allowing longer ABHs to be accommodated within the same area and thereby achieving a lower cut-on frequency. Compared to a conventional 2D embedded ABH occupying the same installation area, the proposed design achieves significantly lower cut-on frequencies and demonstrates superior vibration damping, as validated by comparing numerical and experimental results. Moreover, its performance is further enhanced by breaking the rotational symmetry of the vortex-shaped ABHs.

Keywords: Flexural wave, Vibration suppression, Acoustic black hole

[1] taehwan.son@kaist.ac.kr
[2] kaistian13@kaist.ac.kr
[3] wonju.jeon@kaist.ac.kr

PROCEEDINGS of the
16th International Conference on Theoretical and Computational Acoustics

August 18 to 22, 2025 BEXCO, Busan Korea

Temperature-dependent sound absorption characteristics of microlattice materials backed with air cavity

Xiaozhen Li[1], Xiaobing Cai[2]

[1] State Key Laboratory for Strength and Vibration of Mechanical Structures, School of Aerospace Engineering, Xi'an Jiaotong University, Xi'an 710049, China

[2] State Key Laboratory for Strength and Vibration of Mechanical Structures, School of Aerospace Engineering, Xi'an Jiaotong University, Xi'an 710049, China

ABSTRACT

Acoustic metamaterials artificially designed for high-temperature applications are essential for noise reduction under extreme conditions, especially for controlling low-frequency noise. Conventional porous materials suffer from material degradation and reduced absorption efficiency at high temperatures. Moreover, the irregularity of their inherent micropores makes it difficult to precisely control and optimize their sound absorption performance. In this paper, we utilized advanced additive manufacturing technology to fabricate geometrically regular microlattice materials with tunable sound absorption. By integrating this microlattice material with a backed air cavity, the low-frequency sound absorption is enhanced by multiple energy dissipation mechanisms, including viscous losses, thermal conduction, and resonance effects. We studied the temperature-dependent sound absorption characteristics of this composite absorber through theory, simulation, and experiment. The sound absorption mechanisms and temperature effects are revealed based on numerical simulation results. Finally, the influence of geometrical parameters on sound absorption is discussed and the optimal configuration is found. Compared with traditional porous materials, this composite absorber not only delivers superior low- and broadband sound absorption but also has a unique performance tunability for targeted applications. This study contributes to the development of metamaterials for noise control in high-temperature applications.

[1] 3120106019@stu.xjtu.edu.cn

[2] caixiaobing@xjtu.edu.cn

PROCEEDINGS of the
16th International Conference on Theoretical and Computational Acoustics

August 18 to 22, 2025 BEXCO, Busan Korea

Sound absorption performance of bead-filled honeycomb composite structure

Long Xu[1], Xiaobing Cai[2]

[1] School of Aerospace Engineering, Xi'an Jiaotong University

[2] School of Aerospace Engineering, Xi'an Jiaotong University

ABSTRACT

As a typical porous material, the research on the sound absorption capacity of granular materials has gradually become an active direction. Compared with foam and fibers, granular materials have rheological properties and often need to be cured or inhibited from flowing when used. The numerous small cavities inside honeycomb materials offer the possibility of designing the distribution of granular materials. The honeycomb materials have many advantages such as lightweight and high specific strength, and can also improve the sound absorption performance of the structure. Therefore, in this paper, a honeycomb structure is used to accommodate granular materials and combined with micro-perforated plates to form a composite sandwich structure, in order to further improve its sound absorption capacity. In this paper, the sound absorption curve of the material was obtained through experiments and simulations, and the influence of parameters such as the pore size of the perforated plate and the particle diameter on the sound absorption of the structure was analyzed. The data and conclusions obtained provide support for the application of bead-filled honeycomb sandwich panels in noise reduction.

[1] 3120106019@stu.xjtu.edu.cn

[2] caixiaobing@xjtu.edu.cn

PROCEEDINGS of the
16th International Conference on Theoretical and Computational Acoustics

August 18 to 22, 2025 BEXCO, Busan Korea

Meta-fence for Rayleigh wave isolation

Hongjun Fan[1], Yongquan Liu[2]

[1] State Key Laboratory for Strength and Vibration of Mechanical Structures, School of Aerospace Engineering, Xi'an Jiaotong University

[2] State Key Laboratory for Strength and Vibration of Mechanical Structures, School of Aerospace Engineering, Xi'an Jiaotong University

ABSTRACT

The protection of infrastructures against earthquakes has been a primary objective within the field of civil engineering. The emergence of seismic metamaterials has offered an unprecedented opportunity to design advanced aseismic structures over the last decade. However, a large number of subwavelength resonators that span multiple wavelengths are always required, resulting in a bulky size of existing seismic metamaterials. In this work, we introduce an aboveground single-layer meta-fence to omnidirectionally isolate Rayleigh waves. Based on mode conversion and reflection, the meta-fence can be configured into an enclosed region to effectively safeguard inner infrastructures by reducing the amplitude by 85 %. Furthermore, the designed meta-fence can operate across a wide frequency spectrum ranging from 17 to 200 Hz. Compared with previous seismic metamaterials, our scheme of meta-fence exhibits a distinct advantage in compact size, thereby enhancing applicability in earthquake protection.

[1] fanhj@stu.xjtu.edu.cn
[2] liuy2018@xjtu.edu.cn

DAILY PROGRAM

AUGUST 18 (MON)

14:20 – 16:20	Acoustic Scattering and Geoacoustic Inversion in Marine Sediments 324 (3F)
Chair	Hefeng Dong (Norwegian University of Science and Technology)
14:20	Wavelet Scattering Characteristics and Range Correction of Spherical Targets in Random Boundary Shallow Water Wave-guide
	Jiani Wen, Haisen Li, Jian Wang, Tianyao Xing
14:40	The inference of Deep-Sea Sedimentary Layer Properties in the Northwest Subbasin of the South China Sea
	Yongchao Guo, Haigang Zhang, Minghui Zhang
15:00	Numerical comparison of the frequency response function of different poroelastic models in stratified sediments using pointwise sources
	Andrés Prieto Aneiros
15:20	Numerical Analysis of Acoustic Scattering from Buried Targets under Varying Sediment and Target Conditions
	Yeon-Seong Choo, Sung-Hoon Byun, Sea-Moon Kim
15:40	Geoacoustic Inversion of Continuous Sediment Layer in Deep Water Using Ship Noise
	Houxuan Jiang, Minshuai Liang, Junjie Shi, Dajun Sun
16:00	A Physics-Based Deep Learning Method for Rigid Scattering
	Arzhang Angoshtari

PROCEEDINGS of the
16th International Conference on Theoretical and Computational Acoustics

August 18 to 22, 2025 BEXCO, Busan Korea

Wavelet Scattering Characteristics and Range Correction of Spherical Targets in Random Boundary Shallow Water Wave-guide

Jiani WEN[1,3,4]; Haisen LI[1,3,4]; Jian WANG[2]; Tianyao XING[1,3,4]
[1] National Key Laboratory of Underwater Acoustic Technology, Harbin Engineering University, China
[2] Shenzhen Institute for Advanced Study, University of Electronic Science and Technology of China
[3] Key Laboratory of Marine Information Acquisition and Security, Harbin Engineering University, China
[4] College of Underwater Acoustic Engineering, Harbin Engineering University, China

ABSTRACT

Key challenges in ocean acoustics include modeling, analysis, and the efficient computation of shallow water acoustic fields. Traditional methods for solving non-homogeneous partial differential equations struggle with the complexities introduced by random interfaces, medium absorption, target scattering, and multi-path effects from sea surface and seabed fluctuations, leading to issues in calculation complexity, accuracy, and efficiency. This study introduces a Wavelet Jointless-Weighted Nodal Method to address non-grid partial differential equations for shallow water wave-guides with random boundary. A nonlinear scattering cylindrical coordinate model for spherical targets in shallow water is developed, accounting for random interfaces. During model resolution, correlation terms are expanded using quasi-Shannon wavelet transforms for distance, modified quasi-Shannon wavelet transforms for depth and fourier wavelet transforms for angle. A dense point and weighted least squares iterative strategy is then employed. Simulations demonstrate that this method significantly enhances numerical accuracy with the same iteration count. Finally, the Brent method is utilized to refine the detection range of active sonar, enhancing the accuracy of ranging calculations in active sonar detection systems.

Keywords: Non-homogeneous partial differential equation, Randomly fluctuating boundary, Quasi-scale function

1. INTRODUCTION

The computation and analysis of complex coupled acoustic fields in shallow water have long been one of the key challenges in computational ocean acoustics(1). In shallow marine environments, the coupling effects of scattering from upper and lower randomly fluctuating boundaries cannot be simply approximated as a linear superposition, a phenomenon often described as "the sea surface begins to sense the presence of the seabed."(2). Analytical solutions are nearly impossible for such complex boundary problems. Currently, effective numerical methods include the finite element method (FEM), boundary element method (BEM)(3), perturbation method, and meshless collocation method. Among these, the meshless collocation method holds greater research significance due to its superior adaptability to complex boundaries and lower computational complexity(4).

Based on wavelet transform theory, this paper proposes a 3D Wavelet Jointless-Weighted Nodal Method for approximating and computing the scattered acoustic field from spherical targets in shallow water with random interfaces. Simulations solutions verify that this method possesses high computational accuracy and efficiency. Building on this conclusion, we also propose a range correction method for single-hydrophone horizontal ranging. Simulation experiments demonstrate that this method can theoretically limit ranging errors to within 10% for active sonar. The Wavelet Jointless-Weighted Nodal Method shows significant potential for theoretical research and applications in nonlinear acoustics and wavelet transform-based engineering signal processing.

2. Acoustic Field Model for Spherical Targets with Random Interfaces in Shallow Water

An acoustic field model for a spherical target with random interfaces in a shallow-water waveguide is

[1] hsli@hrbeu.edu.cn

established in the cylindrical coordinate system under excitation. The model considers only the steady-state acoustic field of a linear monochromatic signal (i.e., time-independent). The origin is set at the depth denoted as z_f. The target is a sphere of radius a centered at coordinates (R_0, θ_0, z_0).The ocean depth is H, with both sea surface and seabed modeled as random interfaces. The schematic diagram of the acoustic field model is shown by Figure 1.

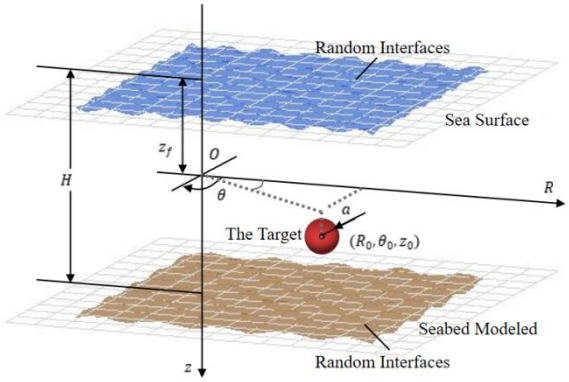


Figure 1 –The schematic diagram of the acoustic field model

Based on the above model setup, the non-homogeneous Helmholtz equation for the velocity potential function $\phi(R, z, \theta)$(abbreviated as ϕ) can be formulated as follows:

$$\nabla^2\phi(R, z, \theta) + K^2\phi(R, z, \theta) = 0 \tag{1}$$

$$\nabla^2 = \frac{1}{R}\frac{\partial}{\partial R}(R\frac{\partial}{\partial R}) + \frac{1}{R^2}\frac{\partial^2}{\partial \theta^2} + \frac{\partial^2}{\partial z^2} \tag{2}$$

the boundary conditions are determined by three aspects: the undulating sea surface, the undulating seabed, and the spherical target.Based on Airy wave theory,the sea surface elevation η and seabed fluctuation h follow a superposition of cosines with random phases, expressed as:

$$\eta(R) = \sum_{n=1}^{\infty} \eta_n \cos(k_{\eta n}R + \phi_n) \ ; \ h(R) = \sum_{n=1}^{\infty} H_n \cos(k_{hn}R + \phi_n) \tag{3}$$

where $k_{\eta n}$、k_{hn} are constants，ϕ_n、$\eta(R)$ and $h(R)$satisfy the dispersion relations：

$$\phi_n \sim U[0,2\pi] \ ; \ E[\eta(R)] = 0 \ ; \ E[h(R)] = 0 \tag{4}$$

For the spherical target boundary condition, the target surface is treated as a perfectly rigid boundary.The boundary condition of the target at (R_0, z_0, θ_0) are as follows:

$$\frac{\partial\phi}{\boldsymbol{n}} = \nabla\phi \cdot \boldsymbol{n} = 0 = (R - R_0\cos(\theta - \theta_0))\frac{\partial\phi}{\partial R} + \frac{R_0\sin(\theta - \theta_0)}{R}\frac{\partial\phi}{\partial\theta} + (z - z_0)\frac{\partial\phi}{\partial Z} \tag{5}$$

In summary, the governing acoustic waveguide model for a spherical target with random interfaces in shallow water, excited by a point source, is defined by：

$$\nabla^2\phi + K^2\phi = 0 \tag{6}$$

$$\begin{cases} \phi|_{z=\eta(R)-z_f} = 0 \\ \dfrac{\partial\phi}{\partial z}|_{z=H-z_f+h(R)} = 0 \\ (R - R_0\cos(\theta - \theta_0))\dfrac{\partial\phi}{\partial R} + \dfrac{R_0\sin(\theta - \theta_0)}{R}\dfrac{\partial\phi}{\partial\theta} + (z - z_0)\dfrac{\partial\phi}{\partial Z} = 0 \end{cases} \tag{7}$$

3. WAVELET JOINTLESS-WEIGHTED NODAL METHOD

Based on the framework of the Wavelet-Based Meshless Collocation Method, we solve this 3D non-homogeneous partial differential equation. The solution assumes that the radial R, axial z, and angular θ variables are mutually independent, and the velocity potential function ϕ can be decomposed as a product of terms containing these three variables:

$$\phi(R, \theta, z) = \varphi(R)\rho(\theta)\psi(z) \tag{9}$$

The core concept of the Wavelet-Based Meshless Collocation Method is to expand the theoretical solution using wavelet bases, thereby transforming the non-homogeneous partial differential equation into the task of approximating the wavelet coefficients for each term. These coefficients are then recombined to obtain the solution of the non-homogeneous PDE. For this 3D acoustic model, each term in the expression

above can be separately expanded using wavelets Eq. (10-12), i.e., expressing ϕ as an expansion in a 3D tensor-product basis Eq. (13). By discretizing the solution domain with collocation points along each dimension, the target PDE is recast into a system represented by the Kronecker product Eq. (14) of collocation matrices.

$$\varphi(R) \approx \sum_{m=1}^{M_R} R_m^{(R)} \varphi_m(R) \tag{10}$$

$$\rho(\theta) \approx \sum_{n=1}^{N_\theta} \Theta_n^{(\theta)} \rho_{\mathrm{n}}(\theta) \tag{11}$$

$$\psi(z) \approx \sum_{q=1}^{Q_z} Z_q^{(z)} \psi_{\mathrm{q}}(z) \tag{12}$$

$$\phi(R,\theta,z) = \sum_{\mathrm{m,n,q}}^{M_R \times N_\theta \times Q_z} R_{\mathrm{m}}\, \Theta_{\mathrm{n}} Z_{\mathrm{q}} \varphi_{\mathrm{m}}(\mathrm{R}) \rho_{\mathrm{n}}(\theta) \psi_{\mathrm{q}}(z) \tag{13}$$

$$\begin{aligned} &[\phi_{1,1,1} \quad \phi_{1,1,2} \quad \cdots \quad \phi_{m,n,q}]^T \\ &= \left(\begin{bmatrix} \varphi_{11} & \cdots & \varphi_{1M_R} \\ \vdots & \ddots & \vdots \\ \varphi_{\mathrm{m1}} & \cdots & \varphi_{mM_R} \end{bmatrix} \begin{bmatrix} R_1 \\ \cdots \\ R_{M_R} \end{bmatrix} \right) \otimes \left(\begin{bmatrix} \rho_{11} & \cdots & \rho_{1N_\theta} \\ \vdots & \ddots & \vdots \\ \rho_{\mathrm{n1}} & \cdots & \rho_{nN_\theta} \end{bmatrix} \begin{bmatrix} \Theta_1 \\ \cdots \\ \Theta_{N_\theta} \end{bmatrix} \right) \otimes \left(\begin{bmatrix} \psi_{11} & \cdots & \psi_{1Q_z} \\ \vdots & \ddots & \vdots \\ \psi_{\mathrm{q1}} & \cdots & \psi_{qQ_z} \end{bmatrix} \begin{bmatrix} Z_1 \\ \cdots \\ Z_{Q_z} \end{bmatrix} \right) \end{aligned} \tag{14}$$

We fully account for the physical properties of each component by selecting:

1. Radial basis (φ_m): Quasi-Shannon wavelet bases derived from Shannon wavelets(5,6);

2. Axial basis (ψ_{q}): Modified quasi-Shannon wavelet bases adapted to random interfaces, inherently satisfying the sea surface boundary condition to simplify stochastic boundary computations;

3. Angular basis (ρ_{n}): Fourier bases.

The expressions are formulated as:

$$\varphi_m(\mathrm{R}) = \frac{\sin(\pi R/\Delta)}{\pi R/\Delta} e^{-\frac{R^2}{2K^2}} \tag{15}$$

$$\psi_{\mathrm{q}}(z) = \frac{\sin(\pi(z-\eta(R)+z_f)/\Delta)}{\pi z/\Delta} e^{-\frac{z^2}{2K^2}} \tag{16}$$

$$\rho_{\mathrm{n}}(\theta) = e^{-jn\theta} \tag{17}$$

Combining the Helmholtz equation with seabed and spherical target boundary conditions yields the following homogeneous linear system for solution.We employ the weighted least squares method for solution.

$$\begin{cases} \nabla^2\phi + K^2\phi = 0 \\ \phi|_{\,z=\eta(R)-z_f} = 0 \\ (R - R_0\cos(\theta-\theta_0))\dfrac{\partial\phi}{\partial R} + \dfrac{R_0\sin(\theta-\theta_0)}{R}\dfrac{\partial\phi}{\partial\theta} + (z-z_0)\dfrac{\partial\phi}{\partial Z} = 0 \\ \Rightarrow \begin{bmatrix} K_{1,1,1} & \cdots & K_{m,n,q} \\ D_{1,1,1} & \cdots & D_{m,n,q} \\ H_{1,1,1} & \cdots & H_{m,n,q} \end{bmatrix} \begin{bmatrix} R_1\Theta_1 Z_1 \\ \cdots \\ R_{\mathrm{m}}\Theta_{\mathrm{n}} Z_{\mathrm{q}} \end{bmatrix} = 0 \end{cases} \tag{18}$$

where:

$$\begin{cases} K_{m,n,q} = \dfrac{\partial^2\varphi_{\mathrm{m}}}{\partial R^2}\rho_{\mathrm{n}}\psi_{\mathrm{q}} + \dfrac{1}{R^2}\varphi_{\mathrm{m}}\dfrac{\partial^2\rho_{\mathrm{n}}}{\partial\theta^2}\psi_{\mathrm{q}} + \varphi_{\mathrm{m}}\rho_{\mathrm{n}}\dfrac{\partial^2\psi_{\mathrm{q}}}{\partial z^2} + \dfrac{1}{R}\dfrac{\partial\varphi_{\mathrm{m}}}{\partial R}\rho_{\mathrm{n}}\psi_{\mathrm{q}} + \dfrac{\omega^2}{{c_0}^2}\varphi_{\mathrm{m}}\rho_{\mathrm{n}}\psi_{\mathrm{q}} \\ D_{m,n,q} = \varphi_{\mathrm{m}}\rho_{\mathrm{n}}\dfrac{\partial\psi_{\mathrm{q}}}{\partial z} \\ H_{m,n,q} = (R - R_0\cos(\theta-\theta_0))\dfrac{\partial\varphi_{\mathrm{m}}}{\partial R}\rho_{\mathrm{n}}\psi_{\mathrm{q}} + \dfrac{R_0\sin(\theta-\theta_0)}{R}\varphi_{\mathrm{m}}\dfrac{\partial\rho_{\mathrm{n}}}{\partial\theta}\psi_{\mathrm{q}} + (z-z_0)\varphi_{\mathrm{m}}\rho_{\mathrm{n}}\dfrac{\partial\psi_{\mathrm{q}}}{\partial z} \end{cases} \tag{19}$$

4. HORIZONTAL RANGE CORRECTION METHOD FOR SINGLE-

HYDROPHONE SYSTEMS

During active sonar operations, assuming time synchronization between transmitter and receiver with transmitted signal interference suppressed, the echo arrival time can be measured via techniques like envelope detection. With sound speed treated as constant, the slant range Δx between the transducer and target can be derived.However, a single-hydrophone system lacks elevation resolution, preventing direct measurement of the depth difference and true horizontal range ΔR between transducer and target. In shallow-water long-range detection, approximating slant range Δx as horizontal range ΔR is common in engineering practice which is an approach that lacks rigor.

We propose a rapid horizontal range correction method for single-hydrophone systems based on the Wavelet-Based Meshless Collocation Method results. Let Δz denote the true depth difference between the transducer at (R_f, θ_f, z_f) and the target at (R_0, θ_0, z_0).The geometric relationship satisfies:

$$\begin{aligned} z_0 &= z_f - \Delta z \\ \Delta x^2 &= \Delta z^2 + \Delta R^2 \end{aligned} \tag{20}$$

From Eq. (18), the decomposition via wavelet bases yields $\begin{bmatrix} R_1\Theta_1Z_1 \\ \cdots \\ R_m\Theta_nZ_q \end{bmatrix}$.At this stage, substituting kreference points into $\begin{bmatrix} H_{1,1,1}{}^i & \cdots & H_{m,n,q}{}^i \end{bmatrix}$ should yield $z_0 = z_f - (\Delta x^2 - \Delta R^2)^{1/2}$.The Brent method is then applied to scan the region within the interval $\Delta R \in [0, \Delta x]$, where:

$$\min \left\{ \sum_{i=1}^{k} \left\| \begin{bmatrix} H_{1,1,1}{}^i & \cdots & H_{m,n,q}{}^i \end{bmatrix} \begin{bmatrix} R_1\Theta_1Z_1 \\ \cdots \\ R_m\Theta_nZ_q \end{bmatrix} \right\|^2 \right\} \tag{20}$$

This yields ΔR, from which Δz and z_0 is derived,completing the correction.

5. SIMULATION EXPERIMENTS

5.1 Simulation Experiment of Wavelet Jointless-Weighted Nodal Method

To better demonstrate the capabilities of our algorithm, the selected parameters and computational results are presented by Table1.

The velocity potential amplitude distribution reveals that while dominant trends in the shallow-water waveguide align with established patterns of regular-boundary spherical target scattering, irregular energy convergence zones emerge due to stochastic boundary interactions. Furthermore, as clearly evidenced by the computational results, Wavelet Jointless-Weighted Nodal Method delivers a significant speed advantage ,which compressing calculation times for complex boundary problems to the order of seconds.

Table 1 –Parameters and Computational Results

Origin Depth z_f	0m
Spherical Target Center (R_0, θ_0, z_0)	(20m,0°,25m)
Target Radius a	$a = 1$m
Sea Depth H	50m
Calculation Region	R ∈ [0,30] $\theta \in [-90°, 90°]$ z ∈ [0,50m]
Sampling Distribution	Total Samples 1000 Volume Samples 500 Spherical Target Samples 100 Seabed Boundary Samples 400
Wavelet Expansion Orders	Radial Basis 10th-order Axial Basis 10th-orde Angular Basis 10th-orde
Computation Time	45.89s

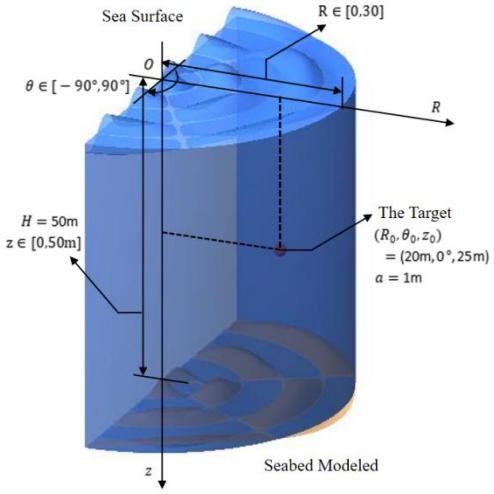


Figure 2 – Diagram of calculation area

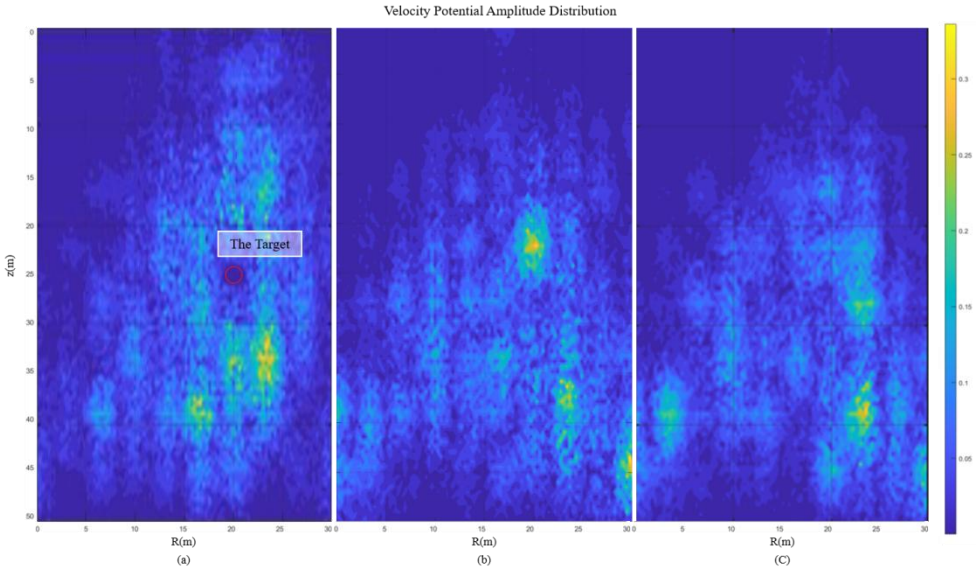


Figure 3 – Vertical Cross-section of Scattered Field: (a) $\theta = 0°$ (b) $\theta = 45°$ (c) $\theta = 90°$

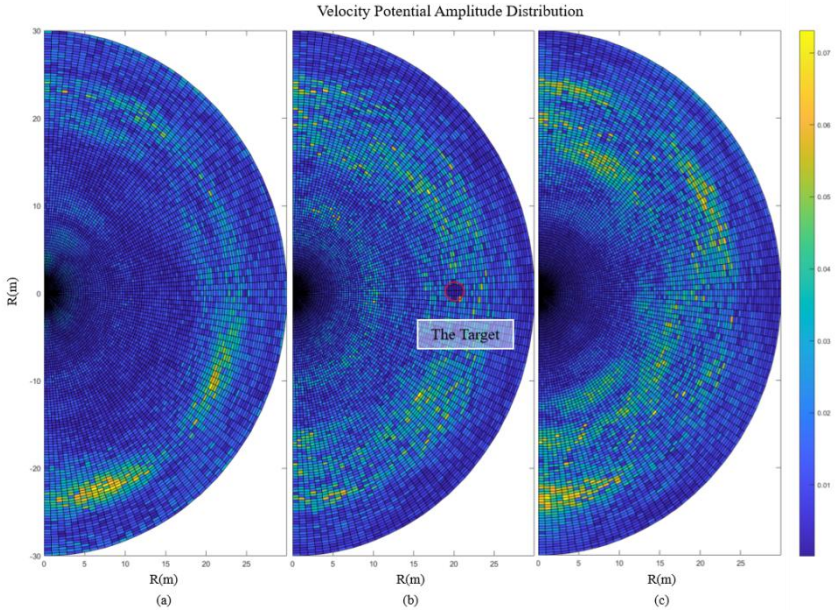


Figure 4 – Horizontal Cross-section of Scattered Field: (a) $z = 10$m (b) $z = 25$m (c) $z = 40$m

5.2 Simulation Experiment of Horizontal Range Correction Method for Single-Hydrophone Systems

Within the computational domain, we discretize the positive radial axis (R-axis) at 0.1 m intervals. To mitigate stochastic boundary effects and align with practical hydrophone deployment constraints, values at depths shallower than 5 m or deeper than 45 m are excluded. This generates 300 × 400 discrete points as single-hydrophone deployment locations. The Horizontal Range Correction Method for Single-Hydrophone Systems is applied iteratively to solve the field, with the resulting error distribution shown in Figure 6.

We observe that errors at most points remain below 10%. A distinct correlation exists between error distribution and hydrophone position: errors increase significantly near the sea surface, while estimates remain robust near the seabed. We conjecture this relates directly to stochastic scattering interference at the sea surface.

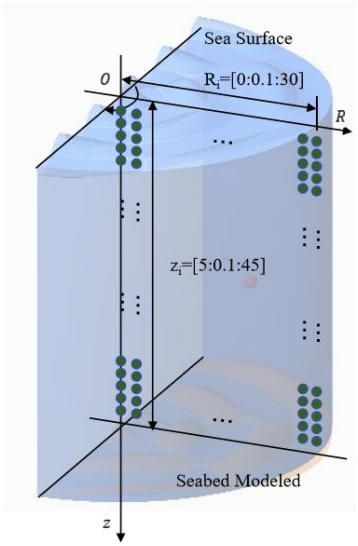


Figure 5 – Schematic diagram of sampling location

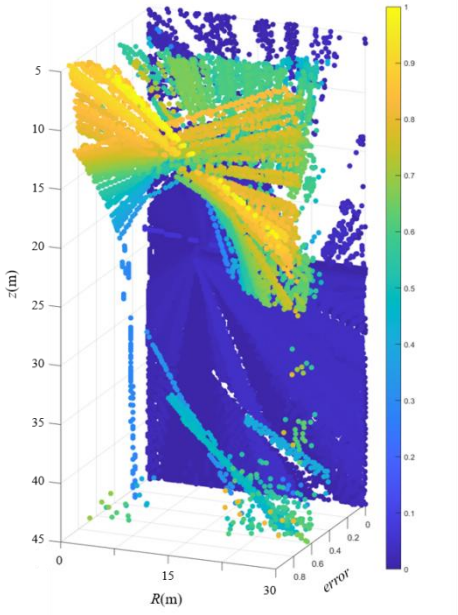


Figure 6 – Error distribution plots

6. CONCLUSIONS

This paper introduces a Wavelet Jointless-Weighted Nodal Method for modeling acoustic scattering from spherical targets in shallow-water waveguides with stochastic upper and lower interfaces. The algorithm effectively resolves complex boundary problems while achieving significant computational

efficiency improvements. Building on this method, we designed a Horizontal Range Correction Method for Single-Hydrophone Systems that theoretically enables precise correction of horizontal ranging measurements.

Owing to its enhanced flexibility and computational efficiency, this approach holds substantial potential for addressing more complex acoustic problems and nonlinear sound field simulations. Furthermore, the wavelet decomposition framework offers high applicability in engineering signal analysis. The wavelet bases selected in this study serve as a representative example; future optimization should prioritize investigating the mathematical properties of wavelet bases and their physical parameterization in acoustic field characterization.

ACKNOWLEDGEMENTS

This research was funded by National Key Research and Development Program of China, under grant no. 2021YFC3101803; in part by Marine Economy Development Special Project of 2024 Guangdong Province, under grant no. GDNRC[2024]43 .

REFERENCES

1. Xuefeng L, Qi L, Rui T, Yilin W, Dajing S, Zhenxing Z. A combined noise source model based on vertical coherence to quantify the proportions of two types of noise power. J Acoust Soc Am. 2025;157(1):84–100.
2. Xu L, Zhang HG. Training a deep operator network as a surrogate solver for two-dimensional parabolic-equation models. J Acoust Soc Am. 2023;154(5):3276-3284.
3. Kirkup S. The boundary element method in acoustics: A survey. Appl Sci. 2019;9(8):1642.
4. Long G, Yang H, Liu LB. A fast wavelet collocation method with compression techniques for Steklov eigenvalue problems of Helmholtz equations. Appl Math Lett. 2025;166:109532.
5. Fu ZX, Lin Q, Deng CX. Wavelet Function Properties Based on Shannon Function. In: Proc 2024 Int Conf Wavelet Anal Pattern Recognit (ICWAPR); September 2024; IEEE. p. 1-6.
6. De Marchi S, Wendland H. On the convergence of the rescaled localized radial basis function method. Appl Math Lett. 2020;99:105996.

PROCEEDINGS of the
16th International Conference on Theoretical and Computational Acoustics

August 18 to 22, 2025 BEXCO, Busan Korea

The inference of Deep-Sea Sedimentary Layer Properties in the Northwest Subbasin of the South China Sea

Yongchao GUO[1]; Haigang ZHANG[2]; Minghui ZHANG[3]
[1] National Key Laboratory of Underwater Acoustic Technology, Harbin Engineering University, Harbin 150001,China;
[2] Key Laboratory of Marine Information Acquisition and Security(Harbin Engineering University),Ministry of Industry and Information Technology;Harbin 150001,China
[3] College of Underwater Acoustic Engineering, Harbin Engineering University, Harbin 150001,China

ABSTRACT
The seabed is an important boundary of underwater acoustic waveguide, but the influence of layered seabed structures is often ignored in the study of deep-sea acoustic field characteristics. During the deep-sea acoustic propagation experiment in the northwest subbasin of the South China Sea, the acoustic intensity striations were observed on the spectrogram of a merchant ship recorded by an Ocean Bottom Seismometer, and the waveguide invariant β shows a trend of initially increasing and then decreasing across the range of 4 to 10 km. This phenomenon was not observed in previous experiments; we have established the relationship between the interference characteristics of deep-sea sound fields and multi-layer seabed parameter models and explained the reasons for the generation of acoustic intensity striations using sound ray tracing codes for multi-layer media. Through analysis and inference, the seabed in the experimental area is found to have a three-layer structure comprising a low-acoustic-velocity sediment layer, a layer with a positive acoustic velocity gradient, and a substrate. The positive velocity gradient layer leads to the generation of refracted waves from the seabed, causing the waveguide invariant β to display a trend of initially increasing and then decreasing within the near range.

Keywords: Multi-layered seabed, Deep sea, Acoustic intensity striations

1. INTRODUCTION
The seabed is an important boundary of underwater acoustic waveguide, but the influence of layered seabed structures is often ignored in the study of deep-sea acoustic field characteristics. This multi-layered sediment architecture generates multiple acoustic paths to receivers, including reflections at sediment interfaces (1), refractions through sediment layers (2), and substrate reflections that occur after full penetration of the sediment column (3). Due to these reflections and refractions from multi-layered seabed sediment, significant interference patterns emerge in acoustic propagation loss (4,5,6,7), pulse travel times exhibit complexity (8) , and anomalous dispersion structures may occur (9).

The present work identifies the feature of striation patterns resulting from interactions with a multi-layer seabed in deep water, which may not have been noticed in previous studies. Specifically, bottom-interacting acoustic intensity striations were observed on the spectrogram of a merchant ship recorded during the Deep-Water Acoustic Propagation Experiment in the Northwest Sub-Basin of the South China Sea. The feature of acoustic intensity striations resulting from the multi-layered seabed was discussed using ray tracing methodology, and this feature was used to infer the geoacoustic structure of the multi-layered seabed.

The remainder of this paper is organized as follows. The experimental configuration in the Northwest Sub-Basin of the South China Sea, as well as the acoustic intensity striations observed on the spectrogram of a merchant ship recorded during the experiment, are described in Sec. 2. The feature of acoustic intensity striations resulting from the multi-layer seabed is discussed and analyzed in Sec. 3.

[1] guoyongchao@hrbeu.edu.cn
[2] zhanghaigang@hrbeu.edu.cn
[3] zhangminghui@hrbeu.edu.cn

2. EXPERIMENT DATA DESCRIPTION

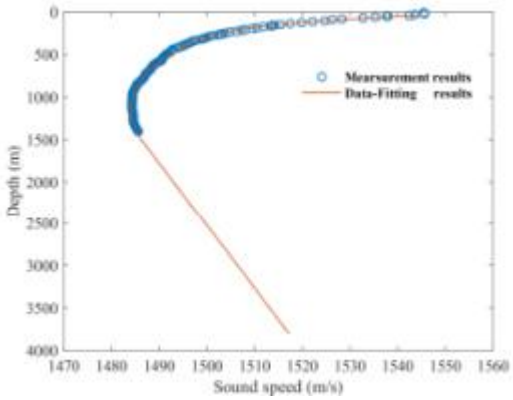


Figure 1 - Water-column sound-speed profile near the receiver, measured during the experiment

In August 2022, a deep-water acoustic experiment was conducted in the northwestern sub-basin of the South China Sea. The primary purpose of the experiment was to investigate the influence of seabed interactions on deep-sea sound propagation characteristics. During the trial, ocean-bottom seismometers (OBS) were deployed on the seafloor to record broadband noise data generated by large cargo ships. The particle velocity components at the sea bottom and the acoustic pressure in the water column are measured by the three-component geophone and hydrophone. All channels were simultaneously sampled at 250 Hz, and the receiver sensitivity of the hydrophone was -190 dB re 1V/μPa. During the experiment, the water-column sound-speed profile near the receiver was measured by CTD, as shown in Fig.1. This is a typical incomplete deep sound channel with the depth of the sound-channel axis (minimum sound speed) at approximately 1140 m, and the maximum sound speed(1550m/s) at the sea surface.

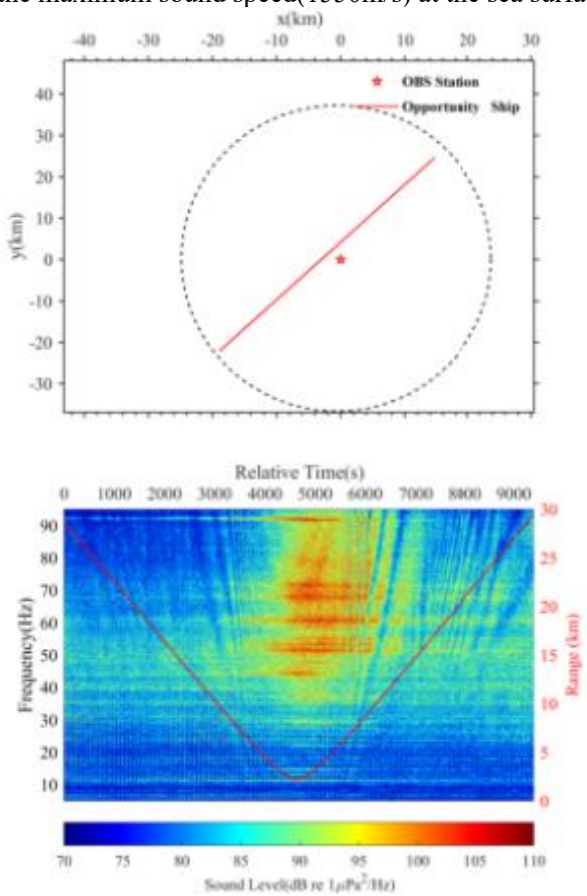


Figure 2 - (a) The trajectory of the opportunity ship and the location of the OBS. (b) Spectrogram from time period of passage of container ship *M/V Mq Delphi*, acoustic data recorded on hydrophone of the OBS on the seabed.

The OBS was located near the shipping lane, and the data sets were collected, including broadband noise

from large commercial ships. The ship, *M/V Mq Delphi*(IMO 9491238), travels in a straight-line course along the northeast-southwest direction (red line in Fig. 2(a)). The AIS message data, including the geographic positions of latitude and longitude, for *M/V Mq Delphi* was derived from the Marine-Traffic website and used to calculate the distance between the ship and the receiver. Fig. 2(b) shows the spectrogram of the significant bulk carrier ship passage and the distance between the ship and the receiver; the closest point of approach (CPA) is approximately 2.25 km. The striations are approximately symmetric about the CPA time over a wide frequency band. The symmetry is not exact due to a stern/bow aspect dependence on the broadband source level and the variability of the characteristics of the seabed.

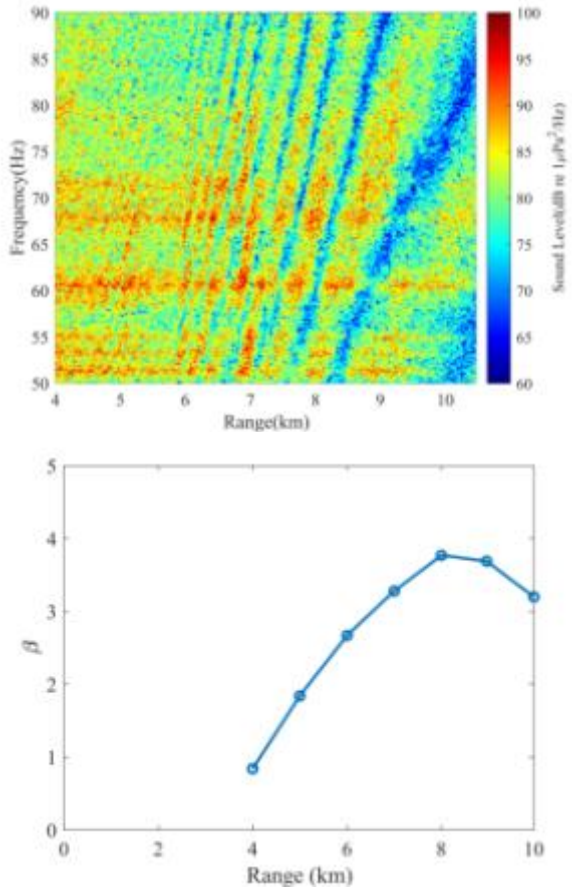


Figure 3 - (a) Range-frequency spectrogram $I(r, f)$ of 4-10km. (b) The estimated values of β for various center ranges.

The time-frequency spectrogram $I(t, f)$ in Fig. 3 can be converted to range-frequency spectrogram $I(r, f)$ and the data segment with sufficient signal-to-noise ratio is used to estimate the waveguide invariant β, as shown in Fig. 3(a). The two-dimensional Fast Fourier Transform (2D-FFT) method, as described in references (10), is utilized to estimate the waveguide invariant β, which corresponds to the slope of the acoustic intensity striations in the range-frequency spectrogram $I(r, f)$. The configuration settings of the 2D-FFT are as follows: central frequency f_0 = 70 Hz, bandwidth B = 40 Hz, and range aperture L = 1 km. Fig. 3(b) displays the estimated values of β for various center ranges. It was observed that the waveguide invariant β shows a trend of initially increasing and then decreasing across the range of 4 to 10 km. This phenomenon has not been observed in previous experiments, and it is inferred that this is related to the properties of the multi-layer seabed. To verify this speculation, the feature of acoustic intensity striations resulting from several typical multi-layered seabed is discussed and analyzed in Sec. 3.

3. THEORY AND NUMERICAL SIMULATION

This section investigates the characteristics of acoustic intensity striations under the four typical multi-layered seabed types listed in Table 1. Furthermore, the relationship between the waveguide invariant and multipath arrivals is established using ray theory. The acoustic intensity striations are calculated using the numerical KRAKEN code (11), and the Eigenray Tracing code, developed in reference to the work of Porter (12) and Westwood (13), is employed to analyze the dependence of β on range.

Table 1 – Geoacoustic parameters of four typical multi-layered seabed

		Type I	Type II	Type III	Type IV
Layer 1	H (m)	-	400	400	50
	c (m/s)	-	1600	1600-2400	1450
	ρ (g/cm^3)	-	1.6	2.0	1.5
	α (dB/λ)	-	0.01	0.01	0.01
Layer 2	H (m)	-	-	-	350
	c (m/s)	-	-	-	1600-2400
	ρ (g/cm^3)	-	-	-	2.0
	α (dB/λ)	-	-	-	0.01
Basement	c (m/s)	2500			
	ρ (g/cm^3)	2.5			
	α (dB/λ)	0.25			

By the ray theory, the total acoustic pressure is dominated by various types of multipath arrivals, and the interference between any pair of these arrivals gives rise to distinct acoustic intensity striations, for a point source at depth z_s, a receiver at depth z_r and range r may be written as follows:

$$p(r,z;f) \propto a_1 e^{i2\pi f t_1} - a_2 e^{i2\pi f t_2} \tag{1}$$

where t_1 and t_2 are travel time along two types of paths, a_1 and a_2 are the amplitude of the corresponding ray.

The acoustic intensity at the receiver can be written in the form

$$I(r,z;f) \propto a_1^2 + a_2^2 - 2a_1 a_2 \cos\left(2\pi f\left(t_2 - t_1\right)\right) \tag{2}$$

Thus, the coherent term of Eq. (2) causes the acoustic intensity to exhibit structured cosine striations in the *r-f* plane. Let arrival time difference $\Delta t = t_2 - t_1$,when f satisfies the condition that

$$f = \left(n - \frac{1}{2}\right)\frac{1}{\Delta t}, n = 1,2,3,\mathrm{L} \tag{3}$$

the acoustic intensity is minimum, which corresponds to destructive interference striations in the *r-f* interference pattern. According to the definition of waveguide invariants

$$\beta = \frac{r}{\omega}\frac{\delta\omega}{\delta r} \tag{4}$$

and Eq. (3), the waveguide invariant can be defined as a range dependent expression

$$\beta = \frac{r}{f}\frac{\delta f}{\delta r} = \frac{r}{\left(n-\frac{1}{2}\right)/\Delta t}\frac{\delta\left(\left(n-\frac{1}{2}\right)/\Delta t\right)}{\delta r} = -\frac{r}{\Delta t}\frac{\delta(\Delta t)}{\delta r} \tag{5}$$

For long-range acoustic propagation in deep water, it is generally appropriate to consider the seabed as a fluid half-space. As shown in Fig. 4(a), the total acoustic pressure at the receiver under the fluid half-space seabed is dominated by the direct and surface-reflected path arrivals. Fig.4(b) shows the arrival time difference between the direct path and the surface-reflected path, and the arrival time difference decreases with the increase in the propagation range. Fig. 4(c) shows the destructive interference striations calculated using Eq. (3), which are in good agreement with the *r-f* interference pattern simulated by KRAKEN. Within the same frequency bandwidth, the simulated *r-f* interference pattern did not display the acoustic intensity striations observed in the experiment. In addition, the waveguide invariant β increases with range (as shown in Fig. 4(d)), which is inconsistent with the experimental results. It can be inferred that the acoustic intensity striations observed in the experiment are not formed by interference between the direct path and the surface-reflected path. Thus, it is unreasonable to assume the seabed is a fluid half-space.

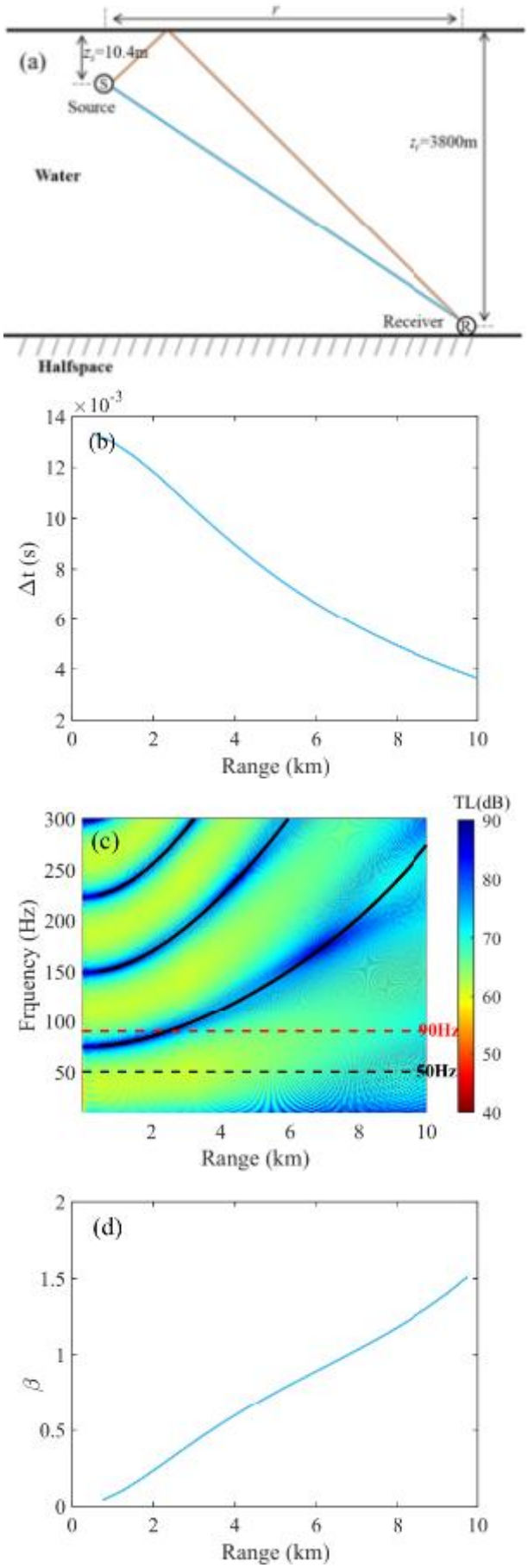
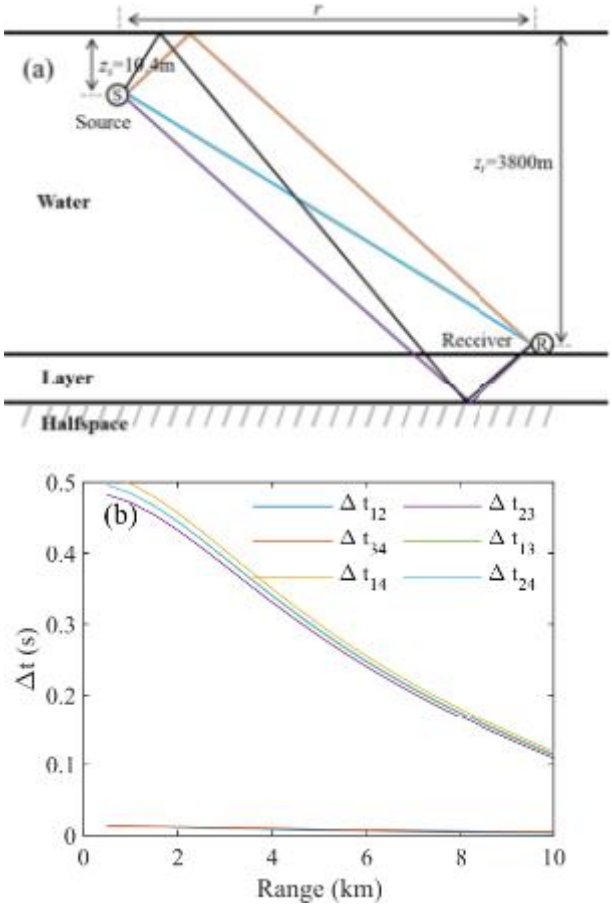


Figure 4 - (a) The main sound path of fluid half-space. (b) The arrival time difference between the direct

path and the surface-reflected path. (c) The *r-f* interference pattern simulated by KRAKEN. (d) The relationship between the waveguide invariant and range variation.

But extensive geological surveys and geoacoustic parameter inversion studies confirm the ubiquitous presence of stratified seabed structures. Consequently, experimentally observed acoustic interference patterns may be perturbed by reflected or refracted waves propagating through seabed sediment layers. Under the assumption of a homogeneous sediment layer configuration, the acoustic pressure at the receiver position is predominantly governed by four distinct propagation paths, as shown in Fig. 5(a). Fig. 5(b) shows the arrival time difference between any pair of arrival paths, resulting in two types of acoustic intensity fringes, as shown in Fig. 5(c). Therefore, the arrival path can be categorized into two groups: (i) the direct and surface-reflected path arrivals, and (ii) the paths (reflections through seabed sediment layers) of interaction with submarine sedimentary layers. The observed acoustic intensity striations arise from both intra-group ray interference and inter-group ray interference between the two path categories. Although the simulated range-frequency interference pattern replicates the acoustic intensity striations observed experimentally within the identical frequency bandwidth, the waveguide invariant β exhibits an increasing trend with range (as shown in Fig. 5(d)), which is lower than the experimental result. Consequently, numerical simulations of homogeneous sediment layer confirm that seabed stratification constitutes the primary mechanism generating acoustic interference patterns. However, actual seabed conditions deviate significantly from this homogeneous model, likely exhibiting far more complex geoacoustic structures.



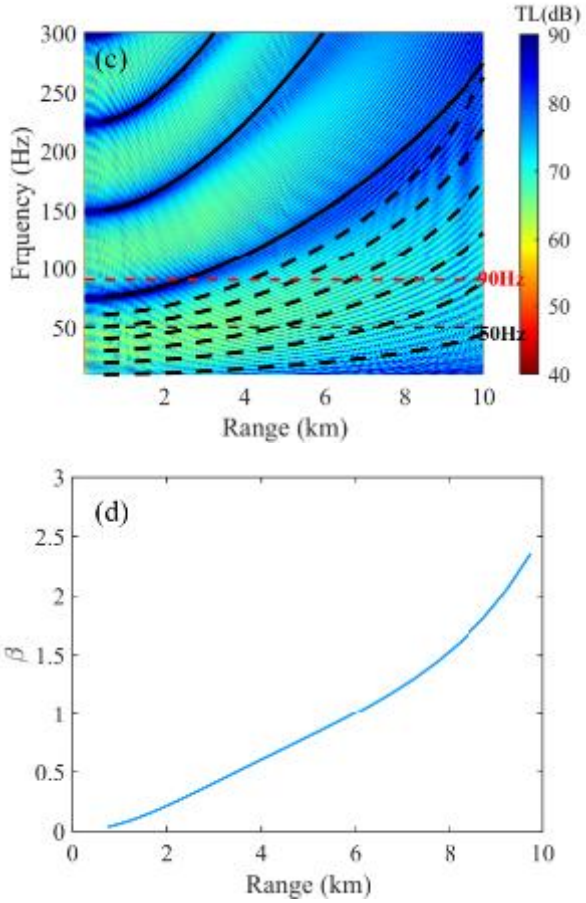
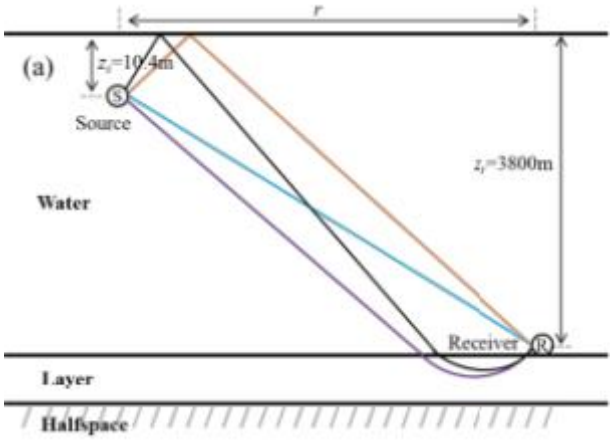


Figure 5 - (a) The main sound path of a homogeneous sediment layer. (b) The arrival time difference between t any pair of arrival paths (c) The r-f interference pattern simulated by KRAKEN. (d) The relationship between the waveguide invariant and range variation.

Due to compaction effects within seabed sediments, sound velocity in sedimentary layers typically increases with depth. Unlike homogeneous sediments, sound interacting with these layered seabed sediments primarily follows refracted paths through the layers, as illustrated in Fig. 6(a). Similar to the homogeneous case, acoustic intensity striations appear in the simulated range-frequency interference pattern in Fig. 6(c). However, the waveguide invariant (as shown in Fig. 6(d)) calculated increases rapidly with range, exceeding the experimentally observed values. Thus, it can be inferred that a multi-layered structure likely exists beneath the seafloor, including sediment layers exhibiting a positive sound speed gradient.



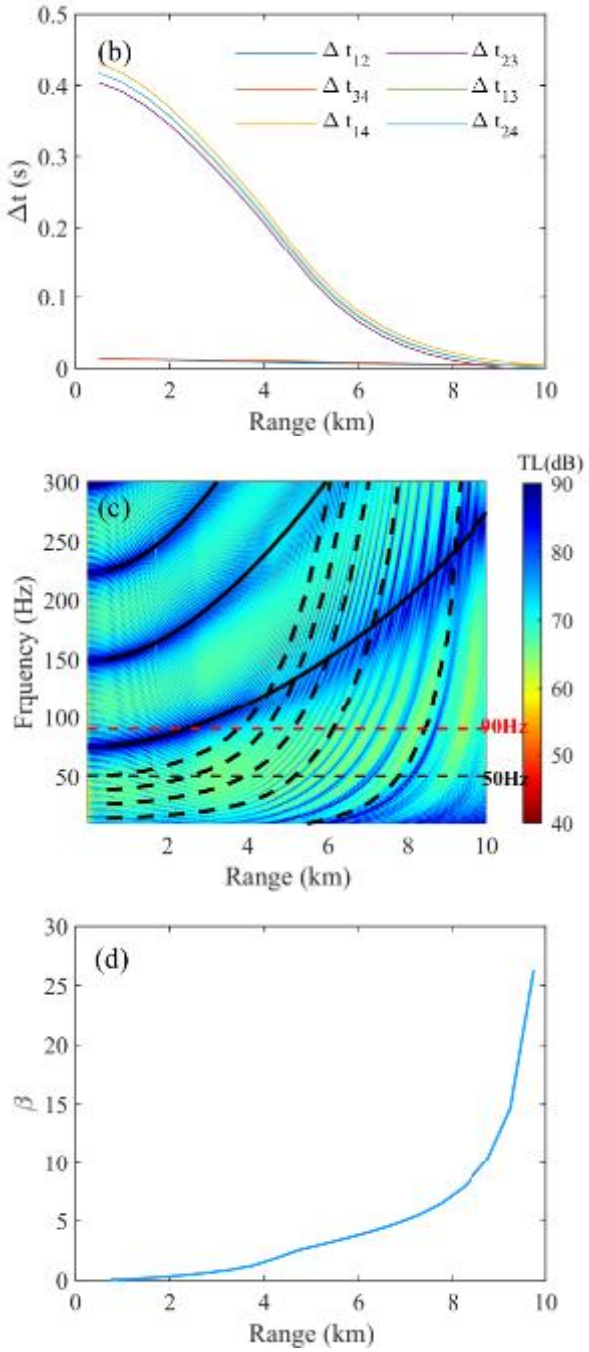


Figure 6 - (a) The main sound path of a positive sound speed gradient sediment layer. (b) The arrival time difference between any pair of arrival paths. (c) The *r-f* interference pattern simulated by KRAKEN. (d) The relationship between the waveguide invariant and range variation.

In the study area, anomalous dispersion effects caused by a low-sound-speed layer in the seabed were observed in references (9). The seabed sediment layers are considered as a low-sound-speed layer overlying a sediment layer with a positive sound speed gradient, as depicted in Fig. 7(a). Fig. 7(b) shows the arrival time difference between any pair of arrival paths, resulting in two types of acoustic intensity fringes, as shown in Fig. 7(c). But sound interacting with these layered seabed sediments primarily follows refracted paths through the low-sound-speed layer and positive sound speed gradient layer. Nevertheless, the waveguide invariant β values in Fig. 7(d) exhibit an initial increase followed by a decrease, matching the variation pattern observed experimentally. Therefore, it can be inferred that the seabed in the experimental area exhibits a three-layer structure comprising a low-sound-speed sediment layer, a layer with a positive sound speed gradient, and a substrate.

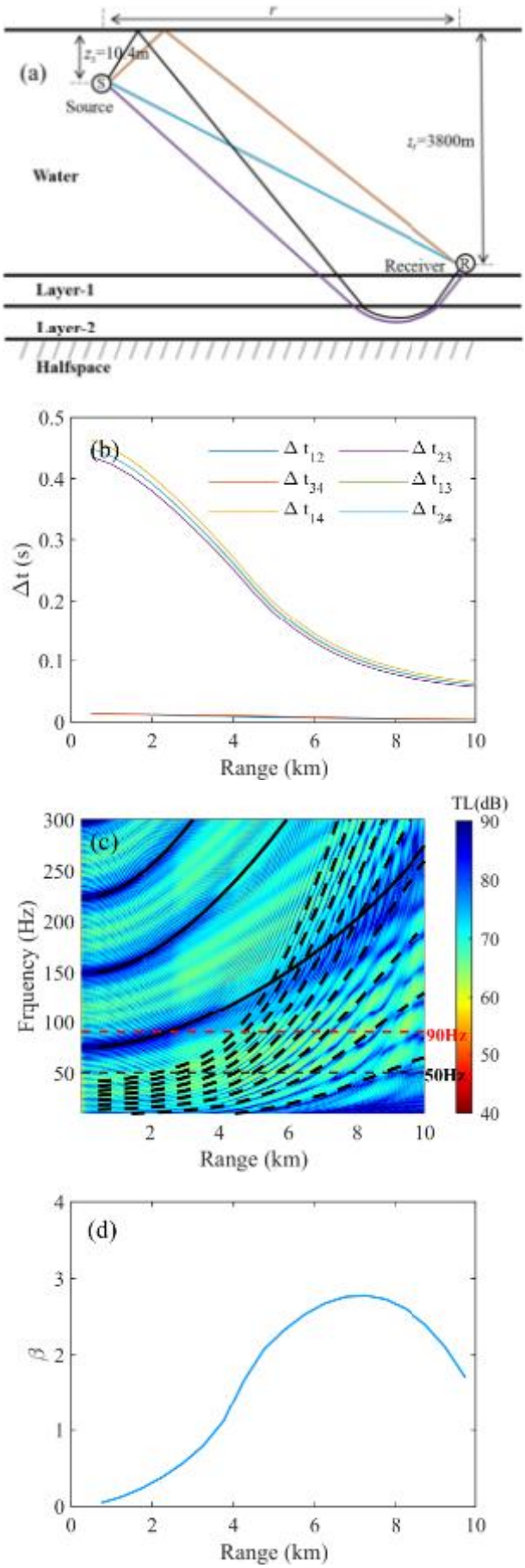


Figure 7 - (a) The main sound path in a low-sound-speed layer over a sediment layer with a positive sound

speed gradient. (b) The arrival time difference between any pair of arrival paths. (c) The *r-f* interference pattern simulated by KRAKEN. (d) The relationship between the waveguide invariant and range variation.

4. CONCLUSIONS

The merchant ship spectrogram measured in the Northwest Sub-Basin of the South China Sea revealed acoustic interference striations resulting from interactions with deep-sea seabed sediment layers. The waveguide invariant β shows a trend of initially increasing and then decreasing across the range of 4 to 10 km. The relationship between the interference characteristics of deep-sea sound fields and four typical multi-layer seabed parameter models was established, and the reasons for the generation of acoustic intensity striations were explained using sound ray tracing codes for multi-layer media. Through analysis and inference, the seabed in the experimental area is found to have a three-layer structure comprising a low-acoustic-velocity sediment layer, a layer with a positive acoustic velocity gradient, and a substrate. The primary cause of the experimentally observed acoustic intensity striations is inter-group interference between (i) arrivals via the direct and surface-reflected paths and (ii) refracted paths traversing the low-sound-speed layer and the layer with a positive sound speed gradient.

REFERENCES

1. Brown M V, Ricard J H. Interference Pattern Observed in Reflections from the Ocean Bottom[J]. The Journal of the Acoustical Society of America, 1965, 37(6): 1033-1036.
2. Hanna J S. Short-range transmission loss and the evidence for bottom-refracted energy[J]. The Journal of the Acoustical Society of America, 1973, 53(6): 1686-1690.
3. Udovydchenkov I A, Stephen R A, Duda T F, et al. Bottom interacting sound at 50 km range in a deep ocean environment[J]. The journal of the acoustical society of America, 2012, 132(4): 2224-2231.
4. Christensen R E, Frank J A, Geddes W H. Low-frequency propagation via shallow refracted paths through deep ocean unconsolidated sediments[J]. The Journal of the Acoustical Society of America, 1975, 57(6): 1421-1426.
5. Frisk G V, Doutt J A, Hays E E. Bottom interaction of low-frequency acoustic signals at small grazing angles in the deep ocean[J]. The Journal of the Acoustical Society of America, 1981, 69(1): 84-94.
6. Miller J F, Nagl A, UBERALL H. Influence of bottom refraction on the propagation of underwater sound[J]. Acta physica Austriaca, 1985, 57(3-4): 165-173.
7. Knobles D P, Wilson P S, Neilsen T B, et al. Influence of seabed on very low frequency sound recorded during passage of merchant ships on the New England shelf[J]. The Journal of the Acoustical Society of America, 2021, 149(5): 3294-3300.
8. Stephen R A, Bolmer S T, Dzieciuch M A, et al. Deep seafloor arrivals: An unexplained set of arrivals in long-range ocean acoustic propagation[J]. The Journal of the Acoustical Society of America, 2009, 126(2): 599-606.
9. Cao J, Qi Y, Zhou S, et al. Anomalous dispersion observed in signal arrivals at a deep-sea floor receiver[J]. JASA Express Letters, 2021, 1(7).
10. Emmetière R, Bonnel J, Géhant M, et al. Understanding deep-water striation patterns and predicting the waveguide invariant as a distribution depending on range and depth[J]. The Journal of the Acoustical Society of America, 2018, 143(6): 3444-3454.
11. Porter M B. The KRAKEN normal mode program[J]. Unknown, 1992.
12. Porter M B, Bucker H P. Gaussian beam tracing for computing ocean acoustic fields[J]. The Journal of the Acoustical Society of America, 1987, 82(4): 1349-1359.
13. Westwood E K, Vidmar P J. Eigenray finding and time series simulation in a layered-bottom ocean[J]. The Journal of the Acoustical Society of America, 1987, 81(4): 912-924.

PROCEEDINGS of the
16th International Conference on Theoretical and Computational Acoustics

August 18 to 22, 2025 BEXCO, Busan Korea

Numerical comparison of the frequency response function of different poroelastic models in stratified sediments using pointwise sources

Andrés Prieto[1]

[1] CITMAGA, University of A Coruña

ABSTRACT

The accurate prediction of wave propagation phenomena in the digital geophysics model is crucial not only for understanding earthquake dynamics but also for detecting and monitoring oil reservoirs and aquifers. A common geological feature of the underground framework consists of the stratified character of the soil. Each layer can be modelled with a variety of viscoelastic and poroelastic models, such as the classical Biot model or, more recently, the Biot-Stoll and Biot-Allard models. This work presents a numerical comparison of the frequency response function of those models in a complex real-world geological stratified scenario, where the numerical response of different poroelastic models is excited from a short pulse generated at a specific point source located on the top surface. The numerical comparison is performed in terms of the post-processing of different frequency response functions, which enable the evaluation of kinematic and energy-based physical quantities of interest. In all cases, a mixed pressure-velocity formulation with a finite element discretization, combined with implicit time-marching schemes, is used to solve all the poroelastic wave propagation problems.

[1] andres.prieto@udc.es

PROCEEDINGS of the
16th International Conference on Theoretical and Computational Acoustics

August 18 to 22, 2025 BEXCO, Busan Korea

Numerical Analysis of Acoustic Scattering from Buried Targets under Varying Sediment and Target Conditions

Yeon-Seong Choo[1], Sung-Hoon Byun[2], Sea-Moon Kim[3]

[1] KRISO(Korea Research Institute of Ships & Ocean engineering)

[2] KRISO(Korea Research Institute of Ships & Ocean engineering)

[3] KRISO(Korea Research Institute of Ships & Ocean engineering)

ABSTRACT

Experimental studies on acoustic scattering from buried seafloor targets—considering burial depth, target size, and shell thickness—are often limited by time and cost. Numerical simulation provides an efficient way to predict and analyze scattering under diverse conditions. When acoustic waves encounter the water–sediment interface, they typically refract into the sediment. At angles below the critical angle, interface-guided waves are generated. This angle depends on the density and sound speed of both media, so sediment properties strongly affect energy transmission and scattering. This study uses numerical simulations to examine how burial depth, target size, and shell thickness influence acoustic scattering under subcritical sonar conditions. Simulations were conducted using COMSOL, a commercial FEM software. The scenario assumes a fixed acoustic source and a moving receiver measuring scattered signals across the target region. These signals are analyzed to assess how their characteristics vary with the selected parameters. The goal is to extract features from scattered responses that are sensitive to changes in burial conditions and target geometry, supporting effective detection of buried underwater objects. [This research was supported by KRIT grant funded by the Korea government(DAPA) (KRIT-CT-23-035, Multi AUV opertaion Technology for Mine Detection (‘23-’28)).]

[1] chooys@kriso.re.kr

[2] byunsh@kriso.re.kr

[3] smkim@kriso.re.kr

Geoacoustic Inversion of Continuous Sediment Layer in Deep Water Using Ship Noise

Houxuan JIANG [1,2,3]; Minshuai LIANG [1,2,3]; Junjie SHI [1,2,3]; Dajun SUN [1,2,3]
[1] 1. National Key Laboratory of underwater Acoustic Technology, Harbin Engineering University, Harbin 150001, China;
[2] Key Laboratory of Marine Information Acquisition and Security (Harbin Engineering University), Ministry of Industry and Information Technology; Harbin 150001, China;
[3] College of Underwater Acoustic Engineering, Harbin Engineering University, Harbin 150001, China

ABSTRACT
A stepwise inversion method is developed to estimate seabed geoacoustic parameters in deep-water environments, based on a continuous sediment structure modeled by the reflection properties of ship noise. The measurements received by vertical array from the surface ship are analyzed to extract multipath reflections on both the surface and the bottom of the sediment. The time delay differences and grazing angle of multipath reflections are employed to infer the sediment velocity profile, and the continuous sediment layer is further modeled using Hamilton's empirical formula. Simulation results of reflection properties based on the continuous sediment model show a good agreement with the data observed.

Keywords: Geoacoustic inversion, seabed model, ship-noise inversion

1. INTRODUCTION
In past decades, there has been growing interest in the acoustic application in deep water environments, with the conduct of long-range ocean acoustic experiments. However, due to the large depth of experiments in deep water, estimating seabed properties through geoacoustic inversion becomes an effective in situ alternative to direct measurements such as coring, and is now widely recognized as an effective approach for acoustic applications in deep water environments.

Most of the research on geoacoustic inversion of seabed models has focus on shallow water. In contrast to deep water, sound propagation in shallow water is significantly influenced by interactions with the seabed, thus existing geoacoustic inversion methods are developed for applications in shallow water. Conventional methods generally compare the measured acoustic data with numerical simulations to estimate model parameters, such as reflection coefficient (1), normal mode (2) and vertical phase gradient (3). However, in deep water environment, only the seabed reflection paths are influenced by the seabed, and the frequency of interactions is greatly reduced. As a result, it is important to consider the applicability of inversion methods in deep water, such as the normal mode method becomes failure due to the lack of clear modal dispersion curves in deep water. The knowledge of seabed could significantly affect the applications in underwater (4-8).

A number of research have utilized seabed bottom loss data for geoacoustic inversion in deep water, with seabed models generally parameterized as layered sediment overlying a half-space basement, considering the sound is reflected from seabed and subsequently reach the hydrophone. Based on this model, sound from seabed is generally assumed to follow a single energy path. However, during a vertical array observation experiment conducted in South China Sea in 2023, we discover both weak upper energy path and strong lower energy path from the seabed. Based on simulation results, a continuous seabed model was selected as the equivalent seabed model. Furthermore, we propose a stepwise inversion method that employs ship noise to estimate geoacoustic parameters in deep water.

The paper is structured as follows. Second B presents the experimental data and the methods used for data processing. Second C provides a simulation study for model selection. Second D presents geoacoustic

inversion methods and inversion results.

2. ACOUSTIC MEASUREMENTS AND DATA PROCESSING

2.1 Geometry of Acoustic Measurements

The deep-water vertical array observation experiment was conducted in April, 2023, in the South China Sea, using a bottom-moored receiver, as illustrated in Fig. 1. The vertical array has a total length of 480 m, and is consisting of 64 equally spaced elements. The water depth at the measurement site was approximately 4360 m. The sound-speed profile, shown in Fig. 1(a), was measured by a CTD profiler during the experiment. Based on data from the ship's Automatic Identification System (AIS), a large oil tanker (with a draft of 11 m, according to the AIS data) passed through the array during the experimental period between 11:12 and 12:20 (UTC+8), following a path with a closest point of approach (CPA) of 3.3 km. Fig. 1(b) shows the ship track.

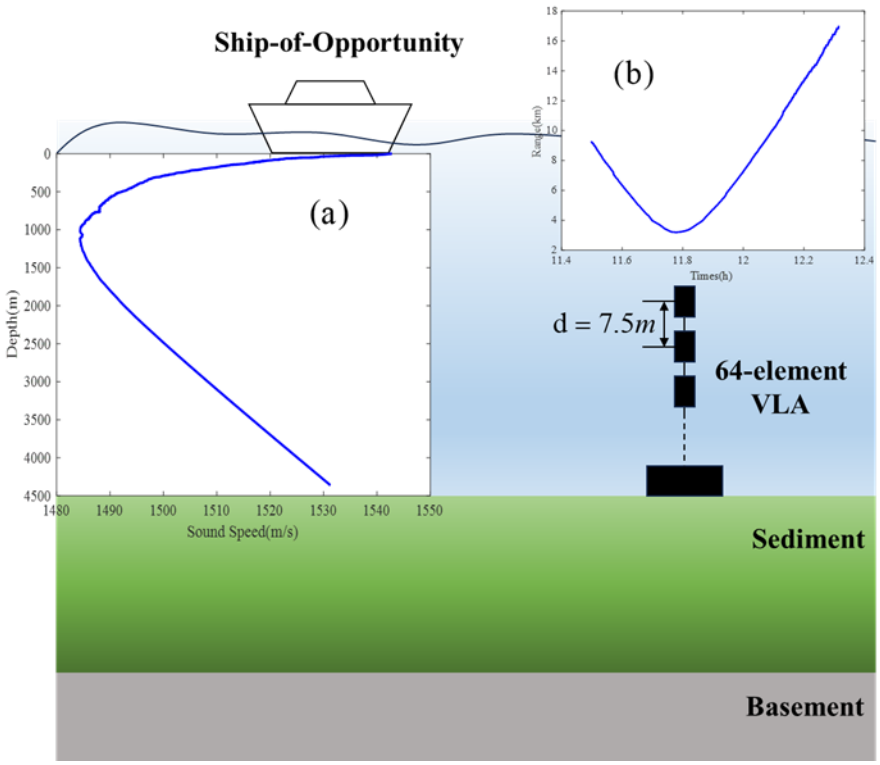


Fig. 1 Geometry of the acoustic measurements using a bottom-moored 64-element vertical array. Insert (a) shows the sound-speed profile in the water column. (b) shows the range of ship

2.2 Data Processing

The broadband low-frequency (20-100 Hz) noise data recorded by the vertical array, were processed by broadband beamforming with a time duration of 1 s. The output of the broadband beamforming is shown in Fig. 2. During the experimental period, the activity of the oil tanker was clearly observed, with high-energy track detected within the angular range of approximately 10° to 50°.

The sea-surface wave clusters were tracked over time on the beamforming output time series, and the corresponding elevation angle values of the direct arrival wave clusters were extracted, as indicated by the blue line. Using the measured sound speed profile of the experimental area and the real-time AIS distance of the oil tanker, the elevation angle corresponding to the upper-seabed (4360 m) path was simulated using the Bellhop based on ray theory, as shown by the pink dashed line in Fig. 2. Additionally, a high-energy lower-seabed path was observed, indicated by the black dashed line. The energy changes at approximately 46° on the path, and the energy is approximately 5-10 dB higher than the upper-seabed path for angles smaller than 46°. Considering the propagation characteristics of very low-frequency waves, this high-energy wave may be reflected from hard basement of the seabed.

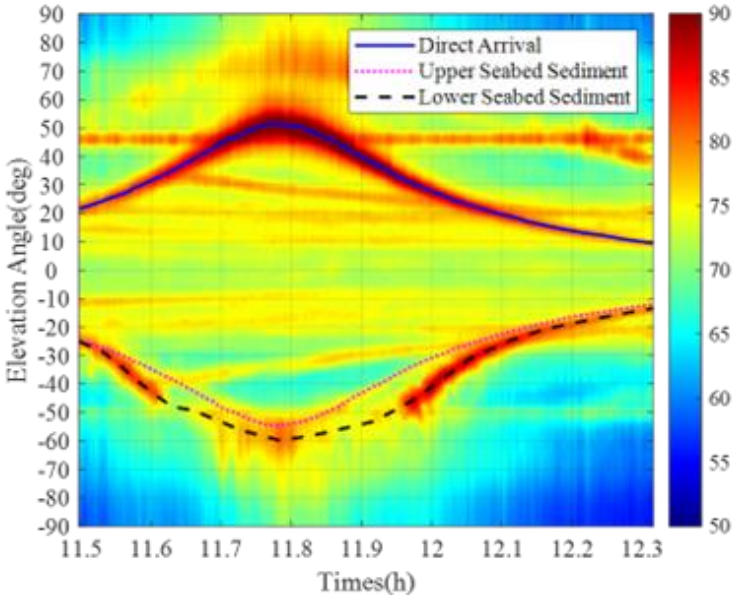


Fig. 2 Broadband beamforming output (20-100Hz)

Based on above inference, the signal's time delay information was further extracted to explain the occurrence of the high-energy lower-seabed path. Following broadband beamforming, the broadband Fourier synthesis method was used to obtain the time-domain waveform of the signal. The time-domain signal corresponding to the surface was cross-correlated with the signal from the seabed to obtain the time delay results. The pink line represents the time delay information for the upper-seabed path, determined through simulations with measured sound-speed profile and real-time AIS distance of the oil tanker. As shown in Fig. 3, a signal correlation peak appeared at the simulated time delay for the upper-seabed path. Additionally, a time delay correlation peak corresponding to the deep seabed signal was extracted, as indicated by the black line. The results suggest the presence of two reflection interfaces in the seabed of the experimental area. It can be inferred that the seabed consists of a hard basement covered with soft sediment.

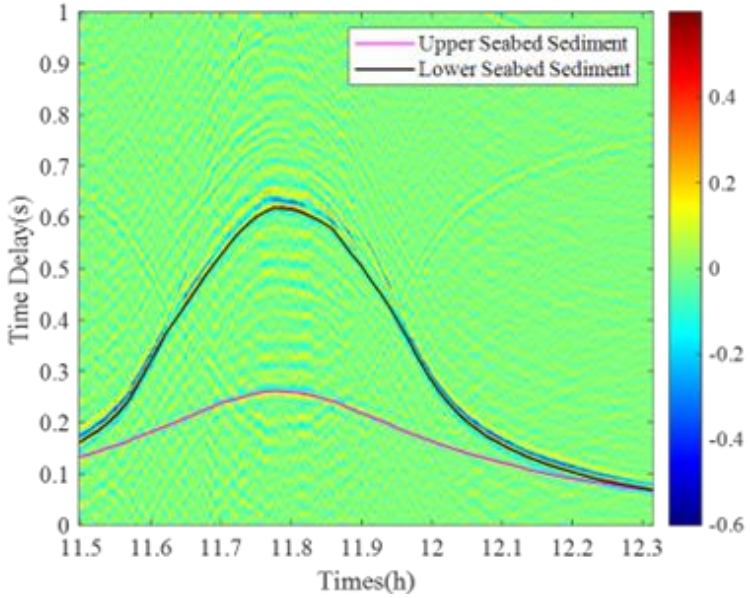


Fig. 3 Time delay estimation results (cross-correlation)

3. NUMERICAL SIMULATION AND MODEL SELECTION

3.1 Simulation Results

The section presents numerical simulations based on the observations of two distinct seabed energy paths, proposing two seabed models: the layered homogeneous sediment and the continuous sediment, and the

process for selecting the equivalent seabed model is also provided.

Based on the analysis results and typical seabed models, two underwater environmental models were developed, as illustrated in Fig. 4. The sound speed profile of water column was measured in the experimental area. Fig. 4(a) employs a conventional parameterized seabed model, in which the sediment is considered as a homogeneous medium, with the sediment divided into two layers for simplicity. However, Fig. 4(b) presents a more realistic model with a sediment that has a gradient in properties, where both sound speed and density increase with depth above the basement.

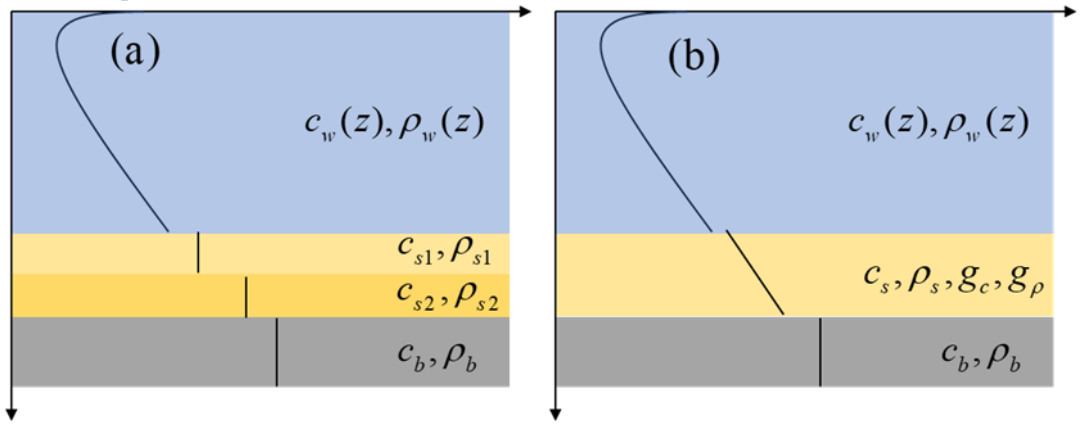


Fig. 4 Two assumed seabed model for (a) layered sediment and (b) continuous sediment

According to the deep-water seabed sampling data (9), the upper sediment is density contrast interface, with an acoustic impedance similar to that of seawater, leading to very low reflection efficiency of acoustic waves. By analyzing the characteristics of the energy path from seabed shown in Fig. 2, it is observed that the energy along the upper-seabed path is weak, indicating that most of the energy penetrates the surface of seabed. The sound speed in the sediment is almost identical to the sound speed at the water-sediment interface, leading to the assumption that the surface sediment sound speed is set at 1530 m/s. Furthermore, due to the broadband beamforming processing employed in this research, interference effects on the beam output caused by the sound field are excluded. The energy change observed near 46° in the lower path is likely due to the interaction of sound waves with a high-speed substrate interface, where total reflection occurs at the critical angle. Based on Snell's Law, the basement sound speed is estimated to be approximately 2190 m/s.

Numerical simulations were carried out for both equivalent seabed models, considering the draft of the oil tanker and the frequency bandwidth of noise data. In the simulation, the target depth was set to 11 m, the frequency bound was 20-100 Hz, with a frequency interval of 1 Hz. A vertical array with 64 elements was used for reception, were set the same depth as experiment. The seabed parameters used in both models are summarized in Table 1.

Table. 1 Seabed parameters for assumed seabed model

Model	Sediment			Basement	
	Depth, m	Sound speed, m/s	Density, g/cm^3	Sound speed, m/s	Density, g/cm^3
Layered Sediment	200 250	1600 1800	1.5 2.0	2190	2.4
Continuous Sediment	460	1530-2190	1.2-2.4	2190	2.4

Following the numerical computations, the received sound pressure from each element was extracted, and broadband beamforming was applied to obtain the elevation angle time series for different models, as shown in Fig. 5. In Fig. 5(a), the model with layered sediment only shows a single energy path from the seabed, with no significant variation in energy along the path. In contrast, Fig. 5(b) shows two distinct energy paths from the seabed: a weak upper path and a stronger lower path. The lower path exhibits significantly higher energy within the small angle (20°-46°). An energy change occurs near 46°, which is consist with the observation of experimental data. Therefore, the seabed is inferred as a continuous sediment underlying a basement.

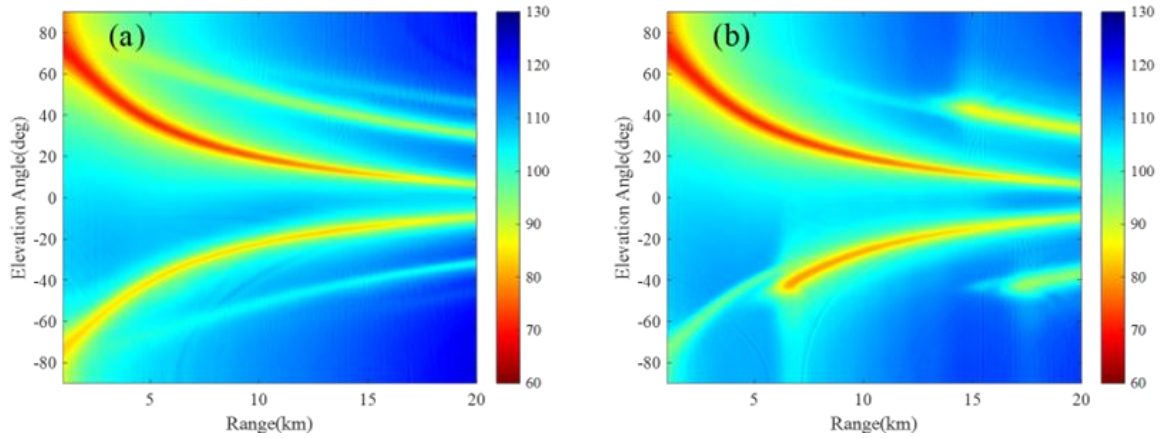


Fig. 5 Simulation results of broadband beamforming for (a) layered sediment and (b) continuous sediment

3.2 Seabed Model

The above analysis demonstrates that the model with continuous sediment effectively characterizes the properties of the seabed in the experimental area. The simulation results are in high agreement with the measured data, particularly in terms of the energy path characteristics and the observed energy distributions, thereby validating the model's applicability. Fig. 6 shows the seabed model based on the characteristic observed in the broadband beamforming results from the model with continuous sediment.

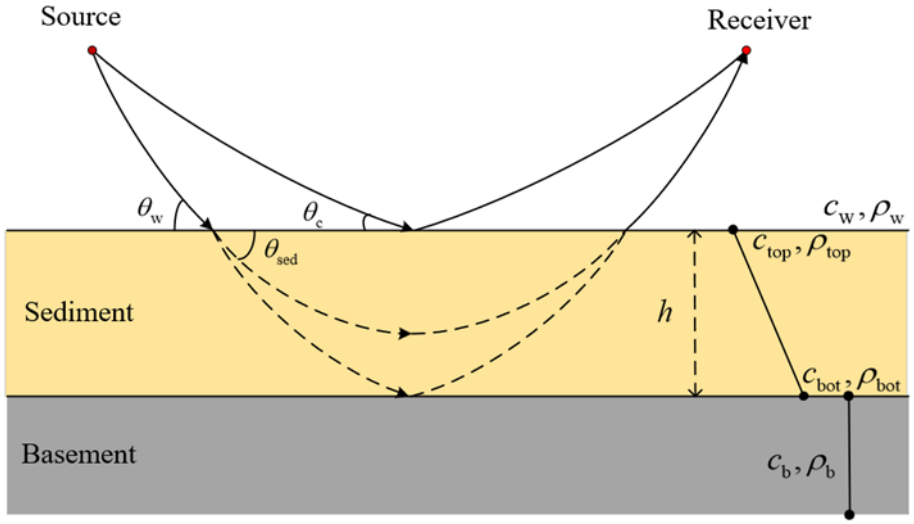


Fig. 6 Ray path for the seabed model with continuous sediment

The weak energy path is attributed to the reflection of sound rays from the upper sediment, when the depth of the seabed is set to 4360 m in the simulation, aligns with the upper seabed path observed in Fig. 2. However, the strong deep energy path results from the reversal and reflection of acoustic paths within the sediment. When the grazing angle is small, the thickness of sediment in deep water, is generally several hundred meter, causes sound rays to reverse before reaching the high-density basement. As the grazing angle increases, sound waves penetrate deeper into the sediment until they are reflected from the basement. Due to the high sound speed and density of the basement, the reflected energy is weaker than that of the upper path. The grazing angle at which reflection occurs corresponds to the angle where the lower energy path exhibits an abrupt change in energy.

4. INVERSION AND RESULTS

The section presents the method and results from inversions of ship-noise data. Fig. 7. shows the spectrogram from a single hydrophone recording of the oil tanker passage. Distinct frequency lines with the frequency band 20-100 Hz can be seen over the extent of the ship passage. Selected acoustic array data for inversion, sampled at 5 kHz, at four frequency bins centered on 30, 50, 70, and 90 Hz, representing broadband ship noise. Based on the assumption of acoustic wave propagation in the sediment, noise data corresponding to angles below the critical angle (46°) were selected. The ship-array distances ranged from 6.5 km to 16.2

km, corresponding to times between 11.97 h and 12.3 h, as indicated by the red boxes.

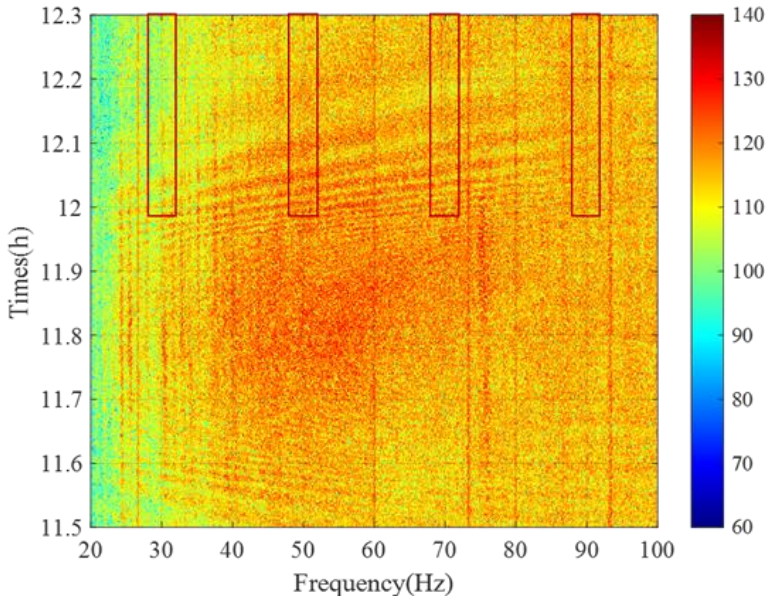


Fig. 7 Spectrogram from time period of passage of oil tanker, acoustic data recorded on hydrophone 1 of the vertical array. Red box denotes data selected for inversion

The data considered in this paper consist of complex (frequency-domain) acoustic fields measured at an N-sensor array and F frequencies corresponding to J distinct ship-array ranges of the recorded acoustic pressure time series (10). Since the data received by the array consist of ship-radiated noise with unstable source levels, the acoustic pressure received by the array was normalized. The replica field generated by inversion was computed only at discrete processing frequencies, and the normalized pressure of array is expressed as:

$$p = \frac{p_N(r,h,c_s)}{p_{\max}(r,h,c_s)} \tag{1}$$

where the denotes the ship-array ranges, represents the maximum depth of sound arrive in the sediment, indicates the corresponding sound speed of the depth of sediment and is the element index in the array. In represents the maximum penetration depth of acoustic waves in the sediment. In order to obtain the minimum misfit between prediction results and measured data, we utilize the normalized correlation coefficient to set the cost function, which is expressed as:

$$E = \frac{1}{F}\sum_{i=1}^{F}\frac{R_{\mathrm{cc}}(p_{replica}, p_{measure})}{\sqrt{R_{ac}(p_{replica}) * R_{ac}(p_{measure})}} \tag{2}$$

where represents the count of selected frequencies for replica field computation, and denote the cross-correlation and autocorrelation, respectively.

Fig. 8 presents estimated sediment depth from the range of ship-array. Specially, Fig. 8(a) shows the results when the range was 6.93 km, corresponding to the critical angle. And the Fig. 8 (b) - (d) denotes the results from 8.48 km, 10.1 km and 16.6 km respectively.

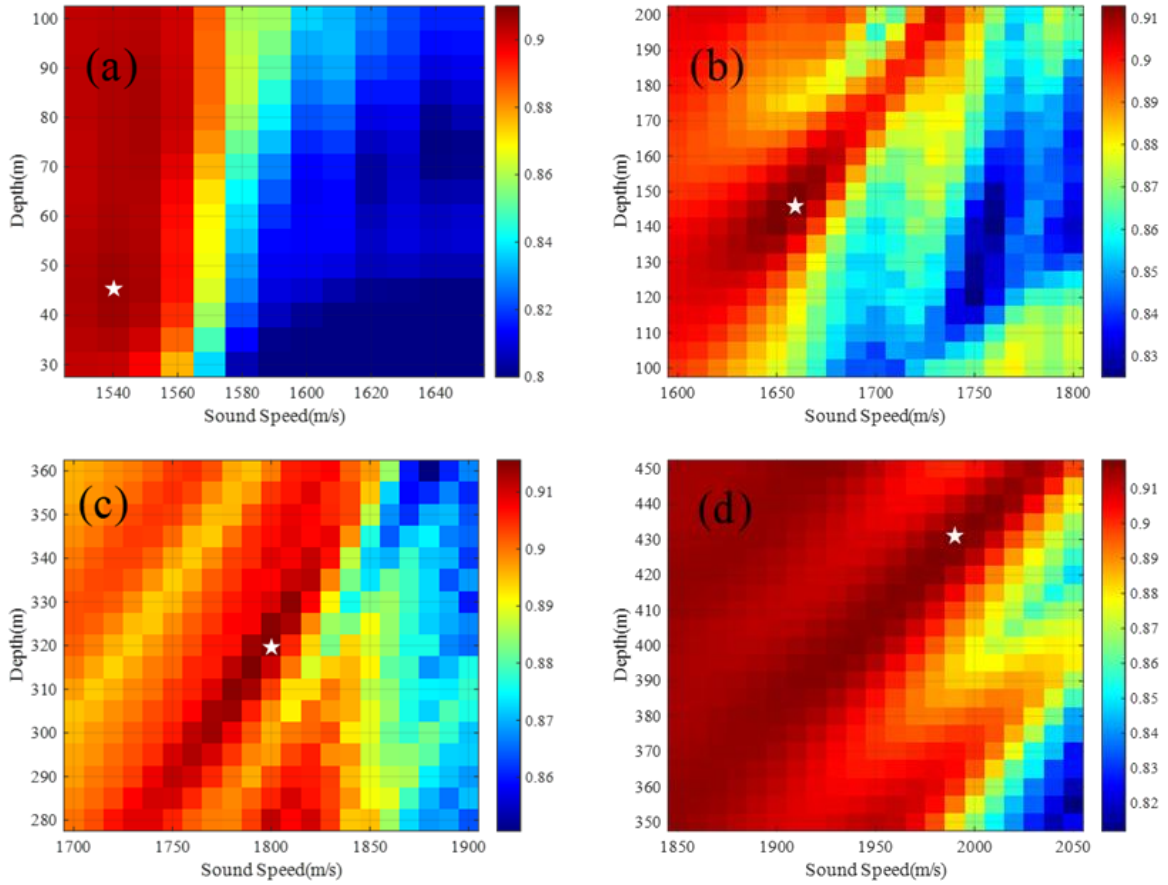


Fig. 8 Estimated sediment depth and sound speed in various ship-array range. The ranges of (a)-(d) are 16.6 km, 10.1km, 8.48km, 6.93km. White stars denotes the inversion results

The estimated sound speed profile within the sediments is presented in Fig. 9, where the solid curve is inverted from the measured data. Due to limitations in grazing angle, the properties of the upper sediment corresponding to small grazing angles could not be directly estimated. Therefore, the upper sound speed profile (dashed curve) was obtained by fitting the solid curve, while constraining the upper sound speed to 1530 m/s.

As illustrated in Fig. 9, the deep-water sediment exhibits a sound speed range of 1530-1990 m/s with a maximum thickness of 438 m. The sediment sound speed demonstrates a gradual increase throughout the medium, showing no distinct transitional stratification. Following the inversion of the sediment sound speed profile, the density profile was further estimated using Hamilton's empirical formula, as shown in Fig. 10. Comparison with geological survey data from the South China Sea (11) confirms the reasonableness of the estimated results.

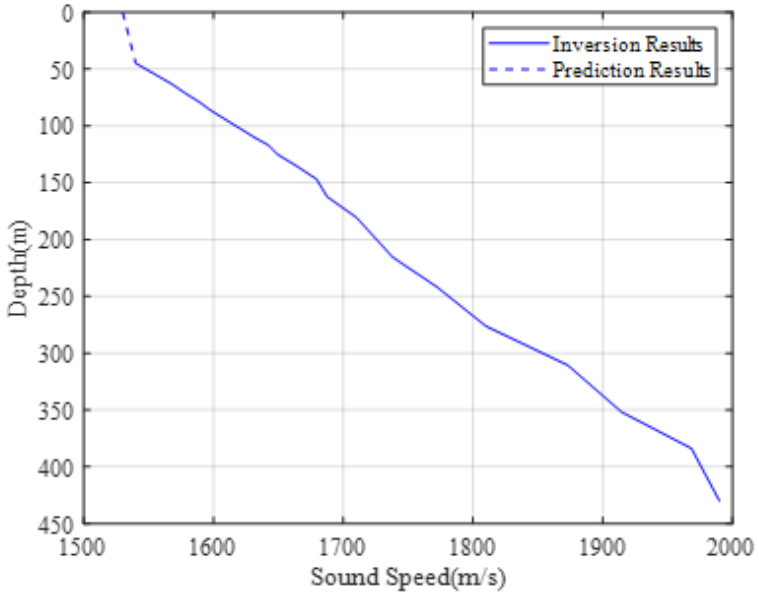


Fig. 9 Estimated sound speed profile in continuous sediment

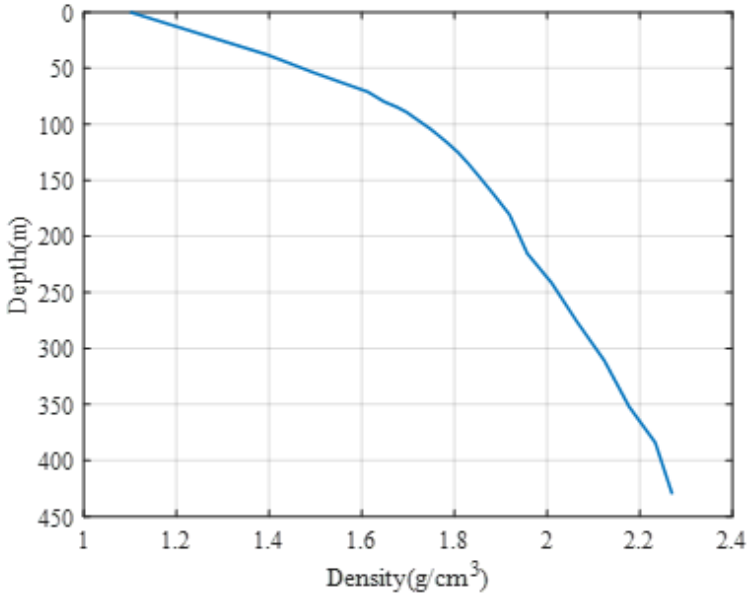


Fig. 10 Estimated density profile in continuous sediment

5. CONCLUSIONS

The paper proposes a stepwise inversion method based on a continuous seabed model for estimating deep-water geoacoustic parameters. Unlike conventional approaches that assume layered homogeneous sediment, the method establishes a continuous sediment seabed model through simulation studies, specifically designed to explain the observed high-energy deep reflection paths in the South China Sea experiment.

Utilizing the seabed model with continuous sediment, we successfully inverted both sound speed and density profiles from normalized pressure data. The density inversion was further constrained by Hamilton's empirical formula. The estimated parameters show strong consistency with classical geoacoustic measurements in the South China Sea, demonstrating the validation of the proposed method.

ACKNOWLEDGEMENTS

This work was supported by the project 25-1-1-2-zyyd-jch supported by Qingdao Natural Science Foundation and Young Scientists Cultivation Foundation of Harbin Engineering University, China (Grant

No. 79000017).

REFERENCES

1. Jiang YM, Holland CW, Dosso SE, Dettmer J. Depth and frequency dependence of geoacoustic properties on the New England Mud Patch from reflection coefficient inversion. J ACOUST SOC AM. 2023 2023-1-1;154(4):2383-97.
2. Bonnel J, Lin YT, Eleftherakis D, Goff JA, Dosso S, Chapman R, et al. Geoacoustic inversion on the New England Mud Patch using warping and dispersion curves of high-order modes. J ACOUST SOC AM. 2018 2018-1-1;143(5): EL405-11.
3. Shi JJ, Dosso SE, Sun DJ, Liu QY. Geoacoustic inversion of the acoustic-pressure vertical phase gradient from a single vector sensor. J ACOUST SOC AM. 2019 2019-1-1;146(5):3159-73.
4. Wang MH, Zheng C, Sun DJ, Xu J. Sonar micro-Doppler effect induced by ship oscillatory motion in waves: Phenomenon, model, and experimentation. JASA EXPRESS LETT. 2025 2025-1-1;5(2).
5. Li XS, Sun DJ, Cao ZY. Generation Mechanism of Acoustic Doppler Velocity Measurement Bias. IEEE J OCEANIC ENG. 2025 2025-1-1;50(1):1-12.
6. Hou KY, Sun DJ, Teng TT, Hu JJ. A coherent integration method of an active sonar for maneuvering turning target detection. J ACOUST SOC AM. 2024 2024-1-1;155(5):2973-89.
7. Sun DJ, Li ZY, Zheng C, Cheng CY. A calibration method of high-precision transducer acoustic phase center under outfield experiments. APPL ACOUST. 2023 2023-1-1;210.
8. Sun DJ, Lu MY, Mei JD, Wang SC, Pei YQ. Generalized Radon transform approach to target motion parameter estimation using a stationary underwater vector hydrophone. J ACOUST SOC AM. 2021 2021-1-1;150(2):952-68.
9. Barclay DR, Buckingham MJ. On the spatial properties of ambient noise in the Tonga Trench, including effects of bathymetric shadowing. J ACOUST SOC AM. 2014 2014-1-1;136(5):2497-511.
10. Tollefsen D, Dosso SE, Knobles DP. Ship-of-Opportunity Noise Inversions for Geoacoustic Profiles of a Layered Mud-Sand Seabed. IEEE J OCEANIC ENG. 2020 2020-1-1;45(1):189-200.
11. Zhou QJ, Li XF, Zheng JL, Li XS, Kan GM, Liu BH. Inversion of Sub-Bottom Profile Based on the Sediment Acoustic Empirical Relationship in the Northern South China Sea. REMOTE SENS-BASEL. 2024 2024-1-1;16(4).

PROCEEDINGS of the
16th International Conference on Theoretical and Computational Acoustics

August 18 to 22, 2025 BEXCO, Busan Korea

A physics-based deep learning method for rigid scattering

Arzhang Angoshtari[1]
[1] Naval Surface Warfare Center – Carderock Division, USA

ABSTRACT

We introduce a deep learning method to solve the governing equation of rigid acoustic scatterers. In particular, we use feedforward neural networks to approximate solutions of a weakly-singular Burton-Miller formulation of the Kirchhoff-Helmholtz integral. A suitable non-local loss function based on the Burton-Miller formulation is defined to train deep networks. Preliminary results for a spherical scatterer suggest accuracy and scalability of this method.

Keywords: Scattering, Burton-Miller Formulation, Deep Learning

[1] arzhang.angoshtari.civ@us.navy.mil

DAILY PROGRAM

AUGUST 18 (MON)

14:20 – 16:00	Model-Based and Data-Driven Approaches in Underwater Acoustics	325 (3F)
Chair	Donghyeon Kim (Scripps Institution of Oceanography)	
14:20	Model-guided deep learning in underwater acoustics: achieving reliable performance via synthetic training data	
	Jongkwon Choi, Youngjoo Kim, Keunhwa Lee	
14:40	Impact of ship density on underwater acoustic detection range	
	Youngseung Kim, Dong-Gyun Han, Wonjun Yang, Sungho Cho, Jee Woong Choi	
15:00	Doppler shift frequency estimation using time-frequency variation analysis	
	Ji-Hyun Lee, Dan-Bi Ou, Ki-Man Kim	
15:20	Coefficient of Variation Analysis in Passive Time-spectral Features for Ship Identification	
	Youngjoo Kim, Hongsung Jeong, Jongkwon Choi, Keunhwa Lee	
15:40	Array Invariant-Based Passive Localization Using Deconvolved Beamforming	
	Wonjun Yang, Dong-Gyun Han, Jee Woong Choi	

PROCEEDINGS of the
16th International Conference on Theoretical and Computational Acoustics

August 18 to 22, 2025 BEXCO, Busan Korea

Model-guided deep learning in underwater acoustics: achieving reliable performance via synthetic training data

Jongkwon Choi[1], Youngjoo Kim[2], Keunhwa Lee[3]

[1] Sejong University

[2] Sejong University

[3] Sejong University

ABSTRACT

In underwater acoustics, the lack of high-quality data makes it difficult to implement artificial intelligence techniques with confidence. This study proposes a model-guided deep learning approach for automatically extracting key features from sonogram in ocean acoustics. This methodology can address the data quality issue and ensure robust performance even with a small amount of actual training data. In this study, two applications of model-guided deep learning based on a U-Net architecture are introduced for automatically extracting key features from sonogram images: extracting modal dispersion curves from spectrograms and detecting time-quefrency curve from cepstrograms. Validation shows that the deep learning networks, trained using only model data, accurately detect features from real data, even though this data was not used for training. This indicates that model-guided learning is effective in extracting features from sonogram images. This work was supported by the Korea Research Institute for defense Technology planning and advancement (KRIT) grant funded by the Korea government (DAPA (Defense Acquisition Program Administration)) (No. 20-106-B00-003 (KRIT-CT-22-052), Physics-guided Intelligent Sonar Signal Detection Research Laboratory, 2025).

[1] jkchoi.sju@gmail.com

[2] submakim@gmail.com

[3] nasalkh2@sejong.ac.kr

IMPACT OF SHIP DENSITY ON UNDERWATER ACOUSTIC DETECTION RANGE

Youngseung KIM[1]; Dong-Gyun HAN[2]; Wonjun YANG[3]; Sungho CHO[4]; Jee Woong CHOI[5]

[1,2,3,5] Hanyang University ERICA

[4] Korea Institute of Ocean Science & Technology

ABSTRACT

Underwater acoustic detection performance is affected by the variability of ocean noise level, which fluctuates with factors such as ship density and weather. However, measuring ocean noise level in real time is time-consuming and often impractical in time-sensitive operations such as anti-submarine warfare. This study investigates ocean noise level variability associated with ship density and its impact on detection range, aiming to support reliable detection range estimation when direct ocean noise level measurements are unavailable

The predicted received level at a single receiver was calculated by summing acoustic intensities radiated from ships within the defined area. For this, the source level of each ship was modeled using the Ross power-law based on its specifications. Transmission loss from each ship to the receiver was computed using a parabolic equation-based acoustic propagation model. To evaluate the reliability of the predicted received level, its similarity to the measured received level was assessed by comparing the two, which showed a high degree of similarity. Therefore, it was used to assess detection range variation as a function of ship density. Detailed discrepancies and limitations will be discussed in the presentation.

Doppler Shift Frequency Estimation Using Time-Frequency Variation Analysis

Ji-Hyun Lee[1]; Dan-Bi Ou; Ki-Man Kim
Korea Maritime and Ocean University, South Korea

ABSTRACT

Doppler shift frequency estimation is one of the critical issues in underwater acoustic communication and active sonar systems. The Doppler effect is typically caused by the mobility of platforms, and if the receiver in an underwater acoustic communication system does not compensate for the Doppler-shifted frequency, the performance degradation may occur. Due to this importance, various methods have been studied, among which the most widely known approach is measuring the cross-correlation between the received signal and a replica of the transmitted signal. In this study, we propose a Doppler shift frequency estimation method based on the time and frequency characteristics of the received signal. Through simulations, we compare the proposed method with existing approaches and verify its effectiveness.

Keywords: Doppler shift frequency estimation, Spectrogram, Deep learning, Convolutional neural network

1. INTRODUCTION

The estimation of Doppler shift frequencies or the radial velocity of a platform is a critical challenge in underwater acoustic communications and active sonar systems. Doppler shifts are usually caused by the movement of platforms (e.g., transmitter and receiver), and many methods have been proposed to address these difficulties. A conventional method in underwater acoustic communication is transmitting a preamble signal which includes an up-LFM connected to a down-LFM[1]. Here, up-LFM means a LFM (linear frequency modulation) signal whose frequency increases linearly over time, and down-LFM means the opposite. At the receiver, the Doppler shift is estimated by calculating the time difference between the two peaks of the correlation results with each LFM signal in the previous work. However, this method can be affected by noise or multipath propagation, which may cause the correlation peaks to appear at an incorrect position, making the estimation difficult to be accurate.

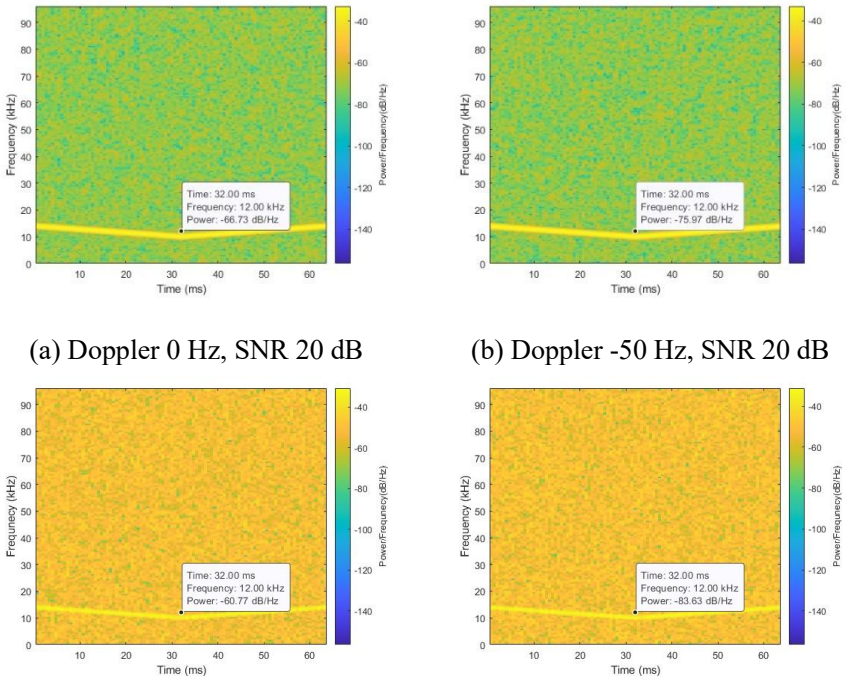
To overcome this problem, we propose a supervised learning-based algorithm using recent achievements with deep learning in this field. Deep learning has enhanced the accuracy of estimating not only in doppler shift estimation but also in time synchronization tasks [2]. However, in this work, we focused on deep learning methods using spectrograms as its training data set [3], [4]. Unlike the previous work done with non-data aided methods [5], we propose an algorithm using LFM signals as preamble data to increase the accuracy of the estimated Doppler shifts. In the proposed method, we create spectrograms for the received up- and down-LFM signals and convert them into image data, which are then used as input data for a convolutional neural network (CNN) to estimate the doppler shift frequency. Simulation results are presented to compare the performance of the proposed algorithm demonstrating its effectiveness and robustness compared to the conventional method.

2. PROPOSED ALGORITHM

When transmitting a signal, the preamble section composed of predetermined signals known from both transmitter and receiver can be used to estimate channel characteristics or Doppler effects, which leads to reliable demodulation of the signal. In our approach, we used a preamble structure where an up-LFM and a down-LFM signal are concatenated. An example of the resulting spectrogram is shown in Figure 1,

[1] jigui99@naver.com

illustrating variations caused by different Doppler shifts and signal-to-noise ratio (SNR) values. When a Doppler shift occurs, a compression or expansion of the signal appears in the time axis, and its spectrum is shifted along the frequency axis. Based on the spectrogram image below, we propose a Doppler estimation algorithm that utilizes CNN to learn variations in power patterns from the spectrogram images and identify the changes caused by the various Doppler effects, enabling the accurate estimation of the Doppler frequency.



(a) Doppler 0 Hz, SNR 20 dB

(b) Doppler -50 Hz, SNR 20 dB

(c) Doppler 0 Hz, SNR 0 dB

(d) Doppler -50 Hz, SNR 0 dB

Figure 1 – Spectrograms of the received signals according to various SNR values and Doppler shifts

The overall structure of the proposed algorithm is presented in Figure 2. Considering that the input spectrogram images have a size of (513×95), the convolutional layers were designed to preserve as much information as possible. For this purpose, zero-padding was applied to the edges of each layer input data, and a stride of (1,1) was applied throughout the network. The ReLU activation function, which is widely used in deep learning models was used as the activation function with all the layers. Thus, to reduce the volume of the data, average pooling with a (2×2) kernel and a stride of (2,2) was used after each convolutional layer. The output of the final convolutional layer was directly flattened and used as the input data for the fully-connected layer, bypassing and pooling process. To enhance the algorithm's ability to extract increasingly complex and abstract features in deeper layer, the number of filters were doubled at each convolutional layer, following the expression $2^{(2+n)}$ filters per layer, where n represents the depth index of the layer.

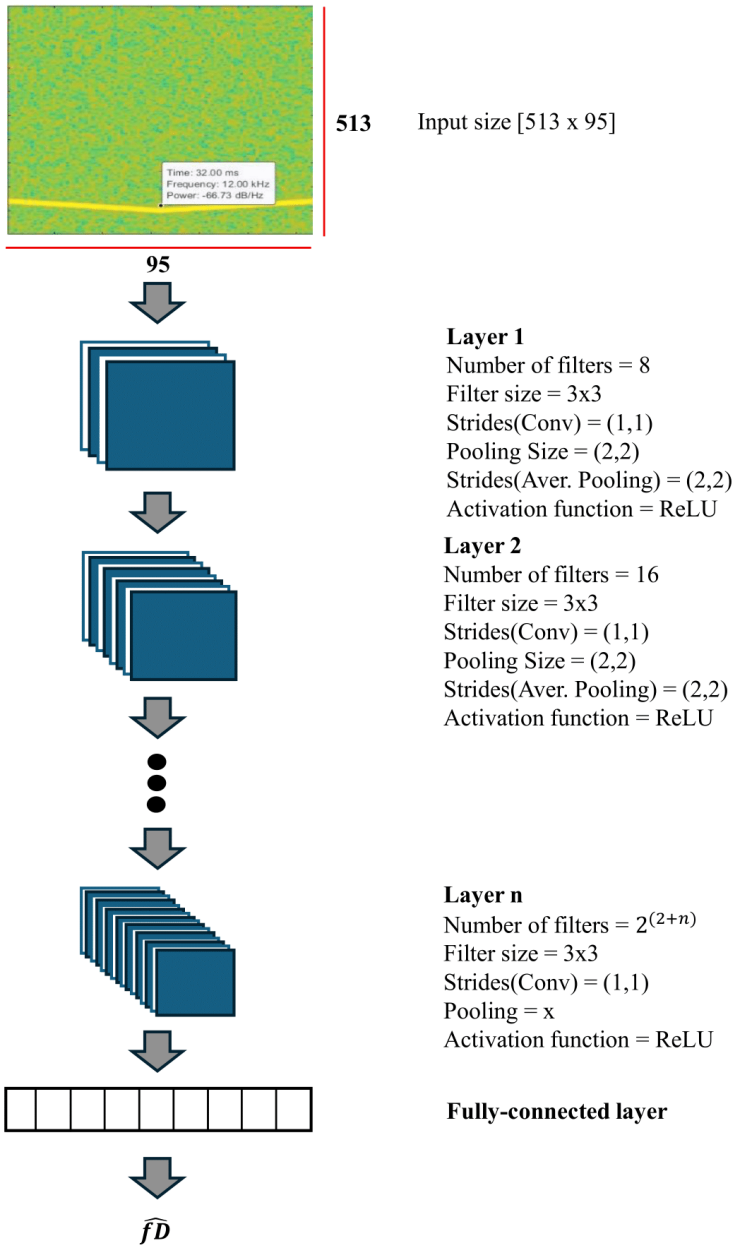


Figure 2 – Overall structure of the proposed algorithm

3. SIMULATION RESULTS

The training dataset was generated using LFM signals composed of two consecutive sweeps: a down-chirp followed by an up-chirp. Each sweep had a bandwidth of 4 kHz centered at 12 kHz and a duration of 32 msecs, resulting in a total preamble length of 64 msecs. The signal was normalized and sampled at 192 kHz.

In the simulations, we analyzed the optimal number of the convolutional layers and kernel size for the proposed algorithm, then compare the results with those of the conventional method. To optimize the proposed algorithm, simulations were conducted to analyze the error rate according to the number of convolutional layers and kernel sizes under varying SNR conditions. To ensure statistical robustness, each configuration was tested over multiple repetitions with randomized noise realizations. The proposed framework can also be extended to other types of preamble structures, depending on the diversity of the training data sets. Table 1 summarizes the computer specifications used to perform the simulations.

Table 1 – Computer specifications for simulation

OS	Windows 11
RAM	32 GB
CPU	Intel(R) Core(TM) i7-10700 @2.90 GHz
GPU	NVIDIA GeForce GTX 970
Toolkit	MATLAB 2024b Deep Learning Toolbox

For the training session, input data was prepared by generating spectrograms corresponding to Doppler shift frequencies ranging from -50 Hz to 50 Hz in 5 Hz increments. For each sample, a spectrogram affected by one of the 20 Doppler shift values and a reference spectrogram without Doppler effects were provided, along with the corresponding Doppler shift magnitude as the target label. This setup enabled the network to learn the amount of frequency shift from the reference spectrogram to the shifted one. To ensure data diversity, the SNR for each Doppler-shifted spectrogram varied from 0 dB to 20 dB in 0.4 dB steps, resulting in total of 1071 image-label pairs used for training. The network was trained for 10^2 iterations to enable robust estimation of Doppler shifts under various SNR conditions. An estimation of the algorithm was considered to be correct if the Doppler shift estimated by the CNN was within ± 0.5 Hz of the true value; otherwise, it was treated as incorrect, and the classification accuracy was calculated accordingly. Figure 3 shows the classification accuracy of the proposed method with respect to the number of convolutional layers.

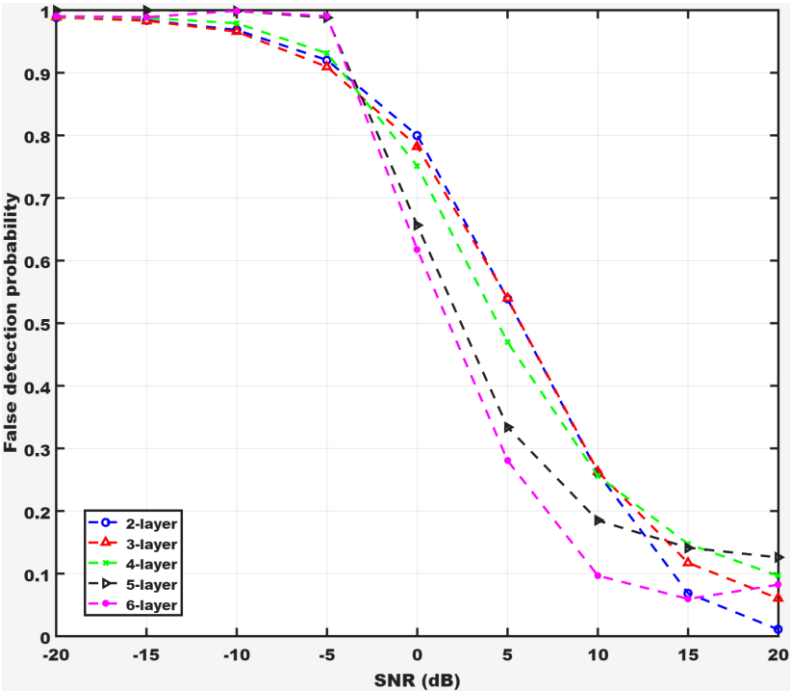


Figure 3 – False detection probability according to the number of convolutional layers

As mentioned earlier, considering the input spectrogram size of (513×95) and assuming that each convolutional layer preserves the majority of the information, the maximum possible convolutional depth was determined to six convolutional layers. Figure 3 shows the relationship between network depth, in terms of the number of convolutional layers, and the resulting classification accuracy of the proposed algorithm. In this analysis, the kernel size was fixed at 3×3, which is widely used in convolutional neural networks. As shown in Figure 3, increasing the number of layers made the algorithm more sensitive to the training data, indicating a higher degree of dependency. When six convolutional layers were used, the algorithm exhibited low error rates within the trained SNR range of 0 dB to 20 dB. However, at 20 dB, where data diversity becomes limited, the algorithm showed signs of overfitting. In contrast, the algorithm with only two convolutional layers had higher error rates in that area. Training was limited to 0 dB to 20 dB range, excluding SNRs below 0 dB, since the amount of data required for effective learning increases exponentially as SNR decreases, which also raises the likelihood of overfitting. Figure 4 shows the classification accuracy of the proposed algorithm according to each method's kernel size. This analysis was conducted using the six-layered CNN structure, which achieved the highest accuracy in the previous simulation show in Figure 3.

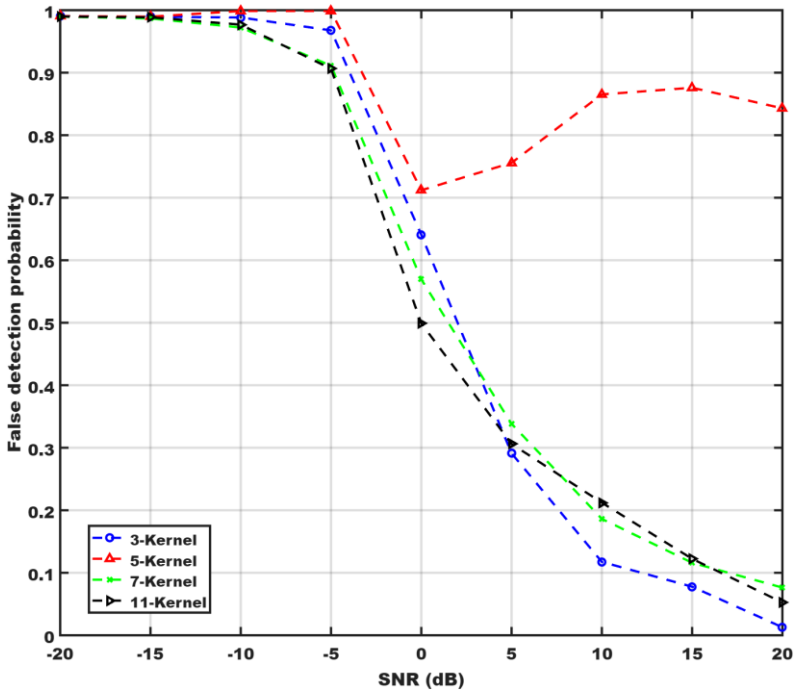


Figure 4 – False detection probability according to the kernel size

As shown in Figure 4, when the SNR environment is over 5 dB, the algorithm using the smallest kernel size (3×3) showed the lowest error rate. Based on the analysis of the training time for each configuration, the (3×3) type algorithm required approximately 800 minutes, the (5×5) type algorithm required approximately 900 minutes, the (7×7) type algorithm required approximately 1,000 minutes, and the (11×11) type algorithm required approximately 1,400 minutes.

Increasing the kernel size helps the algorithm to observe and learn from a more extended area of the input signal. However, since the proposed algorithm applies average pooling with a stride of (1,1) at every convolutional layer, the use of larger kernels results in more extensive spatial averaging. This increases both the training time and the likelihood of over-smoothing, which ultimately leads to higher error rates compared to algorithms trained with smaller kernel sizes.

These observations indicates that the most effective way to utilize larger kernel sizes may be to increase the pooling stride, prioritizing a substantial reduction in training time at the cost of minor accuracy drops. Nonetheless, in this study, the kernel size was optimized under the same conditions used for determining the optimal number of convolutional layers. Considering both the training time and error rate, the (3×3) kernel size was found to be the most appropriate type of model.

Based on the results shown in Figure 3 and 4, we compared the Doppler shift estimation performance of the proposed method, using a six-layer convolutional network with 3×3 kernel sizes, with that of the conventional approach. The test dataset was constructed similarly to the training dataset by generating spectrograms for Doppler shifts ranging from -50 Hz to 50 Hz in steps of 5 Hz and varying the SNR from -20 dB to 20 dB in steps of 5 dB. To obtain statistically reliable results, each simulation was repeated 10^4 times, and the average accuracy was computed. The same evaluation criterion used during the training session was applied, where a prediction was considered correct if the estimation error was within 0.5 Hz of the true Doppler shift value.

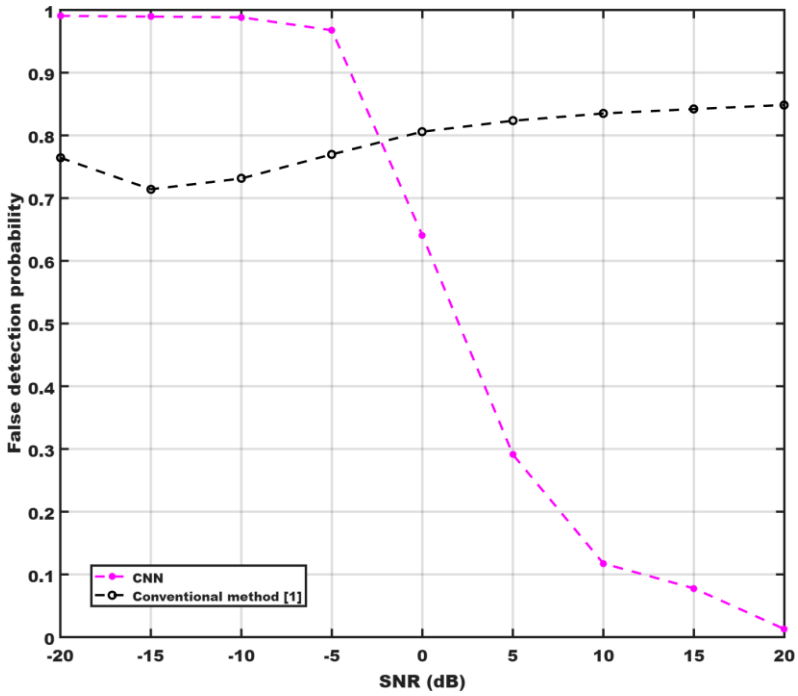


Figure 5 – Comparison between the proposed algorithm and conventional method

Figure 5 compares the proposed algorithm with conventional Doppler shift estimation method [1]. As shown in Figure, the deep learning algorithm designed using the optimal number of convolutional layers and kernel size identified in figure 3 and 4, demonstrated significant performance compared to the conventional method within the bounds of the SNR levels used during training sessions. In the trained SNR range from 0 dB to 20 dB, the proposed algorithm showed a decreasing error rate as the SNR increased. Compared to the conventional method, the proposed algorithm achieved an error rate improvement of approximately 16% at 0 dB SNR, and over 80% at 20 dB SNR. For a fair comparison, the error rate of the conventional method was also computed based on 10^4 simulation iterations, distinguishing an estimate as an error if it deviated by more than ± 0.5 Hz.

4. CONCLUSIONS

In this study, we proposed a CNN-based model that directly estimates the Doppler shift frequency from the spectrogram of the received LFM signals. The proposed algorithm was trained on Doppler shifts ranging from -50 Hz to 50 Hz under varying SNR conditions, using spectrogram images of size (513×95) as input data. Simulation results showed that, in environments with SNRs of 0 dB or higher, the proposed method achieved lower error rates compared to conventional Doppler estimation techniques. Moreover, the model demonstrated greater robustness to both noise and Doppler frequency variations. These results confirm that the CNN effectively learns the Doppler-sensitive features of time-varying signals. In future work, we will study more about the applicability of the proposed method in real-underwater environments, including scenarios with rapidly changing Doppler shift and multipath propagations.

ACKNOWLEDGEMENTS

This work is supported by Korea Research Institute for defense Technology planning and advancement(KRIT) grant funded by the Korea government(DAPA) (KRIT-CT-23-035, Multi AUV operation Technology for Mine Detection ('23~'28))

REFERENCES

1. S.A. Bayan. “Computationally efficient Doppler compensation system for underwater acoustic

communications". IEEE J. Oceanic Eng, vol. 25, no. 1, pp. 52-61, Jan 2000.
2. E. Jeong, E. Lee. "Convolutional neural network (CNN)-based frame synchronization method". Appl. Sci, 2020, vol. 10, no. 20, pp. 7267, Oct. 2020.
3. S. Kojima, Y. Goto, K. Maruta, S. Sugiura, C.J. Ahn. "Timing synchronization based on supervised learning of spectrogram for OFDM systems". IEEE Trans. Cognitive Communications and Networking, vol 9, no. 5, pp. 1141-1154, Oct. 2023.
4. Z. Yuan, Z. Meng, H. Fei, G. Yi. "Spectrum analysis and convolutional neural network for automatic modulation recognition". IEEE Wireless Communication Letters, vol. 8, no. 3, pp. 913-916, Jun. 2019.
5. S. Kojima, K. Maruta, Y. Feng, C.J. Ahn, V. Tarokh. "CNN-based joint SNR and Doppler shift classification using spectrogram images for adaptive modulation and coding". IEEE Trans. Communications, vol. 69, no. 8, pp. 5152-5167, Aug. 2021.

PROCEEDINGS of the
16th International Conference on Theoretical and Computational Acoustics

August 18 to 22, 2025 BEXCO, Busan Korea

Coefficient of Variation Analysis in Passive Time-spectral Features for Ship Identification

Youngjoo Kim[1]; Hongsun Jeong[2]; Jongkwon Choi[3]; Keunhwa Lee[4]
[1,2,3,4]Sejong University

ABSTRACT

Identifying ships from acoustic signals is challenging due to their nonstationary nature, mainly caused by environmental variations and transmission-path effects. In this study, we propose an averaged log-spectrum analysis combined with the coefficient of variation (CV) to improve ship identification accuracy using the ShipsEar dataset measured in waterways, which provides cleaner acoustic signatures than harbor environments. Acoustic signals were segmented into 1-second intervals, normalized through min-max scaling, and averaged in the frequency domain to mitigate environmental variability. Optimal frequency bands for reliable ship identification were selected by analyzing the CV. Correlation analysis demonstrated that using averaged log-spectra significantly enhances identification accuracy compared to individual spectra. These findings confirm that averaged log-spectrum analysis, guided by CV-based frequency band selection, effectively captures stable and distinct acoustic characteristics of ships. This work was supported by the Korea Research Institute for defense Technology planning and advancement (KRIT) grant funded by the Korea government (DAPA (Defense Acquisition Program Administration)) (No. 20-106-B00-003 (KRIT-CT-22-052), Physics-guided Intelligent Sonar Signal Detection Research Laboratory, 2025).

PROCEEDINGS of the
16th International Conference on Theoretical and Computational Acoustics

August 18 to 22, 2025 BEXCO, Busan Korea

Array Invariant-Based Passive Localization Using Deconvolved Beamforming

Wonjun Yang, Dong-Gyun Han, and Jee Woong Choi
Hanyang university ERICA

ABSTRACT

The array invariant is a passive localization method that estimates source range by exploiting the dispersive propagation characteristics of multipath arrivals in shallow water waveguides. Array invariant typically applies conventional beamforming to extract multipath arrivals, which are separated in beam angle and travel time to form a beam-time migration pattern used for range estimation. In this presentation, we apply deconvolved beamforming to the beam-time domain transformation, replacing conventional beamforming to obtain a more distinct representation of multipath arrivals. The output of conventional beamforming can be expressed as a convolution between the source angular distribution and the array beampattern. Accordingly, deconvolved beamforming estimates the source direction by deconvolving the beampattern from the beamformer output. This preserves the robustness of conventional beamforming while achieving the high-resolution characteristics of adaptive beamformers. We compare the beam-time migration and source-range estimates obtained using deconvolved beamforming and conventional beamforming, and the differences between the two approaches will be discussed.

This work was supported by Korea Research Institute for defense Technology planning and advancement (KRIT) -Grant funded by Defense Acquisition Program Administration (DAPA), South Korea (KRIT-CT-23-026, Integrated Underwater Surveillance Research Center for Adapting Future Technologies, 2023 2029).

DAILY PROGRAM

AUGUST 19 (TUE)

09:00 – 10:20	Beamforming and DOA Estimation in Underwater Acoustics	323 (3F)
Chair	Jeung-Hoon Lee (Changwon National University)	
09:00	Effects of Amplitude-phase Deviation on Azimuth Estimation and Beamforming in Vector Subbottom Profiling Signal Processing	
	Yanchao Li, Jianjun Zhu, Zhou Meng, Yu Chen, Guoqing Zhao, Korochentsev V.I.	
09:20	Research on near-field hyper beamforming for large aperture arrays	
	Yanbo Wang, Jidan Mei, Shuhui Wang	
09:40	A joint azimuth and radial velocity method for underwater passive multi-target tracking	
	Yuhang Yi, Bin Qi, Guolong Liang	
10:00	Broadband Target Direction of Arrival Estimation Based on a Single Pressure-difference Vector Hydrophone	
	Zehua Wang, Junjie Shi, Minshuai Liang, Dajun Sun	

PROCEEDINGS of the
16th International Conference on Theoretical and Computational Acoustics

August 18 to 22, 2025 BEXCO, Busan Korea

Effects of Amplitude-phase Deviation on Azimuth Estimation and Beamforming in Vector Subbottom Profiling Signal Processing

Yanchao Li[1], Jianjun Zhu[2], Zhou Meng[3], Yu Chen[4], Guoqing Zhao[5], Korochentsev V.I.[6]

[1] College of Underwater Acoustic Engineering, Harbin Engineering University, Harbin 150001, China

[2] 1 National Key Laboratory of Underwater Acoustic Technology, Harbin Engineering University, Harbin 150001, China; 2 Key Laboratory of Marine Information Acquisition and Security(Harbin Engineering University), Ministry of Industry and Information Technolo

[3] College of Meteorology and Oceanology, National University of Defense Technology, Changsha, 410073, China

[4] College of Meteorology and Oceanology, National University of Defense Technology, Changsha, 410073, China

[5] Sanya Nanhai Innovation and Development Base of Harbin Engineering University, Sanya 572000, China

[6]School of Engineering, Far Eastern Federal University, Vladivostok 690950, Russia

ABSTRACT

Aiming at the problem that the amplitude and phase deviation of vector hydrophones will affect the processing of subbottom profiling vector signals, the signal model of a two-dimensional pressure differential vector hydrophone for subbottom profiling detection in the presence of amplitude and phase deviation is given. Further theoretical formulas for the average acoustic intensity method and the complex acoustic intensity method of azimuth estimation in the presence of amplitude-phase bias, as well as for the conventional beamforming (CBF), the multiple signal classification method of beamforming (MUSIC), and the minimum variance distortion-free response method of beamforming (MVDR), are derived. Theoretical analyses and computer simulation studies have shown that the effect of phase deviation on azimuth estimation of the acoustic intensity flow is consistent in all directions. Still, the impact on azimuth estimation of the formation of each beam is nonlinear. The sensitivity bias, although it does not introduce azimuth estimation bias to the MUSIC and MVDR algorithms, increases the sensitivity of the two algorithms to the signal-to-noise ratio. The results of the thesis research are instructive for practical engineering applications.

[1] liyanchao_heu@163.com
[2] zhujianjun@hrbeu.edu.cn
[3] zhoumeng6806@163.com
[4]chenyu@nudt.edu.cn
[5] zhaoguoqing1001@hrbeu.edu.cn
[6] vkoroch@mail.ru

Research on near-field hyper beamforming for large aperture arrays

Yanbo WANG[1]; Jidan MEI1.[1,2,3]; Shuhui WANG[1,2,3]
[1] College of Underwater Acoustic Engineering, Harbin Engineering University, Harbin 150001,China
[2] National Key Laboratory of underwater Acoustic Technology, Harbin Engineering University, Harbin 150001, China
[3] Key Laboratory of Marine Information Acquisition and Security(Harbin Engineering University), Ministry of Industry and Information Technology, Harbin 15000F, China

ABSTRACT
Large-aperture arrays such as towed arrays and shore-based arrays have been widely used in underwater acoustic detection. Due to the large aperture of the array, these systems exhibit a large near-field region coverage, and the targets are often located within the near-field region of the array during detection operations. For the beam broadening problem of conventional beamforming (CBF) in the near-field region of large-aperture arrays, this paper investigate the near-field hyper beamforming (HBF) method. In this paper, theoretical models and simulation analyses are presented for near-field HBF under both plane wave and spherical wave models. Results demonstrates that: (1) Under the plane wave model, HBF can effectively counteract the beam broadening effect of CBF in the near-field region in the range beyond half of the array aperture for single targets, which significantly reduces the computational complexity compared with the near-field beamforming method; (2) Under the spherical wave model, HBF can reduce main beam width and suppress side lobe simultaneously, is proposed to improve near-field multi-target azimuth estimation and detection ability.

Keywords: Large-aperture arrays, Near field, Hyper beamforming, High-resolution

1. INTRODUCTION

Low-frequency, broadband, and large-aperture linear arrays represent the predominant developmental trend in sonar receiving arrays. However, in the field of underwater acoustics, a sound source is generally considered to be in the far field only when its distance from the array exceeds $r_s \geq 2D^2/(4\lambda)$, where D denotes the array aperture and λ is the wavelength of the radiated signal(1). For near-field targets, the nonlinear phase between receiving array elements render far-field beamforming methods ineffective(2,3). Specifically, this manifests as spectral peak broadening and spurious peaks at non-target directions in the spatial spectrum.Consequently, achieving high-resolution detection of near-field targets represents a fundamental problem for large-aperture array systems.

Hyper beamforming (HBF) is a high-resolution algorithm based on split array. By performing nonlinear operations on the beam output from left and right subarrays, it achieves mainlobe sharpening and sidelobe suppression simultaneously(4,5).When using the central position of a uniform linear array as the reference point for beamforming, planar-wave-compensated HBF can effectively offset quadratic phase errors induced by near-field targets, thereby enabling targets resolution in the non-extreme near-field region.To address the high-resolution requirement for extreme near-field targets, this paper propses a spherical-wave-compensated near-field HBF algorithm,which can significantly enhance direction-finding resolution for near-field scenarios.

2. The basic principles of HBF

The HBF algorithm processes the Conventional beamforming (CBF) outputs from two symmetric subarrays, as illustrated in Figure 1. The receiving array is divided into symmetrical left and right subarrays, with the array center (0, 0) serving as the reference point for CBF, the corresponding beam outputs are denoted as $\boldsymbol{R}_l$ and $\boldsymbol{R}_r$. Subsequently, the sum pattern and diff pattern are computed through the following

[1] yan_@hrbeu.edu.cn
[2] meijidan@hrbeu.edu.cn
[3] 1206312052@qq.com

operations.

$$\boldsymbol{R}_S=|\boldsymbol{R}_l|+|\boldsymbol{R}_r| \quad (1)$$

$$\boldsymbol{R}_D=|\boldsymbol{R}_l-\boldsymbol{R}_r| \quad (2)$$

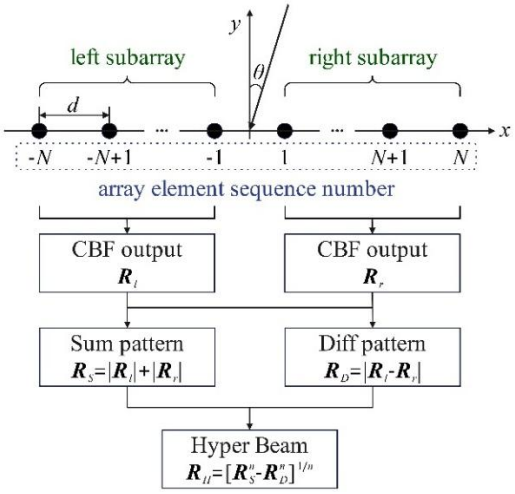


Figure 1 – Split beam system and HBF principle block diagram

The HBF output is obtained through a weighted subtraction operation between the sum and difference beams, mathematically expressed as

$$\boldsymbol{R}_H=[\boldsymbol{R}_S^n-\boldsymbol{R}_D^n]^{1/n} \quad (3)$$

In this equation, n represents the hyper-beam index, which is utilized to adjust the main lobe width and sidelobe level of the hyper-beam output. Its value typically ranges within [0.3, 1].

3. Performance of Far-field HBF for Near-field Targets

3.1 Direction-Finding Model of Far-field HBF for Near-field Targets

For near-field targets, the far-field plane wave model is no longer applicable. In this scenario, the time delay differences of the received signals across array elements depend not only on the target direction but also on the target distance. If the far-field beamforming method is still applied to near-field targets, the resulting spatial spectrum will exhibit peak broadening and spurious peaks.Figure 2 illustrates the received signals of a uniform linear array under near-field target conditions.

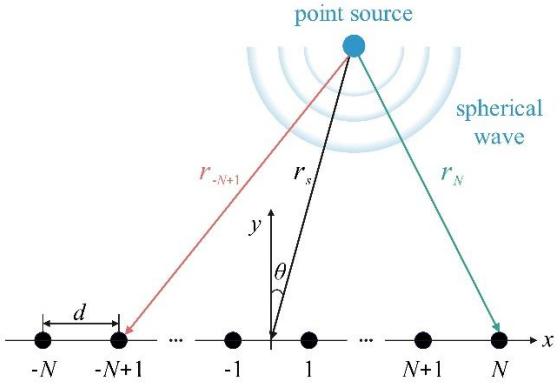


Figure 2 – The received signal of uniform linear array under the near-field target

In the figure, the x-coordinate of the nth array element is given by $x_n=nd+1/2$. Consequently, the distance between the target and the nth array element can be expressed as

$$r_n=\sqrt{r_s^2+x_n^2-2r_sx_n\sin\theta} \quad (4)$$

When $x_n \ll r_s$, the Taylor series expansion of r_n can be performed while neglecting higher-order infinitesimal terms, yielding

$$r_n\approx r_s-x_n\sin\theta+x_n^2\cos\theta/(2r_s) \quad (5)$$

Therefore, the phase difference between the nth array element and the reference point is given by

$$\varphi_n=(r_n-r_s)2\pi f/c\approx(x_n\sin\theta-\frac{x_n^2\cos\theta}{2r_s})\frac{2\pi f}{c} \quad (6)$$

Here, $\Delta\varphi_n=2\pi f x_n^2\cos\theta/2r_sc$ corresponds to the quadratic phase term of spherical waves in near-field

conditions, which constitutes the phase error induced by near-field targets, while $2\pi f x_n \sin\theta / c$ represents the linear phase term under far-field conditions.

Since $x_{-n}^2 = x_n^2$, we have $\Delta\varphi_{-n}=\Delta\varphi_n$, indicating that the beam output exhibits identical phase errors on both sides of the true target bearing. Furthermore, as derived from Equation (6), for the left subarray where $x_n<0$, $\Delta\varphi_n$ causes $|2\pi f x_n \sin\theta / c|$ to be overestimated; conversely, for the right subarray where $x_n>0$, $\Delta\varphi_n$ leads to an underestimation of $|2\pi f x_n \sin\theta / c|$. Consequently, if conventional beamforming (CBF) is still performed under the plane wave assumption for the aforementioned near-field targets, the spurious peaks generated in the left and right patterns due to nonlinear phase errors manifest symmetrically about the true bearing. Under such conditions, the sum beam calculated via Equation (1) will manifest symmetric spurious peaks flanking the true target direction on both sides.

Furthermore, it can be demonstrated that when a signal with frequency f impinges on the array at angle θ, the diff pattern can be expressed as

$$R_D(f,\theta_0)=2A|\sin[\phi(\theta,\theta_0)]| \tag{7}$$

$$A=|R_r(f,\theta_0)|=|R_l(f,\theta_0)| \tag{8}$$

$$\phi(\theta,\theta_0) = 2\arctan\left\{\frac{\sum_{n=1}^{N}\sin\left[\frac{2\pi f}{c}\left(n+\frac{1}{2}\right)d(\sin\theta_0-\sin\theta)+\Delta\varphi_n\right]}{\sum_{n=1}^{N}\cos\left[\frac{2\pi f}{c}\left(n+\frac{1}{2}\right)d(\sin\theta_0-\sin\theta)+\Delta\varphi_n\right]}\right\} \tag{9}$$

In this equation, θ_0 represents the preformed beam direction. As the target direction θ deviates from θ_0, ϕ increases with $(\sin\theta_0-\sin\theta)$, leading to growth in $|\sin\phi|$ and consequently an increase in $R_D(f,\theta_0)$. Therefore, $\boldsymbol{R}_D$ forms a null in the target direction while exhibiting peaks on both sides of this null. After hyper-beam operation, the spurious peaks of $\boldsymbol{R}_S$ can be significantly suppressed by the flanking peaks of $\boldsymbol{R}_D$'s null, while the HBF output produces a distinct peak at the true target direction. This mechanism enables the plane-wave-compensated HBF to achieve resolution of near-field targets.

3.2 Simulation Rsults and Analysis

Considering a 64-element uniform linear array, we simulate the scenario where a 90Hz monochromatic signal impinges on the array from 0° direction, with inter-element spacing set to half the wavelength of the target signal. The target distance is denoted as r_s, and the hyper-beam index is set to 0.5. Under two different near-field conditions (r_s = 0.4km and 0.2km), we conduct comparative simulations of spatial spectrum between far-field CBF and HBF, with the results presented in Figure 3.

Figure 3(a,b) reveals that for near-field targets, far-field CBF exhibits beam broadening and spurious peaks. In contrast, the HBF's diff pattern forms a distinct null at the true target direction, whose flanking peaks effectively cancel the spurious peaks in the sum beam. This mechanism enables the far-field HBF to resolve near-field targets with superior angular resolution. Compared with spherical-wave-compensated near-field beamforming methods, this approach significantly reduces computational complexity.

As shown in Figure 3(d,e), when the target distance decreases below a critical threshold, the null in the difference beam broadens, causing HBF failure. This occurs because higher-order terms in the Taylor expansion of Equation (4) become non-negligible on this occasion, introducing asymmetric nonlinear phase errors. Consequently, the phase relationship between the left and right subarray outputs breaks down, resulting in null widening. Simulation results demonstrate that far-field HBF maintains effective target resolution for near-field regions beyond approximately half the array aperture.

4. Spherical-wave-model-based Near-field HBF

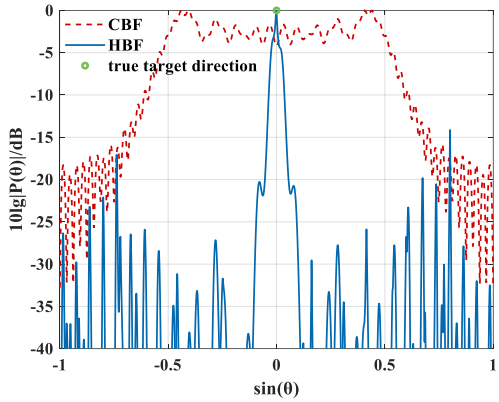
4.1 Principles of Near-field HBF

For the near-field acoustic source depicted in Figure 2, if near-field CBF is implemented, the weight vector can be expressed as

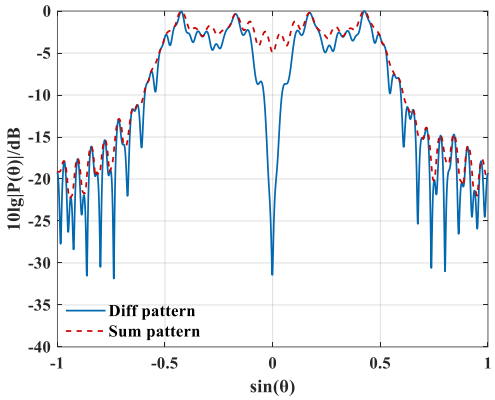
$$\boldsymbol{w}_c=\frac{\boldsymbol{p}_s(r_s,\theta)}{\sqrt{\boldsymbol{p}_s{}^{\mathrm{H}}(r_s,\theta)\boldsymbol{p}_s(r_s,\theta)}} \tag{10}$$

Here, $\boldsymbol{p}_s$ denotes the steering vector in the source direction. Under spherical-wave compensation conditions, the steering vector for the θ_k direction is generally expressed as

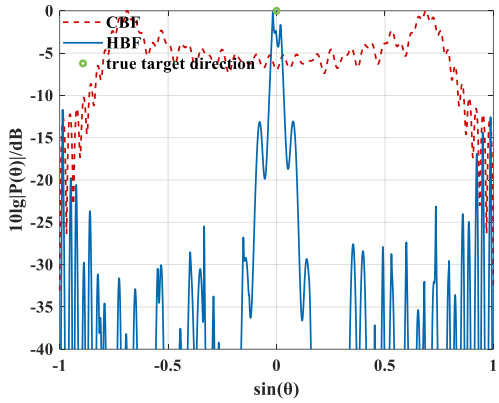
$$\boldsymbol{p}(r_s,\theta_k)=\begin{bmatrix}\exp\{-j2\pi f[r_{-N}(r_s,\theta_k)-r_s]/c\}\cdot[r_s/r_{-N}(r_s,\theta_k)]\\ \exp\{-j2\pi f[r_{-N+1}(r_s,\theta_k)-r_s]/c\}\cdot[r_s/r_{-N+1}(r_s,\theta_k)]\\ \cdots\\ \exp\{-j2\pi f[r_N(r_s,\theta_k)-r_s]/c\}\cdot[r_s/r_N(r_s,\theta_k)]\end{bmatrix} \tag{11}$$



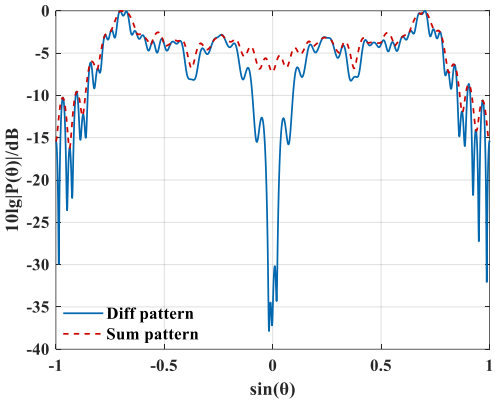
(a) r_s=0.4km, spatial spectrum of CBF and HBF



(b) r_s=0.4km, diff pattern and sum pattern



(c) r_s=0.2km, spatial spectrum of CBF and HBF



(d) r_s=0.2km, diff pattern and sum pattern

Figure 3 – The spatial spectrum of far-field CBF and HBF under near-field target conditions

By performing near-field CBF separately on the left and right subarrays and subsequently computing the sum and difference patterns, the following relationship can be established.

$$R_S=2|R_R|=2|R_L|=2A \tag{13}$$

$$R_D=2A|\sin[\phi(\theta,\theta_0,r_s,r_0)]| \tag{14}$$

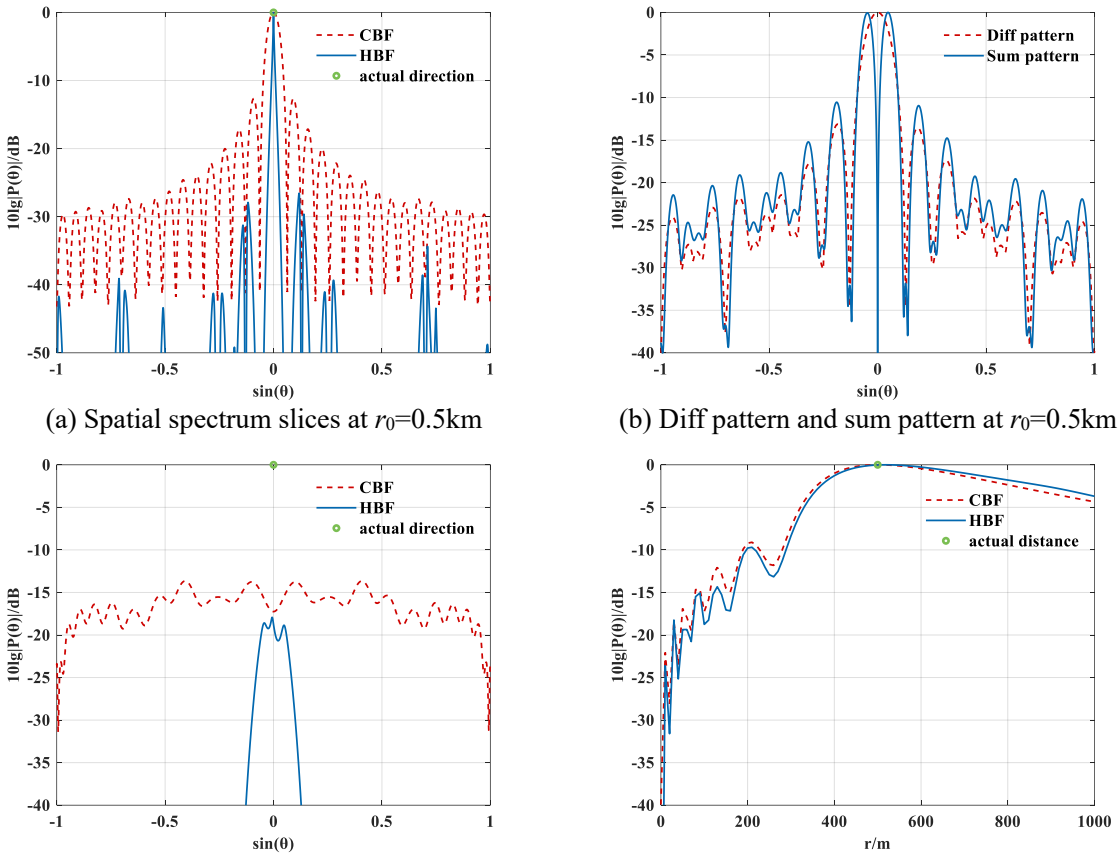
$$\phi(f,\theta,\theta_0,r_s,r_0)=2\arctan\left\{\frac{\sum_{n=1}^{N}\sin\left[2\pi f\left(r_{0,n}-r_{s,n}\right)/c\right]}{\sum_{n=1}^{N}\cos\left[2\pi f\left(r_{0,n}-r_{s,n}\right)/c\right]}\right\} \tag{15}$$

In the equations, (r_0,θ_0) represent the scanning coordinates, while (r_s,θ) denote the true target coordinates. Analysis of these expressions reveals that when $\theta_0=\theta$ and $r_0=r_s$, R_S reaches its peak value while R_D=0. As θ_0 deviates from θ or r_0 diverges from r_s, R_S decreases whereas R_D increases. Consequently, the HBF output simultaneously reduces both the mainlobe width and sidelobe levels of near-field CBF.

4.2 Simulation Rsults and Analysis

Considering a 32-element uniform linear array with a 90 Hz monofrequency signal impinging from the 0° direction, where the inter-element spacing equals half the wavelength of the target signal and the target distance is r_s=500m, we set the hyper-beam index to 0.5. Simulations of the spatial spectrum for near-field CBF and HBF were conducted, with Figure 4 presenting several representative two-dimensional slices.

Figures 4(a) present the spatial spectrum slices for the case where $r_0=r_s$=500m, while figures 4(b) present the corresponding sum pattern and diff pattern. The results demonstrate that near-field HBF achieves significantly narrower mainlobe width and lower sidelobe levels compared to near-field CBF. Figure 4(c) shows the spatial spectrum slice when r_0=100m≠r_s, where a obvious reduction in beam output power is observed. Figure 4(d) displays the spatial spectrum slice at θ_0=0°, revealing that both near-field CBF and HBF produce maximum output power at the the expective focus piont, just though with nearly identical mainpeak widths. Based on the above analysis, it can be concluded that the near-field HBF algorithm can provide effective enhancement of angular resolution for near-field targets, but without significant improvement of range resolution capability.



(a) Spatial spectrum slices at r_0=0.5km

(b) Diff pattern and sum pattern at r_0=0.5km

(c) Spatial spectrum slices at r_0=0.1km

(d) Spatial spectrum slices at θ_0=0°

Figure 4 – The 3D spatial spectrum slices of near-field CBF and HBF

5. CONCLUSIONS

This paper investigates the HBF algorithm in the context of near-field direction finding for large-aperture arrays, presenting an applicability analysis of far-field HBF in near-field scenarios and proposing a near-field HBF algorithm. Through theoretical analysis and simulation paper verify the adaptability and resolution performance of both approaches. The following conclusions are drawn: For non-extreme near-field targets, far-field HBF can achieve DOA with relatively low computational complexity when the target is within the range beyond approximately half of the array aperture. Based on near-field CBF, the proposed near-field HBF algorithm can achieve significant improvement in angular resolution without application range limitations.

ACKNOWLEDGEMENTS

This research was supported in part by the National Key Laboratory of Underwater Acoustic Technology under (Grant No.KY10500240094).

REFERENCES

1. Beranek LL. Criteria for noise and vibration in buildings and communities. In: Beranek LL, Ver IL, editors. Noise and Vibration Control Engineering - Principles and Applications. 2nd ed. New York, USA: Wiley; 2006. p. 887-909.
2. Sherman J. Properties of focused apertures in the fresnel region. IRE Transactions on Antennas and Propagation. 1962;10(4):399-408.
3. Friedlander B. Localization of signals in the near-field of an antenna array. IEEE Transactions on Signal Processing. 2019;67(15):3885-93.
4. Schlieter H, Eigenbrod H. Method for the Formation of Radiated Beams in Direction Finder Systems. US Patent US6,021,096.
5. Schlieter H. Passive sonar detection improved by hyperbeamtm technique. Europe: UDT. 2006:7A-2.

PROCEEDINGS of the
16th International Conference on Theoretical and Computational Acoustics

August 18 to 22, 2025 BEXCO, Busan Korea

A Joint Azimuth and Radial Velocity Method for Underwater Passive Multi-Target Tracking

Yuhang YI[123]; Bin QI[123*]; Guolong LIANG[123]
[1] National Key Laboratory of Underwater Acoustic Technology, Harbin Engineering University, China
[2] Key Laboratory of Marine Information Acquisition and Security (Harbin Engineering University), China
[3] College of Underwater Acoustic Engineering, Harbin Engineering University, China

ABSTRACT

The underwater passive multi-target tracking technique is widely utilized in the field of underwater information sensing due to its stealthiness. The Rao-Blackwellized particle filter-based data association algorithm (RBMCDA) is widely utilized due to its excellent tracking performance. However, this algorithm only leverages the target angle information, which requires to ensure the target state observability through platform maneuvering, and it is prone to inaccuracies in trajectory delineation in multi-target intersection scenarios. To address these issues, this paper proposes a joint azimuth-radial velocity enhanced RBMCDA algorithm (BV-RBMCDA). By fusing the target azimuth and radial velocity information, a multi-dimensional observation model is developed, enabling effective multi-target tracking through the Rao-Blackwellized Monte Carlo data association algorithm. Simulation results demonstrate that this approach outperforms the original approach and significantly enhances the accuracy of multi-target tracking in trajectory-crossing scenarios. This research proposes a novel approach for covert underwater multi-target tracking.

Keywords: Passive multi-target tracking, Data association, Radial velocity

1. INTRODUCTION

As land resources continue to be overexploited and depleted, countries are increasingly focusing on the development of marine resources. Underwater multi-target tracking technology is a core component of marine strategy, widely employed to obtain status information on underwater targets, defend against enemy vessels, conduct underwater search and rescue operations, and protect marine biological resources. Underwater target tracking technology involves using sonar-derived measurement data for data correlation and state estimation to generate continuous motion trajectories for each target. Based on the various operating modes of sonar, underwater multi-target tracking can be divided into two modes: active tracking and passive tracking. Active tracking relies on active sonar to transmit signals and receive echoes for detection, which inevitably reveals its own position. In contrast, passive tracking utilizing passive sonar offers better stealth, enabling measurement information to be obtained without revealing its own position. Therefore, passive sonar is widely utilized in military applications. Passive tracking involves processing noise emitted by targets, typically yielding only azimuth information; thus, it is also referred to as bearing-only tracking (BOT) problem (1). However, due to environmental noise and reverberation in the marine environment, the signal-to-noise ratio of signals received by passive sonar is typically low. Additionally, due to the physical resolution limitations of sonar arrays, when the azimuths of different targets intersect relative to the array, measurement gaps often occur, leading to discontinuous target trajectories or even erroneous measurement associations. Therefore, enhancing the tracking performance of the system under conditions of target missed detection or false alarms is essential.

The process of passive multi-target tracking primarily involves two key components: state filtering

* qibin@hrbeu.edu.cn

and data association. State filtering includes filter types and motion models, whereas data association includes trajectory initiation, trajectory maintaining, and trajectory termination.

Filter technology is one of the key techniques in multi-object tracking problems. In the 1960s, R.E. Kalman proposed the recursive Kalman filter theory under the linear Gaussian model, which represents the optimal solution for Bayesian estimation in linear systems. This theory continues to be widely applied in estimation filtering for linear systems to this day (2). Uhlmann and Durrant-Whyte further extended the KF to nonlinear, Gaussian spaces, proposing the enhanced Extended Kalman Filter (EKF). The EKF calculates the Jacobian matrix of the observation and state matrices, transforming nonlinear problems into linear ones, thereby enabling filtering estimation using the Kalman filter algorithm. Uhlmann and Durrant-Whyte proposed the unscented Kalman filter, which utilizes the UT transformation for sigma-point sampling to obtain an approximate distribution of the nonlinear function, yielding at least second-order Taylor accuracy solutions while eliminating the need for Jacobian matrix calculations, thereby achieving higher filtering accuracy at a lower computational cost (3). However, the above filters assume that the target distribution follows a Gaussian distribution. Gordon et al. proposed the particle filter (PF), which extends the state filter to a non-Gaussian model. By approximating the posterior probability density using a set of randomly sampled particles, the model function can be approximated with any precision, unrestricted by the system model or noise type (4). Doucet et al. proposed the Rao-Blackwellized particle filter, which combines KF and PF based on the idea of marginalization, reducing computational costs without affecting estimation accuracy, making it a commonly utilized filter.

The core task of data association technology is to achieve the initiation, maintenance, and termination of trajectories. In 1971, Singer proposed the nearest neighbor data association technique, which selects the observation data with the smallest statistical distance between the current measurement value and the current predicted value as the target measurement to update the target's state. This method is computationally efficient, simple, intuitive, and easy to implement. However, it is based on the assumption of single-target association and can lead to trajectory clustering in environments with high levels of clutter. To address this issue, Taenzer proposed Global Nearest Neighbor Data Association (GNNDA), which restricts each measurement to associate with only one target. However, it still cannot adapt to cluttered and multi-target scenarios. Bar-Shalom and Jaffer proposed probabilistic data association, which utilizes the probability density function of target features to calculate the similarity between targets, thereby achieving more accurate data association and target tracking. Bar-Shalom extended probabilistic data association from single-target tracking to propose joint probability density data association (JPDA) for multi-target tracking. However, since JPDA does not integrate target birth and death processes, it is only applicable to tracking scenarios with a fixed number of targets. (5). Reid proposed the Multiple Hypothesis Tracking (MHT) algorithm, which assumes that measurements originate from targets, clutter, or new targets. By calculating the probability of each hypothesis and performing random sampling, it achieves target birth and death decisions, making it a multi-target association algorithm suitable for dense clutter backgrounds. Rao-Blackwellized Monte Carlo Data Association (RBMCDA) based on the Rao-Blackwellized particle filter (RBPF) can be regarded as an enhanced multiple hypothesis target tracking method, with association based on the posterior distribution in Bayesian theory (6,7). The RB particle filter combines data association and state filtering while reducing computational load. However, when two target trajectories intersect or appear simultaneously at the same azimuth, the tracking accuracy of the current azimuth-only measurement-based RBMCDA algorithm (BO-RBMCDA) is low, and there is an issue of incorrect target batch division when trajectories intersect.

To address these issues, this paper proposes a joint azimuth and radial velocity passive underwater multi-target tracking method (BV-RBMCDA), which utilizes radial velocity information to enhance tracking accuracy while also improving the accuracy of trajectory intersection correlation, thereby resolving the issue of incorrect target batch division during trajectory intersections. This paper utilizes the Minimum Variance Distortionless Response (MVDR) beamforming algorithm to obtain the azimuth estimate of the target, and utilizes spectral analysis of the cross-correlation signal of the beam domain sound field to obtain the radial velocity of the target. The structure of this paper is as follows. Section 2 introduces the joint azimuth and radial velocity passive underwater multi-target tracking algorithm. Section 3 conducts simulation analysis, comparing and analyzing the tracking performance and target batch classification accuracy during trajectory crossovers between the original algorithm and the enhanced algorithm. Section 4 summarizes the entire paper.

2. PROBLEM FOMULATION

2.1 Discrete state space model of targets

The underwater multi-target tracking problem typically refers to a type of tracking problem based on the analysis of target motion states within a Bayesian framework. It is generally assumed that the motion of the targets and the measurements of the detection platform can be accurately described by certain specific mathematical models. The target state space can be further divided into the target motion state space and the target measurement space. The discrete-space target motion state model and the measurement model are defined as follows

$$\begin{cases} \mathbf{x}_k = \mathbf{F}_{k-1}\mathbf{x}_{k-1} + \mathbf{w}_{k-1} \\ \mathbf{z}_k = \mathbf{H}_k\mathbf{x}_k + \mathbf{v}_k \end{cases} \tag{1}$$

where $\mathbf{x}_k \in \mathbb{R}^n$ represents the state vector at time k; $\mathbf{z}_k \in \mathbb{R}^m$ represents the measurement vector at time k; $\mathbf{F}_{k-1}$ represents the target state transition matrix; $\mathbf{H}_k$ represents the number of observation matrices at time k; $\mathbf{w}_{k-1}$ and $\mathbf{v}_k$ represent process noise and observation noise, respectively, corresponding to sensor errors and random noise interference in the passive detection process,, respectively satisfying $\mathbf{w}_{k-1} \sim \mathcal{N}(0, \mathbf{Q}_{k-1})$ and $\mathbf{v}_k \sim \mathcal{N}(0, \mathbf{R}_k)$.

In traditional multi-target tracking algorithms, measurements are limited to azimuth angles. This paper introduces radial velocity to assist azimuth angle tracking. The target state motion model adopts a constant velocity model (CV). The target state vector $\mathbf{x}_k$ at time k is represented as $\mathbf{x}_k = \begin{bmatrix} \theta_k & \dot{\theta}_k & v_k & \dot{v}_k \end{bmatrix}^{\mathrm{T}}$, the measurement vector $\mathbf{z}_k$ at time k is represented as $\mathbf{z}_k = \begin{bmatrix} \theta_k & v_k \end{bmatrix}^{\mathrm{T}}$, and the state transition matrix $\mathbf{F}_{k-1}$ is represented as

$$\mathbf{F}_{k-1} = \begin{bmatrix} 1 & dt & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & dt \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{2}$$

where dt is the sampling period, the covariance matrix $\mathbf{Q}_{k-1}$ of process noise is expressed as

$$\mathbf{Q}_{k-1} = q \times \begin{bmatrix} dt^3/3 & dt^2/2 & 0 & 0 \\ dt^2/2 & dt & 0 & 0 \\ 0 & 0 & dt^3/3 & dt^2/2 \\ 0 & 0 & dt^2/2 & dt \end{bmatrix} \tag{3}$$

The measurement matrix $\mathbf{H}_k$ at time k is expressed as

$$\mathbf{H}_k = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{4}$$

2.2 Bayesian filtering

Bayesian filtering is an estimation method utilized to infer the state of a system over time, and is widely applied in the field of underwater target tracking. The core approach of Bayesian filtering is to calculate the likelihood function of the target based on the measurement sequence, and then combined with the prior probability of the target to update the posterior probability of the target. The probability density function is described by equation (1), expressed as

$$\begin{cases} \mathbf{x}_k \sim p\left(\mathbf{x}_k \middle| \mathbf{x}_{k-1}\right) \\ \mathbf{z}_k \sim p\left(\mathbf{z}_k \middle| \mathbf{x}_k\right) \end{cases} \tag{5}$$

The system defined above is a first-order Markov process. Given the state at time $k-1$, the state at time k is only related to the state at time $k-1$, i.e.

$$p\left(\mathbf{x}_k \left| \mathbf{x}_{1:k-1}, \mathbf{y}_{1:k-1}\right.\right) = p\left(\mathbf{x}_k \left| \mathbf{x}_{k-1}\right.\right) \tag{6}$$

Given the state at time k, the measurement value at time k is independent of historical state values and historical measurement values, i.e.

$$p\left(\mathbf{z}_k \left| \mathbf{x}_{1:k}, \mathbf{y}_{1:k-1}\right.\right) = p\left(\mathbf{z}_k \left| \mathbf{x}_k\right.\right) \tag{7}$$

The Bayesian filtering algorithm consists of two steps: prediction and update. In the prediction portion, the target transfer probability density function is utilized to perform recursion on the target state at time $k-1$, resulting in the predicted prior distribution $p\left(\mathbf{x}_k \left| \mathbf{z}_{1:k-1}\right.\right)$ of the target at time k, expressed as

$$\begin{aligned} p\left(\mathbf{x}_k \left| \mathbf{z}_{1:k-1}\right.\right) &= \int p\left(\mathbf{x}_k, \mathbf{x}_{k-1} \left| \mathbf{z}_{1:k-1}\right.\right) \mathbf{d}\mathbf{x}_{k-1} \\ &= \int p\left(\mathbf{x}_k \left| \mathbf{z}_{1:k-1}, \mathbf{x}_{k-1}\right.\right) p\left(\mathbf{x}_{k-1} \left| \mathbf{z}_{1:k-1}\right.\right) \mathbf{d}\mathbf{x}_{k-1} \\ &= \int p\left(\mathbf{x}_k \left| \mathbf{x}_{k-1}\right.\right) p\left(\mathbf{x}_{k-1} \left| \mathbf{z}_{1:k-1}\right.\right) \mathbf{d}\mathbf{x}_{k-1} \end{aligned} \tag{8}$$

In the update step, the prediction prior distribution is corrected using the observation data $\mathbf{z}_k$ to obtain the likelihood $p\left(\mathbf{z}_k \left| \mathbf{x}_k\right.\right)$ of the measurement for the prediction target. According to Bayes theorem, the posterior distribution $p\left(\mathbf{z}_k \left| \mathbf{x}_{1:k}\right.\right)$ of the target is obtained, thereby obtaining the optimal estimate of the target state, expressed as

$$p\left(\mathbf{x}_k \left| \mathbf{z}_{1:k}\right.\right) = \frac{p\left(\mathbf{z}_k \left| \mathbf{x}_k\right.\right) p\left(\mathbf{x}_k \left| \mathbf{z}_{1:k-1}\right.\right)}{\int p\left(\mathbf{z}_k \left| \mathbf{x}_k\right.\right) p\left(\mathbf{x}_k \left| \mathbf{z}_{1:k-1}\right.\right) \mathbf{d}\mathbf{x}_k} \tag{9}$$

Equations (8) and (9) can be easily solved analytically under linear Gaussian conditions, however, under nonlinear and non-Gaussian conditions, some approximation methods are necessary to solve them.

3. A JOINT RBMCDA ALGORITHM FOR AZIMUTH AND RADIAL VELOCITY

The Rao-Blackwellized Particle Filter (RBPF) is an enhanced particle filter based on Rao-Blackwellized theory, offering a framework that effectively combines data association and state filtering. It divides the state space into two portions: one portion utilizes a Kalman filter to estimate the state, while the other utilizes sequential Monte Carlo methods for probabilistic density estimation of the state. The RBMCDA algorithm based on RBPF can be regarded as an enhanced multi-hypothesis target tracking method, utilizing latent variables to further represent target states and association conditions. The specific steps of the BV-RBMCDA algorithm proposed in this paper are as follows.

First, the particle state at time $k-1$ can be represented as $S_{k-1}^{(n)} = \left\{w_{k-1}^{(n)}, u_{k-1}^{(n)}, \mathbf{m}_{k-1}^{(n)}, \mathbf{P}_{k-1}^{(n)}\right\} \quad n = 1, \cdots, N$, where N is the number of particles, $w_{k-1}^{(n)}$ is the weight of particle n in the distribution, and $\mathbf{x}_k^{(n)} = \{\mathbf{m}_{k\text{-}1}^{(n)}, \mathbf{P}_{k\text{-}1}^{(n)}\}$ represent the mean and covariance of the state estimate of particle n at time $k-1$. $u_{k-1}^{(n)}$ is the latent variable of particle n, $u_k^{(n)} = \{e_k, c_k\}$, where $e_k = \{e_k^1, e_k^2, ..., e_k^i\}, i = 1, 2, ..., n_k$ is a binary matrix representing the visibility status of n_k targets at time k. c_k is an association indicator that represents the association status of surviving targets at time k. $c_k = 0$ represents measurement data-related noise, $c_k = j$ represents measurement data-related target j, and c_k satisfies a first-order Markov process, i.e. $p(c_k \mid c_{1:k-1}) = p(c_k)$.

Second, the trajectory prediction process. Perform a Kalman prediction process on each target j, in each particle to obtain the prior transition probability of the target. The prediction process is expressed as

$$\begin{cases} \mathbf{m}_{k|k-1}^{(n)} = \mathbf{A}_{k-1}\mathbf{m}_{k-1}^{(n)} \\ \mathbf{P}_{k|k-1}^{(n)} = \mathbf{A}_{k-1}\mathbf{P}_{k-1}^{(n)}\mathbf{A}_{k-1}^{\mathrm{T}} + \mathbf{Q}_{k-1} \end{cases} \tag{10}$$

Prior distribution represented as

$$p(u_k^{(n)} \mid u_{1:k-1}^{(n)}) = \begin{cases} (1-pb)\cdot cp, \text{correlation with noise} \\ (1-pb)(1-cp)/n_{k-1}, \text{correlation existing targets} \\ pb, \text{correlation new target} \end{cases} \tag{11}$$

Third, the data association process. Utilizing a multi-hypothesis data association strategy, list all the association situations, calculate the hypothesis likelihood probability, and then obtain the posterior probability. The measurement values may arise from noise, existing targets, and new targets. The hypothesis likelihood probability expression is

$$p(\mathbf{z}_k \mid u_k^{(n)}, \mathbf{z}_{1:k-1}, u_{1:k-1}^{(n)}) = \begin{cases} 1/V, \text{ correlation with noise} \\ \mathcal{N}\left(\mathbf{z}_k \left| \mathbf{A}_{k-1}\mathbf{m}_{k-1}^{(n)}, \mathbf{A}_{k-1}\mathbf{P}_{k-1}^{(n)}\mathbf{A}_{k-1}^{\mathrm{T}} + \mathbf{Q}_{k-1} \right.\right), \text{correlation existing target} \\ \mathcal{N}\left(\mathbf{z}_k \left| \mathbf{H}_k\mathbf{m}_0\mathbf{H}_k\mathbf{P}_0\mathbf{H}_k^T + \mathbf{R}_k \right.\right), \text{ correlation new target} \end{cases} \tag{12}$$

The posterior probability can be expressed as

$$\begin{aligned} p(u_k^{(n)} \left| \mathbf{z}_{1:k}, u_{1:k-1}^{(n)} \right.) &= \frac{p(\mathbf{z}_k \mid u_k^{(n)}, \mathbf{z}_{1:k-1}, u_{1:k-1}^{(n)}) p(u_k^{(n)} \mid \mathbf{z}_{1:k-1}, u_{1:k-1}^{(n)})}{p(\mathbf{z}_k \mid \mathbf{z}_{1:k-1}, u_{1:k-1}^{(n)})} \\ &\propto p(u_k^{(n)} \mid u_{1:k-1}^{(n)}) \cdot p(\mathbf{z}_k \mid u_k^{(n)}, \mathbf{z}_{1:k-1}, u_{1:k-1}^{(n)}) \end{aligned} \tag{13}$$

Fourth, sample the optimal importance distribution, which is the posterior distribution of data associations, expressed as follows

$$u_k^{(n)} \sim p\left(u_k \left| u_{1:k-1}^{(n)}, \mathbf{z}_{1:k} \right.\right) \tag{14}$$

Update the particle weights and normalize them using the following expression

$$\tilde{w}_k^{(n)} \propto w_{k-1}^{(n)} \frac{p\left(\mathbf{z}_k \left| u_{1:k}^{(n)}, \mathbf{z}_{1:k-1} \right.\right) p\left(u_k^{(n)} \left| u_{k-1}^{(n)} \right.\right)}{p\left(u_k^{(n)} \left| u_{k-1}^{(n)}, \mathbf{z}_{1:k} \right.\right)} \tag{15}$$

$$\tilde{w}_k^{(n)} = \frac{w_k^{(n)}}{\sum_{k=0}^{N} w_k^{(n)}} \tag{16}$$

Fifth, under the conditions determined by $u_k^{(n)}$, perform the Kalman filter update. The update results are as follows

$$\begin{cases} \mathbf{v}_k^{(n)} = \mathbf{z}_k - \mathbf{H}_k\left(u_k^{(n)}\right)\mathbf{m}_{k|k-1}^{(n)} \\ \mathbf{S}_k^{(n)} = \mathbf{H}_k\left(u_k^{(n)}\right)\mathbf{P}_{k|k-1}^{(n)}\mathbf{H}_k^T\left(u_k^{(n)}\right) + \mathbf{R}_k\left(\lambda_k^{(n)}\right) \\ \mathbf{K}_k^{(n)} = \mathbf{P}_{k|k-1}^{(n)}\mathbf{H}_k^T\left(u_k^{(n)}\right)\mathbf{S}_k^{-1} \\ \mathbf{m}_k^{(n)} = \mathbf{m}_{k|k-1}^{(n)} + \mathbf{K}_k^{(n)}\mathbf{v}_k^{(n)} \\ \mathbf{P}_k^{(n)} = \mathbf{P}_{k|k-1}^{(n)} - \mathbf{K}_k^{(n)}\mathbf{S}_k^{(n)}\left[\mathbf{K}_k^{(n)}\right]^T \end{cases} \tag{17}$$

Finally, calculate the number of effective particles. If it is less than a certain threshold value, resampling is required. The estimated result is then obtained as follows

$$p\left(\mathbf{x}_k, u_k \left| \mathbf{z}_{1:k} \right.\right) \approx \sum_{n=1}^{N} \left(w_k^{(n)} \delta\left(u_k - u_k^{(n)}\right) \cdot \mathcal{N}\left(\mathbf{x}_k \left| \mathbf{m}_k^{(n)}, \mathbf{P}_k^{(n)} \right.\right)\right) \tag{18}$$

4. SIMULATION RESULTS

To validate the effectiveness of the enhanced RBMCDA algorithm (BV-RBMCDA), the following simulation experiments were designed. First, an underwater passive multi-target tracking scenario was constructed. The observation platform was stationary, located at the origin (0, 0) of the coordinate system. The scenario included two moving targets, and azimuth data was obtained by setting the target motion trajectories. Both targets started at time 0, with one measurement processed per second, for a total processing time of 1000 seconds.

4.1 Non-intersecting trajectory scenario

The target motion parameters and simulation parameters are shown in Table 1.

Table 1 – Target motion parameters

Target Number	Trajectory equation		Processing time,s
	X-direction,m	Y-direction,m	
1	2000-2T	-1000+4T	1000
2	1500-2500cos(0.003×T)	500-2500sin(0.003×T)	1000

The simulation parameters used in this scenario are shown in Table 2.

Table 2 –Experimental simulation parameters

Simulation parameters	Numerical values
Time step - dt	1
Measure noise - R	8
Process noise - Q	1
Particle number - N	20
Noise density - cd	0.01
Noise prior probability - cp	0.01
Target probability of new births - pb	0.01

The motion status diagrams for the two targets are shown in Figure 1.

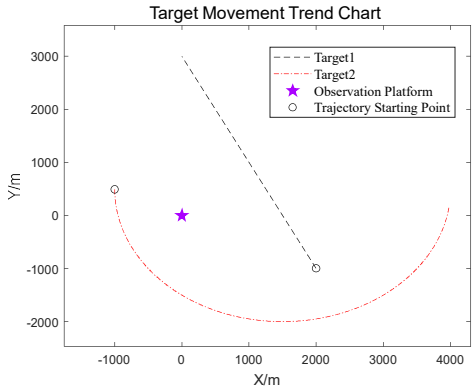
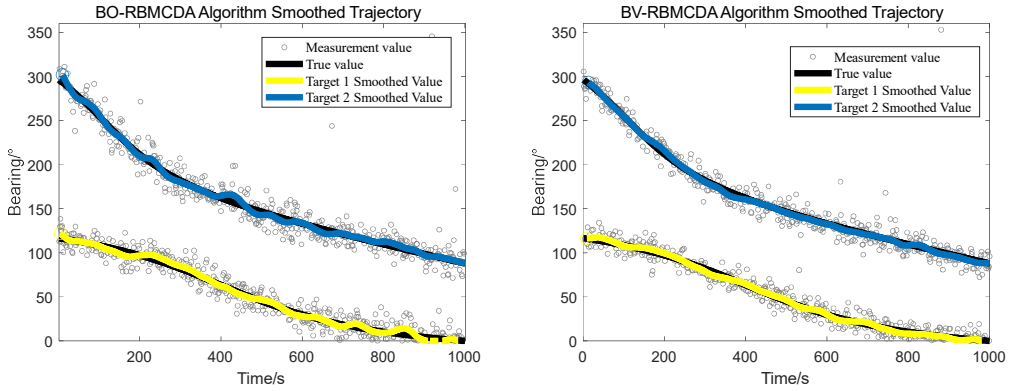


Figure 1 – Target movement pattern chart results

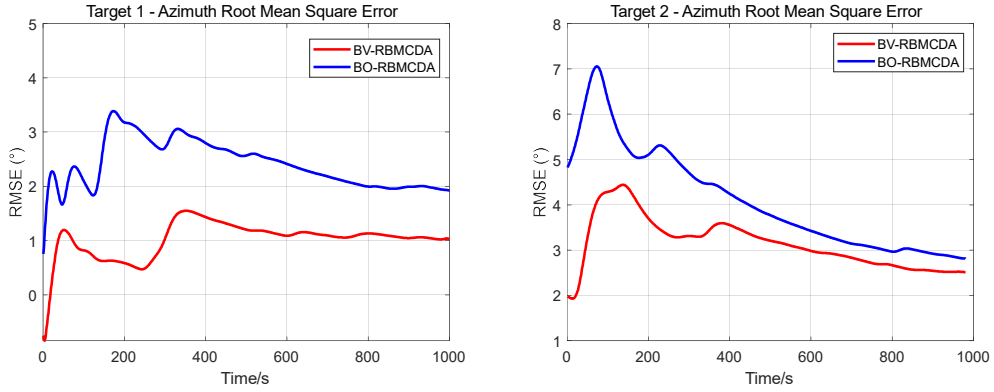
The tracking results based on the BO-RBMCDA algorithm and BV-RBMCDA algorithm after smoothing are shown in Figure 2.



(a) BO-RBMCDA algorithm tracking results (b) BV-RBMCDA algorithm tracking results

Figure 2 – Smoothed tracking results

As shown in Figure 2, both the BO-RBMCDA and BV-RBMCDA algorithms can achieve target trajectory tracking with relatively good tracking performance. Compared with the BO-RBMCDA algorithm, the BV-RBMCDA algorithm produces smoother trajectories that are overall closer to the true values, demonstrating superior stability and accuracy in trajectory tracking, as well as better filtering of measurement noise. The Root Means Square Error (RMSE) of the azimuth angle for both algorithms are shown in Figure 3.



(a) Azimuth RMSE of Target 1 (b) Azimuth RMSE of Target 2

Figure 3 – Target tracking error

Figure 3(a) shows the Root Mean Square Error (RMSE) results for the azimuth of Target 1. It can be clearly observed that the total error level of the BV-RBMCDA algorithm is lower than that of the BO-RBMCDA algorithm. The BV-RBMCDA algorithm converges quickly and maintains a low error level, with minimal subsequent fluctuations; while the BO-RBMCDA algorithm's error decreases, it remains higher than the BV-RBMCDA algorithm's, with more noticeable fluctuations. This indicates that the BV-RBMCDA algorithm processes measurement information more effectively and achieves higher tracking accuracy. Figure 3(b) shows the RMSE results for the azimuth angle of Target 2. It can be seen that the BV-RBMCDA algorithm also demonstrates a significant advantage in terms of RMSE, thereby indicating that the enhanced BV-RBMCDA algorithm proposed in this paper achieves higher tracking accuracy than the original BO-RBMCDA algorithm.

4.2 Trajectory intersection scenario

The target motion parameters and simulation parameters are shown in Table 3.

Table 3 – Target motion parameters

Target Number	Trajectory equation		Processing time,s
	X-direction,m	Y-direction,m	
1	-1000-2T	-3000+4T	1000
2	1500-2500cos(0.003×T)	500-2500sin(0.003×T)	1000

The simulation parameters used in this scenario are the same as those in Table 2. The motion status diagrams of the two targets are shown in Figure 4.

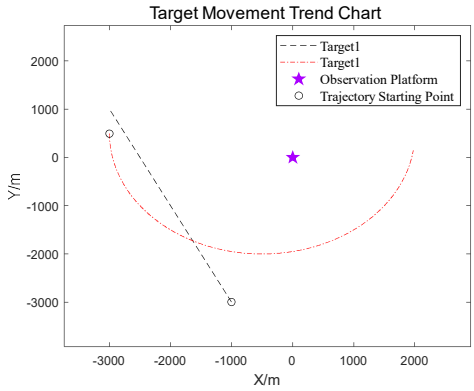
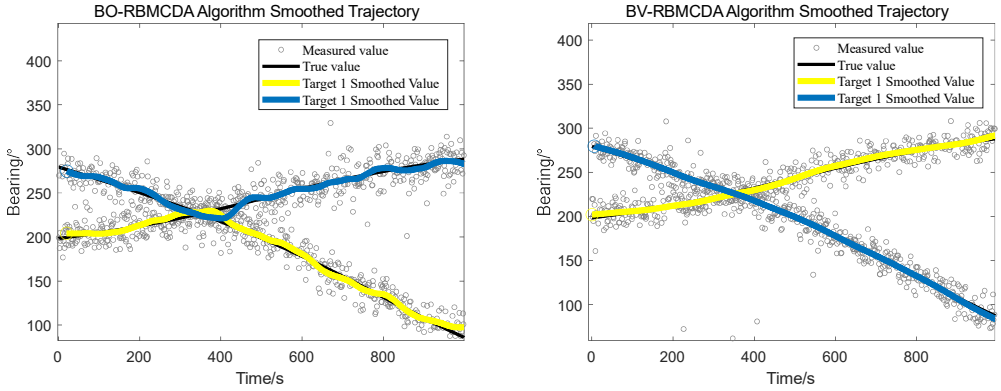


Figure 4 – Target movement pattern chart results

The tracking results based on the BO-RBMCDA algorithm and BV-RBMCDA algorithm after smoothing are shown in Figure 5.



(a) BO-RBMCDA algorithm tracking results (b) BV-RBMCDA algorithm tracking results

Figure 5 – Smoothed tracking results

Figure 5(a) shows the tracking results of the BO-RBMCDA algorithm. It can be seen that before the trajectories intersect, the tracking results of Target 1 and Target 2 are in good accord with the true values, indicating good tracking performance; However, after the trajectories intersect, the tracking trajectories of the BO-RBMCDA algorithm deviate significantly from the true values. The trajectories of Target 1 and Target 2 incorrectly intersect and become confused, making it impossible to accurately distinguish the actual motion paths of the two targets. This results in measurement data association errors, leading to inaccurate tracking results. Under the BV-RBMCDA algorithm, the tracking results for the two targets are almost identical to the actual values, indicating that this algorithm is more effective in handling measurement data association in multi-target trajectory intersection scenarios, with tracking accuracy minimally affected by intersection interference. The BO-RBMCDA algorithm and the BV-RBMCDA algorithm were each simulated 100 times. The batch classification accuracy obtained is shown in Table 4.

Table 4 – Batch classification accuracy rate

Algorithms	Batch classification accuracy rate,%
BO-RBMCDA	73
BV-RBMCDA	99

From the batch classification accuracy data, it is evident that there is a significant difference between the BO-RBMCDA and BV-RBMCDA algorithms in batch classification for multi-target tracking. The BO-RBMCDA algorithm achieves an accuracy rate of only 73%, whereas the BV-RBMCDA algorithm achieves an accuracy rate of 99% in batch classification, indicating that it can reliably and accurately associate measurement data from two targets into batches, effectively distinguishing measurement information from different targets, and demonstrating superior and more accurate tracking performance.

5. CONCLUSIONS

This paper addresses the issue of measurement batch partitioning errors that arise when only azimuth is utilized for tracking in underwater passive single-platform multi-target tracking. It proposes a combined azimuth and radial velocity RBMCDA algorithm (BV-RBMCDA). The linear

Gaussian portion is estimated using Kalman filtering, while the nonlinear portion, i.e., the data association portion, is calculated using particle filtering. Simulation results indicate that incorporating radial velocity significantly enhances tracking accuracy compared to the azimuth-only tracking algorithms. Additionally, in scenarios where target trajectories intersect, the combined azimuth and radial velocity RBMCDA algorithm achieves higher tracking accuracy than the azimuth-only RBMCDA algorithm. This demonstrates that the enhanced RBMCDA algorithm combining azimuth and radial velocity significantly enhances tracking performance, enhances target separability, and possesses strong stealth tracking capabilities and high application value.

ACKNOWLEDGEMENTS

A This work is supported by the National Natural Science Foundation of China under Grant Nos. 62271162.

REFERENCES

1. Nusrat A, Li Y, Cheng C, et al. Underwater Bearing Only Tracking Using Optimal Observer Maneuver Strategies[J]. Journal of Marine Science and Engineering. 2022, 10(5): 576.
2. Kalman R E. A new approach to linear filtering and prediction problems[J]. Journal of Basic Engineering, 1960, 82(1): 35-45.
3. WU Y, HU D, WU M, et al. Unscented Kalman filtering for additive noise case: Augmented versus nonaugmented[J]. IEEE Signal Processing Letters, 2005, 12(5): 357-60.
4. Moghaddasi S S, Faraji N. A hybrid algorithm based on particle filter and genetic algorithm for target tracking[J]. Expert Systems with Applications, 2020, 147(8): 113-88.
5. FORTMANN T E, BAR-SHALOM Y, SCHEFFE M. Multi-target tracking using joint probabilistic data association[C]. Proceedings of the IEEE Conference on Decision and Control, 1980.p.807-12.
6. LIU J, WANG Z, XU M. A Kalman Estimation Based Rao-Blackwellized Particle Filtering for Radar Tracking[J]. IEEE Access, 2017, 5: 8162-74.
7. SäRKKä S, VEHTARI A, LAMPINEN J. Rao-Blackwellized particle filter for multiple target tracking[J]. Information Fusion, 2007, 8(1): 2-15.

PROCEEDINGS of the
16th International Conference on Theoretical and Computational Acoustics

August 18 to 22, 2025 BEXCO, Busan Korea

Broadband Target Direction of Arrival Estimation Based on a Single Pressure-difference Vector Hydrophone

Zehua WANG[1]; Junjie SHI[2]; Minshuai LIANG; Dajun SUN
[1] National Key Laboratory of underwater Acoustic Technology, Harbin Engineering University, Harbin, China
[2] Key Laboratory of Marine Information Acquisition and Security (Harbin Engineering University), Ministry of Industry and Information Technology, Harbin, China
[3] College of Underwater Acoustic Engineering, Harbin Engineering University, Harbin, China

ABSTRACT
A pressure-difference vector hydrophone is capable of acquiring both sound pressure and particle velocity information within the sound field. The particle velocity is typically derived through a differential approximation between axial elements. The direction of arrival (DOA) estimation results obtained using the average intensity method exhibit periodic bias, as particle velocity is derived through an axial element difference approximation, which is subject to limitations imposed by element spacing and signal frequency in practical applications. This paper presents a broadband DOA estimation method based on a single pressure-difference vector hydrophone. The proposed method reconstructs equivalent sound intensity from the axial sound pressure and particle velocity. Additionally, it accurately determines the broadband DOA by utilizing the phase differences between axial elements across various frequencies, which concurrently reduces the impact of element spacing and signal frequency constraints. Consequently, this approach not only avoids DOA estimation deviations but also relaxes engineering constraints on element spacing and extends the operational frequency range of the pressure-difference vector hydrophone. Furthermore, the method demonstrates tolerance to the effects of high-frequency strong interference without the necessity of prior removal. Numerical simulations and sea trial data have demonstrated the feasibility and effectiveness of the proposed method.

Keywords: DOA estimation, Broadband target, Pressure-difference vector hydrophone

1. INTRODUCTION

As underwater autonomous mobile observation platforms become more miniaturized and intelligent, it has become increasingly significant to accomplish tasks like underwater target detection and parameter estimation on small size platforms (1-4). The vector hydrophone has become a crucial area of research for underwater target detection due to its compact size, convenient deployment, and low energy consumption. The vector hydrophone can simultaneously acquire information such as sound field pressure, pressure gradients, or particle velocity. Through the joint processing of this information, it can realize the unambiguous direction-of-arrival (DOA) estimation, demonstrating the potential of deployment on small platforms (5-7).

Vector hydrophones are primarily categorized into co-vibration type and pressure-difference type. Co-vibration vector hydrophones maintain a neutrally buoyant state in water via an elastic suspension system, directly measuring particle velocity vector information in the sound field. However, their complex structure limits installation flexibility (8). Pressure-difference vector hydrophones consist of multiple sound pressure sensors. They indirectly obtain vector information by deriving sound pressure gradients from pressure differences, offering a simpler structure and installation. However, their DOA estimation performance is affected by the spacing between individual elements. The average sound intensity method for DOA estimation was initially proposed for the co-vibration vector hydrophone. When applied to a pressure-difference vector hydrophone, it exhibits a periodic bias that increases with frequency (9-12). This paper introduces an improved sound intensity method for broadband DOA estimation. This method reconstructs

[1] wangzehuamail@163.com
[2] shijunjie@hrbeu.edu.cn

equivalent sound intensity using axial sound pressure and pressure gradients, and then accurately solves for broadband DOA across various frequencies by leveraging the axial element phase difference. This significantly expands the operational frequency range of the pressure-difference vector hydrophone.

Section 1 outlines the potential of vector hydrophones for small platform applications and explains the limitations of applying the average sound intensity method for DOA estimation with a single pressure-difference vector hydrophone. Section 2 details the principle of the improved sound intensity method for broadband DOA estimation. Section 3 provides a comprehensive description of the improved method's unbiased advantage through numerical simulations. Section 4 presents sea trial results demonstrating the improved method's robustness against strong interference. Finally, Section 5 concludes the paper.

2. PRINCIPLE OF DOA ESTIMATION BASED ON A SINGLE PRESSURE-DIFFERENCE VECTOR HYDROPHONE

2.1 Average sound intensity method for DOA estimation

The pressure-difference vector hydrophone is usually composed of four pressure sensors with the same acoustic performance in the plane. The distribution of each channel is shown in Fig. 1. The distance between the center point and the element is d, and the angle between the signal incident direction is θ. Assuming that the signal at the center point is $p_0 e^{j\omega t}$. The received signal by the four channels can be denoted as,

$$\begin{cases} p_1 = p_0 e^{j(\omega t+\Phi_1)} \\ p_2 = p_0 e^{j(\omega t+\Phi_2)} \\ p_3 = p_0 e^{j(\omega t+\Phi_3)} \\ p_4 = p_0 e^{j(\omega t+\Phi_4)} \end{cases}, \tag{1}$$

where $\Phi_1 = -\Phi_3 = kd\cos\theta$, $\Phi_2 = -\Phi_4 = kd\sin\theta$ are the phase difference compensation of the sound field at the relative center point of each channel, wavenumber is $k = \omega / c$, p_i is pressure of channel i.

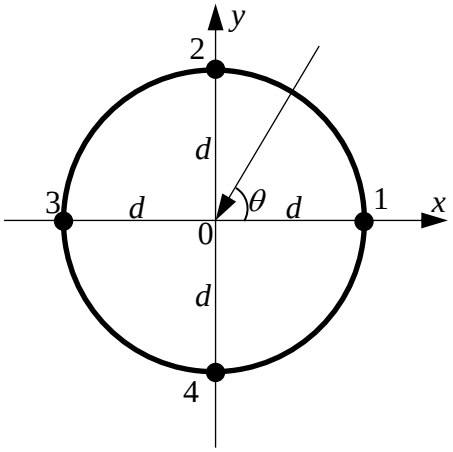


Figure 1. The two-dimensional pressure-difference vector hydrophone.

Co-vibration vector hydrophones typically use the average sound intensity method to measure horizontal components, thereby determining the direction of sound intensity, which corresponds to the target direction. Without loss of generality, assuming the amplitude of the center point signal of a pressure-difference vector hydrophone is 1, and neglecting noise, the sound pressures in the x and y directions, p_x and p_y, are respectively,

$$p_x = \frac{p_1 + p_3}{2} = \frac{p_0}{2}\left(e^{j\omega t}e^{j\Phi_1} + e^{j\omega t}e^{j\Phi_3}\right) = \cos\left(kd\cos\theta\right) p_0 e^{j\omega t}, \tag{2}$$

$$p_y = \frac{p_2 + p_4}{2} = \frac{p_0}{2}\left(e^{j\omega t}e^{j\Phi_2} + e^{j\omega t}e^{j\Phi_4}\right) = \cos\left(kd\sin\theta\right) p_0 e^{j\omega t}. \tag{3}$$

The sound pressure at the center point is the average of the x and y directions,

$$p = \frac{p_x + p_y}{2} = \frac{\cos\left(kd\cos\theta\right) + \cos\left(kd\sin\theta\right)}{2} p_0 e^{j\omega t}. \tag{4}$$

Similarly, the sound pressure gradients in the x and y directions, v_x and v_y, can be equivalently expressed as particle velocities:

$$v_x = \frac{p_1 - p_3}{2} = \frac{p_0}{2}\left(e^{j\omega t}e^{j\Phi_1} - e^{j\omega t}e^{j\Phi_3}\right) = j\sin\left(kd\cos\theta\right) p_0 e^{j\omega t}, \tag{5}$$

$$v_y = \frac{p_2 - p_4}{2} = \frac{p_0}{2}\left(e^{j\omega t}e^{j\Phi_2} - e^{j\omega t}e^{j\Phi_4}\right) = j\sin\left(kd\sin\theta\right) p_0 e^{j\omega t}. \tag{6}$$

Define I_x and I_y as the sound intensities in the x and y directions, respectively. After compensating for the phase of the particle velocity, the results are,

$$I_x = \overline{p(v_x e^{-j\pi/2})^*} = \frac{p_0^2}{2}\sin\left(kd\cos\theta\right)\left(\cos\left(kd\cos\theta\right) + \cos\left(kd\sin\theta\right)\right), \tag{7}$$

$$I_y = \overline{p(v_y e^{-j\pi/2})^*} = \frac{p_0^2}{2}\sin\left(kd\sin\theta\right)\left(\cos\left(kd\cos\theta\right) + \cos\left(kd\sin\theta\right)\right). \tag{8}$$

When the $kd \ll 1$ with $kd\sin\theta$ and $kd\cos\theta$ approaching 0, by utilizing the equivalent infinitesimal $\lim_{x\to 0}\sin x = x$, we have the ratio of sound intensity components,

$$\frac{I_y}{I_x} = \frac{\sin\left(kd\sin\theta\right)}{\sin\left(kd\cos\theta\right)} \overset{kd\ll 1}{\approx} \frac{kd\sin\theta}{kd\cos\theta} = \tan\theta. \tag{9}$$

Then the signal direction can be approximated as,

$$\theta \approx \tan^{-1}\left(\frac{I_y}{I_x}\right). \tag{10}$$

However, for the pressure-difference vector hydrophone, there is a certain spacing between elements. This leads to the average sound intensity method, originally proposed for the co-vibration vector hydrophone, yielding inaccurate DOA estimation results when applied to the pressure-difference vector hydrophone. The average sound intensity method is valid under the condition of $kd \ll 1$, meaning that accurate estimation is only possible for low-frequency signals or extremely small element spacing. Furthermore, the error of this method varies periodically with the DOA of the target.

2.2 Improved sound intensity method for DOA estimation

Here, we define new equivalent sound intensities in the x and y directions as $I_x' = \overline{p_y v_x^*}$ and $I_y' = \overline{p_x v_y^*}$ respectively, and the equivalent sound energy as $L' = \overline{p_x p_y^*}$. By utilizing these newly defined axial components, the approximation in DOA estimation can be avoided, theoretically eliminating periodic bias. The phase terms in each channel are retained here, and the results are directly given as follows.

$$I_x' = \overline{p_y(v_x e^{-j\pi/2})^*} = p_0^2 \sin\left(\frac{\Phi_1 - \Phi_3}{2}\right)\cos\left(\frac{\Phi_2 - \Phi_4}{2}\right), \tag{11}$$

$$I_y' = \overline{p_x(v_y e^{-j\pi/2})^*} = p_0^2 \cos\left(\frac{\Phi_1 - \Phi_3}{2}\right)\sin\left(\frac{\Phi_2 - \Phi_4}{2}\right), \tag{12}$$

$$L' = \overline{p_x p_y^*} = p_0^2 \cos\left(\frac{\Phi_1 - \Phi_3}{2}\right)\cos\left(\frac{\Phi_2 - \Phi_4}{2}\right), \tag{13}$$

where Φ_i is the phase difference of the element i relative to the center point. According to the above

formulas, the target DOA can be estimated as,

$$\frac{I'_x}{L'} = \tan\left(\frac{\Phi_1 - \Phi_3}{2}\right) = \tan\left(kd\cos\theta\right), \tag{14}$$

$$\frac{I'_y}{L'} = \tan\left(\frac{\Phi_2 - \Phi_4}{2}\right) = \tan\left(kd\sin\theta\right), \tag{15}$$

$$\theta = \tan^{-1}\left(\frac{\tan^{-1}\left(I'_y / L'\right)}{\tan^{-1}\left(I'_x / L'\right)}\right). \tag{16}$$

It can be seen that the axial sound intensity ratio in the above equation is related to the phase difference of the axial elements. Without loss of generality, the phase difference will be rewritten in the form of wavenumber multiplied by distance below. By solving a system of linear equations, the estimated DOA of the target can be obtained. Here, the axial phase differences in the two equations above are expressed using the wavenumber components k_x and k_y and the element position difference components Δx and Δy. The system of equations is as follows,

$$\begin{cases} k_x\Delta x_{13} + k_y\Delta y_{13} = 2\tan^{-1}\left(\dfrac{I'_x}{L'}\right) \\ k_x\Delta x_{24} + k_y\Delta y_{24} = 2\tan^{-1}\left(\dfrac{I'_y}{L'}\right) \end{cases}. \tag{17}$$

The above system of equations can be written and solved in matrix form,

$$\begin{bmatrix} \Delta x_{13} & \Delta y_{13} \\ \Delta x_{24} & \Delta y_{24} \end{bmatrix}\begin{bmatrix} k_x \\ k_y \end{bmatrix} = \begin{bmatrix} 2\tan^{-1}\left(\dfrac{I'_x}{L'}\right) \\ 2\tan^{-1}\left(\dfrac{I'_y}{L'}\right) \end{bmatrix}. \tag{18}$$

where k_x represents the target signal wavenumber component in the x direction, k_y represents the target signal wavenumber component in the y direction. Δx and Δy are respectively the position differences of the axial elements in the x and y directions. The DOA can be expressed as $\theta = \tan^{-1}\left(k_y / k_x\right)$.

Based on the DOA estimation principle for single-frequency signals described above, an improved sound intensity method for broadband DOA estimation is proposed. This method obtains a single DOA estimate by decoupling direction information with frequency. The equation is rewritten into a broadband form for all frequency points as follows,

$$\begin{bmatrix} f_{low} \\ \cdots \\ f_{high} \end{bmatrix}\begin{bmatrix} h_x & h_y \end{bmatrix}\begin{bmatrix} \Delta x_{13} & \Delta x_{24} \\ \Delta y_{13} & \Delta y_{24} \end{bmatrix} = \begin{bmatrix} b_{13f_{low}} & b_{24f_{low}} \\ \cdots & \cdots \\ b_{13f_{high}} & b_{24f_{high}} \end{bmatrix}, \tag{19}$$

where $h = k / f$, $b_{13f} = 2\tan^{-1}\left(I'_x / L'\right)\big|_f$, and $b_{24f} = 2\tan^{-1}\left(I'_y / L'\right)\big|_f$. Left-multiplying both sides of the equation by the frequency row vector $\begin{bmatrix} f_{low} & \cdots & f_{high} \end{bmatrix}$,

$$\left(\sum f^2\right)\begin{bmatrix} h_x & h_y \end{bmatrix}\begin{bmatrix} \Delta x_{13} & \Delta x_{24} \\ \Delta y_{13} & \Delta y_{24} \end{bmatrix} = \begin{bmatrix} \sum f b_{13f} & \sum f b_{24f} \end{bmatrix}. \tag{20}$$

Rewriting in column vector representation,

$$\left(\sum f^2\right)\begin{bmatrix}\Delta x_{13} & \Delta y_{13}\\ \Delta x_{24} & \Delta y_{24}\end{bmatrix}\begin{bmatrix}h_x\\ h_y\end{bmatrix}=\begin{bmatrix}\sum f b_{13f}\\ \sum f b_{24f}\end{bmatrix}. \tag{21}$$

By solving for h_x and h_y in the system of equations, the DOA can then be obtained as $\theta=\tan^{-1}\left(h_y/h_x\right)$.

3. NUMERICAL SIMULATION

Based on the single-frequency and broadband DOA estimation theories described above, numerical simulations are conducted to verify the performance of the pressure-difference vector hydrophone. The simulated frequencies include single-frequency 50 Hz, single-frequency 320 Hz, and broadband 50-320 Hz. The spacing between elements of the pressure-difference vector hydrophone is 0.9m, meaning $d=0.45$ m. Under ideal noise-free conditions, the comparative analysis of the DOA estimation error is shown in Fig. 2.

From the comparison in the figure, it can be observed that the DOA estimation error of the average sound intensity method varies periodically, whereas the theoretical error value for the improved sound intensity method is zero. Furthermore, it is evident that for the single-frequency 50 Hz signal, the error of the average sound intensity method is very small, but this error gradually increases as the frequency increases. For the broadband signal, compared with the single-frequency 320 Hz signal, the error of the average sound intensity method is reduced due to fusion and averaging.

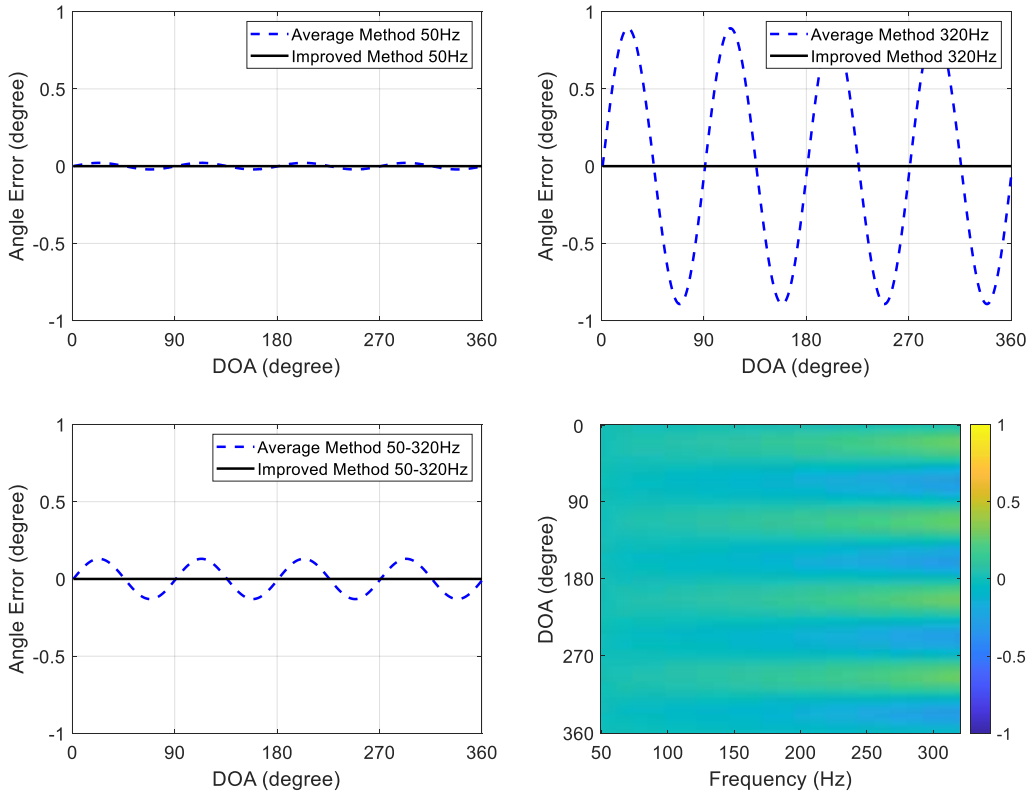


Figure 2. Comparison of DOA estimation angle error for the two methods. (a) Single-frequency 50 Hz; (b) Single-frequency 320 Hz; (c) Broadband 50-320 Hz; (d) DOA estimation error at various frequencies from 50-320 Hz.

4. TRIAL DATA ANALYSIS

The sea trial data was collected in the Yellow Sea, China, on June 28, 2019. A pressure-difference vector hydrophone receiver was deployed at a depth of approximately 50 m, sampling ambient ship radiated noise signals at a frequency of 3 kHz. The broadband continuous spectrum radiated noise from surface ships, acting as the passive sound source, can be used to estimate the DOA using the pressure-difference vector hydrophone.

During the experiment, a ship designated "Jingfeng3" passed by, sailing from northeast to southwest. The relative spatial relationship between the passing ship and the receiver is shown in Fig. 3. The time-frequency spectrum of the signal received by a single element of the pressure-difference vector hydrophone

is presented in Fig. 4. This figure reveals broadband interference fringes and several line spectra, as well as a strong interference appearing at 311 Hz. This strong interference might be caused by the receiver system and can be regarded as self-noise interference.

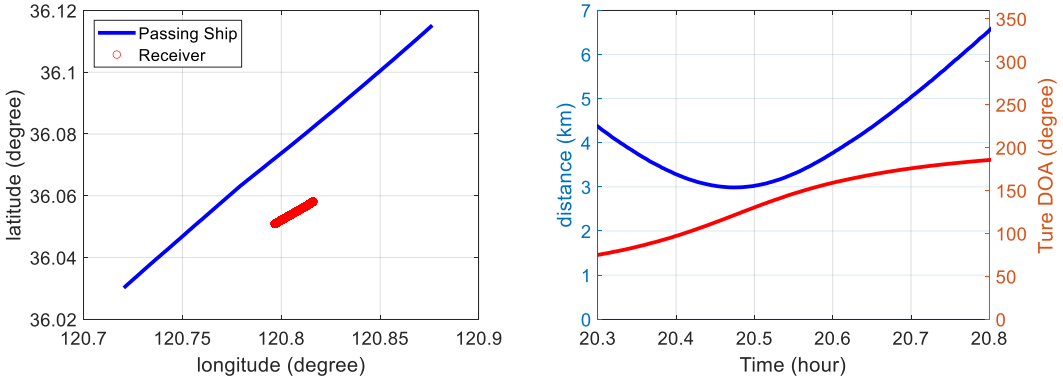


Figure 3. The relative spatial relationship between the passing ship and the receiver. (a) Relative position; (b) Relative range and relative DOA.

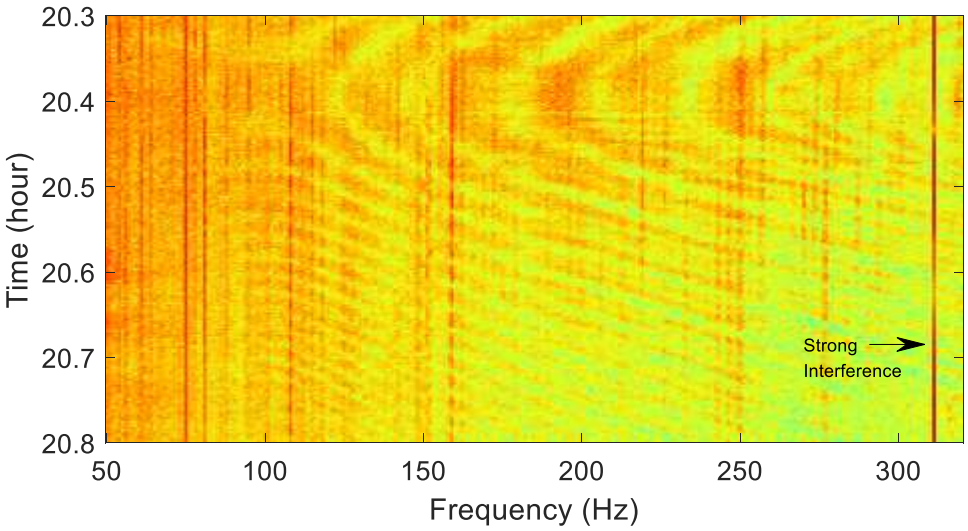
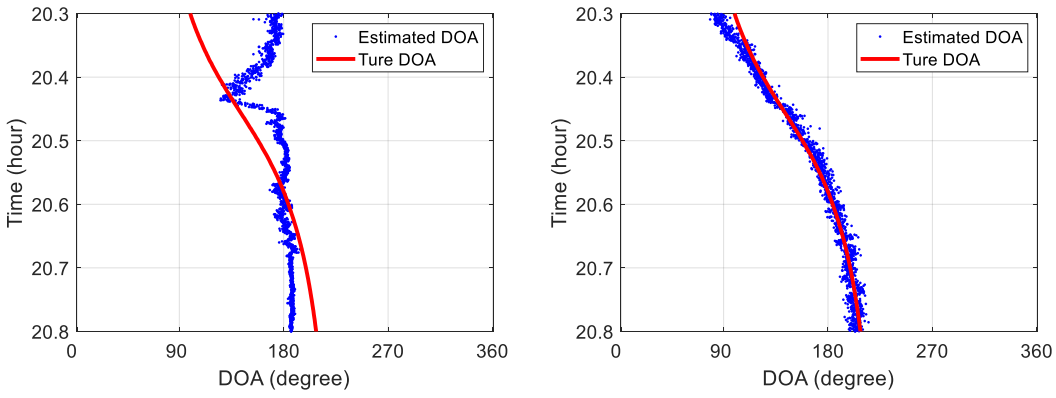


Figure 4. The time-frequency spectrum of the passing ship.

The DOA estimation results for the ship radiated broadband noise received by the pressure-difference vector hydrophone are shown in Fig. 5. When analyzing the full-band signal, the average sound intensity method is severely affected by strong high-frequency interference, causing the estimated bearing to significantly deviate from the true DOA value. After removing the strong interference, the DOA estimation results become more accurate, although there are still acceptable errors. However, the improved sound intensity method consistently maintains estimation accuracy, even in the presence of strong high-frequency interference. This is because it is theoretically unaffected by signal frequency and integrates direction information from various frequencies, allowing it to tolerate the DOA error caused by the strong high-frequency interference.



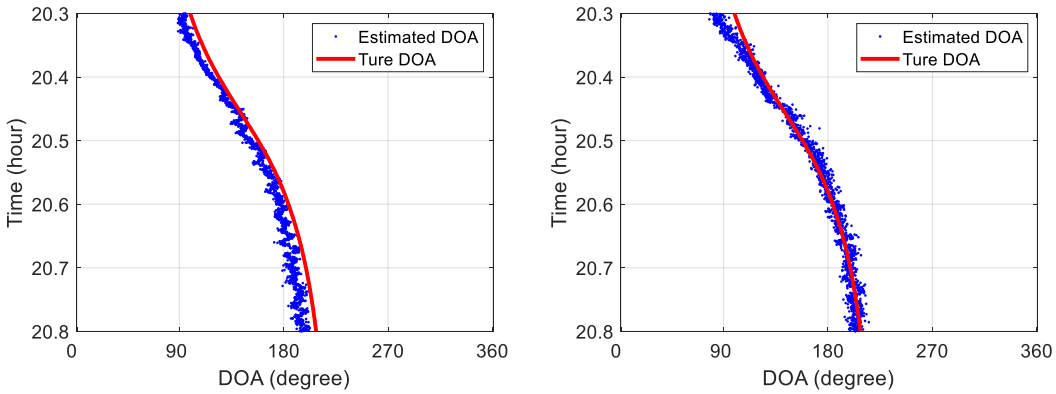


Figure 5. Comparison of DOA estimation results between the average sound intensity method and the improved sound intensity method. (a) Average method with full band; (b) Improved method with full band; (c) Average method after removing 311 Hz strong interference; (d) Improved method after removing 311 Hz strong interference.

5. CONCLUSIONS

This paper proposes a broadband DOA estimation method based on a single pressure-difference vector hydrophone. This method can theoretically avoid the periodic bias in high-frequency signals caused by element spacing, thereby relaxing engineering constraints on element spacing and expanding the operational frequency range of the pressure-difference vector hydrophone. Furthermore, compared to traditional methods, this approach can tolerate the effects of strong high-frequency interference without the need for prior removal of strong interference frequencies, therefore enhancing the robustness of DOA estimation. Numerical simulations and sea trial data have demonstrated the feasibility and effectiveness of the proposed method.

REFERENCES

1. Sun D, Lu M, Mei J, et al. Generalized Radon transform approach to target motion parameter estimation using a stationary underwater vector hydrophone. J. Acoust. Soc. Am., 2021; 150(2): 952-968.
2. Sun D, Ding J, Zheng C, et al. Array geometry calibration for underwater compact arrays. Applied Acoustics, 2019; 145: 374-384.
3. Li X, Sun D, Cao Z. Generation Mechanism of Acoustic Doppler Velocity Measurement Bias. IEEE Journal of Oceanic Engineering, 2024.
4 Wang M, Zheng C, Sun D, et al. Sonar micro-Doppler effect induced by ship oscillatory motion in waves: Phenomenon, model, and experimentation. JASA Exp. Lett., 2025; 5(2).
5. Wilson O B. Acoustic particle velocity sensors: Design, performance, and applications[J]. 1996.
6. Lan H, White P R, Li N, et al. Coherently averaged power spectral estimate for signal detection. Signal Processing, 2020; 169: 107414.
7. Tichavsky P, Wong K T, Zoltowski M D. Near-field/far-field azimuth and elevation angle estimation using a single vector hydrophone[J]. IEEE Transactions on Signal Processing, 2001, 49(11): 2498-2510.
8. Kumar B, Kumar A, Bahl R. Comparative study of PP and PU based acoustic vector sensors for underwater application[C]//2018 OCEANS-MTS/IEEE Kobe Techno-Oceans (OTO). IEEE, 2018: 1-9.
9. Shi J, Sun D. An improved DOA method for single acoustic vector sensor[C]//2009 5th International Conference on Wireless Communications, Networking and Mobile Computing. IEEE, 2009: 1-5.
10. Cheng B, Yang S. Quantitative Analysis of Error of Estimated Bearing of Acoustic Source Location with TDPG (Two-Dimensional Pressure-Gradient) Vector Hydrophone [J] Journal of Northwestern Polytechnical University, 2006, (06): 800-803.
11. Zhang G, Liu M, Shen N, et al. The development of the differential MEMS vector hydrophone[J]. Sensors, 2017, 17(6): 1332.
12. Yu Y, Wang Y F, Wang Y Z, et al. A Study of Orientation Estimation Method Based on Two-Dimensional Differential Pressure Vector Hydrophone[C]//2023 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC). IEEE, 2023: 1-5.

DAILY PROGRAM

AUGUST 19 (TUE)

09:00 – 10:20	Underwater Acoustic Source Localization and Tracking	324 (3F)
Chair	In-Jee Jung (Korea Research Institute of Standards and Science)	
09:00	Range localization of a broadband moving source using the field differencing method	
	Daehwan Kim, Sung-Hoon Byun	
09:20	Underwater Multi-node Fusion Tracking Method with Integrated Target Motion Characteristics	
	Pengchao Liu, Xinyi Guo	
09:40	Localization of whale's call using a single underwater acoustic vector sensor	
	Hefeng Dong, Guosong Zhang, Victor Espinosa Rosello, Isabel Pérez Arjona, Josephine Nell Schulze	
10:00	Multi-range sparse Bayesian learning for broadband modal phase velocity estimation and shallow-water geoacoustic inversion	
	Shanru Lin, Haiqiang Niu, Zhenglin Li, Yonggang Guo	

PROCEEDINGS of the
16th International Conference on Theoretical and Computational Acoustics

August 18 to 22, 2025 BEXCO, Busan Korea

Range localization of a broadband moving source using the field differencing method

Daehwan Kim[1], Sung-Hoon Byun[2]
[1]Korea Research Institute of Ships & Ocean engineering
[2]Korea Research Institute of Ships & Ocean engineering

ABSTRACT

The field differencing method, based on the cross-correlation of pressure fields at adjacent ranges has been proposed as a model-independent approach for passive ranging using a single hydrophone. While this method provides robust velocity estimation for narrowband sources, its application to broadband sources introduces ambiguities due to multipath propagation. In this study, a two-dimensional discrete Fourer transform (2D-DFT) of the cross-correlated field is applied to estimate the true source velocity. Simulation results demonstrates that the method identifies the true source velocity in complex environments with multipath interference.

[1] dkim@kriso.re.kr
[2] byunsh@kriso.re.kr

PROCEEDINGS of the
16th International Conference on Theoretical and Computational Acoustics

August 18 to 22, 2025 BEXCO, Busan Korea

Underwater Multi-node Fusion Tracking Method with Integrated Target Motion Characteristics

Pengchao Liu[1,2] ; Xinyi Guo[1,2]
1The Key Laboratory of Underwater Environment, Chinese Academy of Sciences, Beijing, 100190, China
2Institute of Acoustics, Chinese Academy of Sciences, Beijing, 100190, China

ABSTRACT

To address the challenges of association ambiguity and performance degradation arising from dispersed and asynchronous observations in multi-node underwater target tracking, this paper proposes a novel joint tracking method based on the trajectory features of the target's elevation angle. The method begins by parameterizing the time series of elevation angles observed at each node, from which a motion feature vector characterizing the target's Closest Point of Approach (CPA) process is extracted. Subsequently, a joint association and fusion framework is established within this multi-dimensional feature space to achieve feature-level fusion of the multi-node observations. Simulation results demonstrate that the proposed method significantly enhances tracking robustness and localization accuracy, particularly in ambiguous scenarios such as target crossovers.

Keywords: Underwater Target Tracking, Multi-Node Fusion, Feature-Level Fusion, MHT

1. INTRODUCTION

Multi-node collaborative sensing is a critical technique for enhancing underwater target tracking capabilities [1]. However, the complexity of the underwater acoustic channel often leads to challenges such as temporal a synchrony, spatial dispersion, data loss, and significant clutter interference in observations from different nodes. Traditional tracking methods, such as the Joint Probabilistic Data Association (JPDA) filter [2, 3], typically fuse raw measurement data at the measurement level. Such approaches are highly dependent on the target's instantaneous geometric position, and their performance degrades severely in scenarios with dense clutter or multiple target crossovers, where association ambiguity increases sharply, often leading to track loss or identity swapping.

To overcome this limitation, this paper introduces a new feature-level fusion tracking framework. Instead of directly processing raw measurements, this approach first models the target's motion process within each node to extract key features that describe its kinematic trends. Data association and fusion are then performed in this feature space. Indeed, leveraging motion features to refine and augment classical association algorithms is a significant research direction in multi-target tracking for mitigating uncertainty [4]. This "comprehend-then-fuse" strategy aims to exploit a richer information dimension to improve association accuracy and robustness, thereby achieving stable and precise tracking of underwater targets in complex environments.

2. Methodology

This section details the overall framework and core principles of the proposed method. The method is predicated on the assumption that the target transits past the vicinity of a sensor node following a trajectory that approximates a constant velocity, straight-line path, i.e., a Closest Point of Approach (CPA) scenario.

2.1 Feature Extraction

The input data for this method are derived from the beamforming results of each sensor node (array). At each time step, processing the array's received signals yields the elevation angle $\theta(t)$ of the incoming target signal. The time series of these elevation angles, $\theta(t)$, forms the basis for all subsequent processing.

During a CPA event, the elevation angle's temporal evolution follows a distinct curve that can be accurately described by a physical model. We employ a three-parameter model to fit and characterize this elevation angle trajectory.

$$\theta(t) = \arctan\left(\frac{h}{\sqrt{d_{\min^2} + v^2(t - t_c)^2}}\right) \tag{1}$$

This model defines a unique elevation trajectory using three physically meaningful parameters:

- **Minimum Horizontal Range (d_{min})**: The closest horizontal distance between the target and the sensor during the transit.
- **Speed (v)**: The target's horizontal velocity.
- **Center Time (t_0)**: The time instant when the target reaches the CPA point.

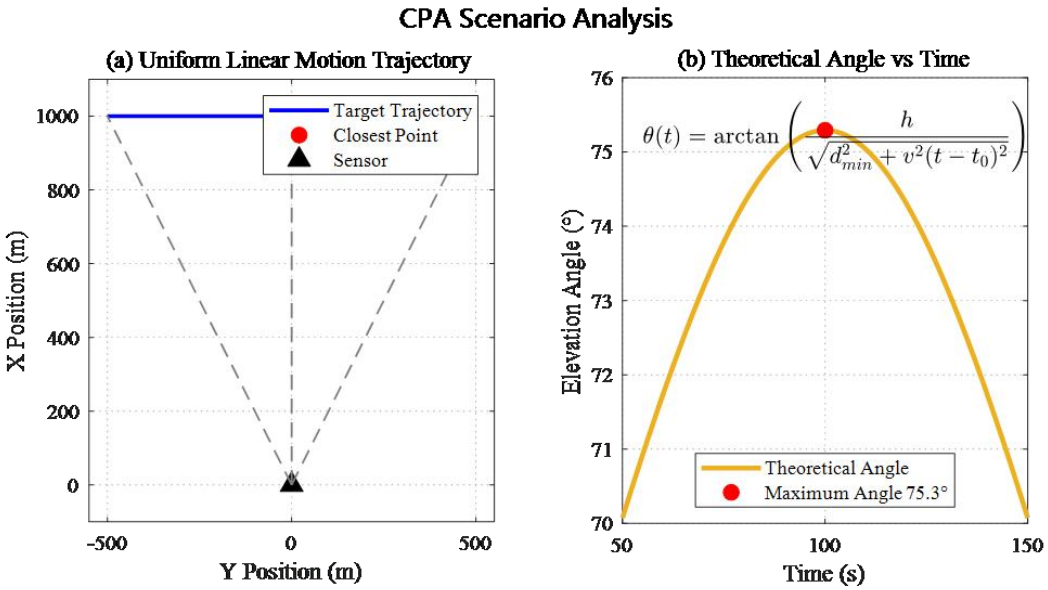


Figure 1 - Illustration of the three-parameter model for a CPA event.

By fitting this model to the elevation angle sequence observed at each node, we can extract a feature vector $f = [d_{min}, v, t_0]T$ for each potential target track.

To extract the feature vector f from an observation sequence, this work utilizes the Levenberg-Marquardt (L-M) algorithm, a non-linear least squares method widely recognized for its robust convergence properties. The initialization of the algorithm is critical for convergence. We adopt a data-driven strategy: the initial value for t_0 is set to the time of the maximum observed elevation angle; the initial d_{min} is estimated from the maximum elevation angle and the known sensor depth h; and the initial speed v is set to a typical empirical value (e.g., 5 m/s).

2.2 Feature Fusion

After obtaining the local track feature vectors from each node, the central fusion task shifts from associating individual, instantaneous measurements to associating these feature vectors, which represent entire motion segments. The fusion process is implemented within a Multiple Hypothesis Tracking (MHT) framework. Studies have shown that MHT generally offers superior performance compared to JPDA in dense clutter and multi-target scenarios, particularly in handling complex association problems [5].

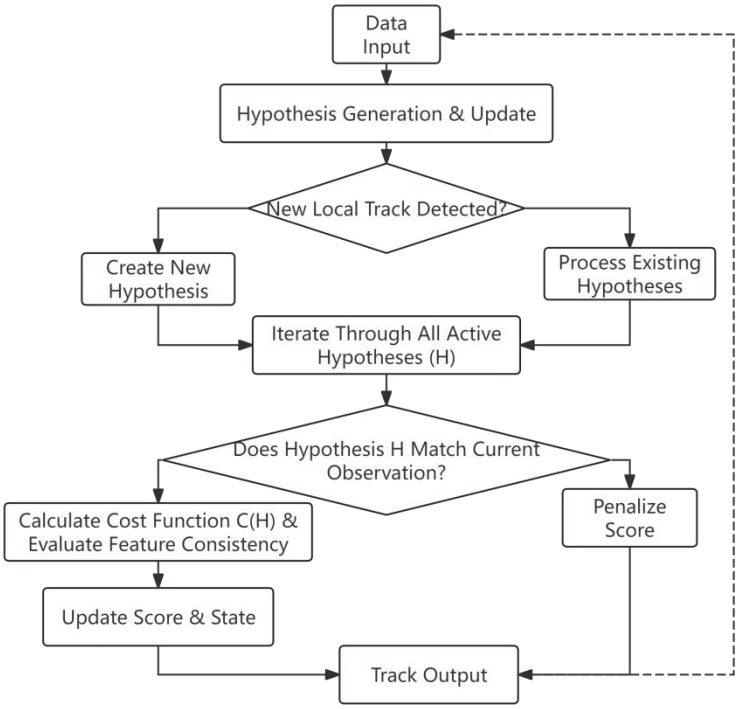


Figure 2 - Flowchart of the MHT-based feature fusion algorithm.

2.2.1 Fusion Cost Function

Within the MHT framework, the cost function C(H) for evaluating the plausibility of a hypothesis H (i.e., a set of local track features from different nodes belonging to the same global target) is defined as:

$$C(H) = w_v \sigma_v^{'} + w_R \sigma_R^{'} + w_t \sigma_t^{'} + w_E E \tag{2}$$

where σ_v , σ_R, and σ_t are the normalized standard deviations of the speed, minimum range, and center time features within the set, respectively. These terms measure the consistency of the motion features. *E* represents the geometric localization residual, calculated after solving for the target's position using the range information from the feature set. w_v , w_R , w_t , w_E are the corresponding weighting coefficients. A lower value of *C(H)* indicates higher internal consistency for the hypothesis.

3. Simulation

To validate the effectiveness of the proposed method, we designed a simulation involving a multi-target crossover scenario and compared its performance against a classical JPDA algorithm.

Scenario Layout: Three passive sensor nodes were arranged in an equilateral triangle with 10 km sides.

Target Trajectories: Two targets were simulated with different speeds and initial positions, set to have their trajectories cross midway through the simulation.

Simulation Parameters: The total duration was 3600 s with a sampling interval of 10 s. The observation noise (standard deviation of elevation angle) was varied among 0.5°, 1.0°, 1.5°, and 2.0° . The probability of detection was set to 95%. Critically, Node 3 experienced complete data loss during the periods 800s–900s and 1600s–1800s.

The baseline for comparison was a JPDA filter combined with an Extended Kalman Filter (EKF). Its state vector was defined as $[x, x^{'}, y, y^{'}]^T$, and it employed a constant velocity (CV) motion model. This configuration is representative of conventional measurement-level fusion techniques

3.1 Tracking trajectory comparison

Figure 3 illustrates the tracking trajectories produced by both algorithms. The proposed method's estimated tracks (left panel) exhibit a significantly better fit to the ground truth compared to the JPDA-EKF method. The JPDA tracks (right panel) show noticeable deviation and jitter, especially near the target crossover region, high- lighting its instability when association ambiguity is high.

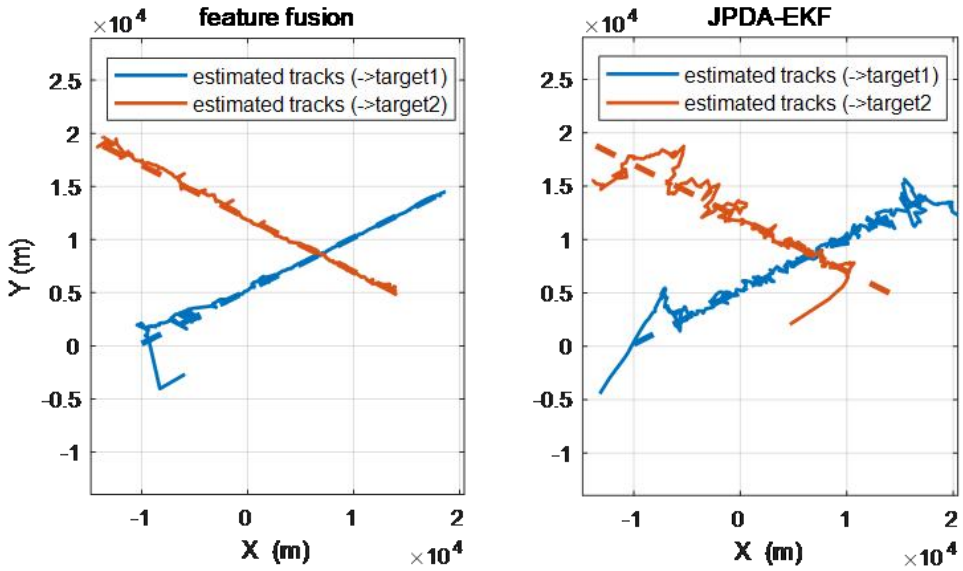


Figure 3 - Tracking trajectories from the proposed feature fusion method (left) and the JPDA-EKF method (right). Dashed lines are ground truth; solid lines with markers are estimated tracks.

To quantify performance, the Root Mean Square Error (RMSE) was calculated for each target. Table 1 clearly shows that the proposed method achieves significantly higher tracking accuracy for both targets across all tested noise levels.

Table 1 – RMSE Statistics for Both Algorithms (in meters)

Obs. Noise (°)	Algorithm	Target 1 RMSE (m)	Target 2 RMSE (m)
0.5	Feature Fusion	578.6	384.7
	JPDA-EKF	943.6	2407.7
1.0	Feature Fusion	876.9	432.6
	JPDA-EKF	1418.5	3369.3
1.5	Feature Fusion	1412.0	1100.3
	JPDA-EKF	3668.6	3829.6
2.0	Feature Fusion	1453.0	1257.9
	JPDA-EKF	2537.3	3640.0

Figure 4 further details the instantaneous localization error over time (for 1° observation noise). The yellow- highlighted region (approx. 1800s–2100s) marks the period of closest spatial proximity between the targets, while the gray regions (800s–900s, 1600s–1800s) denote data loss from Node 3.

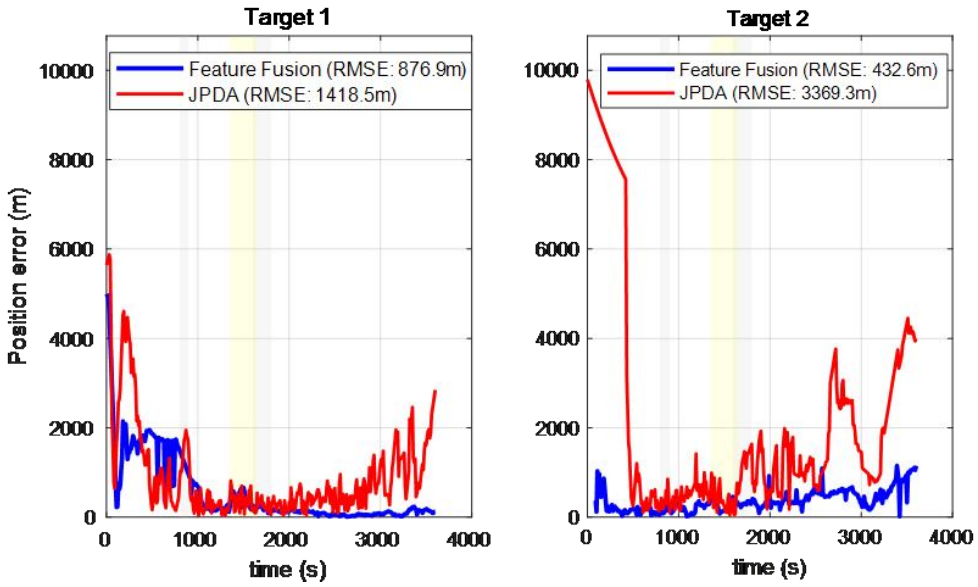


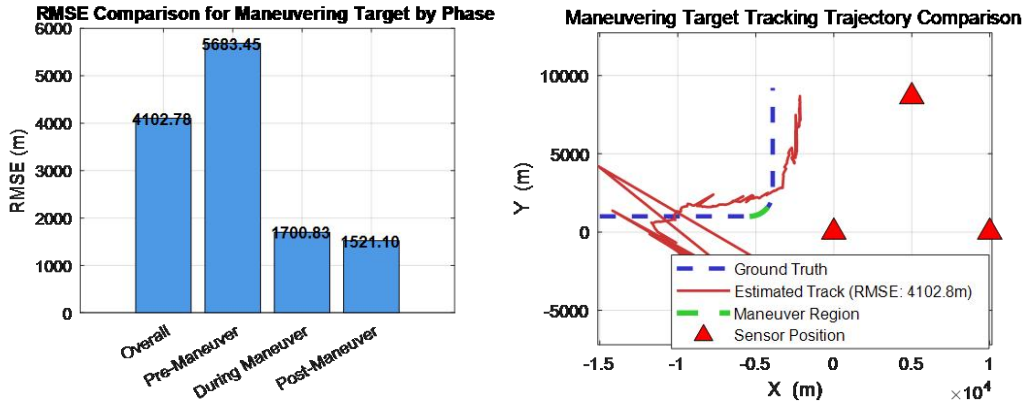
Figure 4 – Instantaneous localization error for Target 1 (left) and Target 2 (right). The proposed method (blue) is compared against JPDA-EKF (red).

3.2 Analysis of Node Data Loss

During the data loss periods for Node 3, the JPDA algorithm's error shows significant jumps due to its reliance on instantaneous measurements. In contrast, our method, which associates feature vectors representing entire motion segments, maintains relatively stable tracking by leveraging information from the remaining two nodes and the inherent history in the feature model, demonstrating superior fault tolerance.

3.3 Analysis of Maneuvering Targets

To investigate performance against maneuvering targets, a scenario with a single target executing a turn was designed. The target performed a 0.3°/s coordinated turn for 300s, starting at t=1200s. As shown in Figure 5, the tracking performance degraded significantly, with the estimated track deviating substantially from the ground truth. This clearly indicates the limitation of the current model for non-linear motion patterns



(a) Tracking trajectory for maneuvering target. (b) RMSE for maneuvering target.

Figure 5 – Tracking trajectory (left) and RMSE (right) for a maneuvering target.

4. Discussion

The core advantage of the proposed method lies in elevating the tracking process from associating unstable, low-information instantaneous measurements to associating stable, information-rich motion features. While a traditional JPDA filter may confront multiple ambiguous measurement points within an association gate, our method associates local tracks endowed with an "identity"—a specific speed, CPA time, and range. This association in a higher-dimensional information space allows for more effective rejection of clutter and discrimination between targets, especially in ambiguous scenarios like crossover.

The efficacy of this method is built upon the key assumption of constant-velocity, straight-line motion. When a target executes significant maneuvers (e.g., turns, acceleration/deceleration), its elevation angle trajectory will deviate substantially from the ideal model. This leads to: 1) large errors in parameter fitting, yielding incorrect feature vectors, and 2) failure of the subsequent fusion cost calculation, potentially resulting in track loss. Therefore, handling maneuvering targets is a critical challenge for future work.

Future research will focus on two primary directions. First, enhancing the model's applicability by incorporating switchable models or adaptive filters to handle maneuvering targets, potentially drawing inspiration from methods like the interactive transient model correction used in active sonar tracking [6]. Second, refining the cost function to include more sophisticated track quality metrics and investigating the impact of sound ray bending corrections to better approximate realistic oceanic environments. Furthermore, exploring the use of deep learning models, such as BiLSTM-Attention networks, presents a promising avenue for enhancing localization robustness, particularly in adversarial environments [7].

5. Conclusion

This paper has proposed and validated a novel multi-node joint tracking method for underwater targets based on elevation angle trajectory features. The key contribution is the successful elevation of the tracking problem from the measurement level to the feature level. This is achieved by first transforming discrete, instantaneous measurements from each node into stable feature vectors that characterize the target's CPA process, and then performing association and fusion within this feature space. Simulation results robustly demonstrate that, compared to the traditional JPDA method, this feature-level fusion strategy significantly improves tracking robustness and localization accuracy in complex scenarios involving target crossovers and data loss. This work indicates that mining and leveraging the deeper motion characteristics embedded within observation sequences is a potent pathway to overcoming the bottlenecks of traditional multi-node tracking.

References

1. Chen R, Li T, Memon I, Shi Y, Ullah I, and Memon SA. Multi-Sonar Distributed Fusion for Target Detection and Tracking in Marine Environment. Sensors (Basel)2022 Apr;22:3335.
2. Bar-Shalom Y and Tse E. Tracking in a cluttered environment with probabilistic data association. Automatica 1975; 11:451–60..
3. Fortmann TE, Bar-Shalom Y, and Scheffe M. Sonar tracking of multiple targets using joint probabilistic data association. IEEE Journal of Oceanic Engineering 1983; 8:173–84.
4. En F, Xie W, Pei J, Hu K, Li X, and Podpe˘an V. Improved Joint Probabilistic Data Association (JPDA) Filter Using Motion Feature for Multiple Maneuvering Targets in Uncertain Tracking Situations. Information 2018;9:322.
5. Wei Z, Duan Z, Han Y, and Mallick M. A New Coarse Gating Strategy Driven Multidimensional Assignment for Two-Stage MHT of Bearings-Only Multisensor-Multitarget Tracking. Sensors (Basel) 2022 Feb; 22:1802.
6. Liu C and Fang S. An Interactive Transient Model Correction Active Sonar Target Tracking Method. Applied Sciences 2024; 14:4865.

7 Jia Y, Lou Y, Zhao Y, Sun S, and Cheng J. BiLSTM-Attention-PFTBD: Robust Long-Baseline Acoustic Localization for Autonomous Underwater Vehicles in Adversarial Environments. Drones 2025; 9:204.

Localization of whale calls using a single underwater acoustic vector sensor

Hefeng Dong[1]; Guosong Zhang[2]; Victor Espinosa Rosello[3]; Isabel Pérez Arjona[3]; Josephine Nell Schulze[1]

[1]Norwegian University of Science and Technology, Trondheim, Norway
[2]Institute of Marine Research, Bergen, Norway
[3]Polytechnic University of Valencia, Valencia, Spain

EXTENDED ABSTRACT

A hydrophone array can be used for localization of underwater sound source determining both direction of arrival and the range using beamforming and matched field processing methods (1). In contrast, a single hydrophone can only be used to estimate the range of sound sources, not the direction (2). However, an underwater acoustic vector sensor (AVS) can simultaneously measure both the acoustic pressure and particle motion in two or three orthogonal directions, allowing for estimating both the bearing and range of sound sources based on active intensity method (3-7) and the normal mode theory and nonlinear warping transform (2). In this study, both active and passive data collected by a single AVS are analyzed to characterize the AVS's performance and estimate the direction of arrival (DOA) and range of a sound source/whale call. A field experiment was conducted in the Norwegian Sea in April 2024 using an underwater AVS system deployed on the seafloor by a ROV on board, along with an underwater sound source suspended from a research vessel, RV *Kronprins Haakon* (8). The area is quite flat with an average water depth of 210 m. A controlled active acquisition was conducted to evaluate the performance of the AVS in resolving the bearing of the sound source (8). Based on the transducer's frequency response, a source signal of 600 Hz single tone was used to obtain a source level of 185 dB re 1 µPa. 640 transmissions at each of 16 locations 500 m from the AVS position covering 360 azimuth angles relative to the AVS were performed. Active intensity method is used to estimate the bearing as a function of time and frequency (7). Figure 1 illustrates the estimated bearing of the 16 transmissions: T01 - T16 and the estimated errors (8). The figure also shows the values of minimum, 25th percentile, median, 75th percentile and maximum. Most of the transmissions have errors around - 20° with fewer exceptions, such as, T01, T09, T13 and T 16. Those transmissions are located close to x-axis, which leads to larger uncertainty. A later test of the AVS system in air shows that the system error is around 20° which matches well with the estimated error shown in Figure 1 (b). Active data analysis demonstrates the AVS's effectiveness in bearing estimation.

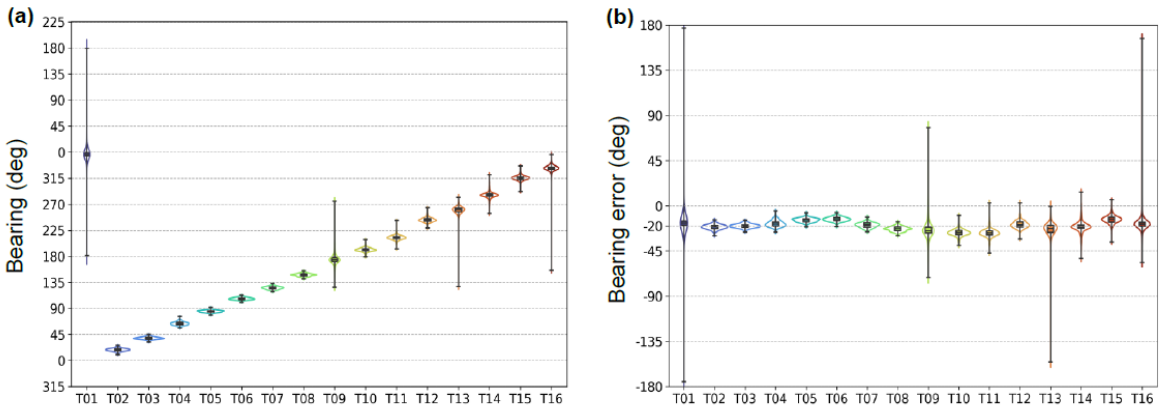


Figure 1 – (a) Estimated bearings of 16 transmissions. (b) Bearing errors. (a) and (b) show minimum, 25th percentile, median, 75th percentile, and maximum.

The AVS system also collected data in a passive mode throughout the experiment. The passive recordings captured several days of whale vocalizations, mainly within a frequency band of 10 - 1000 Hz. Based on spectral characteristics, the vocalizations are identified being from both fin whales and humpback whales, with the majority attributed to humpback whales. The recordings containing whale vocalizations are analyzed using the active intensity method to compute the azigram, which displays the direction of arrival of incoming sounds as a function of time and frequency. The bearing of the whale vocalizations is successfully determined by applying a threshold value to enhance SNR as shown in Figure 2 (b) where two whales calls can be recognized at around 330º and 20º, respectively. Additionally, the range of whale's call is estimated using the method of phase compensation of frequency modulated signal and mode filtering technique as described in (9). The localization results are illustrated in Figure 3. The whale call and its spectrogram are illustrated in Figure 3 (a) and 3 (b). Three dispersion curves are estimated by phase compensation and warping. The dispersion curve differences are shown in Figure 3 (c). The data and replicas of the dispersion curve difference do not fit well. However, the overall behavior of the least square fit is smooth since dispersion curves are very sensitive to range so that the range is well estimated as shown in Figure 3 (d).

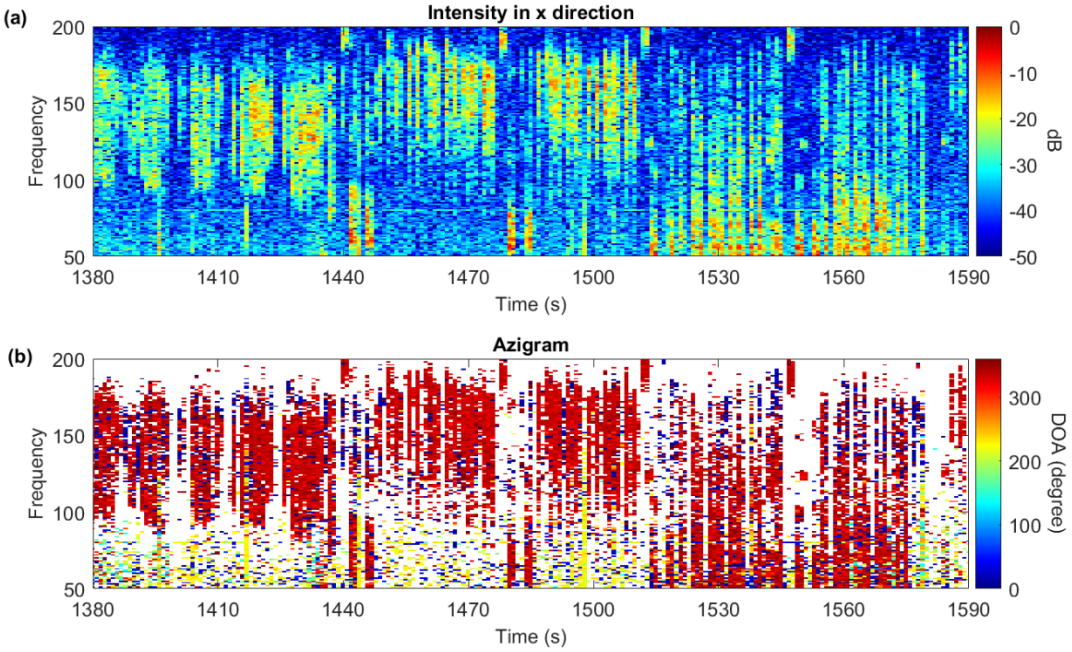


Figure 2 – (a) Normalized intensity in x/east direction. (b) Azigram.

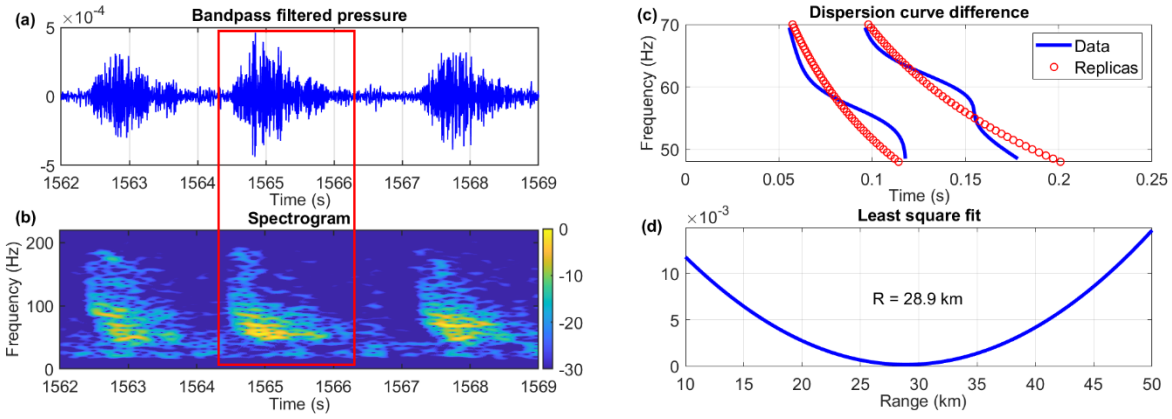


Figure 3 – Humpback whale downsweep localization. (a) Bandpass filtered pressure. (b) Spectrogram. (c) Data and replicas of dispersion curve difference. (d) Estimated range by least square fit.

Keywords: Acoustic vector sensor, Azimuth, Localization

ACKNOWLEDGEMENTS

The authors thank the crew of RV *Kronprins Haakon* and the Norwegian Institute of Marine Research for their assistance of the field experiments, and NTNU engineer Tim Cato Netland for technical support. This work was partially supported by the SFI Centre for Geophysical Forecasting under grant 309960. Hefeng Dong thanks Prof. Julien Bonnel for helpful discussions.

REFERENCES

1. Baggeroer A. B, Kuperman W, Schmidt H. Matched field processing: Source localization in correlated noise as an optimum parameter estimation problem. J Acoust Soc Am. 1988; 83(2): 571-587.
2. Bonnel, J., Thode, A., Blackwell, S., Kim, K., and Macrander, A. (2014). "Range estimation of bowhead whale (Balaena mysticetus) calls in the arctic using a single hydrophone," J. Acoust. Soc. Am. 136(1): 145-155.
3. Felisberto P, Rodriguez O, Santos P, Ey E, Jesus S M. Experimental results of underwater cooperative source localization using a single acoustic vector sensor. Sensors. 2013; 13(7): 8856-8878.
4. Stinco P, Tesei A, Ferri G, Biagini S, Micheli M, Garau B, LePage K D, Troiano L, Grati A, Guerrini P. Passive acoustic signal processing at low frequency with a 3-d acoustic vector sensor hosted on a buoyancy glider. IEEE JOE. 2020; 46(1): 283-293.
5. Dall'Osto D R, Dahl P H, Woong Choi J. Properties of the acoustic intensity vector field in a shallow water waveguide. J. Acoust. Soc. Am. 2012; 131(3): 2023-2035.
6. Thode A M, Sakai T, Michalec J, Rankin S, Soldevilla M S, Martin B, Kim K H. Displaying bioacoustic directional information from sonobuoys using "azigrams". J. Acoust. Soc. Am. 2019; 146(1): 95-102.
7. Zhu X, Dong H, Zhang G. Underwater Acoustic Source Localization Based On A Vector Sensor System: A Field Case Study. Pro 7th. UACE2023, 25-30 June 2023; Kalamata, Greece 2023.
8. Zhang G, Dong H, Cresci A, Browman H I, Pedersen G. Resolving acoustic azimuth using an accelerometer-based underwater acoustic vector sensor. Limnol. Oceanogr.: Methods. 2025; 1-13. doi: 10.1002/lom3.10680.
9. Bonnel J, Thode A, Wright D, Chapman R. Nonlinear time-warping made simple: A step-by-step tutorial on underwater acoustic modal separation with a single hydrophone. J. Acoust. Soc. Am. 2020; 147(3): 1897-1926. doi: 10.1121/10.0000937.

PROCEEDINGS of the
16th International Conference on Theoretical and Computational Acoustics

August 18 to 22, 2025 BEXCO, Busan Korea

Multi-range sparse Bayesian learning for broadband modal phase velocity estimation and shallow-water geoacoustic inversion

Shanru Lin[1], Haiqiang Niu[2], Zhenglin Li[3],Yonggang Guo[4]

[1]Institute of Acoustics, Chinese Academy of Sciences, People's Republic of China
[2]Institute of Acoustics, Chinese Academy of Sciences, People's Republic of China
[3]Sun Yat-sen University, People's Republic of China
[4]Institute of Acoustics, Chinese Academy of Sciences, People's Republic of China

ABSTRACT

This study proposes a broadband modal phase velocity estimation method based on multi-range sparse Bayesian learning (SBL) for shallow-water geoacoustic inversion. Unlike conventional block SBL—which relies on narrowband approximations—the proposed approach estimates local modal depth functions and horizontal wavenumbers directly from multi-range vertical line array data, without a prior knowledge of geoacoustic parameters or source locations. By overcoming the bandwidth limitations of dispersion-based methods, it enables robust phase velocity estimation across a broad frequency range, facilitating accurate inversion via multi-frequency phase-velocity curve matching. Crucially, the method decouples seabed attenuation from other parameters, mitigating mutual interference during inversion. Additionally, it leverages local modes, which depend solely on local seabed properties, making it immune to propagation path or topography mismatches. To enhance efficiency in high-dimensional parameter searches, a Bayesian optimization (BO) algorithm is integrated. The inverted parameters are further applied to construct a sound field model for matched-field processing (MFP) localization. Simulations and experimental data demonstrate superior performance in range and depth estimation compared to conventional matched-field inversion (MFI), validating the method's feasibility and precision.

[1] linshanru@mail.ioa.ac.cn
[2] nhq@mail.ioa.ac.cn
[3] lizhlin29@mail.sysu.edu.cn
[4] guoyg@mail.ioa.ac.cn

DAILY PROGRAM

AUGUST 19 (TUE)

09:00 – 10:40	Inverse Design of Acoustic Metamaterials	325 (3F)
Chair	Jaeyub Hyun (Pukyong National University)	
09:00	Topology Optimization of a Viscoelastic Dynamic Vibration Absorber for Multi-Eigenfrequency Attenuation	
	Hyunggyu Choi, Gil Ho Yoon	
09:20	AI-assisted topology optimization of acoustic mufflers via reinforcement learning	
	Kee Seung Oh, Hayoung Chung, Joo Hwan Oh	
09:40	Deep-subwavelength lightweight acoustic metamaterial for broadband sound insulation	
	Jiwan Kim, Wonju Jeon	
10:00	Low-speed, low-impedance elastic metamaterial for enhancing the mechanical-to-electric energy conversion efficiency of triboelectric harvesters	
	Hyung Jin Lee, Hong Min Seung, Jun Hyeong Park, Hyun Soo Kim, Sunghoon Hur, Hyun-Cheol Song	
10:20	Inverse Design of Electrode Patterns for Multi-Point Acoustic Focusing via Level-Set Topology Optimization	
	Sanguk Park, Jaeyub Hyun	

PROCEEDINGS of the
16th International Conference on Theoretical and Computational Acoustics

August 18 to 22, 2025 BEXCO, Busan Korea

Topology Optimization of a Viscoelastic Dynamic Vibration Absorber for Multi-Eigenfrequency Attenuation

Hyunggyu Choi[1]; Gil Ho Yoon[1]
[1] Department of Mechanical Engineering, Hanyang University, Seoul, South Korea

ABSTRACT

This study presents a topology optimization approach for designing a viscoelastic Dynamic Vibration Absorber (DVA) to effectively attenuate multiple eigenfrequencies. A single DVA system for vibration attenuation of multiple eigenfrequencies provided by the topology optimization method is preferred for many engineering applications rather than a complex Multi-DVA system. To efficiently attenuate split vibration using a DVA system, we present a method for designing a DVA suitable for the eigenmode motion of a main structure. It is modified based on multi-materials consisting of viscoelastic and elastic materials to control the eigenmode motion. An additional constraint is imposed to ensure that the corresponding natural frequencies match based on the modal mass ratio of the main structure and the DVA. Several numerical examples are solved to demonstrate the vibration attenuation of single or multiple frequencies for the topology optimization of Single DVA with viscoelastic material, which minimizes the vibration of the main structure and controls the eigenmode shapes.

Keywords: Topology optimization, Dynamic vibration absorber

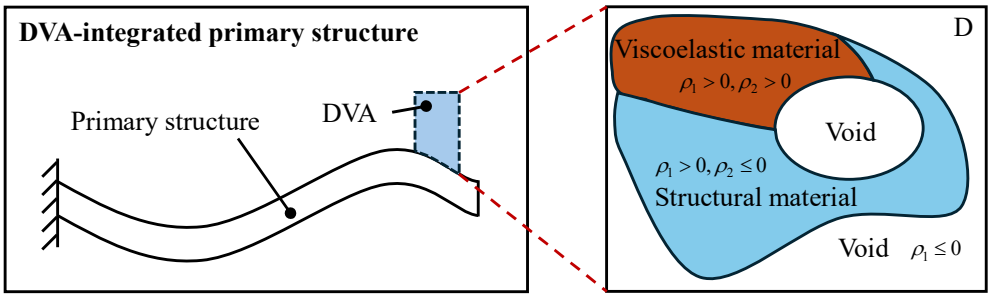


Figure 1 – Schematic of a DVA attached to the primary structure.

ACKNOWLEDGEMENTS

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (RS-2024-00351611).

REFERENCES

1. SVANBERG, Krister. The method of moving asymptotes—a new method for structural optimization. *International journal for numerical methods in engineering*, 1987, 24.2: 359-373.
2. JACQUOT, R. G. Optimal dynamic vibration absorbers for general beam systems. *Journal of Sound and Vibration*, 1978, 60.4: 535-542.
3. HUNT, J. Brn; NISSEN, J.-C. The broadband dynamic vibration absorber. *Journal of Sound Vibration*, 1982, 83.4: 573-578.
4. ALFOUNEH, Mahmoud; TONG, Liyong. Maximizing modal damping in layered structures via multi-objective topology optimization. *Engineering Structures*, 2017, 132: 637-647.

PROCEEDINGS of the
16th International Conference on Theoretical and Computational Acoustics

August 18 to 22, 2025 BEXCO, Busan Korea

AI-assisted topology optimization of acoustic mufflers via reinforcement Learning

Kee Seung Oh[1]; Hayoung Chung [2]; Joo Hwan Oh [1]
[1] Seoul National University
[2] Ulsan National Institute of Science and Technology

ABSTRACT

We propose an AI-based approach to topology optimization for acoustic mufflers. Traditional methods often get trapped in local minima, leading to insufficient performance. We adopt a reinforcement learning (RL) algorithm capable of handling disconnected configurations and multi-objective requirements, addressing critical limitations in previous muffler design studies. Our framework formulates the muffler design as a Markov decision process (MDP), with carefully defined agent actions, states, and rewards. Numerical tests and practical muffler cases confirm that the RL-driven topology optimization significantly outperforms gradient-based methods, providing robust, acoustically efficient solutions. Furthermore, multi-objective optimization is seamlessly incorporated, allowing flexible control over conflicting acoustic requirements. Additionally, the RL process and resulting internal layouts are discussed, illustrating the method's advantage over conventional approaches.

Deep-subwavelength lightweight acoustic metamaterial for broadband sound insulation

Jiwan KIM[1]; Wonju JEON[2]
[1, 2] Korea Advanced Institute of Science and Technology, Republic of Korea

ABSTRACT

This study introduces a deep-subwavelength lightweight acoustic metamaterial, referred to as a meta-panel, for broadband sound insulation in the low-frequency regime. The meta-panel consists of a multi-scale lattice structure sandwiched between two thin membranes, engineered to induce negative effective material properties and achieve high sound transmission loss (STL) over a wide low-frequency range. A theoretical model is formulated to rapidly and accurately predict the meta-panel's STL by incorporating the coupled dynamics of membrane impedances under elastic boundary constraints and wave propagation through the multi-scale lattice structure. This model facilitates an inverse design framework aimed at maximizing STL in the 100-1500 Hz range while minimizing the meta-panel's thickness and weight. Theoretical and numerical predictions are supported by experimental validation using 3D-printed meta-panel samples tested in an impedance tube. Experimental results demonstrate STL improvement of up to 30 dB compared to a same-weight plate, underscoring the meta-panel's superior weight-to-performance efficiency.

Keywords: Acoustic metamaterial, Sound insulation, Multi-scale lattice structure

[1] jiwan.kim@kaist.ac.kr
[2] wonju.jeon@kaist.ac.kr

Low-speed, low-impedance elastic metamaterial for enhancing the mechanical-to-electric energy conversion efficiency of triboelectric harvesters

Hyung Jin Lee [1]; Hong Min Seung [2]; Jun Hyeong Park [3]; Hyun Soo Kim [4]; Sunghoon Hur [5]; Hyun-Cheol Song [6]

[1,2,3] Korea Research Institute of Standards and Science (KRISS)

[4,5] Korea Institute of Science and Technology (KIST)

[6] Korea University

ABSTRACT

In this talk, we will propose a specially configured elastic metamaterial that simultaneously exhibits low wave-propagation velocity and low mechanical impedance in a dynamic manner. The material properties of this metamaterial are ideal for triboelectric harvesters, as it helps the thin constituting layers of the harvesters achieve a fundamental vibration mode at a kHz-level frequency and vibrate sensitively at that frequency with a large amplitude, both of which are essential for achieving high harvesting efficiency. This talk will address two main parts: first, the dynamic mechanical properties of the metamaterial will be demonstrated; second, the application of the metamaterial to a triboelectric harvester will be validated to enhance harvesting efficiency in underwater environments.

Inverse design of electrode patterns for multi-point acoustic focusing via level-set topology optimization

Sanguk Park[1a]; Jaeyub Hyun[1b];
[1] Pukyong National University, Republic of Korea

ABSTRACT

Conventional focusing methods using acoustic lenses have some limitations, such as a low transmittance and crosstalk, which degrade the imaging performance [1]. These issues are attributed to the complex structure and material properties of acoustic lenses, which absorb or scatter sound waves. To overcome these limitations, this study introduces a new focusing concept based on a topology-optimized electrode pattern on a piezoelectric transducer, eliminating the need for an acoustic lens in ultrasonic focusing. To this end, a fully-coupled piezoelectric-acoustic interaction finite element model is developed to simulate the piezoelectric transducer with air-backing conditions, and a systematic inverse design method is proposed to find optimal electrode patterns based on level-set topology optimization (LSTO) [2]. The electric boundary condition is topologically parameterized using the area fractions of design elements cut by zero level-set contour. The electrode pattern layouts are evolved by solving the Hamilton-Jacobi equation, with boundary point sensitivities calculated using a discrete adjoint method. The Kirchhoff–Helmholtz integral is used for computationally efficient far-field evaluation during optimization. Moreover, to achieve fully-connected electrode patterns without floating parts, a fictitious structural compliance constraint is applied. Numerical examples demonstrate that the optimized electrode pattern enables flexible multi-point focusing, with potential for applications such as multi-site stimulation and acoustic holography.

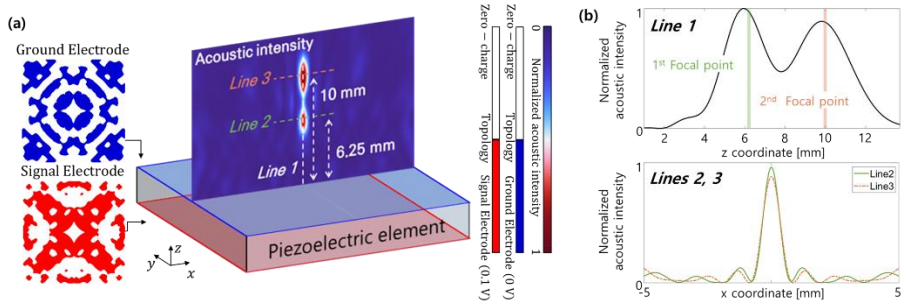


Figure 1 – The optimization results for multi-point (6.25 and 10 mm from the transducer) focusing via the proposed LSTO: (a) Optimized electrode pattern and corresponding acoustic distribution. Blue and red patterns indicate the optimized ground and signal electrodes, respectively. (b) Acoustic intensity profiles: Line 1 along the *z*-axis, Lines 2 and 3 along the *x*-axis at two focal points. Here, these lines are illustrated in Fig. 1(a)

Keywords: Ultrasound focusing, Level-set topology optimization, Coupled piezoelectric-acoustic interaction

ACKNOWLEDGEMENTS

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (RS-2025-00561030)

REFERENCES

1. Maznev AA, Wright OB. Upholding the diffraction limit in the focusing of light and sound. Wave Motion. 2017;68:182-9
2. Dunning PD, Kim HA. Introducing the sequential linear programming level-set method for topology optimization. Structural Multidisciplinary Optimization. 2015;51:631-43

[1a] rhaehf0324@pukyong.ac.kr
[1b] jyhyun00@pknu.ac.kr

DAILY PROGRAM

AUGUST 20 (WED)

09:00 – 10:20	Low-Frequency Acoustic Propagation and Sonar Signal Processing	323 (3F)
Chair	Sung-Hoon Byun (Korea Research Institute of Ships & Ocean engineering)	
09:00	Non-Integer Order Parabolic Cylinder Functions in Very-Low-Frequency sound Propagation: A WKB method for Ocean Boundary Effects	
	Jiankang Zhan, Shengchun Piao, Lijia Gong	
09:20	A Cyclic Overlapping Subpulse Coherent Processing Interval Method with Dual Weighting for Continuous Active Sonar: Uniform Updates and Enhanced Performance	
	Zhe Cao, Long-Hao Qiu, Yan Wang, Guo-Long Liang, Shu-Xian Hao	
09:40	Fast Chirp Rate Estimation Method for LFM Signals in Impulsive Noise Environments	
	Shu-Xian Hao, Long-Hao Qiu, Guo-Long Liang, Zhe Cao	
10:00	Model Order Reduction for Horizontal Refraction Equation Solutions: An Efficient Framework for Directional Three-Dimensional Acoustic Wave Propagation Modeling	
	Tengjiao He	

PROCEEDINGS of the
16th International Conference on Theoretical and Computational Acoustics

August 18 to 22, 2025 BEXCO, Busan Korea

ABS-bs2025c16009

Non-Integer Order Parabolic Cylinder Functions in Very-Low-Frequency Sound Propagation: A WKB Method for Ocean Boundary Effects

Jiankang ZHAN[1,2,3]; Shengchun PIAO[1,2,3]; Lijia GONG[1,2,3]
[1] National Key Laboratory of Underwater Acoustic Technology, Harbin Engineering University, Harbin 150001, China
[2] Key Laboratory of Marine Information Acquisition and Security (Harbin Engineering University), Ministry of Industry and Information Technology; Harbin 150001, China
[3] College of Underwater Acoustic Engineering, Harbin Engineering University, Harbin 150001, China

ABSTRACT

A Wentzel-Kramers-Brillouin (WKB) method has been developed to express normal modes at very-low-frequency in deep ocean, where the solution to the modal equation is described using parabolic cylinder functions (PCF). Numerical results reveal that at frequencies as low as a few Hertz, higher order normal modes are increasingly influenced by the phase shift at turning point caused by ocean boundary, with the phase shift value remaining less than -π. Based on the relationship between the order of the PCF and the phase shift at turning point, it is observed that the order of the PCF no longer be an integer, resulting in changes to the shape of the PCF. Consequently, the spatial distribution of the normal modes is also altered, indicating that the mode shape at very-low-frequency is significantly affected by ocean boundary.

Keywords: Very-low-frequency sound propagation, Normal modes, Parabolic cylinder functions

1. INTRODUCTION

Very-low-frequency (VLF, 1-100 Hz) sound propagation in deep ocean is crucial for underwater acoustic communication, geophysical monitoring, and ocean acoustic tomography. At these frequencies, wavelengths reach hundreds of meters, propagation mechanisms are complex, and the sound field is typically described using orthogonal normal mode decomposition. Traditional solutions like the Wentzel-Kramers-Brillouin (WKB) approximation are often used to solve the modal eigenvalue problem. Classical theory assumes a constant turning point phase shift of $\pi/2$ [1]. However, at VLF, the influence of ocean boundary on the sound field, particularly higher-order modes, significantly intensifies.

This paper addresses the critical issue of how boundary conditions affect mode shape, based on a WKB method using parabolic cylinder functions (PCF) [2]. Theoretical analysis and numerical simulations demonstrate that under VLF conditions, boundary effects reduce the phase shift value below π. This causes the order of PCF to transition from integer to non-integer, subsequently inducing significant alterations in the spatial distributions of higher-order normal modes.

2. NORMAL MODES EXPRESSED BY PARABOLIC CYLINDER FUNCTIONS IN DEEP OCEAN

For a range-independent horizontally stratified medium with a point harmonic source in a cylindrically symmetric situation, the source is at $r = 0$ and $z = z_s$, as illustrated in Figure 1. The eigenfunctions with eigenvalues ξ_n can be expressed by

[1] zhanjiankang@hrbeu.edu.cn
[2] piaoshengchun@hrbeu.edu.cn
[3] lijia.gong@hrbeu.edu.cn

$$U_n(\xi_n,z)=\frac{F_1(\xi_n,z)-\dfrac{F_1(\xi_n,0)}{F_2(\xi_n,0)}F_2(\xi_n,z)}{\sqrt{W(F_1,F_2)\dfrac{\partial}{\partial\xi_n}\left[\dfrac{F_1(\xi_n,0)}{F_2(\xi_n,0)}-\dfrac{F_1'(\xi_n,H)}{F_2'(\xi_n,H)}\right]}} \tag{1}$$

where the Wronskian W is given by

$$W(F_1,F_2)=F_1(\xi,z_s)F_2'(\xi,z_s)-F_1'(\xi,z_s)F_2(\xi,z_s) \tag{2}$$

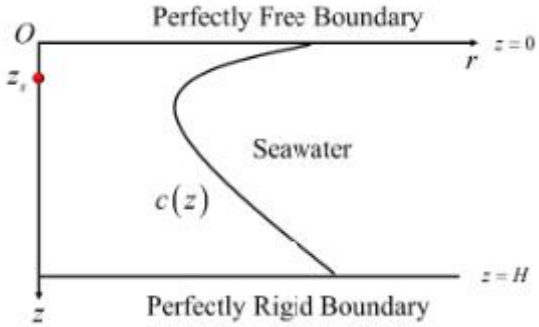


Figure 1 – The ideal waveguide

For a waveguide with a typical deep water sound speed profile, the linearly independent solutions $F_1(\xi,z)$ and $F_2(\xi,z)$ of the modal equation can be expressed using the uniform asymptotic solution based on PCF [2],

$$\begin{aligned}F_1(\xi,z)&=R(E)(S')^{-\frac{1}{2}}D_M(-y)\\F_2(\xi,z)&=R(E)(S')^{-\frac{1}{2}}D_M(y)\end{aligned} \tag{3}$$

where the function $D_M(y)$ is a PCF of order M, and

$$\begin{aligned}y&=\sqrt{2}S(z)\\E(\xi_n)&=\frac{2}{\pi}\int_{z_1(\xi_n)}^{z_2(\xi_n)}\sqrt{k^2(z)-\xi_n^2}\,dz\\M(\xi_n)&=\frac{1}{\pi}\int_{z_1(\xi_n)}^{z_2(\xi_n)}\sqrt{k^2(z)-\xi_n^2}\,dz-\frac{1}{2}\\R(E)&=2^{\frac{1}{4}}\left(\frac{2e}{E}\right)^{\frac{E}{4}}\end{aligned} \tag{4}$$

The values $z_1(\xi_n)$ and $z_2(\xi_n)$ represent the upper and lower turning points of the horizontal wavenumber ξ_n, respectively, and the function $S(z)$ is defined as

$$\begin{cases}\int_{s_1}^{S}\sqrt{\sigma^2-E}\,d\sigma=\int_{z_1}^{z}\sqrt{\xi^2-k^2(z)}dz, & 0\le z<z_1\\ \int_{s_1}^{S}\sqrt{E-\sigma^2}\,d\sigma=\int_{z_1}^{z}\sqrt{k^2(z)-\xi^2}\,dz, & z_1\le z\le z_2\\ \int_{s_1}^{s_2}\sqrt{E-\sigma^2}\,d\sigma+\int_{s_2}^{S}\sqrt{\sigma^2-E}\,d\sigma=\int_{z_1}^{z_2}\sqrt{k^2(z)-\xi^2}\,dz+\int_{z_2}^{z}\sqrt{\xi^2-k^2(z)}dz, & z_2<z\le H\end{cases} \tag{5}$$

where $s_1=-\sqrt{E}$, $s_2=\sqrt{E}$。

Furthermore, from Eq. (4) we obtain the phase shift (which relies on the upper and lower boundary conditions) for the mode as [3]

$$\delta_n=2\left(n-M_n-\frac{3}{2}\right)\pi,\quad n=1,2,\mathrm{K} \tag{6}$$

It can be seen that the phase shift δ_n at the turning points is a function of the order M_n of the PCF.

When analyzing the variation of the mode shape with depth z, according to Eq. (1), the denominator is independent of depth z and can be regarded as a constant. Each term in the numerator contains $R(E)(S')^{-1/2}$, and both can be considered as an overall scaling of the mode

shape. Thus, the eigenfunctions can be simplified to

$$U_n(\xi_n, z) = A_n\left[D_{M_n}(-y) - \frac{D_{M_n}(-y)}{D_{M_n}(y)}\bigg|_{z=0} D_{M_n}(y)\right] \tag{7}$$

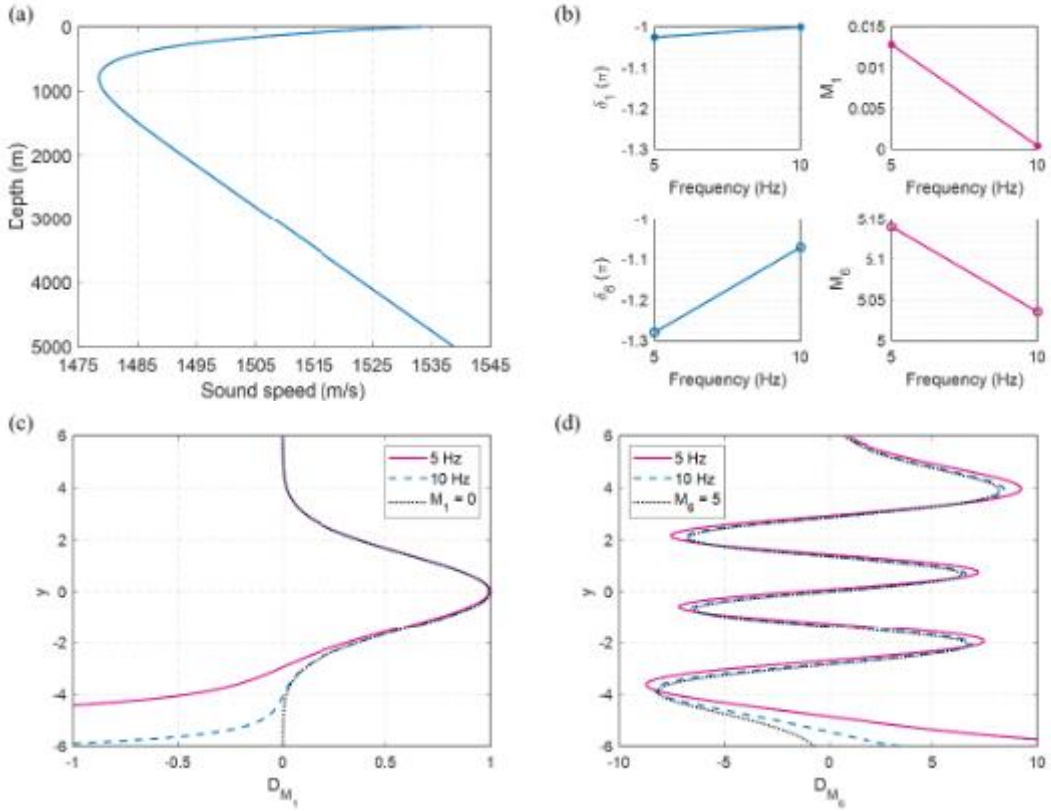
where A_n is a constant, it can be seen from the above equation that the mode shape is only related to the shape of the PCF D_M, and D_M is a function of the order M_n. Therefore, the mode shape can be regarded as a function of the order M_n. In summary, the characteristics of the mode shape under ocean boundary effects can be studied based on the order M_n of the PCF.

3. NUMERICAL SIMULATION

In this section, the Munk profile [4] is selected as shown in Fig. 2(a). When the frequencies are 5 Hz and 10 Hz, the eigenvalues ξ_n are solved according to the dispersion equation, and the phase shifts δ_n at the turning points and the orders M_n of the PCF are obtained by combining Eqs. (4) and (6). (Here, we have calculated the modes 1 and 6, with the mode 6 being the highest-order turning mode at 5 Hz.).

As can be seen from Fig. 2(b), the phase shift δ_1 of mode 1 in the ideal waveguide is close to $-\pi$, causing the order M_1 of the PCF to approach 0 ($M_1 = 0$ in an unbounded scenario [2]). The shape of the PCF $D_{M_1}(y)$ approximates $D_0(y)$, as illustrated in Fig. 2(c). Consequently, whether at 5 Hz or 10 Hz, the shape of mode 1 resembles that in an unbounded environment, as shown in Figs. 2(e) and 2(f).

In contrast, for mode 6, the phase shift δ_6 in the ideal waveguide is less than $-\pi$, with $\delta_6 = -1.07\pi$ at 10 Hz and $\delta_6 = -1.28\pi$ at 5 Hz. This indicates that the phase shift at the upper and lower turning points at 5 Hz is no longer $\pi/2$ due to ocean boundary effects, resulting in the $M_6 = 5.14$, which is no longer an integer order ($M_6 = 5$ in an unbounded case). Compared to 10 Hz, the shape of the PCF $D_{M_6}(y)$ at 5 Hz differs significantly from $D_5(y)$, as shown in Fig. 2(d). By comparing Figs. 2(e) and (f) for mode 6, it is evident that the shape of mode 6 at 5 Hz no longer resembles that in an unbounded environment. Therefore, at frequencies as low as a few hertz, the spatial distribution of higher-order normal modes is significantly influenced by oceanic boundaries.



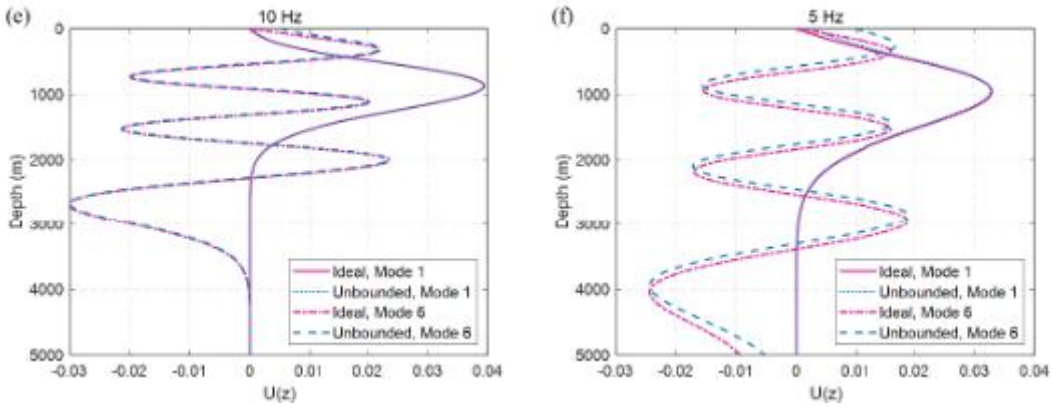


Figure 2 – (a) The Munk profile, (b) phase shift δ and the order M of PCF corresponding to mode 1 and mode 6, (c) $D_{M_1}(y)$, (d) $D_{M_6}(y)$, (e) mode shape at 10 Hz, (f) mode shape at 5 Hz

4. SUMMARY

This paper investigates the mechanism of how ocean boundary influences the mode shape based on the PCF for VLF normal modes in deep ocean. Numerical results indicate that in the ideal waveguide, the phase shift value at the turning points of higher-order normal modes at several hertz is less than π, causing the order of the PCF to no longer be an integer. This alteration in the shape of the PCF leads to a significant impact of ocean boundary on the spatial distribution of normal modes.

ACKNOWLEDGEMENTS

This work is supported by the National Natural Science Foundation of China (Grant Nos. 12174048 and 12204128).

REFERENCES

1. Jensen FB, Kuperman WA, Porter MB, Schmidt H. Computational Ocean acoustics: Springer Science & Business Media; 2011.
2. Zhan J-K, Piao S-C, Gong L-J, Dong Y, Guo Y-C, Zheng G-X. A WKB method based on parabolic cylinder function for very-low-frequency sound propagation in deep ocean. Chinese Physics B. 2025;34(3):034301.
3. Harrison CH. A relation between multipath group velocity, mode number, and ray cycle distance. The Journal of the Acoustical Society of America. 2012;132(1):48-55.
4. Munk WH. Sound channel in an exponentially stratified ocean, with application to SOFAR. The Journal of the Acoustical Society of America. 1974;55(2):220-6.

PROCEEDINGS of the
16th International Conference on Theoretical and Computational Acoustics

August 18 to 22, 2025 BEXCO, Busan Korea

A Cyclic Overlapping Subpulse CPI Method with Shift-Weighting for Continuous Active Sonar: Uniform Updates and Enhanced Performance

Zhe CAO[1]; Long-Hao QIU[2†]; Yan WANG[3]; Guo-Long LIANG[4]; Shu-Xian HAO[5]
1. National Key Laboratory of Underwater Acoustic Technology, Harbin Engineering University, China;
2.Key Laboratory of Marine Information Acquisition and Security (Harbin Engineering University), China;
3.College of Underwater Acoustic Engineering, Harbin Engineering University, China

ABSTRACT

Continuous active sonar (CAS) commonly employs linear frequency modulated continuous wave (LFMCW) signals, where multi-subpulse processing is achieved through multi-channel matched filtering. However, non-overlapping subpulses impose constraints on the achievable detection gain and target revisit rate. To address this issue, a Shift-Weighted Cyclic Overlapping Subpulse Coherent Processing Interval (SW-COSCPI) method is proposed in this paper. Within this method, subpulses are cyclically selected. Non-coherent gain is obtained by applying Doppler-consistency shifting to the detection results of each subpulse. Additionally, noise interference is suppressed by peak-continuity weighting. Through the synergistic effect of these two weightings, non-coherent gain and tracking continuity are simultaneously enhanced. The detection performance and receiver operating characteristic (ROC) curves are compared for the single subpulse CPI (SS-CPI), the non-coherent cyclic overlapping subpulse CPI (NcI-COSCPI), and the proposed SW-COSCPI algorithm via computer simulations. Results demonstrate that the SW-COSCPI method enables continuous detection and exhibits a more robust improvement in detection performance compared to the NcI-COSCPI method. At the same probability of false alarm (Pfa), the detection probability (Pd) for full-band targets is enhanced by up to 2%, and for new targets by up to 3%. This indicates that the SW-COSCPI method outperforms existing methods in terms of both target detection continuity and tracking stability.

Keywords: Continuous Active Sonar, Multi-channel Matched Filters, Sigmoid

1. INTRODUCTION

Continuous Active Sonar (CAS) systems operate with simultaneous transmission and reception, employing continuous-wave probe signals characterized by a High-Duty-Cycle (HDC) instead of traditional pulsed waveforms. Compared to Pulse Active Sonar (PAS), CAS offers several significant advantages as following(1):Enhanced Update Rate & Continuous illumination, reduced Detectability & Environmental Impact and large Time-Bandwidth Product. The feasibility of CAS technology has been demonstrated through numerous experiments. Tests conducted in shallow water using the Interim Removable Low-frequency Active Sonar (IRLFAS), developed by the Netherlands Organization for Applied Scientific Research (TNO), indicated that the target detection performance of CAS is comparable to PAS while simultaneously achieving a lower false alarm rate(2). The TREX13 experiment investigated the impact of sub-band processing bandwidth on CAS performance(3). Results demonstrated that while increasing bandwidth improves the Signal-to-Reverberation Ratio (SRR), the differing propagation characteristics of the spectral components within the broad bandwidth induce coherence losses. Consequently, the actual gain falls below the linear growth predicted by theory under ideal conditions. Furthermore, increasing bandwidth reduces the update rate, necessitating a trade-off between update rate and echo quality. The COLLAB-NGAS14 experiment implemented real-time, autonomous CAS processing for a multistatic, autonomous networked sonar system using proprietary signal processing algorithms(4). Results proved that sub-band matched filtering outperformed PAS in both detection and tracking performance. The LCAS15 experiment utilized a multi-channel matched filter design based on a circular shift method. This approach facilitated

[1] 2354878010@hrbeu.edu.cn

processing with a high overlap rate, effectively boosting the target update rate(5). Additionally, Liu et al. proposed a method to enhance detection performance through the non-coherent (energy) integration of multi-channel matched filter outputs(6).

Employing subpulse decomposition followed by matched filtering does indeed accelerate TRR. This method is particularly well-suited for waveforms exhibiting simple monotonic frequency modulation, such as Linear Frequency Modulation (LFM). While increasing the number of subpulses provides more detection opportunities per PRI, each individual subpulse possesses inherently lower energy and bandwidth compared to the full waveform. This limitation results in degraded Signal-to-Noise Ratio (SNR) and Signal-to-Reverberation Ratio (SRR) for each sub-channel. Therefore, a fundamental trade-off exists between update rate and echo quality (SNR/SRR). The adoption of overlapping sub-pulses in the design of multi-channel matched filters represents a key approach to mitigate this limitation. The corresponding technique is termed the Overlapping Subpulse Coherent Processing Interval (COSPCI) method.

This paper aims to provide a detailed description of the core principles underlying the COSPCI method. Building upon this foundation, and leveraging the inherent Doppler consistency and peak continuity properties observed across multi-channel matched filter outputs, a novel shift-weighting design strategy is proposed. This strategy materializes as the Shift-Weighted Cyclic Overlapping Subpulse Coherent Processing Interval (SW-COSCPI) method. Subsequent sections of this paper will delve into the detailed exposition and systematic analysis of the proposed method.

2. METHODS

1. Receive Signal Model

The conventional CAS system typically employs a Linear Frequency Modulation Continuous Wave (LFMCW) signal as the transmitted waveform, mathematically expressed as:

$$s(t) = \exp(j2\pi(f_0 t + 0.5kt^2)), t \in [-PRI/2, PRI/2] \tag{1}$$

Where k is the frequency modulation slope, f_0 denotes the center frequency of LFM signal, PRI is pulse repetition interval. For a distance R with radial velocity v, the received signal after a propagation delay τ is given by

$$s_r(t) = K_r s(\eta(t-\tau)) + n(t) \tag{2}$$

Where K_r is the amplitude of echo, $\eta = (c-v)/(c+v)$ denotes the Doppler scaling factor due to target radial velocity v. $\tau = 2R/c$ represents the two-way time delay. $n(t)$ is zero-mean complex additive white Gaussian noise (AWGN).

2. Signal Processing Strategy for CAS Systems

target echoes propagate through the underwater channel before being received by the hydrophone array. The received signals are processed through beamforming to generate output signals for subsequent detection. CAS achieves continuous target illumination by transmitting and receiving signals without interruption.

The SW-CPI methodology circumvents non-uniform updates by cyclically extending the original signal, guaranteeing constant update periods across ping boundaries. The processing workflow is structured as follows:

1. Determine T_{up}, T_{sub} and PRI. The actual update period $\tilde{T}_{up}$ is determined by T_{up} and PRI as

$$\tilde{T}_{up} = \mathrm{GCD}(T_{up}, PRI) \tag{3}$$

Where $\mathrm{GCD}(\bullet)$ denotes greatest common divisor. Eq. (7) demonstrates that when $\tilde{T}_{up}$ is an exact divisor of PRI, $\tilde{T}_{up}$ remains unchanged. Conversely, when $\tilde{T}_{up}$ is not an integer divisor of PRI, the actual update period is replaced by their greatest common divisor to ensure continuous updates across ping boundaries.

2. Overlapping sub-pulses are circularly selected from full-band signals as copy signals. The number of Non-Cross-Ping Subpulses (NCPS) $M_{NCPS} = \lfloor (PRI - T_{sub})/T_{up} \rfloor$ and that of Cross-Ping Subpulses (CPS) $M_{CPS} = M_{sub} - M_{NCPS}$.

3. Determine the number of Doppler compensation channels and the target's maximum radial velocity relative to the propagation medium. Construct the multi-channel matched filters by applying Doppler compensation to $h_m(t)$.

4. The COS-CPI method performs target detection by searching for the maximum value among the different Doppler compensation results within each subpulse, expressed as

$$\hat{y}_m(\theta_r,\hat{\eta},\hat{\tau}) = \max_{1\le n\le N}\left\{\left|y_{m,n}(\theta_r,t)\right|^2\right\} \tag{4}$$

When a new target appears in the region, due to the CPS being formed by concatenating signals from two periods, the signal from the second period causes target peaks to appear earlier in the CPS than in the NCPS. The magnitude of these peaks depends on the time-bandwidth product of the second-period signal within the sub-pulse. Reorganize the subpulses sequence as

$$m = \left\{M_{CP}^{1},\cdots,M_{CP}^{M_{CP}},M_{NCP}^{1},\cdots,M_{NCP}^{M_{NCP}}\right\} \tag{5}$$

At this point, the target peaks appear in sequence, facilitating the early capture of the target's initial echo during cyclic transitions.

5. To further achieve non-coherent gain, this paper proposes the Shift-Weighted Cyclic Overlapping Subpulse CPI (SW-COSCPI) method. First, Doppler consistency shifting is applied to perform non-coherent integration of results from identical Doppler compensation factors across different subpulses. To ensure the update cycle, all subpulses are cyclically selected as starting subpulses. With the m -th subpulse channel as the primary channel, a split-window normalization algorithm is used to estimate the local background noise mean for each channel. The Local Peak to Average Ratio (LPAR) at each point is computed to suppress local fluctuation interference and enhance robustness. The LPAR expression for each channel is given by:

$$LPAR_{m,n}(\theta_r,t) = \frac{2y_{m,n}(\theta_r,t)}{\mu_{m,n}^{f}(\theta_r,t)+\mu_{m,n}^{b}(\theta_r,t)+2\varepsilon} \tag{6}$$

where $\mu_{m,n}^{f}$ and $\mu_{m,n}^{b}$ represent the means of the front window and rear window respectively. ε represents the guard background value. Weighting each channel using LPAR enhances detection performance. The weighted result for each channel is expressed as:

$$\overline{y}_{m,n}(\theta_r,t) = \sum_{i=0}^{M_{sub}-1} LPAR_{m',n}(\theta_r,t-k_i\eta_n T_{up})\cdot y_{m',n}(\theta_r,t-k_i\eta_n T_{up}) \tag{7}$$

However, due to the long-tailed nature of the probability density function of LPAR, its performance is suboptimal. Therefore, the Sigmoid function is employed for mapping. The Sigmoid function are given by:

$$S(x) = \frac{1}{1+\mathrm{e}^{-k(x-b)}} \tag{8}$$

The Sigmoid function maps variables into the interval [0,1], exhibiting properties of monotonic increase and a monotonically increasing derivative. The weighted result for each channel becomes:

$$\overline{y}_{m,n}(\theta_r,t) = \sum_{i=0}^{M_{sub}-1} w_{LPAR}^{m'}(\theta_r,t)\,y_{m',n}(\theta_r,t-k_i\eta_n T_{up}) \tag{9}$$

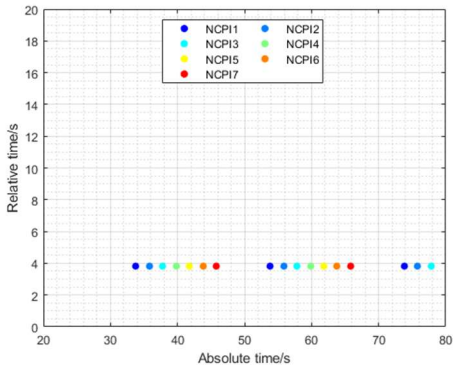
3. SIMULATION and ANALYSIS

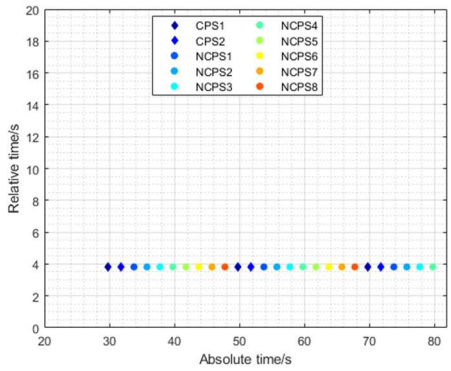
1. Update Cycle Comparison

Simulation Condition 1: The transmitted signal is a LFM signal and the parameters are showed in table 1. Source-Target-Receiver Distance is 5727 m. The expression of transmission loss $TL = -20\lg r$. Background Noise is Additive Gaussian White Noise and Signal-to-Noise Ratio (SNR) is 0dB. Source, target, and receiver are stationary relative to the medium. Detection Results for Each Method are shown in the figure 1.

Table 1 – Parameters of transmitted signal

Signal	Frequency range/ kHz	PRI/s	SL/dB
LFM	1.8-2.6	20	182





(a) (b)

Figure 1 – Detection Results for Each Method. (a) OS-CPI (b) SW-CPI

Figure 1 display the detection results of each method at an update rate of 2s. T_{sub} for OS-CPI and SW-COSCPI is 5s. The horizontal axis in the figures represents the detection time, and the vertical axis shows the relative time delay of the detection result compared to the emission time. Comparison of the detection results reveals that the OS-CPI method produces non-uniform updates when the target echo crosses periodic boundaries. COS-CPI, NcI-COSCPI, and DW-COSCPI effectively achieve uniform target updates when the target echo crosses pings.

2. Full-Band Target Detection Performance Analysis

Full-band target denotes a target which the target echo perfectly matches each channel of COS-CPI. A binary hypothesis test is applied to the received signal on each sub-pulse channel: H_1 means target present, H_0 means target does not exist. The background noise is AWGN, which following $N(0,\sigma^2)$.

In practical applications, k_{LPAR}^m is typically chosen between 3 and 5, TH_{LPAR}^m should be set based on the LPAR distributions under both hypotheses. In this paper, we take half the value at the point where the difference between the CDFs of the two distributions is maximized.

Simulation Condition 2: Number of channels be set as 5. The output SNR after matched filtering in each channel ranges from 3 dB to 11 dB with equally spaced values to simulate real-world SNR fluctuations. Simulations are conducted based on the foregoing results. The ROC curves for each method are shown in the figure 2.

The blue solid line represents the SS-CPI method, the red solid line represents the NcI-COSCPI method, and the purple dashed line represents the SW-COSCPI method. When the target is a full-band target, the SS-CPI method demonstrates the poorest detection performance since it does not utilize inter-channel non-coherent integration and relies on a single channel for detection. The SW-COSCPI method shows marginally superior performance compared to the COS-CPI method. At a false alarm probability of 10^{-4}, the detection probability increases from 0.6 to 0.62 for SW-COSCPI. Both methods significantly outperform the SS-CPI method.

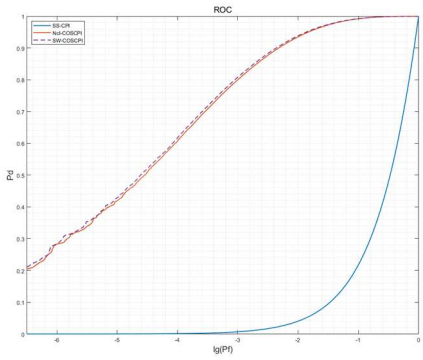


Figure 2 – The ROC curves of full-band target

3. New Target Detection Performance Analysis

Assuming each subpulse signal has a fixed length L, the i-th subpulse S_i can be partitioned into front segment S_i^1 and rear segment S_i^2 with lengths $L_{S_i^1}$ and $L_{S_i^2}$ respectively. When a new target emerges in the surveillance region, the source transmits S_i^2. The target echo only correlates with S_i^2. Under the H_0 hypothesis, the signal model remains unchanged. Under the H_1 hypothesis, since the matched filter employs energy normalization over L, when given a target S_i^2 which echo energy of $E_s^{CPS_i^2}$, the filter output energy is:

$$\breve{E}_s^{S_i^2} = E_s^{S_i^2} \cdot (L_{S_i^2} / L) \tag{10}$$

By replacing E_s with $\breve{E}_s^{S_i^2}$, the normalized energy of the new noise power $\hat{E}_s^{S_i^2}$ and the normalized signal-to-noise ratio of the matched filter output $SNR_{out}^{S_i^2}$ are obtained. When S_i belongs to NCPS, the energy of each subsequent sub-pulse channel equals the full-band energy; while under CPS, the energy of each sub-pulse channel gradually increases until reaching the level of NCPS channels.

Simulation Condition 3: Total channels: 5 (including 3 signal channels and 2 noise channels). The output

SNR of matched filtering SNR_{out} ranges uniformly from 3dB to 7dB, with the remaining channels being noise channels. The ROC curves of each method are illustrated in the figure 3.

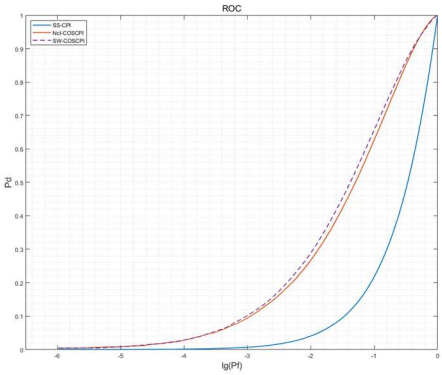


Figure 3 – The ROC curves of new target

The blue solid line in the figure represents the SS-CPI method, the red solid line denotes the NcI-COSCPI method, and the purple dashed line indicates the SW-COSCPI method. When detecting new targets, SW-COSCPI outperforms COS-CPI by increasing the detection probability from 0.26 to 0.29 at a false alarm probability of 10^{-2}. Additionally, it demonstrates improved performance across the false alarm probability range of $10^{-3.5}$ to $10^{-0.5}$.

4. CONCLUSIOIN

This paper analyzes the receiver signal processing method for CAS systems using LFMCW as the transmitted signal. To address the mutual constraint between detection gain and target revisit rate caused by non-overlapping sub-pulses in traditional methods, we propose the OS-CPI method. Detection results demonstrate that this method successfully breaks the mutual constraint. Building upon this, we further present the SW-COSCPI method, which employs shifting based on Doppler consistency and weighting based on multi-channel peak continuity in matched filtering outputs. Simulation results indicate that SW-COSCPI outperforms NcI-COSCPI in detecting both full-band targets and new targets. This proves that SW-COSCPI enables earlier capture of new target echoes while achieving superior detection performance for full-band targets. Optimal configuration of weight mapping parameters remains to be further discussed.

REFERENCES

1. Hines PC, Murphy SM, Abraham DA, Deane GB. The Dependence of Signal Coherence on Sea-Surface Roughness for High and Low Duty Cycle Sonars in a Shallow-Water Channel. Ieee Journal of Oceanic Engineering. 2017;42(2):298-318.
2. van Vossen R, Beerens P, van der Spek E. Anti-Submarine Warfare With Continuousiy Active Sonar TNO Tests the Principle of Continuously Active Sonar With the Interim Removable Low-Frequency Active Sonar System. Sea Technology. 2011;52(11):33-5.
3. Hines PC, Murphy SM, Hicks K, Taylor N. Measurements of Signal Coherence for High and Low Duty Cycle Sonars in a Shallow Water Channel. Oceans 2015 - Genova. 2015.
4. Munafo A, Canepa G, Lepage KD. Continuous Active Sonars for Littoral Undersea Surveillance. IEEE Journal of Oceanic Engineering. 2019;44(4):1198-212.
5. Murphy SM, Zurk LM, Coffin MEW. Processing Continuous Active Sonar Transmissions to Achieve an Arbitrarily Fast Update Rate With Full Instantaneous Bandwidth. Ieee Journal of Oceanic Engineering. 2021;46(3):988-99.
6. Liu DL, Zhao XC, Lu Y. Detection performance analysis of sub-band processing continuous active sonar. Applied Acoustics. 2021;174.

Fast Chirp Rate Estimation Method for LFM Signals in Impulsive Noise Environments

Shuxian HAO[123]; Longhao QIU[123]; Guolong Liang[123]; Zhe CAO[123]
[1] National Key Laboratory of Underwater Acoustic Technology, China
[2] Key Laboratory of Marine Information Acquisition and Security(Harbin Engineering University), China
[3] College of Underwater Acoustic Engineering, Harbin Engineering University, China

ABSTRACT
Underwater environments are often contaminated by intense impulsive noise, which significantly degrades the performance of traditional linear frequency modulated (LFM) signal chirp rate estimation methods. This study proposes a fast chirp rate estimation method for LFM signals, termed Sigmoid-IASR, which integrates Sigmoid filtering and stochastic resonance. In the preprocessing stage, a Sigmoid nonlinear filter is introduced to suppress impulsive noise without distorting the phase of LFM signals. The instantaneous autocorrelation (IA) method is then employed to transform the chirp rate estimation problem into a frequency estimation task for periodic signals, thereby reducing computational complexity. To address frequency estimation challenges in low signal-to-noise ratio (SNR) scenarios, a second-order stochastic resonance (SR) system with a genetically optimized bistable potential model is developed. Additionally, fractional Fourier transform (FrFT) is utilized to resolve the ambiguity in estimating the sign of the chirp rate caused by the SR system. Monte Carlo experiments demonstrate that Sigmoid-IASR achieves a correct estimation probability of 0.8 at an SNR approximately 3 dB lower than traditional IA method under impulsive noise. Numerical simulations confirm the effectiveness of the Sigmoid-IASR method in rapidly estimating LFM signal chirp rates under low SNR conditions with impulsive noise interference.

Keywords: Chip Rate Estimation, Instantaneous Autocorrelation, Stochastic Resonance, Genetic Algorithm

1. INTRODUCTION
Linear frequency-modulated (LFM) signal is a type of signal whose instantaneous frequency varies linearly with time. In underwater pulse detection systems, the rapid and accurate estimation of the chirp rate for LFM pulse signals actively emitted by non-cooperative targets has become a critical step in the data processing workflow.

The Instantaneous Autocorrelation (IA) algorithm is a widely used method for rapid chirp rate estimation.[1][2] However, since the IA algorithm relies on the signal's own delayed conjugate information, its estimation performance is highly susceptible to noise. Current improvements to the IA algorithm primarily focus on two aspects to enhance noise robustness and parameter estimation accuracy: improving the frequency estimation precision of the single-frequency signal, and combining IA with other parameter estimation methods to enhance adaptability under low signal-to-noise ratio (SNR) conditions. The Fractional Fourier Transform (FrFT) is another approach for estimating the chirp rate of LFM signals. Compared to the IA method, FrFT achieves higher estimation accuracy under low SNR conditions. However, when estimating the chirp rate of non-cooperative LFM signals, FrFT requires a two-dimensional peak search in the fractional Fourier domain, leading to high computational complexity and low efficiency. To address this issue, and to enable direct application of the IA algorithm to real-valued signals, IA-FRFT was proposed.[3] Due to the fact that IA processing of real signals causes ambiguity in the sign of the chirp rate, parallel FRFT operations are performed based on two opposite chirp rates. By comparing the maximum magnitudes of the two FRFT output results, the correct sign of the chirp rate can ultimately be determined.

[1] hsx@hrbeu.edu.cn
[2] qiulonghao@hrbeu.edu.cn
[3] liangguolong@hrbeu.edu.cn

Traditional algorithms model background noise as Gaussian-distributed. However, the marine environment is often subject to complex impulsive noise with non-Gaussian characteristics. The symmetric alpha-stable ($S\alpha S$) distribution is widely used to model underwater impulsive noise.[4] Since the alpha-stable distribution has finite moments only for orders lower than its characteristic exponent, algorithms based on second-order or higher-order statistics suffer from performance degradation under such noise conditions. Currently, there are two main approaches for suppressing impulsive noise in LFM chirp rate estimation: one category is fractional low-order (FLOS)-based methods[5]; The other type combines nonlinear filtering techniques. To address the challenge of single-frequency signal estimation under low SNR conditions, stochastic resonance (SR) systems are introduced. SR has been successfully applied in rotating machinery fault detection, and its exploration in underwater acoustic detection is ongoing. Dong et al. proposed a second-order parameter-matched stochastic resonance (PMSR) method, which outperforms first-order matched stochastic resonance (FMSR) in achieving low-noise output with effective bandpass filtering.[6]

To address the aforementioned challenges, this paper proposes a Sigmoid-filtering and stochastic resonance-enhanced instantaneous autocorrelation (Sigmoid-IASR) method for LFM signal chirp rate estimation.

2. SαS Distribution Noise Model

The $S\alpha S$ distribution model has been widely validated in numerous studies as an effective approach for modeling underwater impulsive noise, representing a special case of the more general alpha-stable distribution. The characteristic function of the alpha-stable distribution under the standard parameterization can be expressed as follows:

$$\varphi(t) = \exp\{jat - \gamma^{\alpha}|t|^{\alpha}[1 + j\beta\,\mathrm{sgn}(t)\omega(t,\alpha)]\} \tag{1}$$

where,

$$\omega(t,\alpha) = \begin{cases} -\tan(\pi\alpha/2) & \alpha \neq 1 \\ (2/\pi)\log|t| & \alpha = 1 \end{cases} \tag{2}$$

$$\mathrm{sgn}(t) = \begin{cases} 1 & t > 0 \\ 0 & t = 0 \\ -1 & t < 0 \end{cases} \tag{3}$$

Equation (1) indicates that the alpha-stable distribution model is determined by four parameters: the characteristic exponent α , dispersion coefficient γ , symmetry parameter β , and location parameter a .In the context of alpha-stable distributions, we adopt the generalized signal-to-noise ratio (GSNR), defined as:

$$GSNR = 10\lg(\sigma_s^2/\gamma_n) \tag{4}$$

where, σ_s^2 represents the signal power and γ_n denotes the dispersion parameter of the alpha-stable distribution, which serves as an analog to variance in non-Gaussian noise environments.

3. Sigmoid-IASR LFM Chirp Rate Estimation Method

1. Principle of Sigmoid nonlinear filters

The Sigmoid filter function is defined as follows:

$$\mathrm{Sigmoid}[x] = \frac{2}{1+\exp[-x]} - 1 \tag{5}$$

It is known that if $s(t)$ is a $S\alpha S$ process with $a = 0$ and $1 < \alpha \leq 2$, then Sigmoid[$s(t)$] exhibits finite second-order moments (referred to as a second-order moment process) with zero mean.

Furthermore, using Taylor series, it can be proven that the use of Sigmoid filters has no effect on the LFM modulation frequency slope. The expression for Sigmoid-IA results is as follows:

$$R^{\mathrm{Sigmoid}}(n,\tau) \approx \frac{A^2}{4}\exp[\,j(2\pi\beta\tau n + 2\pi f_1\tau + \pi\beta\tau^2)] + R_{\varepsilon}^{\mathrm{Sigmoid}}(n,\tau) + R_{s,\varepsilon}^{\mathrm{Sigmoid}}(n,\tau) \tag{6}$$

Equation (6) shows that while the Sigmoid filter effectively confines the amplitude of impulsive noise, it may simultaneously attenuate the amplitude of the monochromatic signal component carrying chirp information. This dual effect could lead to a reduction in the output SNR of the Sigmoid-IA estimator.

2. Principle of Genetic Optimization for SMSR

Stochastic resonance (SR) refers to the phenomenon of resonance generated by nonlinear oscillators under the influence of external random forces, which is a type of motion phenomenon caused by the combined effects of external driving forces and internal restoring forces. The following Duffing-Holmes nonlinear differential equation is used to describe the SMSR system:

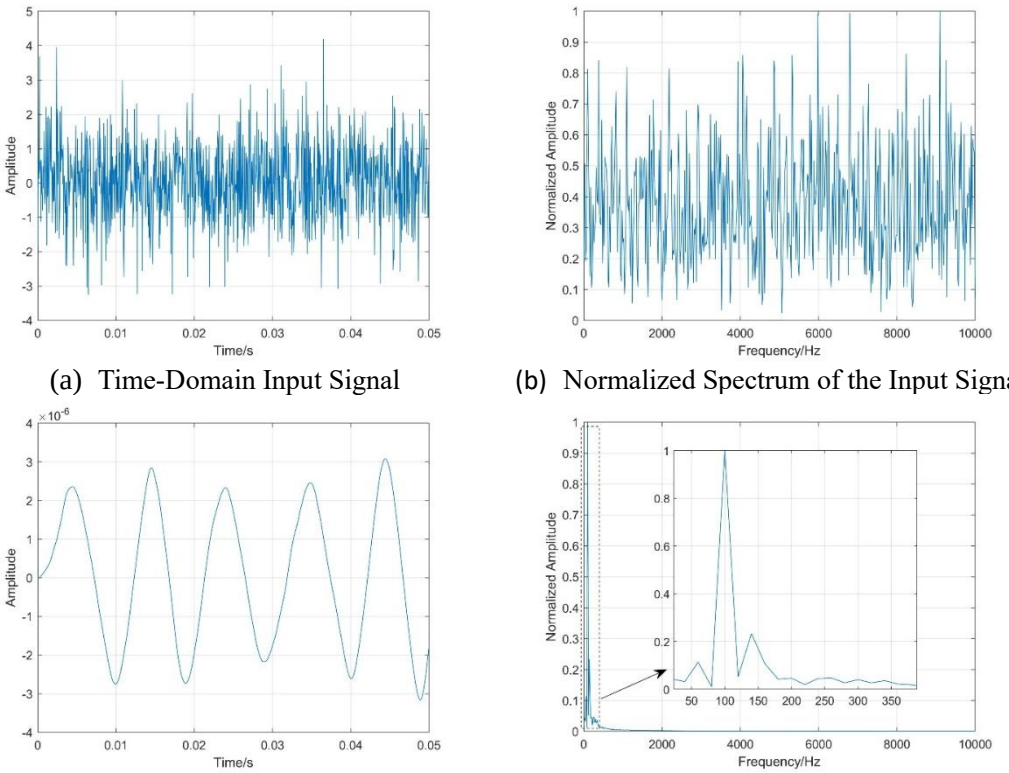
$$\frac{d^2x}{dt^2}+\gamma\frac{dx}{dt}=-\frac{dV(x)}{dx}+s(t)+n(t) \tag{7}$$

In Equation (7), γ represents the damping factor; $s(t)=A\cos(2\pi f_0 t+\varphi)$ denotes the weak monochromatic input signal to the system, where A, f_0, φ correspond to the amplitude, driving frequency, and initial phase of the periodic signal, respectively; $n(t)=\sqrt{2D}\xi(t)$ represents Gaussian white noise with zero mean and variance 2D, which can be generalized under appropriate conditions to additive noise with symmetric distribution and zero mean; $V(x)=-\frac{a}{2}x^2+\frac{b}{4}x^4$ is the bistable potential function.

By employing the genetic algorithm, system parameters can be adaptively selected to drive the system into a state of stochastic resonance. This paper constructs an optimization function based on the maximum output signal-to-noise ratio and the correlation coefficient between the output results and the input signal. The proposed optimization function F is formulated as follow:

$$\mathrm{F}=abs(\rho_{xy})\cdot\exp(SNRI) \tag{8}$$

Here, ρ_{xy} represents the normalized correlation coefficient between the input and output signals; $SNRI$ denotes the output SNR of the bistable stochastic resonance system, where $|Y(K_0)|$ is the maximum spectral envelope of the output signal, K_0 is the target frequency point, and $\sum|Y(K)|^2$ is the frequency-domain energy of the output signal. Consequently, the proposed optimization function yields a larger value when the output signal exhibits greater similarity to the input signal and a higher SNR at the target frequency point.



(a) Time-Domain Input Signal (b) Normalized Spectrum of the Input Signal

(c) GA-SMSR Output Signal (d) Normalized Spectrum of the Output Signal

Figure 1 – Input-Output Performance of the GA-Optimized Stochastic Resonance System

A numerical simulation is performed to validate the described methodology, with the key simulation parameters specified as follows: The simulation adopts a system sampling rate of

$f_s = 20\text{kHz}$. The input signal consists of a 100Hz single-frequency component with amplitude 0.1 and pulse width 0.05s, superimposed with Gaussian white noise (variance is 1.2). The calculated input SNR is approximately -23.8dB. Figures 1(a)-(b) present the time-domain waveform and corresponding frequency spectrum of the input signal, where the single-frequency component is practically indistinguishable due to the strong noise interference. The noisy signal is then processed by a SMSR system, whose parameters a , b and γ are iteratively optimized through the proposed GA. The GA configuration includes: population size is 20, crossover probability is 0.8, mutation probability is 0.3, and maximum generation is 30. Figure 2 illustrates the relationship between the normalized optimization function value of the SR system and GA iteration count. The results demonstrate that the output stabilizes after 6 generations of evolution. Figures 1(c)-(d) display the time-domain waveform and frequency spectrum of the optimized SR system output. Comparative analysis with Figures 1(a)-(b) clearly reveals successful extraction of the 100Hz single-frequency component.

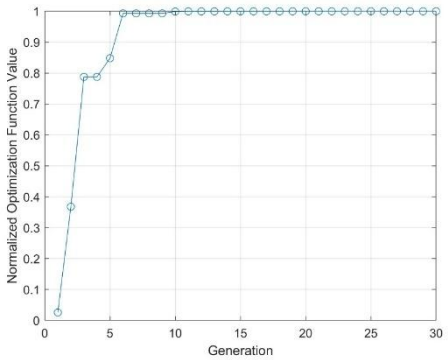


Figure 2 – Convergence characteristics of the normalized optimization function versus GA generation

3. Sigmoid-IASR Estimation of LFM Chirp Rate

Based on Equation (6), the real part of the Sigmoid-IA processed output can be simplified into the following mathematical model:

$$x(t) = s(t) + n(t) = A\cos(2\pi f_0 t) + n(t) \tag{9}$$

where $s(t)$ represents the single-frequency component in the Sigmoid-IA processed result, and $n(t)$ denotes the noise component. Under these conditions, the problem of chirp rate estimation is transformed into a frequency estimation problem for periodic signals in low SNR scenarios, for which SMSR systems can be employed. The SMSR system parameters can be selected adaptively using the method described in Subsection 3.2.

4. Simulation Experiment

The system sampling rate is set to $f_s = 100\text{kHz}$; The LFM signal has an amplitude of 1, a pulse width of 0.05s, and a bandwidth of 1kHz. For the IA method, the delay is configured as 0.4 times the pulse width, resulting in an output signal length of 3000 points. The theoretical single-frequency component of the output signal is 400 Hz. The FFT point number is uniformly set to twice the input signal length. The standard SαS distribution impulsive noise is characterized by a fixed stability index $\alpha = 1.3$. The GSNR ranges from -15 dB to 5 dB, with 1000 Monte Carlo trials conducted for each GSNR level to compare the estimation performance between the traditional method and the proposed approach.

To facilitate a fair comparison between the two algorithms, an estimation is considered accurate if and only if the detected peak frequency of the output signal is located precisely at 400 Hz. Thus, the estimation accuracy probability is defined as the ratio of the number of successful detections at 400 Hz to the total number of Monte Carlo trials under each GSNR condition. Additionally, the normalized root mean square error (NRMSE) is introduced to evaluate the stability of the estimation. The specific formula is given as follows:

$$NRMSE = \sqrt{\frac{1}{N}\sum\nolimits_N (y - \hat{y})^2} \Big/ \max(y) \tag{10}$$

where, N denote the total number of estimation trials, y represent the estimated frequency results, and $\hat{y}$ indicate the expected frequency value (in this experiment, $\hat{y} = 400\text{Hz}$). The term $\max(y)$ corresponds to the maximum estimated frequency value observed.

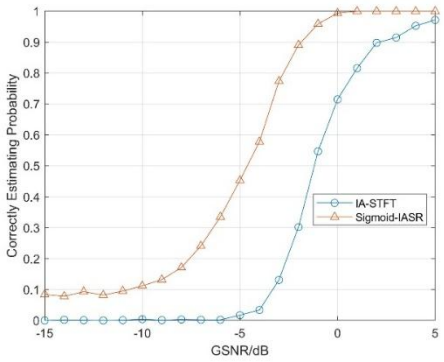


Figure 3 – Sigmoid-IASR Algorithm Process

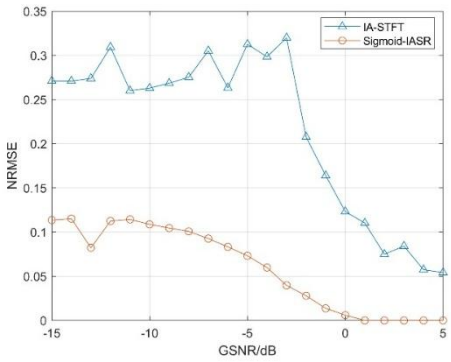


Figure 4 – Sigmoid-IASR Algorithm Process

Figure 3 compares the estimation accuracy rates of both algorithms versus GSNR variation. The results demonstrate that under the configured simulation conditions, the conventional IA-STFT algorithm requires $GSNR > 1\text{dB}$ to achieve an estimation accuracy rate exceeding 0.8, while the proposed Sigmoid-IASR method attains comparable accuracy at $GSNR > 3\text{dB}$. Figure 4 presents the estimation stability comparison between both algorithms as a function of GSNR. In the simulated environment, the proposed Sigmoid-IASR method achieves $NRMSE < 0.05$ when $GSNR > 3\text{dB}$, whereas the IA-STFT method maintains $NRMSE > 0.05$ throughout the entire simulated GSNR range. These results confirm that the proposed method exhibits superior estimation stability compared to conventional IA-STFT in impulsive noise environments.

4. CONCLUSIONS

This paper proposes a novel LFM chirp rate estimation method named Sigmoid-IASR to overcome the performance degradation of conventional algorithms in impulsive noise conditions. The proposed method is an improvement on the IA method: First, the Sigmoid function is selected as a filter to suppress pulse noise, and it is proven that it has no effect on the frequency modulation slope of the LFM signal; Second, an SMSR system optimized by a genetic algorithm is selected to address the problem of estimating the frequency of unknown single-frequency signals under low signal-to-noise ratio conditions. Compared with the traditional IA-FrFT method, under the same GSNR pulse noise conditions, the proposed method significantly improves the accuracy of estimating the frequency modulation slope of LFM signals and provides more stable estimation results than traditional methods. How to further extract the motion state of the target of interest based on the Sigmoid-IASR results is the focus of future research.

REFERENCES

1. Peleg S, Porat B. Linear FM signal parameter estimation from discrete-time observations[J]. IEEE Transactions on Aerospace and Electronic Systems, 1991, 27(4):607-616.
2. Song, J, Sun N, et al. A hybrid algorithm for fast parameter estimation of LFM signal[C]. 2017 IEEE 17th International Conference on Communication Technology (ICCT). IEEE, 2017.
3. Wang M, Li T, Detection Performance of LFM Signal Based on Instantaneous Autocorrelation and FRFT[J]. Electronic Information Warfare Technology, 2013, 28(5). P.1-5.
4. Song Guoli, Guo Xinyi. α stable distribution model in ocean ambient noise. ACTA ACUSTICA. 2019;44(2):177-187.
5. Liu W. Time delay estimation for low SNR signals with impulsive noises using fractional low order covariation sequences. IEEE 2012 International Conference on Audio, Language and Image Processing (ICALIP), Shanghai, 2012. P. 647-650.
6. Dong H T, Wang H Y, Shen X H, et al. Parameter Matched Stochastic Resonance with Damping for Passive Sonar Detection[J]. Journal of Sound and Vibration, 2019, 458(1): 479-496.

PROCEEDINGS of the
16th International Conference on Theoretical and Computational Acoustics

August 18 to 22, 2025 BEXCO, Busan Korea

Model Order Reduction for Horizontal Refraction Equation Solutions: An Efficient Framework for Directional Three-Dimensional Acoustic Wave Propagation Modeling

Tengjiao He[1]
[1]Shanghai Jiao Tong University
[2]Institute of Acoustics, Chinese Academy of Sciences, People's Republic of China
[3]Sun Yat-sen University, People's Republic of China
[4]Institute of Acoustics, Chinese Academy of Sciences, People's Republic of China

ABSTRACT

The modeling of three-dimensional (3D) underwater acoustic propagation has evolved into a mature discipline, achieving significant advancements in computational efficiency, predictive accuracy, and operational capability through collaborative research efforts over the past two decades. While conventional approaches typically assume monopole-like sound sources, this fundamental premise frequently fails in practical applications where various underwater noise sources exhibit pronounced directionality. This study introduces a model order reduction (MOR) framework for solving horizontal refraction equations, enabling derivation of 3D underwater acoustic propagation from directional sources. The MOR formulation decomposes vertical mode amplitudes into orthogonal transverse eigenfunctions with corresponding expansion coefficients. By reformulating the Huygens volume source integral using MOR principles, we derive a semi-analytical solution that circumvents traditional numerical integration requirements. Numerical experiments validate the method's ability to accurately capture directional spatial distributions of the sound field due to a volume source. Remarkably, computational demands are significantly reduced through a single propagation implementation.

[1] hetengjiao@sjtu.edu.cn

DAILY PROGRAM

AUGUST 20 (WED)

09:00 – 10:40	Environmental Effects on Deep Ocean Acoustic Propagation	324 (3F)
Chair	Wei-Chun Hu (National Taiwan University)	
09:00	The effects of Indian Ocean Dipole on the convergence zone in the Eastern Indian Ocean	
	Shuanglin Wu, Jixing Qin, Zhenglin li	
09:20	Mechanisms of Convergence Zone Splitting in Deep-Sea Acoustic Propagation	
	Fujin Yang, Tao Hu, Zhen Wang, Pengxi Zhou	
09:40	Research on sound propagation modeling and characteristics in narrow channels under rough ice layer cover	
	Pengcheng Chai	
10:00	Analysis of Very-Low Frequency Acoustic Signals Obtained by OBS in Deep Ocean	
	Yunuo Jin, Lijia Gong	
10:20	Analysis of surface vessel radiation noise measured by OBS in deep ocean	
	Yuchen Jiang, Lijia Gong, Jiayao Shi	

PROCEEDINGS of the
16th International Conference on Theoretical and Computational Acoustics

August 18 to 22, 2025 BEXCO, Busan Korea

The effects of Indian Ocean Dipole on the convergence zone in the Eastern Indian Ocean

Shuanglin Wu[1], Jixing Qin[2], Zhenglin li[3]

[1]The Institute of Acoustics of the Chinese Academy of Sciences
[2]The Institute of Acoustics of the Chinese Academy of Sciences
[3]School of Ocean Engineering and Technology, Sun Yat-sen University

ABSTRACT

The physical ocean phenomena in the ocean affect the changes of water body, which has an important impact on the sound propagation. A hyperbolic frequency modulation (HFM) signal sound propagation experiment of towed sound source in the East Indian Ocean (EIO) was conducted in summer 2019. The influence of the physical ocean phenomenon—Indian Ocean Dipole (IOD) on sound propagation was observed. The processing and analysis of the sound propagation experimental data through the IOD, the effects of sound source depth and water fluctuation on the deep-sea CZ were analyzed. The results show that the second convergence zone (CZ) is not formed under the influence of the warm water mass formed by the IOD and the depth of the sound source in the incomplete deep channel environment of the EIO. Under the influence of the IOD, the thermocline fluctuation at the location of the second CZ will have an important impact on the formation and location of the third CZ. It is found that the location of the third CZ shifts 2-3 km to the direction of the sound source in the experiment. The formation mechanism of sound propagation phenomenon in the experiment was explained theoretically.

[1] wushuanglin@mail.ioa.ac.cn
[2] qjx@mail.ioa.ac.cn
[3] lizhlin29@mail.sysu.edu.cn

Mechanisms of Convergence Zone Splitting in Deep-Sea Acoustic Propagation

Fujin YANG[123]; Tao HU[12]; Zhen WANG[12]; Pengxi ZHOU[123];
[1] Institute of Acoustics,Chinese Academy of Sciences, China
[2] Key Laboratory of Underwater Acoustic Environment, Chinese Academy of Sciences, China
[3] University of Chinese Academy of Sciences, China

ABSTRACT

Convergence zone (CZ) splitting challenges traditional acoustic propagation models, yet its physical mechanisms and its dependence on sound speed profile remains unclear. This study investigates the splitting phenomenon of CZ observed in low-frequency deep-sea acoustic propagation experiments, where double-peak structures with a 19.5 km separation and a 16.67 dB energy difference between peaks and troughs were detected. Using ray-mode theory, we simulated CZ splitting dynamics with varying propagation distances and receiver depths. Two key mechanisms were identified: (1) For receivers shallower than the source, large and small grazing-angle ray clusters within the critical angle gradually separate with increasing range, exhibiting cumulative distance effects that widen peak spacing. (2) At deeper receiver depths, upward and downward ray clusters form distinct peaks, with peak separation widening as depth increases.

Keywords: Underwater acoustics, Deep-sea propagation, Convergence zone, Ray-mode theory

Research on sound propagation modeling and characteristics in narrow channels under rough ice layer cover

Pengcheng CHAI [1,2,3]; Xiaohan WANG [1,2,3] ; Ge YU [1,2,3] ; Shengchun PIAO [1,2,3]
[1] National Key Laboratory of Underwater Acoustic Technology, Harbin Engineering University Harbin 150001, China
[2] Key Laboratory of Marine Information Acquisition and Security (Harbin Engineering University), Ministry of Industry and Information Technology Harbin 150001, China
[3] College of Underwater Acoustic Engineering, Harbin Engineering University Harbin 150001, China

ABSTRACT
Based on the theory of layered elastic media and ray propagation, this study develops an sound propagation prediction model for narrow channels under rough ice layer cover, incorporating the effects of ice-water interface roughness and complex channel topography. The model treats the ice layer as an elastic layered medium with a rough interface. A one-dimensional Gaussian-distributed random rough ice-water interface is simulated using the Monte Carlo method combined with measured interface data. The ice-water reflection coefficient is computed by a layered elastic finite element model (LEFEM), which handles the rough interface with larger root mean square (RMS) height than the wavelength, validated against the wavenumber integration model. Ray propagation theory is then applied to analyze and predict acoustic characteristics in the rough ice-covered narrow channel. Numerical and experimental results demonstrate that phase and amplitude interference between reflections from the ice layer and channel bottom partially offsets energy loss. This leads to a gentler transmission loss compared to models considering only topography and ice roughness. The predicted sound transmission loss shows small errors relative to measured data, confirming the model's validity.
Keywords: Rough Ice Layer, Narrow Channels, Sound Transmission Loss

1. INTRODUCTION

Recent global warming trends have enhanced opportunities for polar development and Arctic route utilization, elevating the region's strategic significance [1]. Concurrently, advancements in active sonar detection and torpedo acoustic homing technologies underscore the critical need to study mid-to-high-frequency acoustic propagation in narrow channels beneath rough ice cover. The ice-covered narrow channel impacts sound propagation through three primary mechanisms: Ice-induced positive sound speed gradient bends acoustic rays upward, forming a surface duct through reflection and scattering at the rough ice-water interface [2,3]. As an elastic medium, ice causes transmission loss and interface scattering [4]. Complex seabed topography refracts and scatters acoustic energy [5-8]. Repeated interactions with boundaries dissipate energy through reflection, scattering, and absorption, creating intricate sound fields and multipath structures. Thus, acoustic modeling must address reflections at both the rough ice-water interface and complex seabed boundaries.

The rough ice-water interface exhibits complex, spatiotemporally varying random undulations, prompting extensive research. Diachok [9] modeled sea ice as uniformly distributed semi-elliptical rigid cylinders at the free surface, calculating reflection coefficients using the Burke-Twersky scattering theory (BT model) [10]. Zhu Guangping employed the BT model combined with ray theory to study acoustic communication performance in Arctic sub-ice channels [11]. However, the BT model neglects ice elasticity and phase information, potentially causing significant errors. Kuperman [12] integrated wavenumber integration with perturbation methods to derive low-frequency reflection loss at rough elastic interfaces, computing Arctic propagation loss and simulating vertical array receptions—yet omitting mid/high frequencies. Hope [13] utilized the OASES wavenumber integration model to investigate ice roughness effects on long-range

[1] cpc_just@163.com
[2] wangxiaohan@hrbeu.edu.cn
[3] yuge211@hrbeu.edu.cn
[4] piaoshengchun@hrbeu.edu.cn

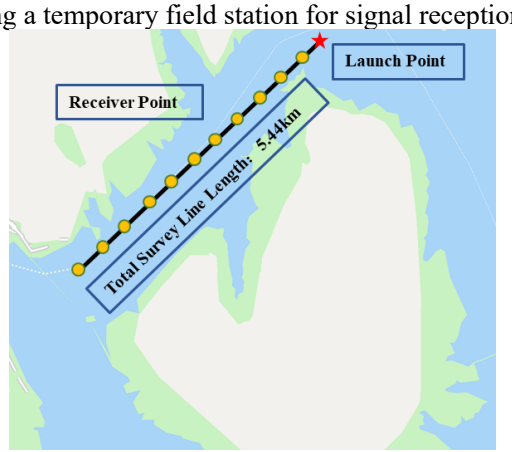
propagation and arrival structures, identifying a roughness limit (<0.6 m RMS height at 900 Hz). While OASES is among the most suitable models for rough sea ice, its perturbation-based implementation ignores large-scale surface features, restricting applicability to small RMS-height-to-wavelength ratios. Simon [14] applied the finite element method (FEM) to model ice as either rough elastic or pressure-release boundaries, revealing greater far-field propagation loss for elastic ice. Li [15] used COMSOL (FEM-based) to solve coherent reflection/scattering coefficients for layered porous ice, validating porous ice models. Although FEM enables acoustic modeling under ice cover—including near-ice variations, marginal ice zones, and porous ice—it becomes computationally prohibitive for mid/high frequencies in large domains.

Although the aforementioned research addresses the impact of scattering at the rough ice-water interface on sound propagation, it neglects the applicability relationship between ice-water interface roughness and frequency. Furthermore, it fails to account for the coupled effects of complex topography in narrow straits and the rough ice-water interface on acoustic propagation. From an applied perspective, this paper transforms the rough ice-water interface into an equivalent layered, uneven elastic rough boundary and establishes a coupled LEFEM-BELLHOP model. In the second section, acquisition of environmental data for sound propagation experiments in narrow channels under rough ice layer cover. In the third section, development of a sound propagation prognostic model for narrow straits under rough ice cover. In the fourth section, the experimental sound transmission data processing are introduced in detail. Finally, the conclusions are given.

2. SOUND PROPAGATION Experiments in NARROW CHANNELS under ROUGH ICE LAYER COVER

2.1 Experiment Overview

To investigate sound propagation characteristics in narrow straits under rough ice cover, field experiments were conducted on Songhua Lake in Jilin City, Jilin Province. Songhua Lake, the largest lake in Jilin Province, has an average depth of 22 meters and a maximum depth exceeding 70 meters, featuring islands. During February, the average outdoor temperature is approximately -15°C, with ice thickness ranging from 0.3 to 0.8 meters in the test area. Figure 1(a) shows the experimental area map, indicating the acoustic source, receiver positions, and measurement line length. The stationary source, housed in a tent (Fig. 1(b)①), transmitted signals. Receivers were deployed at 11 positions from 0.5 km to 5.44 km range at 0.5-km intervals (Fig. 1(b)②-⑫), each requiring a temporary field station for signal reception.



(a) Songhua lake experimental area

(b) Songhua lake measurement points

Figure 1 – Songhua lake field setup

2.2 Sound Velocity Profile, Topography and Sediment of Narrow Channel

During under-ice experiments on Songhua Lake, full-depth CTD casts were performed at 1-km intervals. Due to highly variable bathymetry, sound velocity profiles were calculated using data from the deepest CTD cast at each location. Temperature-depth-salinity data from CTD were processed with Mackenzie's equation to derive sound velocity profiles (Fig. 2). Bathymetry in the test area (Fig. 3) showed significant variations, with depths ranging from 11.27 m to 40.29 m. The lakebed consisted of silt-clay sediment with acoustic properties: sound speed 1510 m/s, density 1.42 g/cm³, absorption coefficient 0.02 dB/λ.

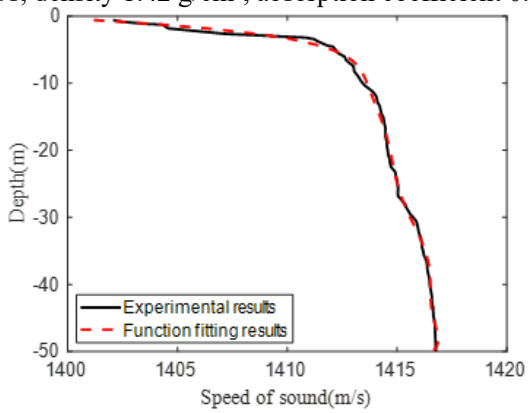


Figure 2 – Experimental area sound velocity profile

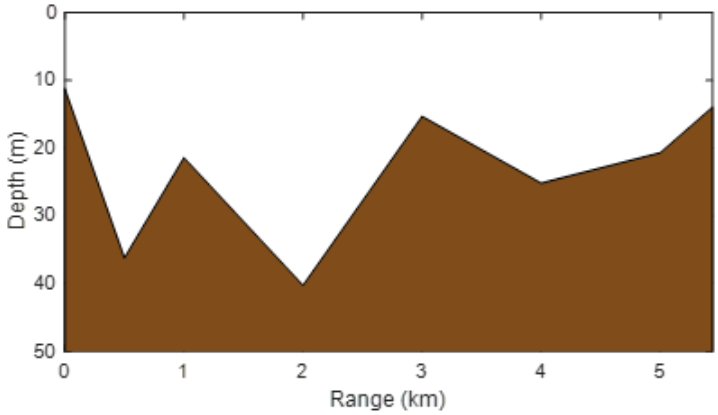


Figure3 – Bathymetry of songhua lake test area

3. ROUGH ICE-WATER INTERFACE and SOUND PROPAGATION MODELs

3.1 Modeling of Rough Ice-Water Interface

Songhua Lake ice is predominantly first-year ice, formed from initial ice growth within a single winter, with thicknesses of 0.3–1 m. Its ice-water interface was thus modeled using a Gaussian-distributed randomly rough surface generated via Monte Carlo methods [16,17].

One-dimensional rough surface samples of length L can be given by Eq. (1):

$$f(x_n) = \frac{1}{L}\sum_{j=-N/2+1}^{N/2} F(k_j) e^{ik_j x_n} \tag{1}$$

where $x_n = n\Delta x \quad (n = -N/2+1, \cdots, N/2)$ denotes the n-th sampling point on the rough surface, $F(k_j)$and $f(x_n)$form a Fourier transform pair defined as:

$$F(k_j) = [2\pi L S(k_j)]^{1/2} \cdot \begin{cases} \frac{[N(0,1)+iN(0,1)]}{\sqrt{2}} & j \neq 0, N/2 \\ N(0,1) & j = 0, N/2 \end{cases} \tag{2}$$

where $k_j = 2\pi j/L$ defines the discrete wavenumber, $N(0,1)$denotes a normally distributed random number with zero mean and unit variance, and $S(k_j)$represents the power spectral density (PSD) of the rough surface. The PSD for the Gaussian-distributed random rough surface is expressed as:

$$S(k_j) = \frac{\delta^2 l}{2\sqrt{\pi}} \exp\left(-\frac{k_j^2}{4}\right) \tag{3}$$

where δ denotes the root-mean-square height deviation, l represents the correlation length. For $j > 0$, $F(k_j)$satisfies the conjugate symmetry condition $F(k_j) = F(k_{-j})^*$. This formulation yields the rough ice-

water interface model.

Based on Songhua Lake ice characteristics, actual ice draft profiles from similar conditions were selected—specifically a segment measured in November 2005 along 73°N, 203.8°W to 73.2°N, 203.8°W (https://nsidc.org/) (shown in Fig.4). Processing this data yielded statistical parameters for the ice draft profile: mean ice thickness, root-mean-square (RMS) height, and correlation length (shown in Table 1).

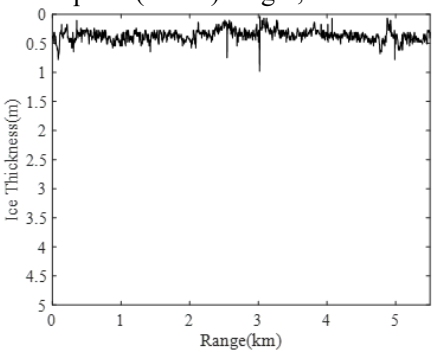
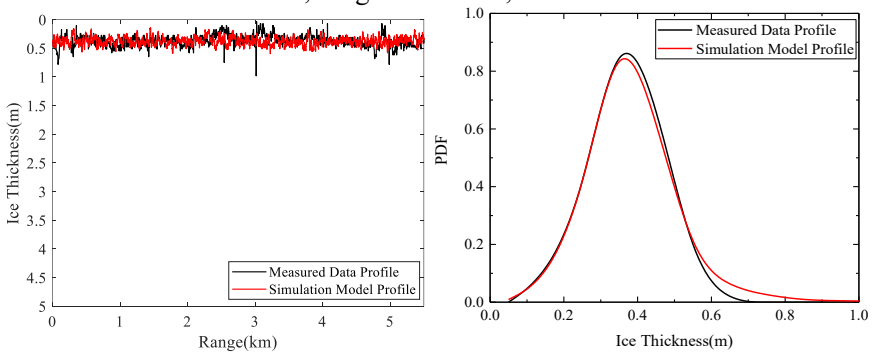


Figure4 – 2005 Ice Draft Profile Segment

Table 1 –Statistical Parameters of the 2005 Ice Draft Profile

Statistical parameter	Mean Ice Thickness t/m	RMS Height δ/m	Correlation Length l/m
Value	0.3800	0.0986	24

To validate the Monte Carlo method for simulating Gaussian-distributed rough ice-water interfaces, measured and simulated ice draft profiles were compared through probability density functions (PDFs) and draft distributions (Fig. 5). The results show close agreement: while exhibiting local differences, the profiles share consistent overall vertical variations, roughness scales, and statistical characteristics.



(a) Ice Thickness (b) PDFs

Figure5 –Measured and Simulated Ice Thickness and PDFs

3.2 LEFEM-BELLHOP Coupling Model

First, the computational capability of the COMSOL software (based on the finite element method) is constrained by both frequency and domain size. Consequently, it cannot predict sound propagation characteristics in narrow straits under rough ice cover for high-frequency, large-domain scenarios. Second, while the OASES-BELLHOP coupled model overcomes frequency/domain limitations, OASES (Ocean Acoustics and Seismic Exploration Synthesis) implements interface roughness via perturbation theory. This restricts applicability to surfaces with RMS heights much smaller than the acoustic wavelength, neglecting large-scale features—a key cause of instability in OASES' rough-interface reflection coefficient calculations (Fig.6). Ice-water interface parameters are listed in Tables 1 and 2 [18]. Fig. 6 demonstrates that coherent reflection coefficients exceeding 1 at certain grazing angles confirm OASES' inapplicability when RMS height is significant relative to wavelength.

Table 2 – Ice Physical-Acoustic Parameters

Parameter Name	Value
Ice Density ρ_{ice}	$0.9 \times 10^3 \mathrm{kg/m^3}$
Compressional Wave Speed c_{p}	$3600\mathrm{m/s}$
Shear Wave Speed c_{s}	$1800\mathrm{m/s}$
Compressional Wave Attenuation α_{p}	$0.6\mathrm{dB/m}$ /kHz （$0.216\mathrm{dB}/\Lambda$）
Shear Wave Attenuation α_{s}	$0.36\mathrm{dB/m}$ /kHz （$0.648\mathrm{dB}/\Lambda$）

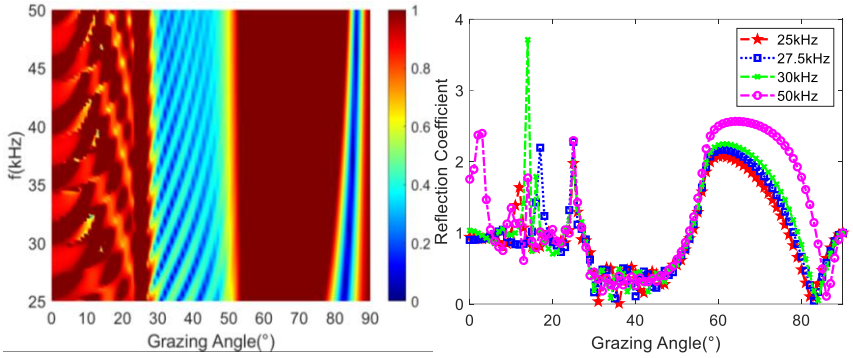


Figure6 –OASES-Coherent Reflection Coefficient versus Grazing Angle and Frequency for Rough Ice Cover

To overcome these limitations, this study employs a coupled LEFEM-BELLHOP model for sound propagation prediction in narrow straits under rough ice cover. First, a one-dimensional Gaussian-distributed rough ice-water interface is simulated via Monte Carlo methods using statistical parameters (ice thickness, RMS height, correlation length). This interface is then imported into COMSOL Multiphysics to construct a 2D periodic domain model comprising PML-Air Layers-Ice Layers -Water Layers -PML (Fig. 7). The domain width W equals the acoustic wavelength, with layer heights H_{air}, H_{ice}, and H_{water}. Perfectly Matched Layers (PMLs) of thickness λ/2 truncate the computational domain. LEFEM-calculated ice reflection coefficients were validated against OASES and the impedance transfer method [19] under uniform ice density, smooth interface, and 3-m ice thickness conditions (acoustic parameters in Table 2). Fig. 8 demonstrates close agreement among all three methods, confirming the validity of the LEFEM approach.

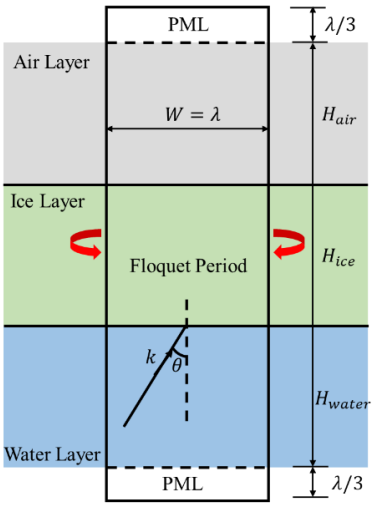
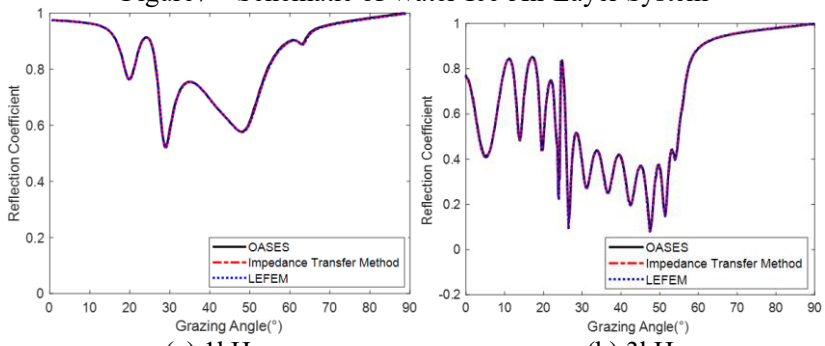


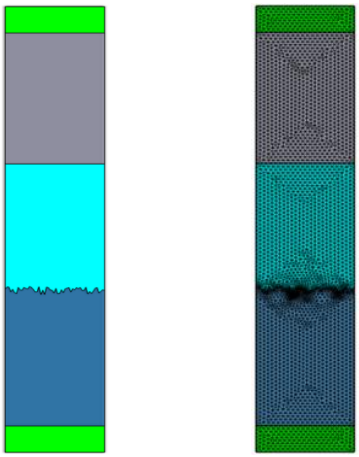
Figure7 – Schematic of Water-Ice-Air Layer System



(a) 1kHz (b) 3kHz

Figure8 –Planar Sea-Ice Reflection Coefficient: Computational Results Comparison

Figure 9 shows the LEFEM geometry and mesh model for the air-rough ice-water system. The computed rough ice-water interface reflection coefficient versus grazing angle and frequency (Fig. 10) exhibits consistent trends across 25 kHz–50 kHz: coefficients initially decrease then increase with grazing angle, asymptotically approaching unity. Higher frequencies reduce fluctuation amplitudes, while reflection coefficients decrease at fixed grazing angles. Critically, all values remain below 1.



(a) Geometry Model (b) Mesh Model

Figure9 – LEFEM of Air-Rough Ice-Water Layers

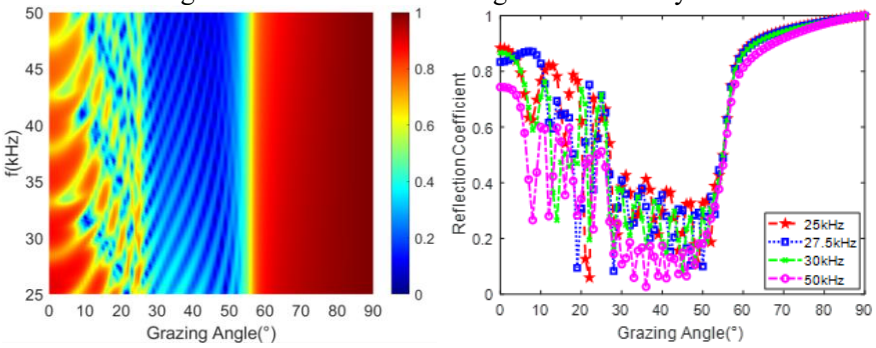


Figure10 –LEFEM-Coherent Reflection Coefficient versus Grazing Angle and Frequency for Rough Ice Cover

The derived rough ice-water interface reflection coefficients, sound velocity profile, and range-dependent bathymetry/sediment parameters were imported into the BELLHOP acoustic model for predicting sound propagation in narrow straits under rough ice cover. BELLHOP is a canonical ray-theory model [20], where ray acoustics approximates the wave equation under high-frequency conditions defined by:

$$f > 10\frac{c}{H} \tag{4}$$

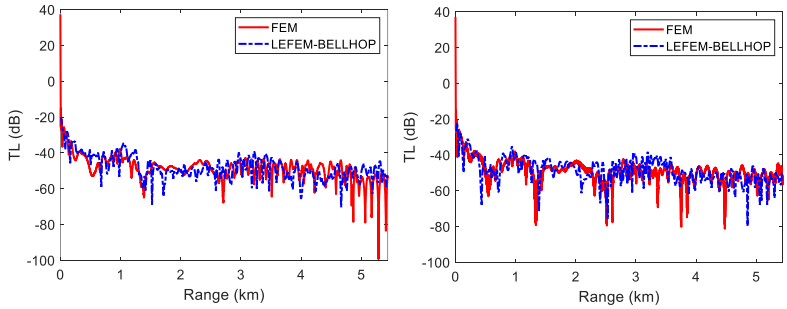
where c is sound speed and $H$is water depth. Ray acoustics treats sound propagation as bundles of rays normal to wavefronts, with each ray perpendicular to its local wavefront. Equations (5) and (6) are the Eikonal and Transport equations—the fundamental equations of ray theory—determining ray direction, trajectory, and travel time.

$$\nabla\varphi^2 = (\frac{c_0}{c})^2 = n^2(x,y,z) \tag{5}$$

$$\nabla\cdot(A^2\nabla\varphi) = 0 \tag{6}$$

where φ is the eikonal, c_0 the reference sound speed, c the local sound speed, ∇ the gradient operator, and A the wave amplitude.

To validate the LEFEM-BELLHOP coupled model for predicting sound propagation in narrow channel under rough ice layer cover, FEM simulations were used for verification. At test frequencies of 300 Hz and 500 Hz with source/receiver depths of 7 m and 10 m, transmission loss predictions from both methods show close agreement (Fig. 11). This confirms the LEFEM-BELLHOP model's validity for forecasting sound propagation characteristics in such environments.



(a) 300Hz (b) 500Hz

Figure11 –Prediction of Transmission Loss in Narrow Channels Under Rough Ice Layer Cover

4. EXPERIMENTAL DATA PROCESSING AND ANALYSIS

Under-ice sound propagation experiments acquired time-domain signals from co-depth hydrophones at multiple ranges. These signals were processed to derive transmission loss values at each range, following these steps:

(1) Time-domain signals received by a co-depth hydrophone were truncated to yield $x(t)$;

(2) The truncated signal $x(t)$underwent Fourier transformation to obtain spectrum $X(\omega)$, followed by normalization $X(\omega)/N$;

(3) Identify amplitudes A_i at target frequencies/bands, then compute the energy spectrum as $(\mathrm{abs}(A_i))^2$;

(4) Transmission loss (TL) is defined as the difference between source level (SL) and received level (RL):

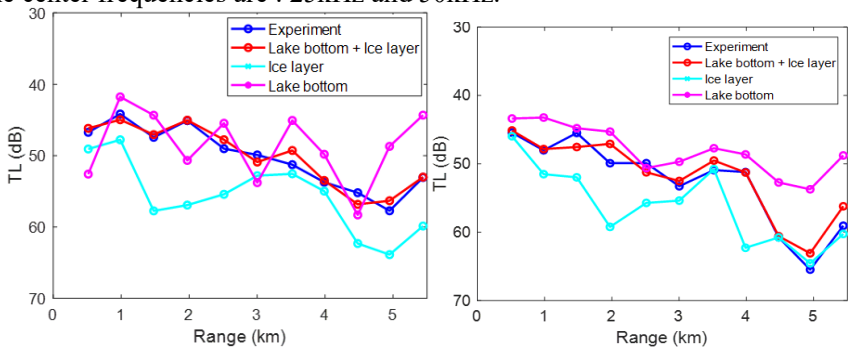
$$\mathrm{TL} = \mathrm{SL} - \mathrm{RL} \tag{7}$$

$$\mathrm{SL} = 10 * \log10(2 * (\mathrm{abs}(A_L))^2) - (\mathrm{G}_L + \mathrm{Me}_L) \tag{8}$$

$$\mathrm{RL} = 10 * \log10(2 * (\mathrm{abs}(\mathrm{A}_R))^2) - (\mathrm{G}_R + \mathrm{Me}_R) \tag{9}$$

Where A_L denotes transmit signal amplitude at the target frequency, $\mathrm{G}_L$the gain of the transmit-side system, and Me_L the sensitivity of the transmit-side hydrophone; A_R denotes received signal amplitude at the target frequency, G_R the gain of the receive-side system, and Me_R the sensitivity of the receive-side hydrophone.

According to Eqs.(7)-(9), Fig.12 shows the variation curves of transmission loss with receiving distance measured at two different center frequencies and sound source depth and receiving depth of 7m and 10m, respectively. The center frequencies are : 25kHz and 50kHz.



(a) 25kHz (b) 50kHz

Figure12 – Comparison of experimental data and simulation results

Combined with the reflection coefficient of the lake bottom topography and the rough ice layer, it can be seen that the transmission loss of the experimental measurement shows a sharp fluctuation with distance and frequency, especially at 3.5km, the transmission loss of all frequencies appears obvious wavelet peaks but quickly tends to be gentle, which may be caused by the combined action of the convex topography and the rough ice layer at 3km.At 4.5km, there are obvious troughs in the transmission loss of all frequencies, which may be caused by the combined action of the concave topography and the rough ice layer at 4km. Based on the sound propagation characteristics of flat terrain and the experimental results of narrow channel and undulating terrain under ice cover, it can be concluded that the undulating lakebed topography and ice roughness induce anomalous transmission loss fluctuations at specific ranges, but phase-amplitude interference between ice and lakebed reflections partially offsets energy dissipation, yielding smoother propagation loss profiles than predicted by isolated terrain or ice models.

5. CONCLUSIONS

Based on the theory of layered elastic media and ray propagation, considering the influence of ice-water interface roughness and the complex terrain of narrow channel, the FEM model is coupled with the ray theory model Bellhop, and the FEM-Bellhop coupling model is established to predict the acoustic propagation characteristics of narrow channel under rough ice layer cover, which is verified with the experimental data of Songhua Lake. the research results demonstrate :

(1) The LEFEM accurately predicts coherent reflection coefficients at rough ice-water interfaces, which handles the rough interface with larger RMS height than the wavelength, and validated by OASES and the impedance transfer method.

(2) The LEFEM-Bellhop coupled model accurately predicts transmission loss in narrow channel under rough ice layer cover, validated against FEM simulations at low frequencies and experimental measurements

at mid-to-high frequencies, confirming model efficacy.

(3) Undulating lakebed topography and ice roughness induce anomalous transmission loss fluctuations at specific ranges, but phase-amplitude interference between ice and lakebed reflections partially offsets energy dissipation, yielding smoother propagation loss profiles than predicted by isolated terrain or ice models.

ACKNOWLEDGEMENTS

We acknowledge all the participants for their hard work and perseverance, which provided extremely valuable experimental data for this article.

REFERENCES

1. LI Q H, WANG N, ZHAO J P, et al. Arctic underwater acoustics: an attractive new topic in ocean acoustics [J]. Journal of Applied Acoustics, 2014, 33(6): 471-483.
2. Dan Hutt; An overview of Arctic Ocean acoustics. AIP Conf. Proc. 5 November 2012; 1495 (1): 56–68.
3. Christie A. O'Hara, Jon M. Collis; Underwater acoustic propagation in Arctic environments. J. Acoust. Soc. Am. 1 October 2011; 130 (4_Supplement): 2529.
4. YIN L, WANG N, YIN J W, et al. Polar underwater acoustic signal processing research[J]. Bulletin of Chinese Academy of Sciences, 2019, 34(3): 306-313.
5. N. Ross Chapman, Gordon R. Ebbeson; Acoustic shadowing by an isolated seamount. *J. Acoust. Soc. Am.* 1 June 1983; 73 (6): 1979–1984.
6. S. E. Dosso, N. R. Chapman; Measurement and modeling of downslope acoustic propagation loss over a continental slope. *J. Acoust. Soc. Am.* 1 February 1987; 81 (2): 258–268.
7. Tappert FD, Spiesberger JL, Wolfson MA. Study of a novel range-dependent propagation effect with application to the axial injection of signals from the Kaneohe source. J Acoust Soc Am. 2002 Feb;111(2):757-62.
8. Liang Xu, Haigang Zhang, Minghui Zhang; Training a deep operator network as a surrogate solver for two-dimensional parabolic-equation models. *J. Acoust. Soc. Am.* 1 November 2023; 154 (5): 3276–3284.
9. O. I. Diachok; Effects of sea-ice ridges on sound propagation in the Arctic Ocean. J. Acoust. Soc. Am. 1 May 1976; 59 (5): 1110–1120.
10. J. E. Burke, V. Twersky; Scattering and Reflection by Elliptically Striated Surfaces. J. Acoust. Soc. Am. 1 October 1966; 40 (4): 883–895.
11. ZHU G P, YIN J W, CHEN W J, et al. Modeling and characteristics the typical under-ice acoustic channel for the Arctic [J]. *Acta Acustica*, 2017, 42(2): 152-158.
12. W. A. Kuperman, Henrik Schmidt; Self-consistent perturbation approach to rough surface scattering in stratified elastic media. J. Acoust. Soc. Am. 1 October 1989; 86 (4): 1511–1522.
13. Gaute Hope, Hanne Sagen, Espen Storheim, Halvor Hobæk, Lee Freitag; Measured and modeled acoustic propagation underneath the rough Arctic sea-ice. J. Acoust. Soc. Am. 1 September 2017; 142 (3): 1619–1633.
14. Blake Simon, Marcia Isakson, Megan Ballard; Modeling acoustic wave propagation and reverberation in an ice covered environment using finite element analysis. Proc. Mtgs. Acoust. 7 May 2018; 33 (1): 070002.
15. Li S, Liu S, Yuan S, Wen J, Zhang Z. Reflection of Acoustic Wave through Multilayered Porous Sea Ice Sandwiched between the Water and Air Half-Spaces. Applied Sciences. 2021; 11(16):7411.
16. GUO L X. Fundamental Theory and Methods of Scattering from Randomly Rough Surfaces[M]. Beijing: Science Press, 2010.
17. Thorsos E I , Jackson D R . The validity of the perturbation approximation for rough surface scattering[J]. The Journal of the Acoustical Society of America, 1987, 81(S1):20.
18. Laible, Henry A . Temporal evolution of under ice reflectivity[J]. Acoustical Society of America Journal, 1996, 99(2):851-865.
19. Diana F. McCammon, Suzanne T. McDaniel; The influence of the physical properties of ice on reflectivity. *J. Acoust. Soc. Am.* 1 February 1985; 77 (2): 499–507.
20. Porter M B. The BELLHOP manual and user's guide: Preliminary draft. Heat, Light, and Sound Research, Inc., La Jolla, CA, USA, Tech.Rep,2011.

Analysis of Very-Low Frequency Acoustic Signals Obtained by OBS in Deep Ocean

Yunuo JIN[1,2,3]; Lijia GONG[1,2,3]
[1] National Key Laboratory of Underwater Acoustic Technology, Harbin Engineering University, Harbin 150001, China
[2] Key Laboratory of Marine Information Acquisition and Security (Harbin Engineering University), Ministry of Industry and Information Technology; Harbin 150001, China
[3] College of Underwater Acoustic Engineering, Harbin Engineering University, Harbin 150001, China

ABSTRACT
This study analyzes very-low frequency (1-100 Hz) acoustic signals using continuous Ocean-Bottom Seismometer (OBS) data collected in the South China Sea from August to October 2022. Power Spectral Density (PSD) analysis was applied through time-frequency methods to evaluate the impacts of disturbance sources including typhoons, earthquakes, and shipping activities. Probability Density Function (PDF) was employed to analyze the statistical distribution characteristics of the ambient ocean noise field. Comparisons were made with the classical Wenz noise spectrum to elucidate the features of ambient ocean noise within the 1-100 Hz band. The results reveal that typhoon-induced microseisms increased the ambient ocean noise level by 6 dB within the 3-10 Hz range; the time-frequency diagrams of shipping noise exhibit periodic line spectra, reflecting vessel motion parameters; and a magnitude 6.5 earthquake located 531 km from the OBS significantly affected ambient ocean noise at frequencies below 15 Hz, with rapid attenuation observed at higher frequencies (>15 Hz).

Keywords: Ocean ambient noise, Ocean-Bottom Seismometer, Power spectral density analysis, Probability density function

1. INTRODUCTION
Ocean Ambient Noise is a fundamental component of the underwater sound field, resulting from a combination of various natural and anthropogenic sound sources, including wind and waves, rainfall, marine biological activities, earthquakes, volcanic eruptions, and ship navigation. In 1948, Knudsen [1] and Wenz [2] conducted extensive research on the frequency and intensity of noise through statistical analysis of a substantial amount of noise data, revealing the primary sources of marine noise across different frequency bands. In 1962, Wenz systematically summarized the spectral characteristics of ocean ambient noise and introduced the renowned 'Wenz curve spectrum.' This spectrum has been widely utilized as a fundamental model for predicting the spectral levels of ocean ambient noise and has illustrated the universal principles of ocean ambient noise in a relatively detailed manner.

However, the intensity and characteristics of ocean ambient noise, particularly within the 1-100 Hz very-low-frequency (VLF) band, are significantly influenced by specific sound sources and may exhibit spatial and temporal variations. A thorough understanding of the source contributions, statistical properties of noise within this frequency band, and its similarities to and differences from classical models (e.g., the Wenz spectrum) is of great significance for refining noise models and evaluating the impact mechanisms of sound sources.

This study is based on continuous passive acoustic monitoring data acquired by an ocean-bottom seismometer (OBS) deployed in the deep ocean at 3800 meters regions in the South China Sea from August to October 2022. It aims to systematically investigate the characteristics of ocean ambient noise and its dominant sources within the 1-100 Hz frequency band. The research employs power spectral density (PSD) time-frequency analysis to identify and quantify the contributions of different noise sources. Concurrently,

[1] jinyunuo@hrbeu.edu.cn
[2] lijia.gong@hrbeu.edu.cn

the probability density function (PDF) is utilized to analyze the statistical distribution properties of the noise field. The objectives are to explore the temporal dynamics of noise in the VLF band, assess the significant impacts of specific events (e.g., typhoon passage, vessel transit, earthquake events) on the noise field, and compare the measured results with the classical Wenz noise spectrum to elucidate the distinctive characteristics of VLF noise in the South China Sea.

2. Describtion of the experiments

During the experimental period, continuous passive acoustic time-series measurements were acquired using an OBS deployed in the deep-sea region of the South China Sea. The OBS was positioned on the seafloor at a depth of 3800 meters. The OBS simultaneously sampled four channels: P (hydrophone pressure), X (horizontal north-south seismometer), Y (horizontal east-west seismometer), and Z (vertical seismometer), with a sampling rate of 250 Hz.

Figure 1 displays the spectrogram of ocean ambient noise within the 1-100 Hz band during the measurement period. It can be observed that the primary sources of ocean ambient noise are wind-noise, ship-generated noise, and earthquake events.

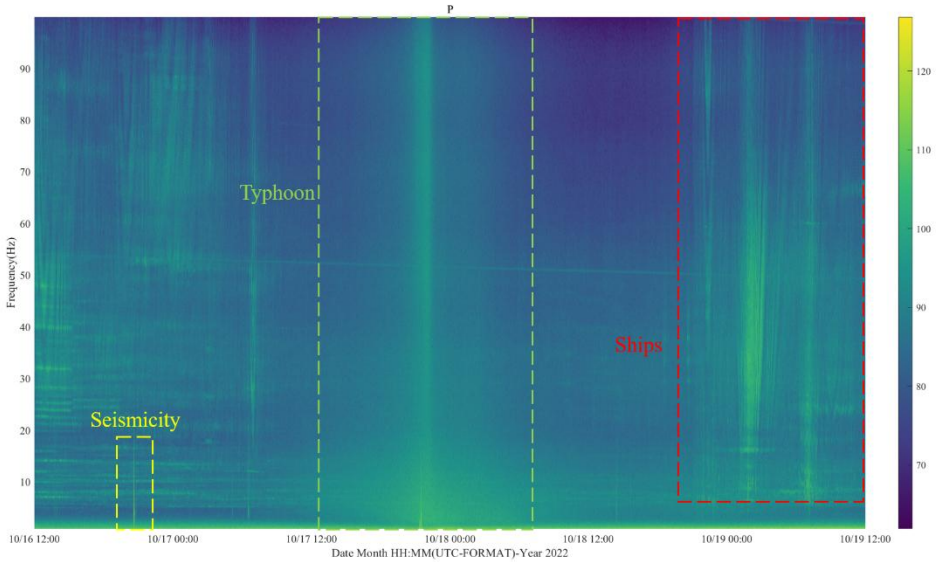


Figure 1 – Time-frequency diagram of ocean ambient noise (1-100 Hz)

3. Wind noise

This section focuses on the noise field characteristics during the passage of typhoon Nesat. Typhoon Nesat was upgraded to a severe tropical storm at 08:00 (local time) on 16 October 2022. Figure 2 (a) displays the typhoon's track and the location of the OBS. The closest point of approach (CPA) between the OBS and the typhoon center was approximately 85.67 km. The typhoon dissipated on 20 October 2022. According to data provided by the Copernicus Marine Environment Monitoring Service (CMEMS), peak wind speeds during landfall reached approximately 40 to 42 m/s. Figure 2 (b) shows the time-frequency diagram of Typhoon Nesat. It can be observed that the VLF noise level significantly increased during typhoon landfall.

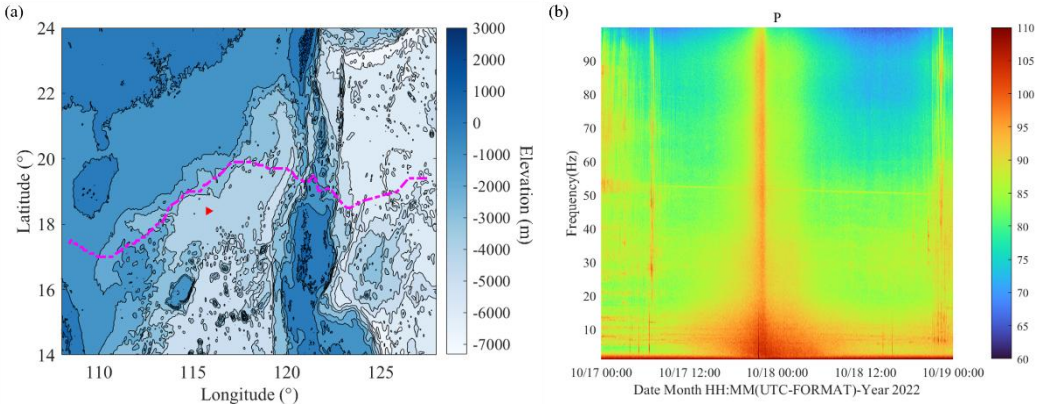


Figure 2 – (a) Typhoon Nesat Track, (b) Time-Frequency Diagram During Typhoon Period

Figure 3 shows a comparison of noise levels before, during, and after the typhoon. Maximum wind speeds occurred from 18 to 19 October. The results indicate that during typhoon landfall, high wind speeds were the primary factor driving noise level increases. Noise levels increased with wind speed but decreased with frequency. Typhoon-induced microseisms elevated ocean ambient noise by 6 dB within the 3-10 Hz frequency range, as shown in the figure.

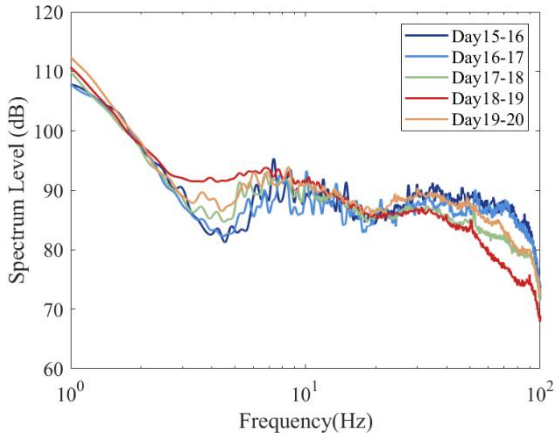


Figure 3 – Comparison of spectral level changes before and after the arrival of the typhoon

4. Ship-generated noise

The ocean ambient noise can be decomposed into a linear combination of wind-induced noise and ship noise, and the contribution of ship noise to the ocean ambient noise at different frequencies can be quantified by using change of the vertical coherence at a specific frequency along a vertical distance [3]. Vessel-radiated noise signatures depend on multiple factors, such as vessel type, size, and speed [4, 5].

This section focuses on shipping noise during the period from 7 to 9 August 2022. According to historical wind speed data obtained from the CMEMS, marine wind speeds remained consistently low throughout this interval. Wind noise exhibited negligible effects on low-frequency ocean ambient noise levels. These conditions provided an optimal environment for isolating and characterizing vessel-radiated noise signatures.

Figure 4 displays the time-series diagrams and corresponding time-frequency diagrams of shipping noise as a vessel transited near the receiving point. Distinct U-shaped striations are observable in the diagrams, indicating a stable range-frequency interference structure within the acoustic field of the study area. The temporal position corresponding to the vertex point of these striations marks CPA as the vessel passed the OBS.

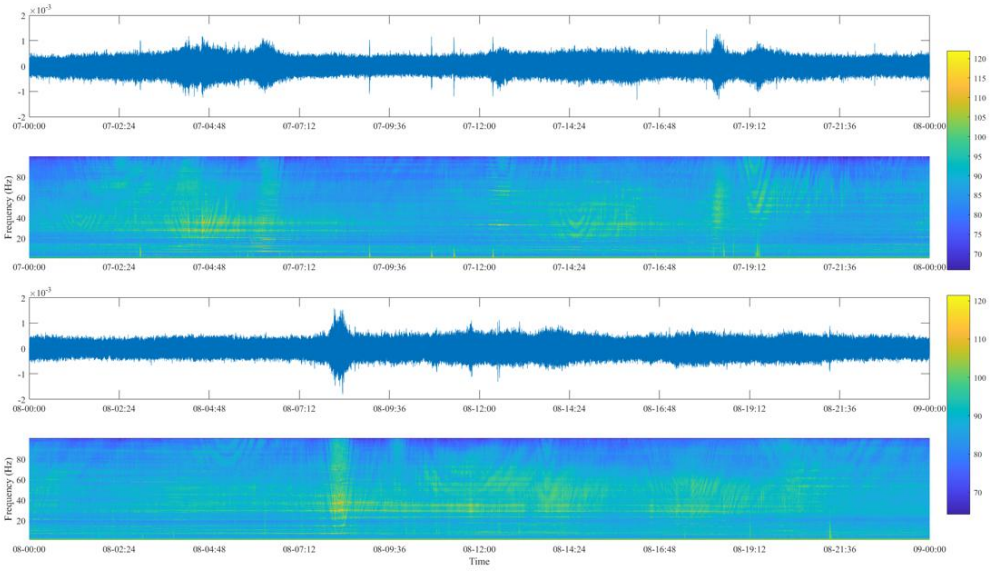


Figure 4 – Time-Series and Time-Frequency Diagrams of Vessel-Radiated Noise

Table 1 – Main information of the ships

Date and time of ship (UTC format-Date Month, HH:MM)	IMO	Ship Type	Length, m	Width, m	Draught, t	Speed, m/s
7 Aug 2022,04:48 Hrs	9605231	Container	300	48	12.3	19.5
7 Aug 2022,04:48 Hrs	9836490	Cargo ship	229	32	12.4	12
7 Aug 2022,14:35 Hrs	9354234	Cargo ship	117	19	8.9	10
7 Aug 2022,18:09 Hrs	9579872	Cargo ship	292	45	18.1	10.8
7 Aug 2022,19:22 Hrs	9632765	Container	221	35	9.6	15.8
8 Aug 2022,08:32 Hrs	9790050	Container	211	33	11.1	18.7
8 Aug 2022,11:51 Hrs	9769635	Cargo ship	128	20	9.5	12.1

Table 1 presents vessel information acquired via the Automatic Identification System (AIS) during the experimental period. Comparative analysis between time-frequency diagrams and vessel data reveals two key relationships: For vessels of identical type with comparable length or beam dimensions, higher transit speeds correlate with narrower U-shaped striations. At similar speeds, larger vessel dimensions correspond to wider striations.

As the Wenz curve applies to frequencies >1 Hz [2], Figure 5 demonstrates remarkable agreement between measured underwater ambient noise and the Wenz spectrum for heavy traffic noise within this frequency band.

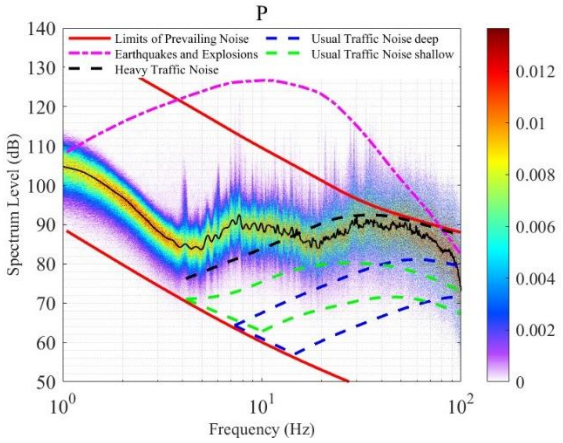


Figure 5 – Comparison of the PDF distribution and Wenz curve of ship-generated noise

5. Earthquake events

This section documents a magnitude 6.5 earthquake recorded by the OBS based on seismic observation data. According to information from the China Earthquake Networks Center (CENC), the earthquake occurred on 25 October 2022 at 22:58:59 UTC (06:58:59 local time, 26 October), with its epicenter located at 17.70°N, 120.80°E (Philippines) and a focal depth of 10 km.

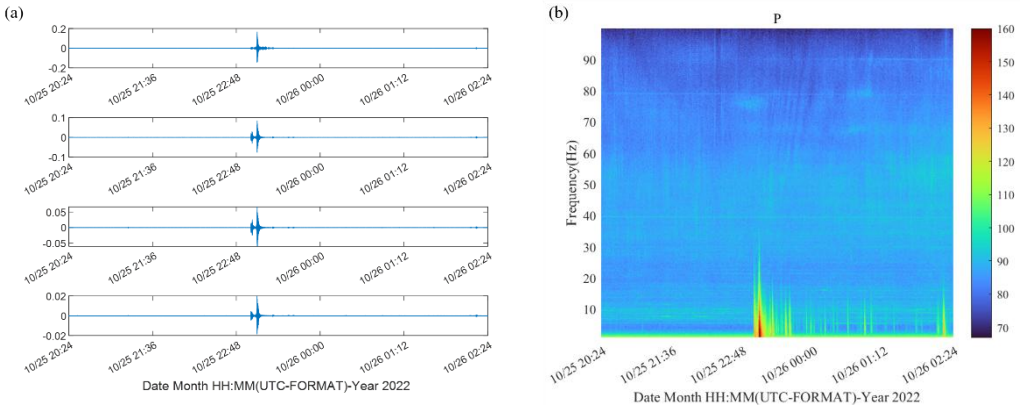


Figure 6 – (a) Seismic time-domain diagram, (b) Seismic time-frequency diagram

According to the Haversine formula

$$d = 2r \cdot \arcsin\left(\sqrt{\sin^2\left(\frac{\Delta\phi}{2}\right) + \cos\phi_1 \cos\phi_2 \cdot \sin^2\left(\frac{\Delta\lambda}{2}\right)}\right) \quad (1)$$

where r denotes Earth's radius, $\Delta\varphi$ represents the latitude difference, and $\Delta\lambda$ denotes the longitude difference. These parameters yield an approximate epicentral distance of 531 km between the OBS and the hypocenter. Given the P-wave propagation velocity of 5-8 km/s, the theoretical travel time is 1 min 29 s; for S-waves propagating at 3-4 km/s, the travel time is 2 min 32 s. These calculated arrival times align precisely with the seismic phases observed in Figure 6(b), confirming this event as the recorded earthquake. Figure 6(b) demonstrates that the earthquake significantly impacted ocean ambient noise at frequencies below 15 Hz. Subsequent aftershocks and post-seismic acoustic energy exhibited rapid attenuation above this threshold frequency.

6. CONCLUSIONS

Based on continuous OBS observations at 3,800-meters depth in the South China Sea, this study employs integrated time-frequency analysis of PSD and PDF statistics to effectively separate and quantify contributions from three primary noise sources—typhoons, shipping, and earthquakes—alongside their dynamic characteristics. The features and dominant sources of VLF ocean ambient noise were systematically investigated, yielding the following conclusions:

Wind noise serves as the primary driver of ambient noise in the VLF band, with noise levels exhibiting significant correlation with wind speed. Higher wind speeds correspond to elevated noise levels, while noise levels decrease with increasing frequency. During typhoon events, ambient noise levels are elevated by approximately 6 dB within the 3-10 Hz frequency band.

Under low-wind conditions, ship-generated noise constitutes the dominant source across the 1-100 Hz frequency band. Its time-frequency representations exhibit characteristic U-shaped striations. Higher vessel speeds yield narrower U-shaped striations. Larger vessel dimensions correspond to wider striations. These regular line spectra providing a basis for vessel motion identification and monitoring.PDF distribution of ship-generated noise within this band shows remarkable agreement with the heavy traffic noise spectrum described by the Wenz curve. This validates the applicability of classical models for shipping noise prediction in the South China Sea.

Submarine earthquakes similarly exert significant influence on ocean ambient noise, predominantly within the VLF band below 15 Hz. Seismic events and their resultant aftershocks or echoes are distinctly identifiable in time-frequency representations but exhibit rapid attenuation characteristics at frequencies exceeding 15 Hz.

ACKNOWLEDGEMENTS

This work is supported by the National Natural Science Foundation of China (Grant No. 12204128).

REFERENCES

1. Knudsen V O R, Alford R S, Emling J W. Underwater ambient noise. Journal of Marine Research, 1948,7(3):410-429.
2. Wenz, Gordon M. Acoustic Ambient Noise in the Ocean: Spectra and Sources. Journal of the Acoustical Society of America, 1962, 34(12):1936-1956.
3. Liu X, Li Q, Tang R, et al. A combined noise source model based on vertical coherence to quantify the proportions of two types of noise power. Journal of the Acoustical Society of America, 2025, 157(1).
4. Gassmann M, Wiggins S M, Hildebrand J A. Deep-water measurements of container ship radiated noise signatures and directionality. Journal of the Acoustical Society of America, 2017, 142(3):1563-1574.
5. Mckenna M F, Ross D, Wiggins S M, et al. Underwater radiated noise from modern commercial ships. Journal of the Acoustical Society of America, 2012, 131(1):92-103.

Analysis of surface vessel radiation noise measured by OBS in deep ocean

Yuchen JIANG[1,2,3]; Lijia GONG[1,2,3]; Jiayao SHI[1,2,3]
[1] National Key Laboratory of Underwater Acoustic Technology, Harbin Engineering University, Harbin 150001, China
[2] Key Laboratory of Marine Information Acquisition and Security (Harbin Engineering University), Ministry of Industry and Information Technology, Harbin 150001, China
[3] College of Underwater Acoustic Engineering, Harbin Engineering University, Harbin 150001, China

ABSTRACT
Vessel noise is an important source of marine ambient noise. An ocean-bottom seismometer (OBS) was used to record noise of different kinds of vessels, which was deployed at a depth of approximately 3800 m in the South China Sea. Vessel trajectories were obtained according to data of the Automatic Identification System (AIS). The signal-to-noise ratio (SNR) of the vessel radiation noise compared to background noise was analyzed. It is found that the SNR calculated from the vertical particle velocity channel data of vessel radiation noise is higher than that of the other channels in the near distance. This phenomenon is due to the combined effect of low-frequency marine ambient noise directivity and the dipole directivity of the vertical particle velocity channel of the OBS.

1. INTRODUCTION
Marine ambient noise is a common background acoustic field in the ocean. The temporal, frequency, and spatial characteristics of marine ambient noise affect the performance of underwater acoustic equipment. The study of marine ambient noise initially focused on its frequency and intensity characteristics, such as the Wenz curve (1). The Wenz curve analyzed various noise sources and their predominant frequency bands in the ocean. With the development of research, marine ambient noise has gradually become the main content in underwater acoustics research. Marine ambient noise is composed of diverse noise sources, such as vessel traffic, human activities, and distant storm-induced waves (2). Due to the varied spatial distributions of noise sources and the different marine ambient parameters, the spatial characteristics of marine ambient noise fields are also heterogeneous. Marine ambient noise is anisotropic and exhibits complex spatial directivity (3,4). Relevant studies have shown that vessel radiation noise is the main source of low-frequency marine ambient noise. The vessel radiation noise has significant influence on sonar performance and underwater monitoring. The noise above 1 kHz is primarily from the sea surface; the noise below 1 kHz, particularly around 100 Hz, is mainly from the horizontal direction. Contributions of noise from the sea surface or seabed are minimal. It indicates that vessel traffic is the dominant source of noise in this spectral band (1,3).

Traditionally, sound pressure hydrophones or hydrophone arrays have been commonly used to measure radiated noise. Because the hydrophone arrays exist problems such as large volume, cumbersome deployment and difficult transportation, vector hydrophones have gradually become an important focus of research. Compared to sound pressure hydrophones, vector hydrophones possess frequency-independent dipole directivity. It can simultaneously measure sound pressure and particle velocity. Therefore, acoustic field information can be acquired completely. As a result, a single vector hydrophone can be used to measure low-frequency vessel noise, and the amplitude variations of sound pressure and particle velocity recorded by vector hydrophone can be analyzed.

We deployed an OBS at approximately 3800 meters depth in the South China Sea and used it to monitor ambient noise for two months. Combined with real-time vessel trajectory obtained according to data of the AIS, the sound pressure and particle velocity amplitude variation tendency of different kinds of vessels are

[1] jiangyuchen501@hrbeu.edu.cn
[2] lijia.gong@hrbeu.edu.cn
[3] Shijiayao@hrbeu.edu.cn

analyzed.

2. EXPERIMENT AND OBSERVABLES

The experimental sea area is located to the north of the Zhongsha Islands and south of Pratas Island in the South China Sea. This area is a typical deep-sea at a depth of approximately 3800 meters, located near major shipping lanes with fewer small vessel traffic. The underwater acoustic characteristics are significant, and it is highly meaningful to carry out marine trials. This study analyzes the noise from August to September 2022 and the LOFARgrams are shown in Figure 1. Typical noise sources can be observed by the OBS in this area, such as easily identifiable vessels, whale calls, and seismic events.

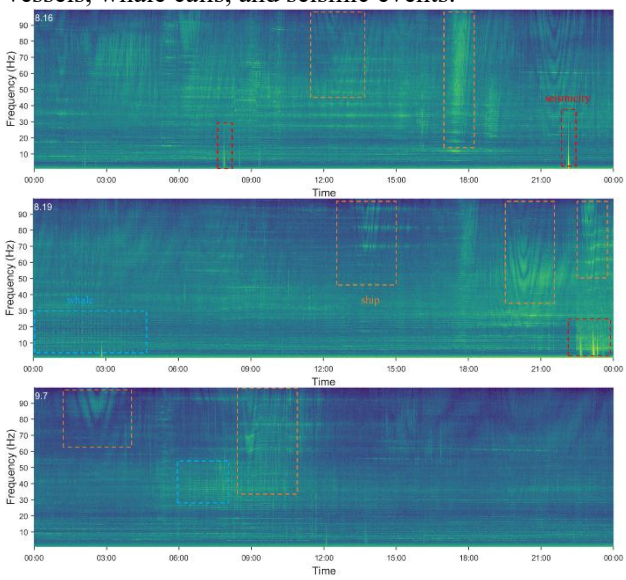


Figure 1 – The LOFARgram of typical noise sources.

3. METHODOLOGY

3.1 Signal Model

According to acoustic theory, the sound field is the vector field, and merely measuring the sound pressure in the sound field cannot analyze various phenomena in the sound field. Sound pressure and particle velocity are the fundamental physical quantities in the acoustic field, and the relationship between them constitutes the basis of vector field characteristics.

According to Euler 's equation, the relationship between sound pressure and particle velocity is

$$\boldsymbol{v}(r,t) = 1/\rho c \cdot p(r,t)\boldsymbol{n} \tag{1}$$

In an acoustic medium defined within a three-dimensional Cartesian coordinate system, when the far-field condition is satisfied between a sound source and a receiver, the acoustic wave radiated by the sound source can be approximated as a plane wave propagating to the receiver. The expressions for sound pressure and the three orthogonal components of particle velocity in the harmonic plane wave field are respectively given as

$$\begin{cases} p(t) = p(t) \\ v_x(t) = p(t)\cos\theta\cos\alpha \\ v_y(t) = p(t)\sin\theta\cos\alpha \\ v_z(t) = p(t)\sin\alpha \end{cases} \tag{2}$$

In the above expression, the constant term $1/\rho c$ is neglected.

3.2 Target Detection

Under the conditions of an ideal homogeneous medium, harmonic sound field, and far-field plane-wave propagation (5), the relationship between pressure sensitivity M_p and velocity sensitivity M_v can be expressed as

$$M_p = (1/\rho c)M_v \tag{3}$$

The relationship among the voltage data received by the four channels of the vector hydrophone is derived from Eqs. (1) and (3) as

$$V_p^2 = V_x^2 + V_y^2 + V_z^2 \tag{4}$$

Here, V_p, V_x, V_y, and V_z denote the voltage data of the four channels of the vector hydrophone respectively. According to Eq. (4), the square of the P channel voltage data is converted to the sum of the squares of the X, Y, and Z channel voltage data.

To analyze the amplitude characteristics of each channel of the vector hydrophone during vessel passages, the experimental data are first processed. The noise data are analyzed in both temporal and frequency domains, as shown in Fig. 2(a) and (b). With a sampling rate of 250 Hz, a temporal resolution of 5 s, and a frequency resolution of 0.2 Hz, the data processing flow is as follows:

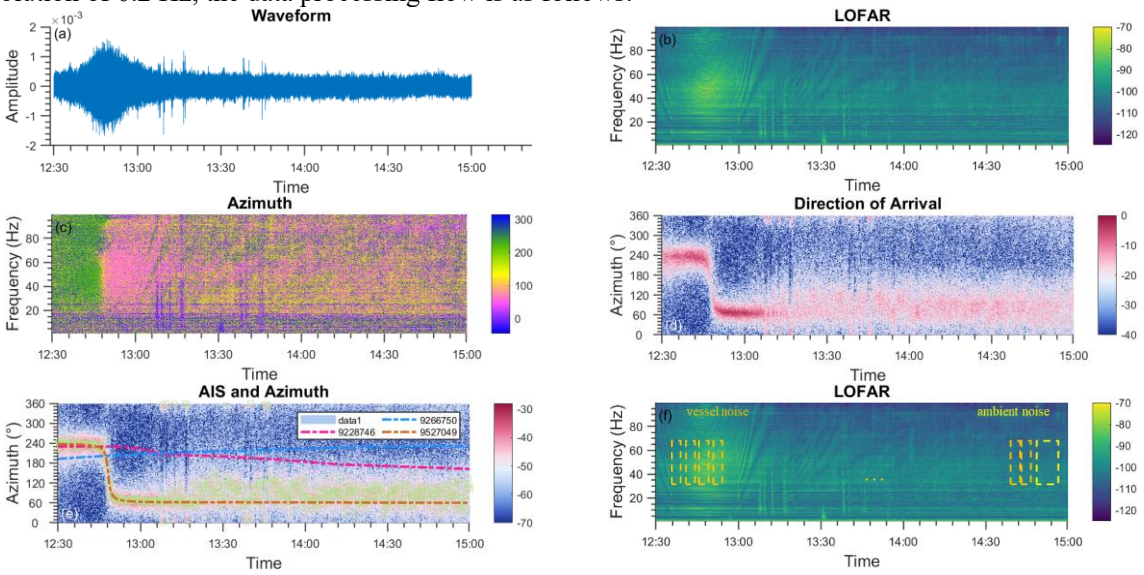


Figure 2 – Data processing flow diagrams. (a) Temporal domain plot. (b) Frequency domain plot. (c) Azimuth angle time series. (d) Azimuth angle estimation plot. (e) Angle estimation and AIS trajectory correspondence diagram. (f) Signal and noise extraction time.

At each discrete moment, the power spectrum of the received signal was analyzed using the pwelch method. The interference fringes of distinct noise sources are observed by plotting the LOFARgram, as shown in Fig. 2(b). To assess the target-detection performance of each vector hydrophone channel, prior knowledge of the target's state is required. This study uses transiting vessels for analysis. The amplitude of each channel change significantly when the ship passes the closest point of approach (CPA). To reduce interference of other vessels, the time period of only a single vessel passing should be judged combined with the target azimuth estimation value. The Bayesian method (6) is briefly introduced herein. Firstly, the complex acoustic intensity is applied to process the noise data and the azimuth can be obtained, as shown in Fig. 2(c). Subsequently, the posterior probability is calculated by the prior probability and likelihood probability of vessel azimuth estimation value, as shown in Fig. 2(d). By correlating vessel trajectories obtained according to data of AIS during this period, the azimuth between the passing vessels and the receiver are plotted. According to azimuth estimation value (Fig. 2(e)), the single-vessel transit period and the vessel type can be identified.

3.3 Definition of SNR

The SNR between Z and P channels is expressed as follows:

$$\begin{aligned} SNR_P &= 10\lg\left(\frac{E_P^{\ 2}}{N_P^2}\right) \\ SNR_Z &= 10\lg\left(\frac{E_Z^{\ 2}}{N_Z^2}\right) \end{aligned} \tag{5}$$

Here, $E_P^{\ 2}$ and $E_Z^{\ 2}$ denote the signal energies of the sound pressure and vertical particle velocity channels, respectively. Marine ambient background noise was characterized by selecting received data from vessel-free periods within the detection area and computing its power spectral density. Received data from periods during which a single vessel was present in the monitoring area were treated as the signal. A fixed-length sliding time window was then employed across the signal period to derive the power spectral density of the noisy signal for each window, as shown in Fig. 2(f). Next, the SNRs of four channels of OBS were calculated according to the standard SNR definition. Finally, the variation trends of SNR for both the sound pressure

channel and the particle velocity channels were analyzed throughout vessel transit.

4. Result Analysis

4.1 SNR Calculation Result

This study used only a single container vessel as an example, as shown in Fig. 2(e). The azimuth of the vessel was determined according to the posterior probability. The vessel was determined to be a container vessel 9527049 combined with the ship trajectory obtained according to data of AIS. We found that, at the time of CPA to the OBS, the vessel was in very close proximity, with the pitch angle near 90°. It indicated that the vessel passed nearly directly overhead. The SNR of the vessel noise received across the four channels at distinct frequency bins was computed.

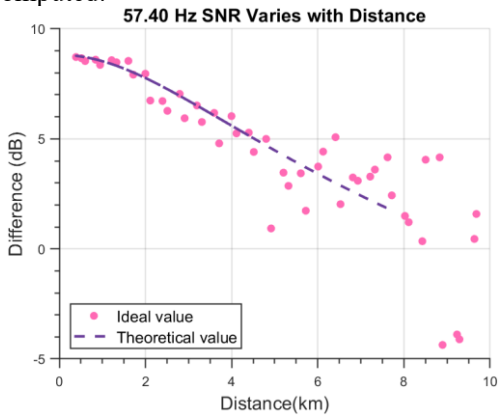


Figure 3– SNR varies with distance.

Figure 3 display the SNR difference between the vertical particle velocity and sound pressure channels during the vessel passed. Circular markers indicate the trend of measured SNR difference versus range, while the dashed curve represents the ideal-condition SNR difference trend (the computational procedure for which is detailed in the following section). It can be seen that within a certain distance, the SNR of the vertical particle velocity channel will be higher than that of the other channels.

4.2 Causal Analysis

As shown in Fig. 4, $\theta \in [0, 2\pi]$ is the azimuth of the arrival signal (the angle between the incident sound wave and the positive direction of the X axis). $\alpha \in [-\pi/2, \pi/2]$ is the pitch angle of the arrival signal (the angle between the incident sound wave and the positive direction of the Z axis). The acoustic sound pressure channel of the vector hydrophone exhibits no directivity, displaying an "O"-shaped pattern, while its particle velocity channels exhibit a dipole directivity, characterized as "figure-of-eight"-shaped.

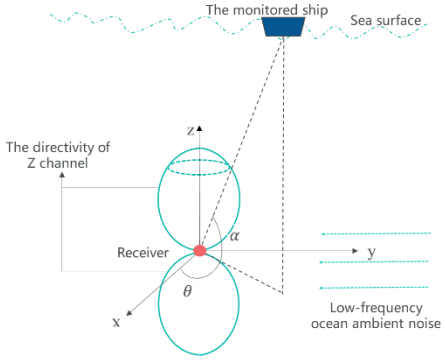


Figure 4 – Schematic diagram of receiving signal of the vector hydrophone.

Owing to the anisotropic nature of ambient noise, noise above 1 kHz predominantly originates from the sea surface. For frequencies below 1 kHz, particularly in the 10–100 Hz range, the noise level generated by ships is higher than that of other sources. This suggests that the predominant source of noise in this frequency band is vessel noise. Marine ambient noise from distant sources relative to the receiver primarily arrives along near-horizontal angles (1,3). The background noise expressions for the pressure and vertical particle velocity channel are as follows:

$$N_p^2 = \int N_P(\varphi) d\varphi$$
$$N_z^2 = \int N_P(\varphi) \sin\varphi d\varphi \quad (6)$$

Here, φ is the incident angle of background noise from all directions, N_p^2 is the background noise energy received by the sound pressure channel, and N_z^2 is the background noise energy received by the vertical particle velocity channel.

According to the definition of SNR, the SNR difference between the two channels is defined as:

$$SNR_Z - SNR_P = 10\lg E_Z{}^2 - 10\lg E_P{}^2 - (10\lg N_Z^2 - 10\lg N_P^2) \quad (7)$$

The energy difference between the vertical particle velocity and pressure channels, resulting from their respective directivity during the reception of shipping noise signals, is calculated by Eq. (2):

$$10\lg E_Z{}^2 - 10\lg E_P{}^2 = 10\lg \sin^2 \alpha \quad (8)$$

Due to the anisotropy of the marine ambient noise field, the background noise during the vessel-passage period was calculated using Eq. (6). The background noise received by the pressure channel was subtracted from that received by the vertical velocity channel. Subsequently, by combining Eqs. (7) and (8), the SNR difference at each frequency bin was computed and termed the ideal-condition value. This ideal value was then compared with the measured SNR difference from actual data to generate Fig. 5.

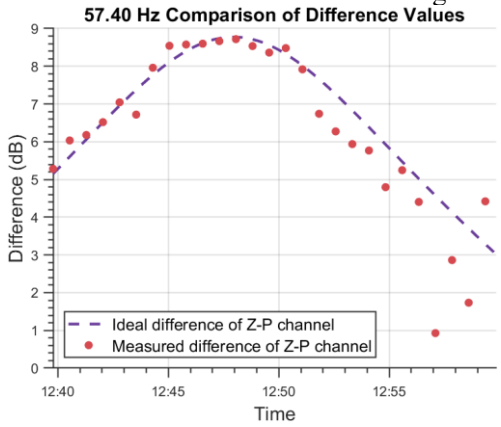


Figure 5 – Comparison of SNR difference.

Fig. 5 demonstrates that the dashed curve, representing the ideal-condition SNR difference computed at each frequency bin by Eq. (7), and the circular markers, denoting measured SNR difference. It can be observed that the SNR difference calculated by the two methods match well.

5. CONCLUSIONS

In this paper, we found that the SNR of the vertical particle velocity channel data of vessel radiation noise is higher than that from the other channels at close range. Firstly, due to the sensitivity difference among the four channels of the vector hydrophone, we corrected for these variations. Then, the SNR of vessel noise at each frequency bin was calculated in the period when a single vessel passed. We found that, within a specific distance range, the SNR difference between the vertical particle velocity channel and the sound pressure channel closely matches that predicted under ideal conditions. This phenomenon is due to the anisotropic nature of the low-frequency marine ambient noise field and the dipole directivity of the vertical particle velocity channel of the OBS. This finding offered a novel approach for close-range vessel detection using vertical particle velocity channel data.

ACKNOWLEDGEMENTS

This work is supported by the National Natural Science Foundation of China (Grant No. 12204128).

REFERENCES

1. Wenz GM. Acoustic Ambient Noise in the Ocean: Spectra and Sources. J Acoust Soc Am. 1962;34(12):1936–56.
2. Hildebrand JA. Anthropogenic and natural sources of ambient noise in the ocean. Mar Ecol Prog Ser. 2009;395:5–20.
3. Shan Y, Lin J, Yi X. Vertical directionality of low frequency ocean ambient noise in areas far away from the shipping lane. Tech Acoust. 2022;41(5):642–7.
4. Huang Y. Review of the spatial correlation modeling of ocean ambient noise. J Appl Acoust. 2019;38(4):729–33.
5. Zhu L, Li F, Chen D. Study of vector noise field characteristics in shallow water by fiber optical vector

hydrophone. Tech Acoust. 2016;35(2):101–8.
6. Trabattoni A, Barruol G, Dréo R, Boudraa A. Ship detection and tracking from single ocean-bottom seismic and hydroacoustic stations. J Acoust Soc Am. 2023;153(1):260–73.

DAILY PROGRAM

AUGUST 20 (WED)

09:00 – 10:40	Acoustic Metamaterials and Metasurfaces for Noise Reduction and Wave Control 325 (3F)
Chair	Hyung Jin Lee (Korea Research Institute of Standards and Science)
09:00	Noise reduction in a flow duct with varying cross-sectional area using ventilated acoustic metamaterials
	Dohaeng Kim, Jiwan Kim, Wonju Jeon
09:20	Ultrasonic metamaterial for enhancing pulse transmission through acoustic barrier
	Junyong An, Chankyu Kim, Wonju Jeon
09:40	Acoustic metasurfaces with complex-valued impedances for reducing noise from an open structure
	Eunjin Yang, Wonju Jeon
10:00	Uniform sound distribution in indoor environments using acoustic metasurfaces with complex-valued impedance
	Eunji Choi, Jiwan Kim, Wonju Jeon
10:20	Deep generative model-based optimization and inverse design of ventilated acoustic metamaterials
	Min Woo Cho, Keon Ko, Sang Min Park

PROCEEDINGS of the
16th International Conference on Theoretical and Computational Acoustics

August 18 to 22, 2025 BEXCO, Busan Korea

Noise reduction in a flow duct with varying cross-sectional area using ventilated acoustic metamaterials

Dohaeng Kim[1]; Jiwan Kim[2]; Wonju Jeon[3]

[1,3] Department of Mechanical Engineering, Korea Advanced Institute of Science and Technology, Republic of Korea
[2] Mechanical Engineering Research Institute, Korea Advanced Institute of Science and Technology, Republic of Korea

ABSTRACT

In mechanical systems that involve fluid-guiding structures, such as ventilation systems in automobiles and aircraft, HVAC systems, and data center cooling systems, it is important to reduce noise propagating along ducts while allowing airflow. Due to the trade-off between noise reduction and airflow when using conventional acoustic materials, acoustic metamaterials have been actively studied to solve the trade-off problem. However, most of existing studies have focused only on ducts with uniform cross sections, limiting their applicability to real-world problems. Therefore, in this study, we present a metaliner—a type of ventilation acoustic metamaterial—that effectively reduces noise without disrupting fluid flow, even in ducts with non-uniform cross sections. A theoretical model is developed to characterize the metaliner's effective impedance under locally varying flow velocity profiles resulting from variations in cross-sectional area. To accurately account for the influence of flow on the effective impedance, the friction velocity is incorporated as a key physical variable, enabling the model to remain valid even in ducts with non-uniform cross sections. Based on this framework, we design metaliners that achieve high insertion loss under various flow conditions and experimentally demonstrate its noise reduction performance in a flow duct with varying cross-sectional area.

Keywords: acoustic metaliner, duct noise reduction, varying cross-sectional area, friction velocity

[1] dohaeng.kim@kaist.ac.kr
[2] jiwan.kim@kaist.ac.kr
[3] wonju.jeon@kaist.ac.kr

PROCEEDINGS of the
16th International Conference on Theoretical and Computational Acoustics

August 18 to 22, 2025 BEXCO, Busan Korea

Ultrasonic metamaterial for enhancing pulse transmission through acoustic barrier

Junyong An; Chankyu Kim; Wonju Jeon[1]

Department of Mechanical Engineering, Korea Advanced Institute of Science and Technology, 291 Daehak-ro, Yuseong-gu, Daejeon 34141, Republic of Korea

ABSTRACT

This study designs metamaterials to enhance the ultrasound transmission through an acoustic barrier that exhibits different impedance from surrounding media. The metamaterial is composed of meta-layers 1 and 2 to alleviate the impedance mismatch between the acoustic barrier and incident medium. When it is attached to the acoustic barrier, transmittance of harmonic wave is enhanced by 20 times across 12% relative bandwidth. The designed metamaterial is fabricated by wire electric discharge machining, and the experiment is conducted in a water tank to validate the transmission performance. By comparing transmitted signals through acoustic barriers with and without metamaterial, we confirm that the energy of 10-cycle Gaussian pulse is enhanced by 7.8 times.

Keywords: ultrasonic pulse, high transmission, acoustic metamaterial, input impedance matching

[1] wonju.jeon@kaist.ac.kr

Acoustic metasurfaces with complex-valued impedances for reducing noise from an open structure

Eunjin YANG[1]; Jiwan KIM[2]; Wonju JEON[3]
[1,2,3] Korea Advanced Institute of Science and Technology, Republic of Korea

ABSTRACT

We propose an acoustic metasurface designed to mitigate noise emanating through openings of open structures. The metasurface consists of subwavelength Helmholtz resonators, and we utilize a theoretical model that accurately predicts their effective impedance by accounting for visco-thermal losses in narrow orifices. Based on the theoretical model, the surface impedances along the inner wall of the open structure are optimally tuned to complex values that minimize acoustic energy emanating through openings. Numerical results demonstrate that tonal noise is effectively controlled by trapping and dissipating the energy inside the structure, thereby suppressing its propagation into the outside of the open structure. The noise reduction performance of the metasurface is validated through fabrication via 3D printing and experiments, demonstrating that λ/31-thick metasurfaces can achieve more than 12.6 dB SPL reduction at the target frequency of 360 Hz.

Keywords: Acoustic metasurface; Complex-valued impedance; Noise reduction

[1] yoasdf@kaist.ac.kr
[2] jiwan.kim@kaist.ac.kr
[3] wonju.jeon@kaist.ac.kr

Uniform sound distribution in indoor environments using acoustic metasurfaces with complex-valued impedance

Eunji CHOI[1]; Jiwan KIM[2]; Wonju JEON[3]
[1,2,3] Korea Advanced Institute of Science and Technology, Republic of Korea

ABSTRACT
In indoor environments, low-frequency sound often becomes unevenly distributed, resulting in poor sound quality. To address this issue, we present an acoustic metasurface composed of subwavelength Helmholtz resonators with complex-valued impedances. The complex-valued impedance of the metasurface is tailored to minimize the standard deviation of sound pressure level within the designated listener region. The acoustic performance of the metasurface is analyzed using full-wave simulations that account not only for low-frequency wave phenomena such as diffraction and scattering, but also for frequency-dependent thermo-viscous losses within narrow orifices. These simulations demonstrate that the metasurface with optimized complex impedance significantly reduces the spatial variation of sound pressure levels at the target frequency, while occupying substantially less volume compared to traditional absorptive wall treatments. This study further extends to a multi-frequency design, achieving improved spatial uniformity across a broad frequency range.

Keywords: Sound field control, Acoustic metasurface, Complex impedance

[1] cej9049@kaist.ac.kr
[2] jiwan.kim@kaist.ac.kr
[3] wonju.jeon@kaist.ac.kr

Deep generative model-based optimization and inverse design of ventilated acoustic metamaterials

Min Woo Cho[1]; Keon Ko[2]; Sang Min Park[3];
[1,2,3] Pusan National University

ABSTRACT

To mitigate noise pollution in urban environments where ventilation is required, ventilated acoustic metamaterials (VAM), a type of acoustic metamaterials (AM), has enables simultaneous noise reduction and ventilation. However, the inverse design of VAM with analytical methods is a challenging task due to the high nonlinearity between structure and acoustic response. Furthermore, existing deep learning-based inverse design methods for AMs mainly rely on parameter-based approaches, which limit the design flexibility, or pixel image-based methods which deteriorate the essential shape required for sound attenuation performance. To address limitations of conventional design methods for AM, we propose a two novel deep generative model-based design methods for the optimization and inverse design of VAM: latent space-based optimization method and data augmented Dual-variational autoencoder (Dual-VAE) with iterative transfer learning. Latent space-based optimization method enables ultra-broadband VAM by performing genetic algorithm optimization on the latent space generated by deep generative model. Data augmented Dual-VAE with iterative transfer learning realizes accurate inverse design of VAM exhibiting the target acoustic response. These frameworks offer potential solution for overcoming the limitations of conventional optimization and inverse design method for AM, paving the way for advanced noise control applications in urban environments.

DAILY PROGRAM

AUGUST 21 (THU)

09:00 – 10:20	Advanced Tracking Algorithms and Acoustic Estimation for Underwater Targets 323 (3F)
Chair	Sea-Moon Kim (Korea Research Institute of Ships & Ocean engineering)
09:00	Improved Underwater Multi-target Track-Before-Detect Algorithm Based on Firefly Algorithm
	Yushi Shen, Bin Qi, Xiang Li, Chenxin Hui
09:20	A Particle Filter Track-Before-Detect Method with Adaptive Diffusion Functions for Proximity Scenarios
	JunFei Zhou, Bin Qi, ChenXin Hui, Xiang Li, Dong Xu
09:40	Estimation of source levels for unmanned underwater vehicles
	GuhnHyeko Ko, ChangSoo Kim, ByoungNam Kim, YoSup Park, CheolSoo Park, DongGuk Paeng
10:00	Research on bearings-only Association Algorithm under Unknown Measurement Noise Conditions
	Jiaxin Zhou, Bin QI, Jinjin Wang, Yilin Wang

PROCEEDINGS of the
16th International Conference on Theoretical and Computational Acoustics

August 18 to 22, 2025 BEXCO, Busan Korea

Improved Underwater Multi-target Track-Before-Detect Algorithm Based on Firefly Algorithm

Bin QI[1,2,3]; Yushi SHEN[1,2,3]; Xiang LI[1,2,3★]; Chenxin HUI[1,2,3]

[1] National Key Laboratory of Underwater Acoustic Technology, Harbin Engineering University, Harbin 150001, China;

[2] Key Laboratory of Marine Information Acquisition and Security(Harbin Engineering University), Ministry of Industry and Information Technology;Harbin 150001, China;

[3] College of Underwater Acoustic Engineering, Harbin Engineering University, Harbin 150001, China

ABSTRACT

Conventional Independent Partition Particle Filter Track-Before-Detect algorithms for multi-target tracking often see most particles' weights approach zero over iterations, with only a few particles retaining significant weights. This leads to a loss of particle diversity and hinders effective state-space exploration. In scenarios where targets have trajectory crossing, particles from different targets can interfere, causing association errors. The algorithm struggles to distinguish targets accurately at trajectory intersections, resulting in trajectory merging or incorrect associations. To address these issues, this paper proposes an improved underwater multi-target pre-tracking algorithm based on the Firefly Algorithm. It mimics fireflies' flashing and attraction mechanisms, selects locally optimal particles at each moment, and guides the particle swarm toward these optimal particles. This enables the swarm to approach the true target trajectory. Experimental results show that this algorithm can accurately estimate target numbers and azimuth angles in near-crossing scenarios. It also enhances multi-target tracking accuracy and robustness.

Keywords: Independent Partition Particle Filter Track-Before-Detect Algorithm, Firefly Algorithm, Trajectory Crossing

1. INTRODUCTION

Passive acoustic detection and tracking[1] is crucial in underwater acoustics and monitoring. It involves monitoring and tracking targets by listening to their radiated noise and is widely used in large-area marine surveillance, ocean resource development, and addressing increasingly intelligent underwater targets. The passive sonar system, a core part of this technology, offers high concealment, low power consumption, and long detection ranges. Using it for underwater detection and tracking is known as the bearing-only tracking problem.[2] By acquiring target-bearing information, it estimates target motion states, like azimuth and velocity.

Traditional detect-before-track (DBT) algorithms divide detection and tracking into separate steps. In the detection phase, a threshold is set to classify single-frame echo data. Points below it are discarded as noise, while those above are retained as potential target data. DBT's detection results are susceptible to this threshold. An improper setting can lead to missed detections and false alarms, and the data loss from thresholding reduces information availability. In passive sonar detection and tracking scenarios, targets often have low signal-to-noise ratios due to inherent and marine environmental factors. This increases the likelihood of missed detections in the initial detection phase and causes a significant loss of tracking information in the subsequent tracking phase. Consequently, DBT algorithms face significant performance limitations in passive sonar scenarios and require stringent parameter settings for the initial detection threshold.

The Track-Before-Detect (TBD)[3] algorithm directly processes raw data, omitting the data thresholding step. This enhances data diversity by preserving amplitude, velocity, and other helpful information. However, utilizing all data increases the computational complexity of multi-target tracking algorithms[4], leading to higher computational demands and slower processing speeds. The Independent

★ lixiang97@hrbeu.edu.cn

Partition Particle Filter Track-Before-Detect (IP-TBD)[5] algorithm is introduced to resolve this. It breaks down the multi-target tracking problem into several single-target tracking tasks, reducing computational complexity and improving processing speed. The algorithm also employs an energy likelihood function to evaluate the probability of target existence by calculating the energy difference between observations and target states.

While the IP-TBD algorithm excels at reducing computational complexity and boosting processing speed, it faces challenges in complex environments. Over iterations, most particles ‘ weights dwindle to near zero, with only a few particles retaining significant weight. This leads to a loss of particle diversity and a decline in estimation accuracy. Although particle filtering's resampling technique can mitigate particle degeneration, it may trigger "sample impoverishment, "further reducing particle diversity. Moreover, in scenarios with closely located target[6], the IP-TBD algorithm's batch management accuracy, which relies on clustering results, is compromised, potentially causing algorithmic failure.

To address the aforementioned issues, this paper abandons the traditional particle filter resampling technique in pre-tracking detection and incorporates the firefly algorithm (FA)[7] to propose an improved underwater multi-target pre-tracking algorithm. After the prediction step of the original independent partitioned particle filter algorithm, the FA's attraction mechanism is introduced to update particles. By mimicking the attracting behavior between fireflies, particles move towards higher-weight ones based on weight differences. This makes low-weight particles gather around locally optimal ones, expanding the search range and avoiding early convergence. Thus, the algorithm can better locate the global optimal matching results, enabling precise particle-target approach for pre-tracking detection.[8]

The structure of this paper is as follows. Section 2 presents the independent partitioned particle filter algorithm with birth and death probability models and the firefly algorithm, including theoretical derivations. Section 3 discusses the improved underwater multi-target pre-tracking algorithm based on the firefly algorithm and its implementation process. Section 4 verifies the improved algorithm through single- and multi-target tracking simulations, demonstrating that it can accurately estimate target numbers and azimuth angles in underwater scenarios. Additionally, in near-crossing multi-target scenarios, the improved algorithm can correctly separate and track crossing-trajectory targets.

2. Track-before-detection Algorithm

2.1 IP-TBD

The independent partition particle filter algorithm is a particle filtering method for multi-target tracking. Its core lies in calculating the weights of particles in each partition and performing independent crossover sampling among the same partitions of different multi-target particles. This ensures better state sampling for each partition and improves the second-order weights of multi-target particles. This algorithm implements crossover sampling by resampling the sub-particles of the same partition based on their first-order weights. This retains high-quality sub-particles and discards poor-quality ones. As a result, the number of particles needed is reduced, and the algorithm's efficiency is enhanced.

Assume the number of targets is N , and the number of particles is M . When applying the independent partition particle filter algorithm, we need to initialize various parameters. The initial state vector of each target is given by $X^j = [x^j \quad v_x^j \quad y^j \quad v_y^j]$, where x^j and y^j represent the target's position in the horizontal and vertical directions, and v_x^j and v_y^j represent the target's velocity in the horizontal and vertical directions. The constant velocity motion model is used for target movement in this experiment.

For each multi-target particle partition j corresponding to target j , the first step is to sample the state at time k based on the probability density $p(X_k^j \mid X_{k-1}^j)$. The given target motion equation determines the probability density function. Then, calculate the partition weights and the likelihood function of the partition can be as follows

$$P^j = p(z_k \mid X_k^j) \tag{1}$$

Repeat the above steps to sample the state and calculate the weight for each multi-target particle partition j . Normalize the weights of partition j . Independent crossover sampling for partition j will be performed based on the weights of all multi-target particles in that partition. After completing independent crossover sampling for each partition, calculate the second-order weights of the multi-target particles. Each target corresponds to a unique partition in this experiment, so second-order weight calculation is unnecessary.

Resample the multi-target particles and estimate the target state based on the known model criteria. Repeat the steps if the current time is still within the tracking duration; otherwise, output the final tracking results.

2.2 Probabilities Model of Birth and Death

The IP-TBD algorithm incorporates a Bayesian probabilistic framework for target birth and death management, utilizing dual-threshold decision mechanisms to dynamically handle target lifecycle transitions. This model synergistically integrates real-time spatial spectrum analysis with historical trajectory evidence to achieve robust multi-target tracking in complex environments.

In the target birth model, potential targets are detected at each time k by identifying local energy peaks in the spatial spectrum and their indices. The spatial spectrum has 361 continuous points, corresponding to azimuth angles from 0° to 180° with a resolution of $b = 0.5$. The azimuth angle difference between the detected peak and existing targets is calculated. If the angle difference exceeds a set birth threshold θ_{birth}, the peak is judged as a new target; otherwise, it is part of an existing target sequence.

For target death determination, a historical window ε_{win} records the particle weights of each existing target for the most recent win moments. When the window is whole, the earliest data is discarded as new data arrives, maintaining a fixed window size. The cumulative weight over the window is calculated and compared to a death threshold θ_{death}. If it is below the threshold, the target is deemed dead, and tracking stops; otherwise, tracking continues.

2.3 Firefly Algorithm

The Firefly Algorithm (FA) is a metaheuristic optimization algorithm inspired by the behavior of fireflies in nature. It is designed to solve optimization problems, particularly global optimization problems. By mimicking fireflies flashing and mutual attraction, the algorithm guides fireflies towards better solutions, gradually approaching the global optimal solution. The specific steps of the Firefly Algorithm are as follows

$$I = I_{\max} \times e^{-\eta r_{mn}^2} \tag{2}$$

where I denotes the relative light intensity of a firefly, $I_{\max}$ represents the maximum light intensity, η is the light absorption coefficient, and r_{mn} stands for the Euclidean distance between firefly m and firefly n. Subsequently, the mutual attraction between fireflies is calculated. The attractiveness between fireflies is inversely proportional to the distance between them and is related to the light intensity. This can be expressed explicitly as

$$\beta = \beta_{\max} \times e^{-\eta r_{mn}^2} \tag{3}$$

where $\beta_{\max}$ represents the maximum attractiveness, typically set to 1, while η denotes the light intensity absorption coefficient, which adjusts the rate at which light intensity decreases with distance; after mutual attraction occurs, each firefly moves closer to other brighter fireflies. The updated position can be expressed as

$$x^m = x^m + \beta \times (x^n - x^m) + \alpha \times \varepsilon \tag{4}$$

where x^m and x^n represent the positions of firefly m and firefly n within the spatial range, respectively, α is the step length factor where $\alpha \in [0,1]$, and ε denotes the random disturbance term, which is used to prevent the algorithm from prematurely falling into a local optimum.

3. Improved Independent Partition Particle Filter with Firefly Algorithm

The conventional IP-TBD algorithm can effectively track multiple targets, especially when the number of targets is known, and their movements are relatively stable. However, it has limitations in underwater multi-target scenarios with neighboring trajectory intersections. For example, particle degeneracy can cause most particles to lose representativeness, and insufficient particle diversity can lead to premature particle clustering in local areas, hindering compelling exploration of the entire state space. Additionally, in underwater multi-target scenarios, particles of different targets may interfere with each other, affecting tracking accuracy.

This paper introduces a firefly algorithm-based improved underwater multi-target detection and pre-tracking algorithm to address these issues. The idea is to liken the weight category of particles to the

brightness of fireflies and adjust particle positions using the attraction and repulsion mechanisms of fireflies. The attraction mechanism is based on the principle that the attraction between particles is proportional to their weights. According to this attraction, particles with smaller weights move toward the particle with the most significant local weight, enabling the model to converge to a locally optimal solution quickly. The repulsion mechanism is not explicit but is indirectly achieved through the exponential decay of attraction. The attraction rapidly decreases when the distance between two particles reaches a certain threshold, even approaching zero. This allows the influence between distant fireflies to be ignored. Moreover, during particle movement, random perturbations are introduced to approximate firefly-like random motion.This enhances local search ability and prevents the algorithm from prematurely converging to local optima.

At time k , the model takes the particle state information from the previous time step as input and combines it with the state transition matrix F to obtain the current particle state prediction result $x_k^{j,i}$. Specifically expressed as

$$x_k^{j,i} = Fx_{k-1}^{j,i} + W \tag{5}$$

where j is the target index, and i is the particle index. F is the state transition matrix, expressed as

$$F = \begin{bmatrix} 1 & \Delta T \\ 0 & 1 \end{bmatrix} \tag{6}$$

where ΔT is the time interval, and W is the process noise. The covariance $\hat{Q}$ of process noise during target motion satisfies the condition, which is specifically manifested as

$$\hat{Q} = \hat{q} \times \begin{bmatrix} \Delta T^3/3 & \Delta T^2/2 \\ \Delta T^2/2 & \Delta T \end{bmatrix} \tag{7}$$

where $\hat{q}$ is the intensity of process noise, the corresponding particle weight $q_k^{j,i}$ is calculated based on the signal amplitude information. The local optimal particle with the highest weight is selected, and its index $\mathrm{I0}_k^{i,\max}$ is recorded. The relative distance between other particles and the local optimal particle is calculated, which affects the attractiveness between particles. The particle attractiveness directly determines the distance particle i moves towards the local optimal particle $\mathrm{I0}_k^{i,\max}$.

$$\hat{x}_k^{j,i} = x_k^{j,i} + \beta_k^{j,i} \times (x_k^{j,\mathrm{I0}_k^{i,\max}} - x_k^{j,i}) + \alpha \times \varepsilon \tag{8}$$

where $\beta_k^{j,i}$ represents the relative attractiveness of the i -th particle of the j -th target at time k to the locally optimal particle, $x_k^{j,i}$ and $x_k^{j,\mathrm{I0}_k^{i,\max}}$ denote the state information of the i -th particle of the j -th target and the locally optimal particle at time k , respectively. α is the step size factor, and ε represents random perturbations, following a normal distribution with a mean of -1 and a variance of 4, are used to prevent the algorithm from prematurely converging to local optima.

Recalculate the weights of the updated particles to obtain $\hat{q}_k^{j,i}$. Combine the predicted and updated state information and weights into X_k^j and Q_k^j . X_k^j and Q_k^j are represented as:

$$X_k^j = \left\{ x_k^{j,1}, \ldots, x_k^{j,M}, \hat{x}_k^{j,1}, \ldots, \hat{x}_k^{j,M} \right\} \tag{9}$$

$$Q_k^j = \left\{ q_k^{j,1}, \ldots, q_k^{j,M}, \hat{q}_k^{j,1}, \ldots, \hat{q}_k^{j,M} \right\} \tag{10}$$

where the lengths of X_k^j and Q_k^j are both $2M$. Sort them by weight magnitude and select the top M particles with larger weights for subsequent target death judgments. Repeat the above steps for all targets at the current time, then proceed to the next time step $k = k + 1$, until k exceeds the preset tracking time T . Stop tracking and output the final results. Figure 1 presents the flow chart of the IP-TBD algorithm's prediction step followed by the FA algorithm for update and optimization.

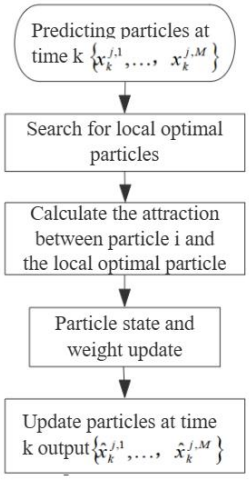


Figure 1 – Flow chart for introducing FA algorithm update optimization

4. Simulation And Experiment

To verify the improved pre-tracking detection algorithm's effectiveness, two sets of experiments are designed. The simulation environment uses a 14-element uniform linear array with 0.2 m spacing. It receives target-radiated noise signals. The signal-processing frequency band is 500-3000 Hz, and the sampling rate is 8 kHz. Within the working frequency band, signals are divided into 51 sub-bands, each with a 50 Hz bandwidth and a 1s integration time. All sub-bands meet the narrow-band signal conditions. The root mean square error (RMSE) evaluation indicator is used to measure the tracking performance in all simulations.

In experimental scenario one, there's only one target moving at a constant speed. The target's amplitude remains constant during the observation period. The observer turns between 23-32 s and 70-76 s, and moves at a constant speed during the rest of the observation time. The experiment obtains bearing angle data by setting the positions and velocity information of targets and observers. The initial information of targets and observers is shown in Table 1, and the start and end time information of each target in the simulation scenario is shown in Table 2. Figure 2 shows the true motion trajectories of the targets and the observer.

Table 1 – Initial Target and Observer Information Table

Designation	X-coordinate, m	Y-coordinate, m	Horizontal velocity, m/s	Vertical velocity, m/s
Target 1	-2000	3000	-12.8	7.3
Observer	0	0	4.3	6.2

Table 2 – Starting and ending times of each target in the simulation scenario

Target Index	T_{birth}	T_{death}
Target 1	1	100

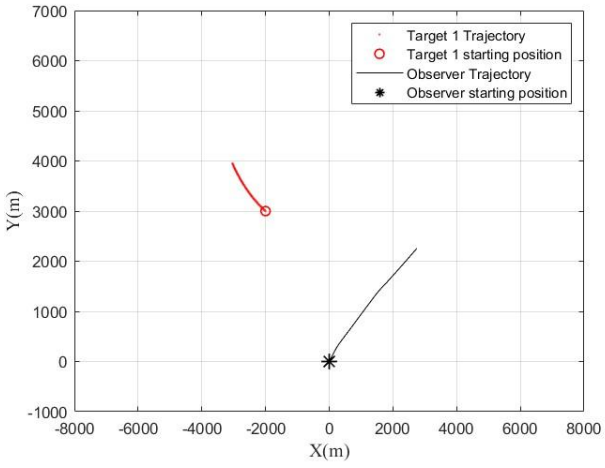


Figure 2 – The real motion trajectory of the target and observer

The simulation time T is 100 seconds, with a sampling interval dt of 1 second and a process noise standard deviation of $\sigma_v = 0.2°/s$. When performing beamforming on the simulated signals, the power spectrum of target 1 is concentrated in the 800-1800 Hz band, and the full-band signal-to-noise ratio is -25 dB. Figure 3 shows the true azimuth information of the target, and Figure 4 shows the time-azimuth trajectory obtained after conventional beamforming.

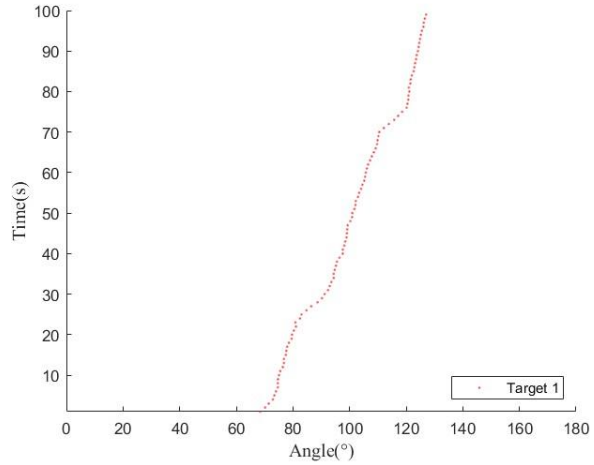


Figure 3 – Real directional information map of the target

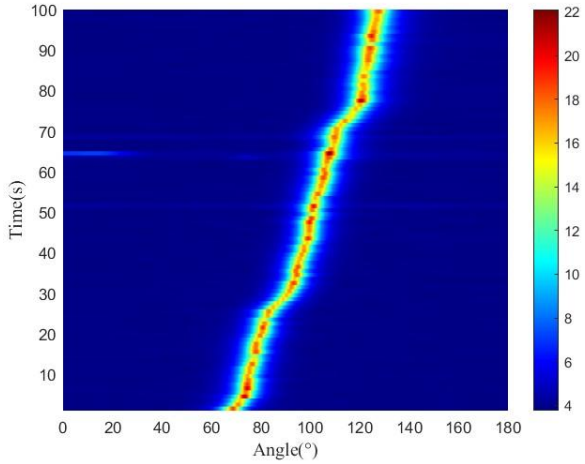


Figure 4 – Time and azimuth history diagram obtained after conventional beamforming

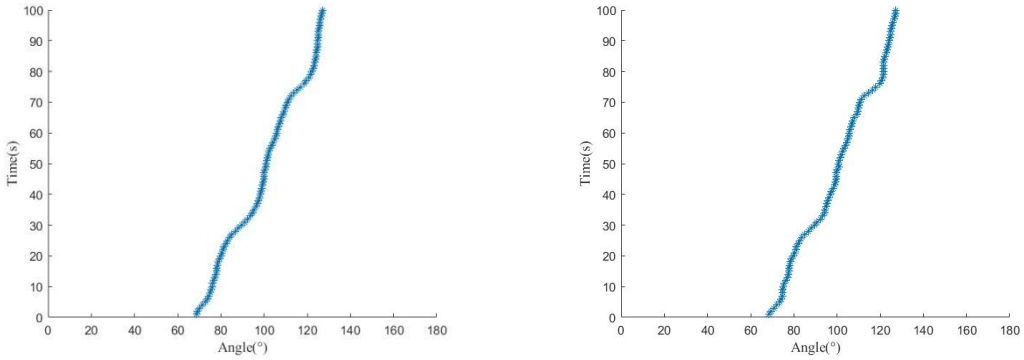
Key parameter settings for the improved IP-TBD algorithm with the FA algorithm in the simulation experiment are shown in Table 3.

Table 3 – Other simulation parameter settings

Simulation Parameters	N	M	$\hat{q}$	Win	θ_{birth}	θ_{death}	$\beta_{\max}$	η	α
Value	4	100	0.013	5	15	0.05	1	0.1	0.1

In the simulation experiment, as shown in Figure 5, the traditional IP-TBD algorithm and the IP-TBD algorithm with FA were used to track and estimate a single target. Figure 5(a) shows the estimation results from the traditional IP-TBD algorithm. Figure 5(b) shows the results from the improved IP-TBD algorithm with the FA. In the single-target scenario, both algorithms could obtain the correct target trajectory. However, the prediction of traditional IP-TBD algorithm relies more on the target state at historical moments, so the target orientation trajectory for maneuvering will wait for a relatively smooth tracking result, and the tracking trajectory is also relatively smooth, especially at maneuvering turns where the turn is large. The improved traditional IP-TBD algorithm with the introduction of FA algorithm can further update and correct based on the current state information, making the estimation closer to the local optimal solution and obtaining a more realistic maneuvering target trajectory. The RMSE of the IP-TBD algorithm

is 0.94°, while the IP-TBD algorithm with FA has a RMSE of 0.56°, indicating better tracking accuracy with FA. This indicates that the FA enhanced IP-TBD algorithm improves tracking accuracy in single target tracking scenarios.



(a) (b)

Figure 5 – Single target tracking estimation results of two algorithms
(a)Estimation of azimuth using traditional IP-TBD algorith (b)Improved IP-TBD algorithm with FA for estimating azimuth

Experimental scenario two involves four moving targets. Among them, target 2 and target 4 have intersecting trajectories, while target 1 and target 3 are not affected by other targets. The initial details of the target and observer are shown in Table 4. The start and end times of each target in the simulation are shown in Table 5.

Table 4 – Initial Target and Observer Information Table

Designation	X-coordinate, m	Y-coordinate, m	Horizontal velocity, m/s	Vertical velocity, m/s
Target 1	-4000	3000	2	-10
Target 2	0	2000	12	-4
Target 3	6185.2	357.5	13.8	2.6
Target 4	-4023.7	6378.3	-3	1.8
Observer	0	0	5	5

Table 5 – Starting and ending times of each target in the simulation scenario

Target Index	T_{birth}	T_{death}
Target 1	40	90
Target 2	10	83
Target 3	20	80
Target 4	1	100

The scenario includes four widely spaced targets moving at a constant speed. Their amplitudes remain constant during the observation period. The observer turns between 10-34 s and 60-66 s, and moves at a constant speed during the rest of the observation time. Figure 6 shows the true motion trajectories of the four targets and the observer.

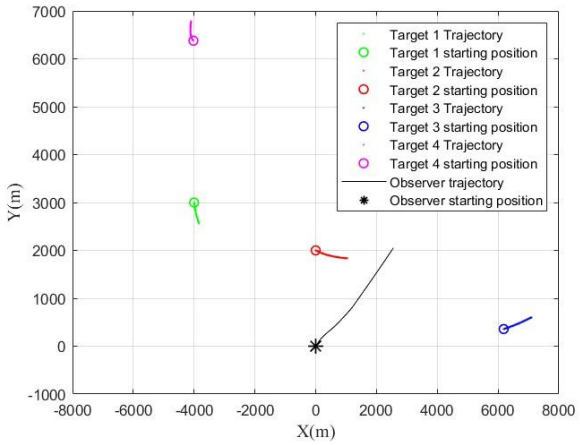


Figure 6 – The real motion trajectory of the target and observer

The simulation time, sampling interval, and process noise standard deviation of Simulation Two are the same as those of Simulation One. When performing beamforming on the simulated signals, the power spectrum of target 1 is concentrated in the 1500-2500 Hz band with a full-band SNR of -13 dB; that of target 2 in the 2000-3000 Hz band with a full-band SNR of -18 dB; that of target 3 in the 500-1500 Hz band with a full - band SNR of -10 dB; and that of target 4 in the 1000-2000 Hz band with a full-band SNR of -12 dB. Figure 7 presents the true azimuth information of the four targets, and Figure 8 illustrates the time-azimuth trajectories obtained after conventional beamforming.

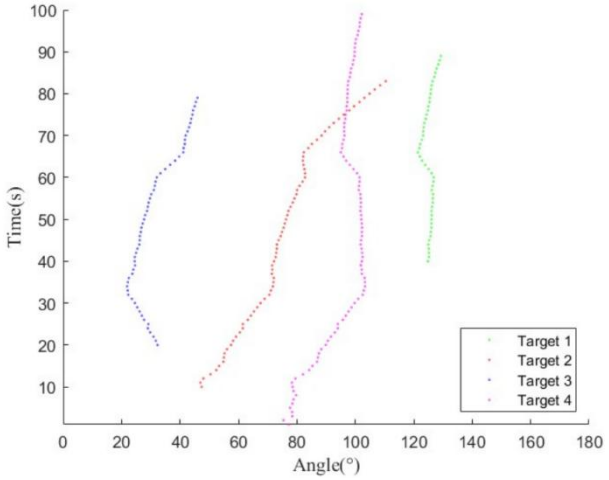


Figure 7 – Real directional information of four targets

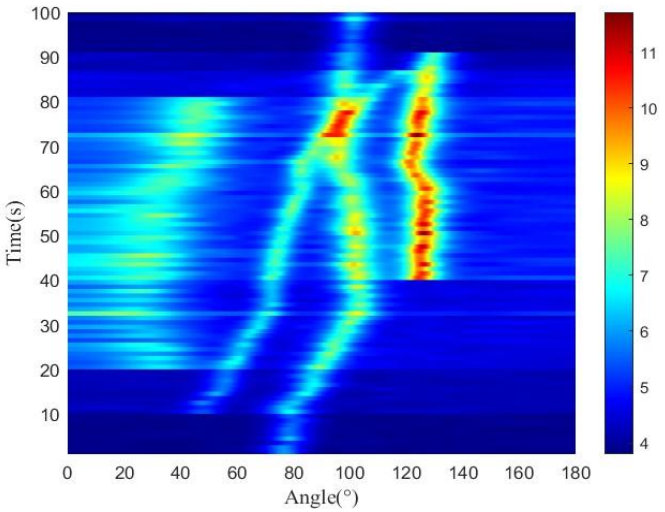
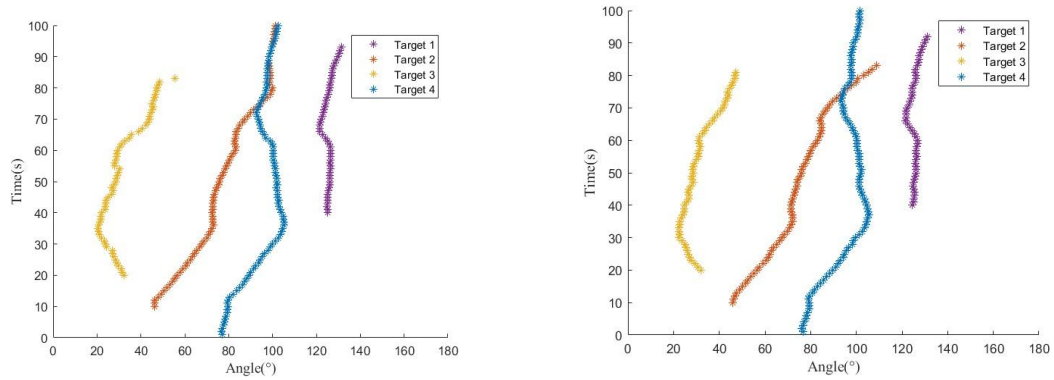


Figure 8 – Time and azimuth history diagram obtained after conventional beamforming

In the simulation, we used the traditional IP-TBD algorithm and the improved IP-TBD algorithm with FA to track four targets, as shown in Figure 9. The key parameters of the improved algorithm remained unchanged from the first simulation. As shown in Figure 9(a), when using the conventional IP-TBD algorithm, at 75 s, when the azimuth trajectories of target 2 and target 4 cross, the particle group for target 2 splits into two parts, distributed around both target 2 and target 4. Over time, during the resampling process, the particles associated with target 2 are gradually eliminated, eventually concentrating all particles around target 4, causing the trajectories of the two targets to overlap. In contrast, Figure 9(b) illustrates that the improved IP-TBD algorithm with the FA can accurately locate the local optimal particles for both targets during the attraction process. This ensures that the particle groups for target 2 and target 4 converge toward their respective local optimal particles and are evenly distributed around their corresponding trajectories. As a result, the improved algorithm achieves accurate tracking trajectories.



(a) (b)

Figure 9 – Multi object tracking estimation results of two algorithms
(a)Estimation of azimuth using traditional IP-TBD algorithm (b)Improved IP-TBD algorithm with FA for estimating azimuth

Table 6 – RMSE error of two algorithms

RMSE	IP-TBD,°	AL-IP-TBD,°
Target 1	0.49	0.41
Target 2	2.65	1.28
Target 3	1.63	1.54
Target 4	1.32	1.19

In this simulation, Table 6 shows the RMSE errors of tracking using the traditional IP-TBD algorithm and the IP-TBD algorithm with the introduction of FA in multi-target scenarios. It can be seen that, except for target 2, the trajectory tracking error of the IP-TBD algorithm with FA introduced is smaller than that of the traditional IP-TBD algorithm for other targets. However, due to the inability of the traditional IP-TBD algorithm to correctly associate the true target trajectory, Target 2 has generated greater errors. The introduction of FA in the IP-TBD algorithm ensures lower estimation errors while accurately associating the true target trajectory. This indicates that the FA enhanced IP-TBD algorithm improves the tracking accuracy in multi-target tracking scenarios.

5. CONCLUSIONS

This paper proposes an improved underwater multi-target pre-tracking algorithm based on the firefly algorithm. The aim is to overcome the limitations of the traditional IP-TBD algorithm in multi-target crossing scenarios. The improved algorithm mimics the flashing and attraction mechanisms of fireflies. It treats particle weights as firefly brightness, moving particles toward better solutions. This expands the search range, prevents early convergence, and enhances tracking accuracy and robustness. Simulation experiments show that the improved algorithm can correctly distinguish and track targets in crossing scenarios compared to the traditional IP-TBD algorithm. It effectively solves the association error problem of the traditional algorithm at trajectory intersections. However, the new algorithm may reduce processing speed in high-computational-complexity situations and is sensitive to parameter selection. Further optimization is needed for complex dynamic environments.

ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (Grant No. 62271162).

REFERENCES

1. Wang C, Zhu G, Yin J, et al. Experimental study of passive detection of underwater targets across ice[J]. Applied Acoustics, 2022, 191: 108672.
2. Li X, Wang Y, Qi B, et al. Long baseline acoustic localization based on track-before-detect in complex underwater environments[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 1-14.
3. Yi W, Kong L, Yang J. Thresholding process based dynamic programming track-before-detect algorithm[J]. IEICE transactions on communications, 2013, 96(1): 291-300.
4. Guolong L, Boyu Z, Bin Q I, et al. Underwater multitarget fusion tracking method for passive sonar[J]. Acta Acustica, 2024, 49(3): 501-512.
5. Lin M T, Zhang J L, Cheng Q, et al. Independent particle filters[J]. Journal of the American Statistical

Association, 2005, 100(472): 1412-1421.
6. Yan W, Xiang L I, Bin Q I, et al. A track-before-detect method for neighboring targets based on auxiliary particle filter in passive sonar scenarios[J]. Acta Acustica, 2023, 48(2): 277-290.
7. Tian M, Bo Y, Chen Z, et al. Multi-target tracking method based on improved firefly algorithm optimized particle filter[J]. Neurocomputing, 2019, 359: 438-448.
8. Zhou W, Liu L, Hou J. Firefly algorithm-based particle filter for nonlinear systems[J]. Circuits, Systems, and Signal Processing, 2019, 38: 1583-1595.

A Particle Filter Track-Before-Detect Method with Adaptive Diffusion Functions for Proximity Scenarios

Bin Qi[123]; Junfei Zhou[123*]; Chenxin Hui[123];Xiang Li[123];Dong Xu[123]
[1] National Key Laboratory of Underwater Acoustic Technology, Harbin Engineering University, China;
[2] Key Laboratory of Marine Information Acquisition and Security(Harbin Engineering University); China;
[3] College of Underwater Acoustic Engineering, Harbin Engineering University, China;

ABSTRACT
In the proximity scenario, the traditional track-before-detection method has the problem of "the failure of linear superposition due to energy overlap when the target pitch is smaller than the main lobe beam width". This paper proposes an adaptive adjustment method for the kernel width of the diffusion function: shrink the kernel width in the proximity target region to suppress energy overlap, and extend the kernel width in the low signal-to-noise ratio or sparse scenario to enhance the accumulation of weak signals. Calculate the observational likelihood probability based on the Gaussian diffusion function; at the same time, introduce the motion continuity constraint, combine the target state transfer prior and the observational likelihood, and construct a composite Bayesian likelihood function to ensure the unified optimization of the trajectory smoothness and the measurement fit. This paper adaptively adjusts the diffusion function, deduces the composite Bayesian likelihood function in the case of adjacent targets, and realizes the batch division of target tracks in adjacent scenarios based on the particle filter (PF) and multi-hypothesis tracking method. Experimental simulations show that the algorithm proposed in this paper can accurately determine the start and end of tracks and achieve batch division of azimuthally adjacent tracks.

Keywords: Track-Before-Detect (TBD), particle filter, proximity scenarios

1. INTRODUCTION

Underwater target tracking is one of the core tasks in sonar signal processing. Through data correlation of sonar probe results, obtain batch information of the target, and provide key support for subsequent target location, recognition, and command decision-making. In traditional underwater target tracking methods, the Detect-before-Track (DBT) method dominates. This method first extracts measurement values through threshold judgment on single-frame data, and then achieves target tracking through filtering processing and data correlation[1]. However, the DBT method has three main problems:

1) Target signals with low signal-to-noise ratios may not pass the detection threshold, resulting in missed target detections.

2) When the measurement values of multiple adjacent targets overlap, only one measurement value can be obtained after detection processing.

3) As the number of targets increases, target correlation becomes extremely complex [2].

To resolve the above problems, the Track-before-Detect (TBD) method has emerged. The TBD method avoids the threshold judgment process for single-frame data. It uses the original observational data collected by sensors (such as sonar echo signal strength, energy, Doppler frequency shift, etc.) , fully mines the data information, and improves the signal-to-noise ratio through the accumulation of multi-scan observational data to achieve the detection and tracking of targets, with better detection and tracking performance [3]. In recent years, in response to the challenges of complex scenarios such as the intersection of adjacent multi-target trajectories, low signal-to-noise ratio, and background clutter interference in underwater sonar observations, the Particle Filter-based Track-before-Detect (PF-TBD) method has gradually become a research highlight due to its robustness to non-Gaussian/nonlinear observational statistical characteristics and multi-feature Fusion ability. This method can flexibly construct a likelihood function to fuse complex observational information such as the texture manifold features of the target azimuth spectrum and the

[1] 2019051327@hrbeu.deu.cn

regional energy distribution density, significantly promoting the trajectory resolution accuracy and tracking stability of adjacent targets. The reference [4] proposed to construct the likelihood function using data fitting. The advantage of this method is that it can more flexibly select the front-facing DOA estimation algorithm. The reference [5] uses the historical cumulative energy of the passive sonar as the likelihood function. The establishment of the likelihood functions of the above methods is based on the independence of the posterior distributions of each target. However, when the targets are adjacent to each other in the sensor space, and when the target pitch is less than the main lobe beam width, the energy responses of the targets in the azimuth spectrum overlap, resulting in the failure of the likelihood function based on linear superposition.This results in the inability to distinguish the direction of the target based on the measurement information, or a relatively large bias occurs when estimating the direction of the target.

To resolve the problem of neighboring target tracking, reference [6] proposed a likelihood function. This likelihood function describes the target response through the point spread function, but it adopts a fixed kernel width model in the neighboring scene. The point spread function (PSF) with a fixed kernel width may cause two types of errors: 1) Energy overestimation: The measured value in the overlapping area is lower than the linearly superimposed value due to coherent cancellation. 2) Peak blunting: The response at the target center is "submerged" by the energy of neighboring targets. Therefore, this paper innovatively proposes a kernel width self - adapting mechanism, which calculates the observational likelihood probability based on the Gaussian divergence function. Meanwhile, it introduces the motion continuity constraint, combines the prior of target state transfer and the observational likelihood, and constructs a composite Bayesian likelihood function to achieve batch division of target tracks in the neighboring scene.

This article can be divided into five parts: The second part introduces the problems existing in the adjacent scenes; the third part introduces a particle filter model, improves the likelihood function existing in the particle filter, which is used to estimate the target batches and their corresponding states in the multi-target scenario, and provides the implementation process of this algorithm. Leverage the MHT to estimate the target batch states and the PF to estimate the target azimuth angles. In the fourth section, under the typical target trajectory intersection scenario, the tracking effects before and after the improvement of the likelihood function are compared, aiming to test the effectiveness of the proposed method. Finally, we end with a brief summary and discussion in the fifth section.

2. Problem Description and Model Modeling

When two or more targets approach each other within the observational range of the sensor (limited by factors such as the performance of the sonar system and beamforming technology), the interaction between the targets becomes non - negligible. This research focuses on the pure - bearing tracking problem. In this scenario, a significant issue caused by the proximity of targets is that it is difficult to accurately distinguish the bearings of each target based solely on the measurement data, or it may lead to a significant increase in bearing estimation errors. Specifically, in the proximity scenario as shown in the figure (such as when the target pitch is too small, the tracks cross, or they are in similar bearings simultaneously), the problem of incorrect batch division of targets due to track crossing is very likely to occur.

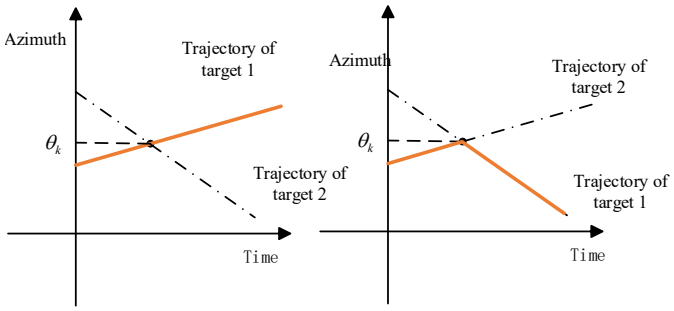


Figure 1– Error in the division of target batches under the adjacent scene

For this problem, this paper tracks multiple targets based on particle filtering. The track-before-detect algorithm forms a Time Bearing Display obtained through beamforming algorithms (such as regular beamforming, MUSIC algorithm, etc.). Assume the current time step is K, there are M sports targets, assuming each target moves independently, then the state variable of the m-th target is $\boldsymbol{x}_k^m = [\theta_k^m, \dot{\theta}_k^m]^{\mathrm{T}}$, the state sequence of multiple targets is $\boldsymbol{x}_{1:k} = [(\boldsymbol{x}_k^1)^{\mathrm{T}}, (\boldsymbol{x}_k^2)^{\mathrm{T}}, ..., (\boldsymbol{x}_k^M)^{\mathrm{T}}]^{\mathrm{T}}$. At this time, the state-space model of the motion system can be described as:

$$\boldsymbol{x}_k = f(\boldsymbol{x}_{k-1}, \boldsymbol{w}_{k-1}) \tag{1}$$

$$z_k = h(\boldsymbol{x}_k, \boldsymbol{v}_k) \tag{2}$$

Where $f(\bullet)$ and $h(\bullet)$ are respectively the system state mapping and measurement mapping of the known system, $\boldsymbol{w}_{k-1}$ and $\boldsymbol{v}_k$ are respectively the state process noise and observation noise, the two groups of noise are independent of each other and independent of the system state, and z_k represents the measurement value at time k.

In the actual tracking scene, a certain target motion model needs to be adopted as the basis. In this paper, we choose a uniform angular velocity motion model. Then the state distribution of multiple targets can be expressed as:

$$p\left(\boldsymbol{x}_k \left| \boldsymbol{x}_{k-1}\right.\right) = \mathrm{N}\left(\boldsymbol{x}_k \left| \mathbf{F}_{k-1}\, \boldsymbol{x}_{k-1}, \mathbf{Q}_{k-1}\right.\right) \tag{3}$$

$$p\left(z_k \left| \boldsymbol{x}_k\right.\right) = \mathrm{N}\left(z_k \left| \mathbf{H}_k\, \boldsymbol{x}_k, \mathbf{R}_k\right.\right) \tag{4}$$

Where $\mathrm{N}(\boldsymbol{x} \mid \mu, \sum)$ represents the Gaussian probability density function for computation, μ represents the mean size, $\sum$ is the covariance matrix. $\mathbf{F}$ represents the state transition matrix, $\mathbf{F} = \begin{bmatrix} 1 & \mathrm{T} \\ 0 & 1 \end{bmatrix}$, $\mathbf{Q}_{k-1}$ is the covariance matrix, $\mathbf{Q}_{k-1} = \mathrm{q} \cdot \begin{bmatrix} \mathrm{T}^3/3 & \mathrm{T}^2/2 \\ \mathrm{T}^2/2 & \mathrm{T} \end{bmatrix}$, q represents the power spectral density of the process noise, T represents the time interval between consecutive frames. $\mathbf{H} = [1 \quad 0]$ represents the measurement matrix, and the covariance is $\mathbf{R} = \sigma_x^2$.

3. Description of the Improved Particle Filter Pre-Detection Tracking Algorithm

3.1 Particle Filter Derivation

The basic idea of particle filter is to use a set of particle sets with weights, replace the integral calculation with the sample mean of the sampling set, and thus approximately obtain the posterior probability density of the target state distribution. Extract sample points from the importance distribution $q(\boldsymbol{x}_{0:k}^i \mid z_{1:k})$, and the importance weight of the i-th particle at time k can be described as:

$$w_k^i = p\left(\boldsymbol{x}_{0:k}^i\right) p\left(z_{1:k} \mid \boldsymbol{x}_{0:k}^i\right) / q\left(\boldsymbol{x}_{0:k}^i \mid z_{1:k}\right) \propto p(\boldsymbol{x}_{0:k}^i \mid z_{1:k}) / q(\boldsymbol{x}_{0:k}^i \mid z_{1:k}) \tag{5}$$

In which $\boldsymbol{x}_{0:k}^n = [\boldsymbol{x}_0^n, \boldsymbol{x}_1^n, \cdots, \boldsymbol{x}_k^n]$ represents the state of the i-th particle from time 0 to time k, $p(\boldsymbol{x}_{0:k}^i \mid z_{1:k})$ represents the target posterior probability density function, and $q(\boldsymbol{x}_{0:k}^i \mid z_{1:k})$ represents the importance density function. Assuming the motion of the target follows a first-order Markov process, using a recursive method to calculate the importance weights of the particles, the importance density function can be decomposed into:

$$q\left(\boldsymbol{x}_{0:k} | z_{1:k}\right) = q\left(\boldsymbol{x}_k | \boldsymbol{x}_{0:k-1}, z_{1:k}\right) q\left(\boldsymbol{x}_{0:k-1} | z_{1:k-1}\right) \tag{6}$$

In the traditional PF algorithm, assuming the target posterior probabilities are independent, this paper selects the prior density $p\left(\boldsymbol{x}_k \mid \boldsymbol{x}_{k-1}^i\right)$ as the importance function, then formula (5) can be modified to

$$\begin{aligned} w_k^i &\propto w_{k-1}^i p\left(z_k \mid \boldsymbol{x}_{0:k}^i\right) \prod_{j=1}^{r} p\left(\boldsymbol{x}_{k,j}^i \mid \boldsymbol{x}_{k-1,j}^i\right) / \prod_{j=1}^{r} q_j(\boldsymbol{x}_{k,j} \mid \boldsymbol{x}_{k-1,j}, z_k) \\ &= w_{k-1}^i \cdot \frac{p(z_k \mid \boldsymbol{x}_{0:k}^i)}{\prod_{j=1}^{r} p(z_k \mid \boldsymbol{x}_{k,j}^i)} \end{aligned} \tag{7}$$

From the formula (7), it can be seen that particle weights are closely related to the likelihood function of the target-measurement values.

3.2 Improved Likelihood Function

3.2.1 Adaptive Diffusion Function

The measured value of target tracking before detection is the signal energy intensity of the sensor in each direction. Suppose there are N resolution units in the observation region, which can be expressed as $j\{j = 1, ..., N\}$, and the length of each resolution unit is Δ_L. Then the measured value received at time k is

$z_k = \{I_k^j, j = 1,...,N\}$, where I_k^j represents the signal strength of the j-th resolution unit at time k. The expression can be modified to

$$I_k^j = \begin{cases} h_k^j(\theta_k) + noise_k^j & E_k = 1 \\ noise_k^j & E_k = 0 \end{cases} \tag{8}$$

In which, $h_k^j(\theta_k)$ is the signal strength of the target at point θ_k, and $noise_k^j$ is the corresponding observational noise. $E_k = 1$ means target exist; $E_k = 0$ means target do not exist.

When a target appears in an observational unit, in addition to the increase in the signal energy of this unit, the signal energy of adjacent observational units will also increase. The degree of increase shows an inverse relationship with that of this observational unit. Define the target response as the contribution of the target to the measurement area, and use the point diffusion likelihood function to describe the target response as:

$$h_k{}^j(\theta_k) = \frac{\Delta_L I_k}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(\Delta_L j - \theta_k)^2}{2\sigma^2}\right] \tag{9}$$

Where σ is the kernel width of the point spread function, representing the degree of energy scattering of the sensor. In the reference [7], σ is a fixed value preset according to the beamforming algorithm and the performance of the sonar system. The beamforming algorithm has directivity, and the value of σ varies for different resolution cells.

According to the above point spread function model, the likelihood function of the m-th target in the k-th frame follows an exponential distribution, expressed as

$$p(z_k \mid \boldsymbol{x}_{k,m}) = \frac{1}{I_k^j} \exp(-z_k^j / I_k^j) \tag{10}$$

Where j represents the resolution unit corresponding to θ_j, $j = \text{floor}(\theta_j / \Delta\theta) + 1$.

When the targets are in close proximity, the precondition in traditional methods that assumes the posterior probabilities of targets are independent of each other lapses. The target energies overlap in the azimuth spectrum, causing the likelihood function based on linear superposition to become void. When using a fixed kernel width, there may be a situation as shown in Figure 2: The estimated central responses of two targets are affected by the energy of neighboring targets and are superimposed into an energy response of one target, resulting in an error in the division of the two targets when their tracks cross. Shrink the kernel width in the area of neighboring targets to suppress energy overlap, and expand the kernel width in scenarios with low SNR or sparse data to enhance the accumulation of weak signals. Therefore, the formula for setting the kernel width of the point spread function expressed as

$$\sigma_j = \sigma_0 e^{-\alpha \cdot \frac{\Delta\theta}{\theta_{ref}}} e^{-\beta \cdot \frac{SNR}{\text{SNR}_0}} \quad (\Delta\theta < \theta_{ref} \text{ or } SNR < \text{SNR}_0) \tag{11}$$

Where σ_0 is the base kernel width, which is a fixed value preset by the system; α is the proximity coefficient, $\Delta\theta$ is the azimuth difference of the proximity target, and θ_{ref} is the angle for determining the proximity of two targets (this algorithm is set to 5°); β is the proximity coefficient, SNR is the current signal-to-noise ratio, and SNR_0 is the threshold for determining low signal-to-noise ratio.

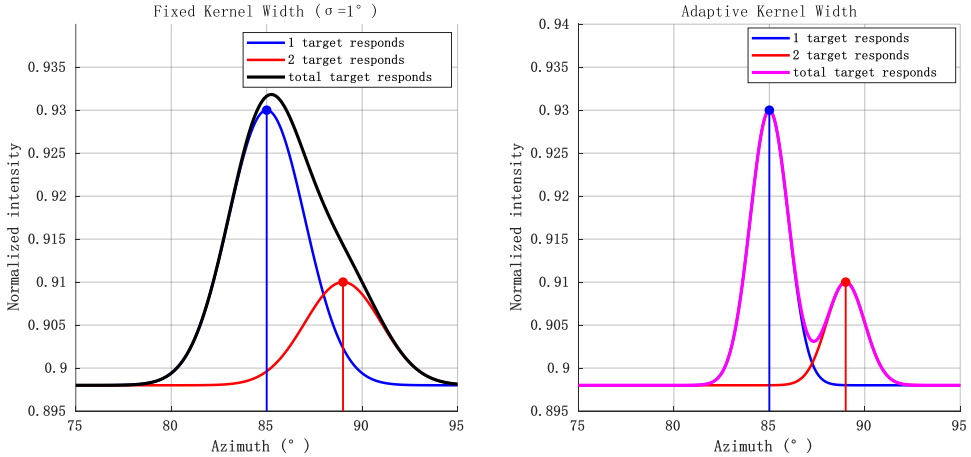


Figure 2– Comparison of Target Responses between Fixed and Adaptive Kernel Width

As shown in Figure 2, there are the response graphs of fixed kernel width and self - adapting kernel width when the targets are approaching: From the graphs, it can be seen that when the target pitch is 4°, the target responses under the fixed kernel width overlap seriously, forming a single peak, and it is impossible to distinguish the two targets, which ultimately leads to trace merger; while the self - adapting kernel width maintains a high likelihood at the target central locations and reduces the likelihood values between the targets, thus reducing the possibility of trace merger. The self - adapting kernel width makes the terrain of the likelihood function more accurately reflect the physical reality by dynamically optimizing the shape of the PSF.

3.2.2 Composite Bayesian Likelihood Function

Generally, the likelihood function is defined as the degree of matching between the measured value and the predicted value, which reflects the difference between the measured value $\boldsymbol{z}_k$ and the predicted value $h(\boldsymbol{x}_k)$ of the target state $\boldsymbol{x}_k$.

The movement of a target in space usually has continuity, that is, the target's state (position, velocity, etc.) does not change abruptly between adjacent frames. Based on this characteristic, the likelihood function can also be constructed from the aspect of target state change: the difference between the target state $\boldsymbol{x}_k$ and the state $\boldsymbol{x}_{k-1}$ at the previous moment.

The change of the target state between adjacent frames can be represented by a Gaussian distribution as

$$p(\boldsymbol{x}_k \mid \boldsymbol{x}_{k-N:k-1}) = \frac{1}{\sqrt{2\pi\sigma_x^2}} \exp\left(-\frac{\sum_{i=1}^{N} (\boldsymbol{x}_{i-k+1} - \boldsymbol{x}_{i-k})^2}{2\sigma_x^2}\right) \tag{12}$$

Combine the matching degree between the measured value and the predicted value and the matching degree of state changes to obtain the joint likelihood function as

$$p_{\text{composite}}(\boldsymbol{z}_k \mid \boldsymbol{x}_k) = p(\boldsymbol{z}_k \mid \boldsymbol{x}_{k,m}) \cdot \frac{1}{\sqrt{2\pi\sigma_x^2}} \exp\left(-\frac{(\boldsymbol{x}_k - f(\boldsymbol{x}_{k-1}))^2}{2\sigma_x^2}\right) \tag{13}$$

By combining these two points, a likelihood function that can not only reflect the matching degree between the measurement data and the target state but also reflect the smoothness of the target motion can be constructed.

3.3 Target Particle Weight Calculation and Resampling

After each target partition in the particle completes independent sampling, partition weight calculation, and resampling, the sampling result of the particle at the current moment is obtained.

Importance sampling, calculating particle weights, and normalizing as follow:

$$w_k^i = \frac{w_{k-1}^i p(\boldsymbol{z}_k \mid \boldsymbol{x}_{0:k}^i)}{\prod_{m=1}^{r} p(\boldsymbol{z}_k \mid \boldsymbol{x}_{k,m}^i)} \tag{14}$$

And perform weight normalization of Formula 14 :

$$\tilde{w}_k^i = w_k^i / \sum_{i=1}^{N} w_k^i \tag{15}$$

To avoid particle degradation, resampling can reduce the impact of particle degradation. However, overly frequent resampling can lead to sample impoverishment because there are identical particles after resampling. Therefore, whether to conduct child particle resampling is decided by calculating the number of effective particles:

$$\hat{N}_{\text{eff},j} = 1 / \sum_{n=1}^{n_p} \left(p_{\text{composite}}(\boldsymbol{z}_k \mid \boldsymbol{x}_{k,j}^n) \right)^2 \tag{16}$$

Conduct resampling of the particle swarm to obtain a particle collection.

Finally, calculate the estimated state of the target.

$$\hat{\boldsymbol{x}}_{k,m} = \sum_{i=1}^{N} \boldsymbol{x}_k^i \tilde{w}_k^i \tag{17}$$

3.4 Target Birth and Death Judgment

The Multiple Hypothesis Tracking (MHT) assumes the observation sources of each observational data fall into three cases: an existing target, a newly emerging target, and clutter outliers. For multiple observational data, there are multiple assumptions of data correlation combinations. If an observational data

has a high correlation with several assumptions, instead of uniquely pairing the observational data to a single assumption, each possible correlation assumption is passed to the next time step for prediction and update. Then, through subsequent observational data, it is determined which of these highly correlated assumptions is correct. The framework of the MHT algorithm is shown in Figure 3.

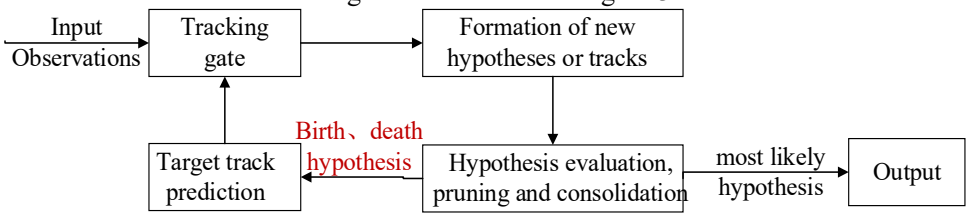


Figure 3– Multi-Objective Hypothesis Algorithm Framework

3.4.1 Target Birth Judgment

The pre-detection tracking algorithm theoretically does not set a threshold for the raw data, meaning that all measurement regions are suspected targets. However, this method requires searching the entire beam space, resulting in an excessively high computational load and is actually not operable. Therefore, the suspected targets are first filtered.

For a certain measurement value $z(b)$, a data window with a window length of $2N+1$ is selected, and the data within the window is $z(b-N), z(b-N+1), \cdots, z(b), \cdots, z(b+N-1), z(b+N)$. Weight the data in the window and take the weighted average to get the threshold size， express as:

$$\Lambda_1 = \varepsilon \frac{1}{2N+1} \sum_{i=b-N}^{b+N} z(i) \tag{18}$$

In which ε is the threshold adjustment parameter. If $z(b) > \Lambda_1$, it is determined as a suspicious target, and finally the set of all suspicious targets is obtained as U_k. Then, a likelihood ratio hypothesis test is used to judge the real target among them as follow:

$$P(B_k \mid U_k, Z_k) = \begin{cases} P_r & birth \quad hypothesis \\ 1-P_r & clutter \quad hypothesis \end{cases} \tag{19}$$

In which B_k represents the newly born real target, and P_r represents the reception probability.

The likelihood ratio of the suspicious target can be written as

$$p(\boldsymbol{z}_k \mid \boldsymbol{x}_{k,m}, U_k) = \frac{p(\boldsymbol{z}_k^i \mid \boldsymbol{x}_{k,m}, E_k = 1)}{p(\boldsymbol{z}_k^i, E_k = 0)} \tag{20}$$

Determine whether a suspicious target is a real new target, which can be determined by comparing the weight of the particle with the set threshold value as follow

$$\sum_{i=1}^{N} p(\boldsymbol{z}_k \mid \boldsymbol{x}_{k,m}, U_k) > \eta \tag{21}$$

In the case where the target signal-to-noise ratio is very low, it is difficult to extract the correct real target through single-frame judgment. Suppose the spatial spectrum of a certain target in a certain frame is very low. At this time, the sum of particle weights is very likely to be lower than the judgment gate η. In this case, the target will be deleted, resulting in intermittent target tracks. At this time, multi-frame joint judgment is required. Since the target follows the law of motion between consecutive frames, while the noise is stochastic between consecutive frames. Therefore, clutter can be removed through the discrimination of consecutive multi-frames to identify the real target.

N_f represents the number of joint frames. Modify Equation 19 as follow:

$$P(B_k \mid U_k, \boldsymbol{Z}_{k-N_f+1:k}) = \begin{cases} P_r & birth \quad hypothesis \\ 1-P_r & clutter \quad hypothesis \end{cases} \tag{22}$$

At the same time, Modify Equation 21 as follow:

$$\sum_{f=1}^{N_f} \sum_{i=1}^{N} p(\boldsymbol{z}_k \mid \boldsymbol{x}_{k,m}, U_k) > \eta_f \tag{23}$$

3.4.2 Target Death Judgment

Similar to the judgment of new targets, determining whether the targets in the previous frame continue to

form a track is also a problem of binary likelihood assumption, and add joint multi-frame judgment as follow:

$$P(\boldsymbol{S}_k \mid I_{k-1}, \boldsymbol{Z}_{k-N_f+1:k}) = \begin{cases} P_r & exist \quad hypothesis \\ 1-P_r & death \quad hypothesis \end{cases} \tag{24}$$

Where the set S_k represents the targets that did not disappear at the previous moment. At this time, the likelihood ratio threshold judgment formula con be express as

$$p(\boldsymbol{z}_k \mid x_{k,m}, I_{k-1}) = \frac{p(\boldsymbol{z}_k^i \mid \boldsymbol{x}_{k,m}, E_k = 1)}{p(\boldsymbol{z}_k^i, E_k = 0)} \tag{25}$$

The weight and judgment formula for target death can be written as

$$\sum_{f=1}^{N_f} \sum_{i=1}^{N} p(\boldsymbol{z}_k \mid \boldsymbol{x}_{k,m}, I_{k-1}) > \eta_{f1} \tag{26}$$

4. Simulation and Performance Evaluation

The parameters of the passive sonar system in the simulation are as follows: The received signal is a broadband stochastic signal with an operating frequency band of a. The signal sampling frequency is a. The number of hydrophones in the uniform linear array is a, and the element spacing is 0.3m. The spatial spectrum azimuth resolution is 0.1 degrees. The number of resolution units is a. The target movement time in the simulation scene is 120 frames.

The simulation environment includes 3 targets that are relatively close to each other. Its simulation situation graph is shown as Figure 4. Beamforming processing is conducted on the simulation signal, and the obtained Time-Bearing Display is shown as Figure 5. From the Time-Bearing Display, it can be seen that when the targets are approaching, the energy values of the targets at the trajectory intersection points increase significantly.

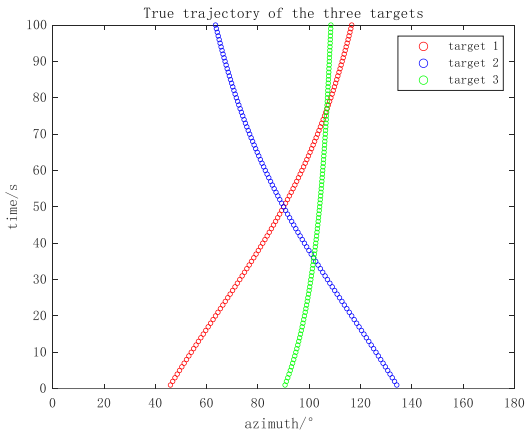


Figure 4– true trajectory of the three targets

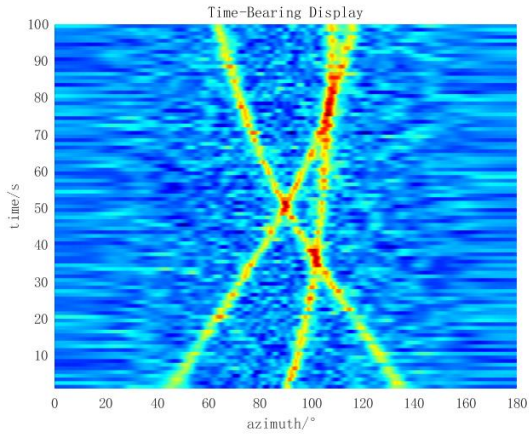


Figure 5– Time-Bearing Display

The following analyzes and compares the particle filter pre-detection tracking algorithm before and after improvement. Before improvement, as shown in the figure 6, the trajectories of two targets intersect at 70s. At this time, figure 7 shows that, the original algorithm has tracking errors and incorrect batch division of targets, while the improved algorithm has carried out good tracking and batch division at the three trajectory intersection points .

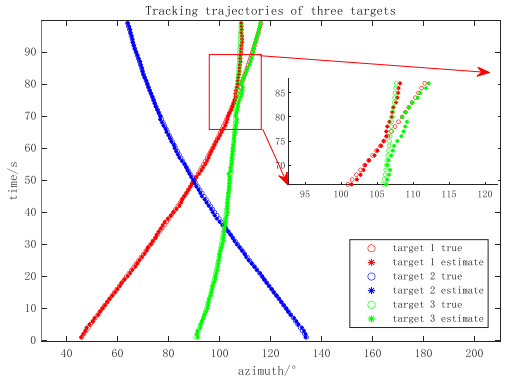


Figure 6– false estimation using the original method

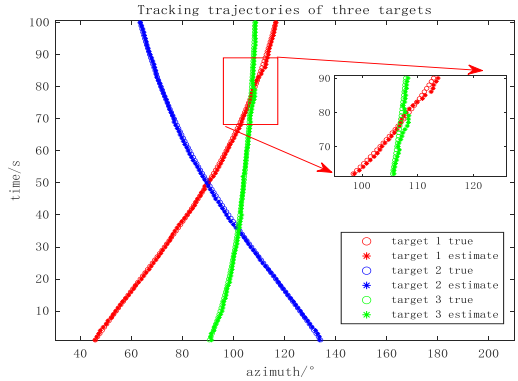


Figure7–ture estimation using the improved method

Among them, the OSPA error comparison is shown in Figure 8(a), and the errors of the estimated and true orientations of the three target tracks are shown in Figure 8(b-d). From the figure, it can be seen that before the trajectories did not cross, the OPSA errors of the improved algorithm were all smaller than those before the improvement. 50s, when the trajectories of target 2 and target 1 crossed, both algorithms were able to differentiate between the two targets due to the different motion trends and directions of the two targets; however, at 70s, when the trajectories of target 1 and target 3 were close to crossing, the original algorithm had an error in the delineation of trajectories, and the estimated bearing angle The error in estimating the azimuth angle increases dramatically, while the algorithm with improved likelihood function tends to 0 after a small increase in the error, and tracks the correct two targets.

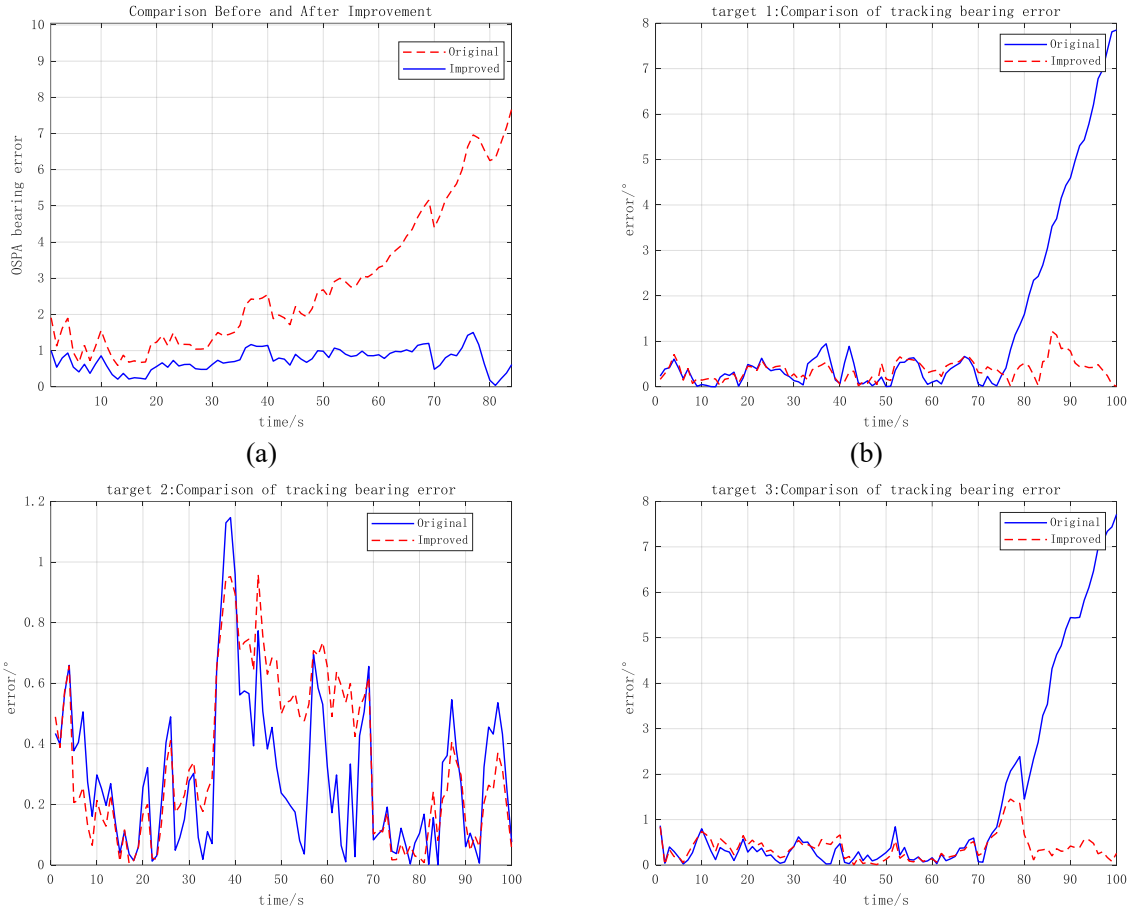


Figure 8–OPSA error and RMS errors of the three targets

5. CONCLUSIONS

Aiming at the target proximity problem of passive sonar underwater multi-target tracking, the improved likelihood function particle filtering pre-detection tracking method is proposed, the adaptive diffusion function quantitative likelihood is deduced, and the joint motion continuity state transfer likelihood improves the tracking performance under the neighbor-crossing scenarios; and the target tracking is realized by multi-hypothesis tracking and particle filtering. The simulation results show that the improved method improves the tracking performance of the target under the neighbor crossing scene, and can effectively judge the trajectory batch division under the neighbor crossing scene.

ACKNOWLEDGEMENTS

This work is supported by the National Natural Science Foundation of China under Grant Nos. 62271162 and the Hainan Provincial Natural Science Foundation under Grant Nos. 624MS081.

REFERENCES

1. Boers Y, Driessen J N. Particle filter based detection for tracking[C]. Proceedings of the 2001 American Control Conference, 2001, 6: 4393-4397.

2. Buzzi S, Lops M, Venturino L, et al. Track-before-detect procedures in a multi-target environment[J]. IEEE Transactions on Aerospace and Electronic Systems, 2008, 44(3): 1135-1150.
3. Brennan M, Kreucher C. Multistatic radar change detection using a sparse imaging approach[C]. 2012 IEEE Radar Conference, 2012: 0045-0050.
4. Yi W, Fu L, García-Fernández Á F,et al. Particle filtering based track-before-detect method for passive array sonar systems. Signal Processing,2019;165: 303–314. doi:10.1016/j.sigpro.2019.07.027
5. Northardt T., Nardone S C. Track-Before-Detect Bearings-Only Localization Performance in Complex Passive Sonar Scenarios: A Case Study. IEEE Journal of Oceanic Engineering, 2019;44(2): 482–491. doi:10.1109/JOE.2018.2811419
6. Wang Y, Li X, Qi B, et al. A track-before-detect method for neighboring targets based on auxiliary particle filter in passive sonar scenarios[J]. Acta Acustica, 2023, 48(2): 277-290.
7. RUTTEN M G, RISTIC B, GORDON N J. A comparison of particle filters for recursive track before-detect[C]. 2005 7thInternational Conference on Information Fusion, IEEE, 2005,165-179..

Estimation of source levels for unmanned underwater vehicles

GuhnHyeko Ko[1]; ChangSoo Kim[2]; ByoungNam Kim[3]; YoSup Park[4]; CheolSoo Park[5]; DongGuk Paeng[6]

[1,2,6] Jeju National University

[3,4] Korea Institute of Ocean Science and Technology

[5] Korea Research Institute of Ships and Ocean Engineering

ABSTRACT

With increasing interest in unmanned underwater vehicles (UUVs), source levels are essential to understand their radiated noises and underwater acoustic propagation characteristics. In this study, a source level of a UUV (Sparus 2, Iqua Robotics) was estimated by fitting a regression curve of measured sound pressure levels (SPL) at various distances in the east sea of Korea. Field measurements were conducted in the Yeongilman area offshore Pohang, Korea, using an underwater hydrophone (ST300 HF, SoundTrap) deployed at a depth of 15 m (water depth: 40 m). SPL at each distance was derived by integrating the power spectral density (PSD) over the 800–1000 Hz frequency band, corresponding to the dominant frequency range of UUV's propeller noise at a constant speed. Assuming a spherical spreading loss model, a regression curve was fitted to these SPL measurements to a 1 m reference distance, yielding an estimated source level of 114 dB with a coefficient of determination (R^2) of 0.45.

Research on bearings-only association algorithm under unknown measurement noise condition

Bin Qi[1,2,3]; Jiaxin Zhou[1,2,3,*]; Jinjin Wang[1,2,3]; Yilin Wang[1,2,3]
[1] National Key Laboratory of Underwater Acoustic Technology, Harbin Engineering University, Harbin 150001,China
[2] Key Laboratory of Marine Information Acquisition and Security(Harbin Engineering University),Ministry of Industry and Information Technology; Harbin 150001,China
[3] College of Underwater Acoustic Engineering, Harbin Engineering University, Harbin 150001,China

ABSTRACT
This paper investigates the target tracking problem of Rao-Blackwellized Monte Carlo Data Association (RBMCDA) algorithm in passive scenarios. Firstly, this paper proposes integrating an adaptive unscented Kalman filter (AUKF) into RBMCDA algorithm, which reduces tracking errors under nonlinear measurement conditions. To address label mismatching during target proximity, a proximity-based adapt label management strategy combined with Density Based Spatial Clustering of Applications with Noise clustering is proposed. This method improves association accuracy at target proximity by sampling particles within clusters according to their target labels.

Keywords: RBMCDA, Label Management, AUKF

1. INTRODUCTION
The history of human societal development is also a history of progressively exploring and understanding the oceans. As the largest ecosystem on Earth, the ocean plays an indispensable role in human life through its natural resources and ecological functions (1). Against the backdrop of global competition for ocean resources, China has actively formulated strategies to strengthen its position through marine science and technology and to build itself into a maritime power. As a critical component of this national strategy, the field of underwater multi-target tracking has long been a central focus of underwater acoustics development.

The key to underwater tracking lies in the use of sonar and other underwater sensors to obtain the target's azimuth, speed, depth and other information to continuously detect the movement of the same target (2), the tracking method is mainly divided into active tracking and passive tracking method (3).Compared to active tracking, passive tracking systems are widely used in submarine reconnaissance and underwater surveillance due to their strong concealment, as they do not actively emit acoustic signals but instead rely on passively receiving noise radiated by the target (4). Since passive tracking usually only obtains the target's azimuth information (5), there are still many shortcomings in practical applications. The complexity of the marine environment and the nonlinear relationship between the actual target and the obtained measurements make it extremely challenging to estimate target positions. Under conditions such as weak target signals, significant changes in motion states, varying numbers of targets, or closely spaced multiple targets, the resulting tracking trajectories are prone to distortion, placing higher demands on the theories and methods of tracking and localization.

Multi-target tracking requires real-time processing of information from multiple targets and faces complex challenges such as missed detections, false alarms, and difficulties in distinguishing data from different targets (6). The traditional Rao-Blackwellized Monte Carlo Data Association (RBMCDA) algorithm, as a hybrid Bayesian filtering method, provides an effective solution to complex multi-target tracking problems (7). By combining Kalman filtering for continuous target state estimation with particle filtering for handling data association uncertainties, it achieves good accuracy and robustness even in challenging environments (8). However, issues inherent to RBMCDA algorithm, such as particle confusion when targets are in close proximity (9) and significant errors in Kalman filtering under nonlinear target

* Corresponding author: zhoujiaxin1@hrbeu.edu.cn

motion (10), continue to limit its applicability in more complex and large-scale scenarios. This paper focuses on the aforementioned issues of the RBMCDA algorithm through experimental studies, aiming to enhance its applicability and tracking accuracy in complex multi-target tracking scenarios.

2. Algorithm Principles

2.1 Rao-Blackwellized Particle Filtering

The Rao-Blackwellized Particle Filtering (RBPF), also known as marginal particle filtering, is commonly used to address target tracking problems in nonlinear and non-Gaussian scenarios due to its combination of analytical filtering and particle filtering. It significantly improves estimation efficiency and accuracy in high-dimensional state spaces.

Let x_k and z_k denote the target state and the measurement at frame k, respectively. Then, the state-space model for the tracked target is given as follows:

$$\begin{cases} x_k = (A_{k-1}(b_{k-1})x_{k-1} + Q_{k-1}(b_{k-1}) \\ z_k = \left(H_k(b_k)x_k + R_k(b_k)\right) \\ b_k \sim p(b_k \mid b_{k-1}) \end{cases} \tag{1}$$

where A_{k-1} is the state transition matrix at frame $k-1$, Q_{k-1} is the process noise covariance at frame $k-1$, H_k is the observation matrix of the target at frame k, R_k is the measurement noise covariance at frame k. b_k is a nonlinear variable, and the values of A_{k-1}, Q_{k-1}, H_k and R_k are all dependent on it.

RBPF applies the Kalman Filter (KF) and Particle Filter (PF) to the linear and nonlinear components, respectively. In target tracking, when the state of a target a is randomly associated with a measurement ($c_k = a$), the measurement likelihood function can be expressed as follows:

$$p\left(z_k \mid x_k^i, c_k = a\right) = \mathcal{N}(z_k \mid H_k, x_k, R_k) \tag{2}$$

Assuming there are N particles, the particle set at frame $k-1$ is given as follows, where $w_k^{\,i}$ denotes the weight corresponding to the i-th particle at frame $k-1$, μ_{k-1}^i and P_{k-1}^i are the mean and covariance corresponding to the linear state estimate of particle i at frame $k-1$.

$$\{w_{k-1}^i, b_{k-1}^i, u_{k-1}^i, P_{k-1}^i, i = 1, 2, \cdots, N\} \tag{3}$$

for each particle i, Kalman prediction is performed to obtain the prior state and covariance of the target. At each frame, the RBPF algorithm proceeds as follows:

$$\begin{cases} \mu_{k|k-1}^i = A_{k-1}(\mathrm{b}_{k-1}^i) \cdot \mu_{k-1|k-1}^i \\ P_{k|k-1}^i = A_{k-1}(\mathrm{b}_{k-1}^i) \cdot P_{k-1|k-1}^i \cdot A_{k-1}^{\mathrm{T}}(\mathrm{b}_{k-1}^i) + Q_{k-1}(\mathrm{b}_{k-1}^i) \end{cases} \tag{4}$$

for each particle i, a Kalman update is performed to obtain the target distribution with the likelihood.

$$\begin{cases} \hat{\mathbf{z}}_k^i = H_k \mu_{k|k-1}^i \\ S_k^i = H_k P_{k|k-1}^i H_k^{\mathrm{T}} + R_k \\ K_k^i = P_{k|k-1}^i H_k^{\mathrm{T}} \left(S_k^i\right)^{-1} \\ \mu_{k|k}^i = \mu_{k|k-1}^i + K_k^i (z_k^i - \hat{z}_k^i) \\ P_{k|k}^i = P_{k|k-1}^i - K_k^i S_k^i \left(K_k^i\right)^{\mathrm{T}} \end{cases} \tag{5}$$

$$p(z_k \mid \mathrm{b}_{1:k}^i, Z^{k-1}) = \mathcal{N}\left(z_k \mid \boldsymbol{H}_k(b_k^i)\mu_{k|k-1}^i, \boldsymbol{H}_k(\mathrm{b}_k^i)\boldsymbol{P}_{k|k-1}^i \boldsymbol{H}_k^{\mathrm{T}}(\mathrm{b}_k^i) + \boldsymbol{R}_k(\mathrm{b}_k^i)\right) \tag{6}$$

where $\hat{\mathbf{z}}_k^i$ is the predicted measurement matrix, S_k^i is the observed covariance, K_k^i is the Kalman gain, $\mu_{k|k}^i$ and $P_{k|k}^i$ correspond to the state update and covariance update, Update particle weights and normalize c_k^i represents the data association history of particle i, which will be introduced later.

$$\begin{cases} \tilde{w}_k^i = w_{k-1}^i \sum_{j=0}^{n_k+1} p\left(c_k = a \mid c_{1:k}^i\right) p\left(z_k \mid c_{1:k}^i, Z^{k-1}\right) \\ w_k^i = \dfrac{\tilde{w}_k^i}{\sum_{n=0}^{n_k+1} \tilde{w}_k^i} \end{cases} \tag{7}$$

2.2 Principles of RBMCDA

After RBPF algorithm is used to estimate the target states, the latent variable $\mu_k^i = \{e_k, c_k\}$ serves as the key factor linking RBPF algorithm and RBMCDA algorithm. The implementation of RBMCDA can be divided into three processes: prediction, update, and resampling. In RBMCDA algorithm, the state indicator $e_k = [e_{1,k}, e_{2,k} \ldots\ldots e_{\xi,k}]$ is a binary vector that indicates the existence of targets, where $e_{j,k} = 1$ denotes that the target exists and $e_{j,k} = 0$ indicates that the target has not yet appeared or has disappeared and it determines the number of targets. Define c_k to characterize the association of data with a target or clutter. If there are a_{k-1} targets at frame k, and at most one new target is born in each cycle, then $c_k \epsilon \left[0, a_{k-1}+1\right]$, where $c_k = 0$ indicates that the data is associated with clutter, and other values indicate association with the corresponding targets. If a target is determined to have disappeared, the corresponding data is deleted. For surviving targets, Kalman prediction is performed to obtain the prior state estimate of the particles. Based on the posterior probability of the association hypothesis and μ_k^i, the state and covariance of the target are obtained, and the probabilities of association with clutter and new target birth are computed. Targets that are not associated with a measurement then only perform the prediction frame, and ultimately determine whether to perform resampling based on the weights. In each cycle of computation, RBMCDA algorithm defines the importance distribution function as $q(u_k \mid Z^k, u_{k-1}) = p(u_k \mid Z^k, u_{k-1})$, Substituting e_k^i gives:

$$\begin{cases} b_k^i \sim p(b_k^i \mid Z^k, b_{1:k-1}^i) \\ \tilde{w}_k^i \propto w_{k-1}^i \dfrac{p(Z_k \mid b_k^i, Z^{k-1}, b_{1:k-1}^i) p(b_k^i \mid b_{1:k-1}^i)}{p(b_k^i \mid Z^k, b_{1:k-1}^i)} \\ w_k^i = \tilde{w}_k^i \,/\, \sum_{i=0}^{n_{k+2}} \tilde{w}_k^i \end{cases} \tag{8}$$

using the particle set at frame $k-1$ k-1, the particle set at frame k can be obtained, and the target state estimate at frame k is finally given by:

$$p(\boldsymbol{x}_k, c_k \mid Z^k) \approx \sum_{i=1}^{N} w_k^i \cdot \delta(c_k - c_k^i) \cdot \mathcal{N}(\boldsymbol{x}_k \mid \mu_{k|k}^i, \boldsymbol{P}_{k|k}^i) \tag{9}$$

2.3 AUKF-RBMCDA Algorithm

The traditional RBMCDA algorithm uses Kalman filtering in the target state estimation phase for linear, Gaussian scenarios, whereas when the scenario is nonlinear and non-Gaussian, such as nonlinear motion of the target, the process generates an error accumulation due to its linear approximation, which is ultimately transferred to the entire framework through Rao-Blackwellized algorithm. Based on this, the Unscented Kalman Filter is introduced to replace the original Kalman Filter for target state estimation.

The core of the Unscented Kalman Filter lies in the Unscented Transform (UT), which constructs a set of deterministically chosen sigma points whose mean and covariance match those of the original distribution. These sigma points are then propagated through the nonlinear transformation, and the transformed points are used to compute a new mean and covariance, thereby avoiding errors caused by linear approximation. Like the Kalman Filter, the Unscented Kalman Filter also consists of prediction and update stages. Let m_k, P_k denote the mean and covariance of the target state at frame k, respectively. First, $2n+1$ sigma points：

$$\begin{cases} \chi_0 = \mathrm{m}_{k-1} \\ \chi_i = \mathrm{m}_{k-1} + \left(\sqrt{(n+\lambda)P_{k-1}}\right)_i, i = 1,\ldots,n \\ \chi_i = \mathrm{m}_{k-1} - \left(\sqrt{(n+\lambda)P_{k-1}}\right)_{i-n}, i = n+1,\ldots,2n \end{cases} \tag{10}$$

where $\lambda = \alpha^2(n+\kappa) - n$. After generating the sigma points, they are propagated through the nonlinear state transition function to obtain the predicted sigma points, which are then used to compute the predicted state mean and covariance:

$$\begin{cases} \chi^i_{k|k-1} = f\left(\chi^i_{k-1}\right) \\ \mu^i_{k|k-1} = \sum_{i=0}^{2n} W^i_m \chi^i_{k|k-1} \\ P^i_{k|k-1} = \sum_{i=0}^{2n} W^i_c \left[\chi^i_{k|k-1} - \mu^i_{k|k-1}\right]\left[\chi^i_{k|k-1} - \mu^i_{k|k-1}\right]^T + \mathbf{Q} \end{cases} \tag{11}$$

In most cases, the parameters are chosen as $\alpha = 1, \mathrm{b} = 2, \kappa = 3 - \mathrm{n}$, Q is the measurement noise covariance, W^i_m, W^i_c are weight coefficients, which are calculated as follows:

$$W^i_c = \begin{cases} \dfrac{\lambda}{n+\lambda} + 1 - a^2 + b, i = 0 \\ \dfrac{1}{2(n+\lambda)}, i = 1,2,.....,2n \end{cases} \tag{12}$$

$$W^i_m = \begin{cases} \dfrac{\lambda}{n+\lambda}, i = 0 \\ \dfrac{1}{2(n+\lambda)}, i = 1,2,.....,2n \end{cases} \tag{13}$$

For each particle i, the predicted sigma points in the update frame are passed through the measurement equation to obtain the particle's state estimate and covariance estimate:

$$\begin{cases} \xi^i_{k|k-1} = h(\chi^i_{k|k-1}) \\ P^i_{xz} = \sum_{i=0}^{2n} w^i_c (\chi^i_{k|k-1} - \mu^i_{k|k-1})(\xi^i_{k|k-1} - \hat{z}^i_k)^T \\ S^i_k = \sum_{i=0}^{2n} w^i_c \left(\xi^i_{k|k-1} - \hat{z}^i_k\right)\left(\xi^i_{k|k-1} - \hat{z}^i_k\right)^T + R \\ K^i_k = P^i_{xz} S^{i^{-1}}_k \\ \mu^i_{k|k} = \mu^i_{k|k-1} + K^i_k (z^i_k - \hat{z}^i_k) \\ P^i_{k|k} = P^i_{k|k-1} - K^i_k S^i_k K^{i\,T}_k \end{cases} \tag{14}$$

where $\xi^i_{k|k-1}$ is the observation sigma points, $\hat{z}^i_k$ is the expected predicted measurement, P^i_{xz} is the state-measurement cross-covariance, S^i_k is the predicted measurement covariance, K^i_k is the Kalman gain, $\mu^i_{k|k}$ and $P^i_{k|k}$ correspond to the state update and covariance update, respectively.

As an improved Kalman filtering approach for nonlinear and non-Gaussian conditions, the Unscented Kalman Filter effectively avoids errors accumulated from linearization. However, in practical applications, many parameters are often unknown and time-varying, which contradicts the assumption in the Unscented Kalman Filter that the process noise covariance and measurement noise covariance are known and fixed. Therefore, a feedback mechanism is incorporated into the Unscented Kalman Filter to dynamically adjust the process noise covariance and measurement noise covariance based on the discrepancy between observed and predicted data, enhancing the accuracy and robustness of the algorithm.

The previously mentioned Unscented Kalman process produces the difference between the actual and predicted measurements, referred to as the innovation residual, which indicates the accuracy of the prediction. Therefore, introducing a real-time feedback mechanism based on the innovation residual can effectively

prevent the accumulation of errors between measurements and predictions during the iterative process. In each update, factors such as changes in noise variance, external environmental disturbances, and modeling errors cause the process noise covariance and measurement noise covariance to vary in real time. By incorporating the innovation residual into the computation of the noise covariance matrices Q and R, the system quantifies model uncertainty based on the variation in the current state covariance and the measurement residual, and accordingly updates the covariance matrices.

$$\begin{cases} Q_k = Q_k \cdot \max\left(1, \dfrac{|\,diag(P_k - P_{k-1})\,|}{diag(Q_k)}\right) \\ R_k = R_k \cdot \max\left(1, \dfrac{|\,z_k - \hat{z}_k\,|}{\mathrm{diag}(R_k)}\right) \end{cases} \tag{15}$$

where P_k and P_{k-1} represent the covariance estimates obtained at the current and previous frames, z_k and $\hat{z}_k$ represent the current measurement and the predicted measurement. Through the above adaptive mechanism, the filter can automatically increase the corresponding covariance estimates when there are sudden changes in the state or abrupt increases in measurement errors, thereby reducing the impact of external disturbances and enhancing the system's adaptability and fault tolerance in complex environments. As discussed above, calculating the probabilities of data association to corresponding targets, clutter, and target birth, target birth model is established as:

$$p\left(e_k \mid e_{1:k-1}\right) = \begin{cases} pb, e_{k-1}^i = 1 \\ 1 - pb, e_{k-1}^i = 0 \end{cases}, i = 0,1,2,\ldots,n_k + 1 \tag{16}$$

where pb is the prior probability of the target. The probabilities of the target associating with each hypothesis are:

$$p(c_k = a \mid e_{1:k-1}, c_{1:k-1}) = \begin{cases} cp, & a = 0 \\ \dfrac{1-cp}{n_{k-1}}, & 1 \le a \le n_{k-1} \\ 1, & a = n_{k-1} + 1 \end{cases} \tag{17}$$

where cp denotes the prior probability of clutter. The prior probabilities in RBMCDA can be obtained by Combining the above two equations:

$$p(u_k^i \mid u_{1k-1}^i) = \begin{cases} (1 - pb) \cdot cp, & clutter\ association \\ \dfrac{(1 - pb) \cdot (1 - cp)}{n_{k-1}}, & surviving\ targets\ association \\ pb, & target\ birth \end{cases} \tag{18}$$

The posterior probability density of target association can be obtained by using the prior probability with the likelihood described in the previous section:

$$p\left(b_k^i \mid Z^k, b_{1:k-1}^i\right) = \frac{p\left(z_k \mid b_k^i, Z^{k-1}, b_{1:k-1}^i\right) p\left(b_k^i \mid Z^{k-1}, b_{1:k-1}^i\right)}{p\left(z_k \mid Z^{k-1}, b_{1:k-1}^i\right)} \tag{19}$$

Table 1 – Improved AUKF-RBMCDA Algorithm

step1	Determine target existence to obtain the number of surviving and disappeared targets.
step2	Generate 2n+1 sigma points $\chi_0, \chi_1, \ldots, \chi_{2n+1}$
step3	Compute the predicted sigma points mapped through the nonlinear state transition function. $\chi_{k\|k-1}^i$
step4	Calculate the predicted state mean $\mu_{k\|k-1}^i$ and covariance $P_{k\|k-1}^i$ based on the obtained sigma points.
step5	Compute the target birth probability and the clutter association probability, assign each measurement to either an existing target $a1$ or a newly initiated target $a2$
step6	Propagate the predicted sigma points through the observation equation to

obtain the updated state $\mu_{k|k}^{i}$ and covariance $P_{k|k}^{i}$ of the particle.

step7 Adaptively adjust Q and R based on the residual.

$$Q_k = Q_k \cdot \max\left(1, \frac{|diag(P_k - P_{k-1})|}{diag(Q_k)}\right)$$

$$R_k = R_k \cdot \max\left(1, \frac{|z_k - \hat{z}_k|}{\mathrm{diag}(R_k)}\right)$$

step8 Update and normalize the particle weights.

step9 Particle Resampling.

step10 Obtain the estimated target output

$$p(\boldsymbol{x}_k, c_k \mid Z^k) \approx \sum_{i=1}^{N} w_k^i \cdot \delta(c_k - c_k^i) \cdot \mathcal{N}(\boldsymbol{x}_k \mid \mu_{k|k}^i, \boldsymbol{P}_{k|k}^i)$$

step11 Delete tracks that have not been assigned observations for an extended period, and assign tracks to newly detected targets.

2.4 Improved Label Management Based on Target Proximity

AUKF-RBMCDA algorithm can be regarded as track management at the particle level, where each iteration involves data association sampling and track state updating. It yields satisfactory tracking results when targets are well separated. But when the target is close and the observations of multiple targets overlap at the sensor level, tracking failure may occur due to the difficulty in distinguishing the targets. In order to obtain better tracking results, many clustering algorithms have been proposed over the years to achieve the correlation between points and trajectories.

Clustering refers to an unsupervised classification method that partitions data into different categories based on similarity. Among them, Density Based Spatial Clustering of Applications with Noise (DBSCAN) is a commonly used density-based clustering method that classifies particles based on two parameters: the radius Eps and the minimum number of points threshold $MinPts$. The core concepts of particles include the following three types:

Core Point: A data point with at least $MinPts$ points within its Eps radius neighborhood.

Border Point: A data point that has fewer than $MinPts$ points within its Eps radius, but lies within the Eps neighborhood of a core point.

Noise Point: A data point that does not belong to either the core point or border point categories.

In RBMCDA algorithm, the particle state set D_k can be obtained, and the estimated state and number of targets under each particle in the set may be different. In this case, DBSCAN can be used to obtain the clustering clusters formed by particles corresponding to the same real target. For example, at frame k , L clusters can be represented as:

$$C_{L,k} = \{C_{1,k}, C_{2,k}, ...C_{l,k}\}, \quad l = 1, 2, ...L \tag{20}$$

due to the information of state mean and corresponding particle weights in C , the state estimation of the corresponding target can be represented by the following equation:

$$\hat{\mu}_{l,k} = \sum_{\mu_{a,k}^i \epsilon C_{l,k}} w_k^i \mu_{j,k}^i \tag{21}$$

Density-based clustering in RBMCDA algorithm can improve trajectory separation when targets are in close proximity. However, when different target trajectories become adjacent or intersect, the clustering process still tends to suffer from “merging” effects. To address this, trajectory labels are incorporated into the data processing procedure, and a Adjust Label Management (ALM) strategy based on target proximity is proposed. Target trajectory labels can be understood as identifiers assigned to each newly born target in chronological order to mark its identity. When two targets are close to each other or occlude one another, distinct labels help prevent target confusion and maintain identity consistency. During the recursive particle computation, the target label for each particle remains consistent when the measurement is associated with the target. Therefore, a unified target label can be assigned to each trajectory by evaluating the degree of correspondence between particles in clusters across adjacent frames and different trajectories. The label management process can be described as follows:

Assuming a total of N particles, the particle label matrix of clusters at frame k , denoted as L_k , is matched with the trajectory label matrix at frame $k-1$, denoted as T_{k-1} , The matching rate is defined as θ ,

d is the number of matching elements between the two matrices. When θ exceeds a predefined threshold, the match is considered successful, and the particle labels in L_k are used to update the trajectory labels T_{k-1}, resulting in the updated trajectory labels T_k. In the cluster particle label matrix L_k at frame k, the rows and columns represent each cluster at the current frame and the particle indices within each cluster, at frame $k-1$, the trajectory label matrix T_{k-1} has rows representing the surviving trajectories at the current frame; the first column contains the identifiers of different trajectories, while columns 2 to $N+1$ indicate the particle indices matched from clusters at frame $k-2$. The matching process between the two matrices involves matching the i-th column of L_k with the $(i+1)$-th column of T_{k-1}.If the number of particles in a cluster exceeds N, it indicates target proximity, where particles corresponding to multiple targets are grouped into a single cluster. In previous label management processes, only the first N particles in this cluster were considered effective for computation, which contradicts the earlier assumption that particles corresponding to the same target form one cluster and may cause significant tracking errors to some extent. To prevent such situations and improve tracking accuracy, it is proposed that when the number of particles in a cluster exceeds N, the cluster should be split into two based on the target identities of the particles from the previous frame. Each resulting sub-cluster is then used for matching, and if the number of particles is less than N, zero-padding is applied.

At frame k, determine whether any cluster contains more particles than the specified threshold. Let $\mathrm{C_{m,k}} = \{\mathrm{i}_1^{(1)}, \mathrm{i}_2^{(1)}, \ldots, \mathrm{i}_{\mathrm{J1}}^{(1)}, \mathrm{i}_1^{(2)}, \mathrm{i}_2^{(2)}, \ldots, \mathrm{i}_{\mathrm{J2}}^{(2)}, \ldots, \mathrm{i}_1^{(\mathrm{n})}, \mathrm{i}_2^{(\mathrm{n})}, \ldots, \mathrm{i}_{\mathrm{Jn}}^{(\mathrm{n})}\}$ denotes the set of particles within that cluster at this frame, where $\mathrm{i}_1^{(\mathrm{j})}, \mathrm{i}_2^{(\mathrm{j})}, \ldots, \mathrm{i}_{\mathrm{Jj}}^{(\mathrm{j})}, (0 < \mathrm{j} \le \mathrm{n})$ represent the states of the Jn particles in the n-th cluster, and each particle contains its corresponding information $\{\mathrm{w}_\mathrm{i}^{(\mathrm{n})}, \mathrm{b}_\mathrm{i}^{(\mathrm{n})}, \mu_\mathrm{i}^{(\mathrm{n})}, \mathrm{P}_\mathrm{i}^{(\mathrm{n})}\}$. If $J1 + J2 + \ldots + Jn > N$ and $Ji > MinPts$, each cluster $\mathrm{C_{m,k}}$ identifies the target label corresponding to each particle, and particles with the same label are grouped into a separate cluster. If the number of particles in a new cluster is less than N, zeros are appended after the last particle to reach N. In this way, each resulting cluster contains particles that correspond to the same target.

$$\begin{cases} C'_{m,k} = \{i_1^{(1)}, i_2^{(1)}, \ldots\ldots, i_{J1}^{(1)}\} \\ C'_{m+1,k} = \{i_1^{(2)}, i_2^{(2)}, \ldots\ldots, i_{J2}^{(2)}\} \\ \ldots \\ C'_{m+n-1,k} = \{i_1^{(n)}, i_2^{(n)}, \ldots\ldots, i_{Jn}^{(n)}\} \end{cases} \tag{22}$$

At this point, there is a new particle label matrix L_k' whose rows and columns are updated, so the clusters and particles within each cluster are matched using the updated particle label matrix and trajectory label matrix.

Table 2 – Label Management Based on Target Proximity

step1	When the number of particles in a cluster exceeds the threshold N, it is considered an indication of target proximity or trajectory crossing.
step2	Classify the particles within the cluster based on their target labels, and retain only those groups whose particle count exceeds MinPts.
step3	Each group of particles is treated as an individual cluster. If the number of particles in a cluster is less than N, zeros are appended to reach N particles.
step4	Match the updated particle label matrix with the trajectory matrix.

3. Simulation Experiments

3.1 Parameter Settings

To validate the effectiveness of the proposed target label management and AUKF-RBMCDA algorithm, simulations are conducted based on passive sonar observations in a multi-target tracking scenario. A total of 400 particles are used, with the measurement angle ranging from [0°, 360°] to ensure full-range real-time observation and tracking of all targets within the region. Three trajectories are initialized as shown in figure 1 and the trajectory equations as shown in table 3:

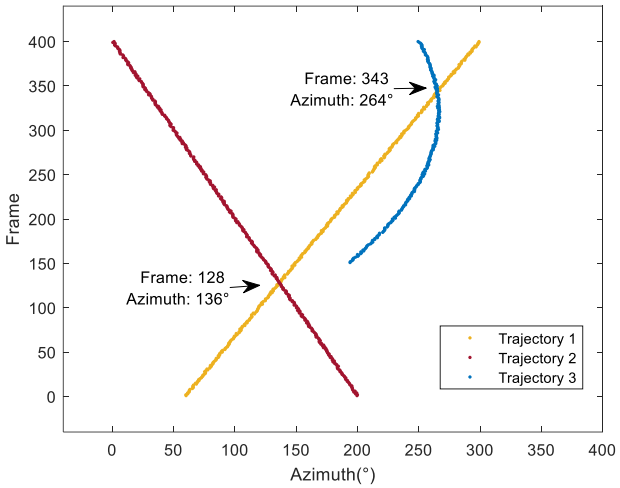


Figure 1 – Target Trajectory with Gassian Noise

Table 3 – Target's Real Trajectory

Target	Trajectory Equation
Target1	$\theta = 1.20t + 60.00$
Target2	$\theta = -0.0101(t-100)^2 + 1.2105(t-100) + 230$
Target3	$\theta = -t + 200$

The tracking duration is set to 200 seconds. Target birth probability $pb = 0.06$ is defined to model the random appearance of new targets, clutter prior probability $cp = 0.10$ is used to simulate false alarms in the measurements. This experiment is based on a linear discrete-time state-space model, with the observation interval defined as $T = 1$ frame, the sampling interval is 0.5 seconds so there are 400 frames in total. The state transition matrix is given by $\mathbf{A} = [1 \quad T; 0 \quad 1]$. The observation matrix $\mathrm{H} = [1 \quad 0]$ is defined to measure the target position and define the target priori mean and priori covariance as $\mu_0 = [90 \quad 0]$, $\mathbf{P}_0 = \mathrm{diag}(90^2, 1)$, respectively. The initial process noise covariance matrix and initial measurement noise covariance matrix are denoted as $Q_0 = diag(1.5, 10^{-3})$ and $R_0 = 3$, representing the disturbances of process noise on the state model and the system measurement errors. The clustering algorithm parameters in the particle filter are set as follows: the neighborhood radius $Eps = 1$, and the particle count threshold $MinPts = 60$ to ensure the accuracy and stability of the clustering process.

3.2 Results Analysis

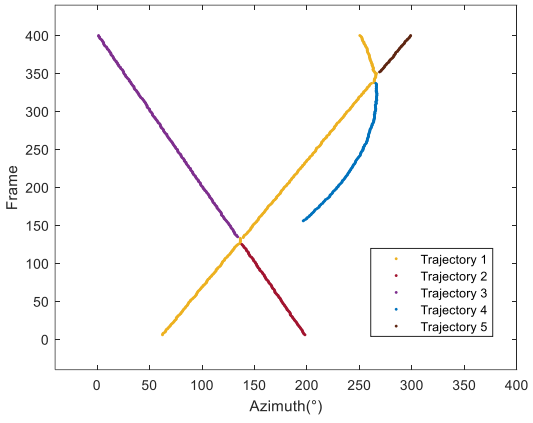


Figure 2 (a) – Track Trajectory with AUKF- RBMCDA

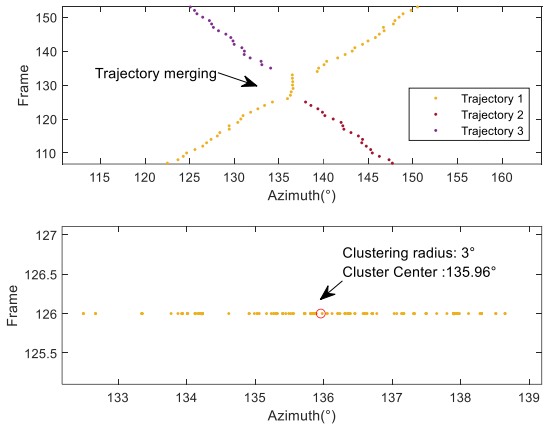


Figure 2 (b) –Particle clusters at the point of association

Figure 2 –Track Result with AUKF- RBMCDA

Figure 2(a) shows the trajectory tracking results obtained by the AUKF-RBMCDA algorithm, while

Figure 2(b) illustrates the particle cluster distribution at frame 126 from Figure 2(a). As shown in Figure 2(a), the AUKF-RBMCDA algorithm produces five tracking results during trajectory crossing, indicating severe trajectory misidentification. From figure 2, a trajectory association error occurs at frame 128 when Trajectory 1 and Trajectory 2 are in close proximity. Trajectory 2 terminates at frame 125, and a new Trajectory 4 appears at frame 135. Trajectories 1 and 3 intersect at frame 343. However, due to clustering errors caused by target proximity, Trajectory 3 terminates prematurely at frame 337. After frame 337, the trajectory is mistakenly associated with Trajectory 1, and a new Trajectory 5 is incorrectly generated from the cluster at frame 352. By analyzing Figures 2(a) and 2(b), it can be observed that at frame 128, during the trajectory crossing, particles with similar estimated values are grouped into the same cluster. As a result, Trajectory 2, which originally belonged to the second cluster, is interrupted.

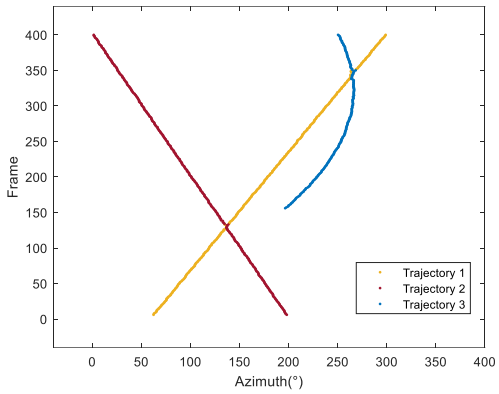


Figure 3 (a) – Track Trajectory with AUKF-ALM-RBMCDA

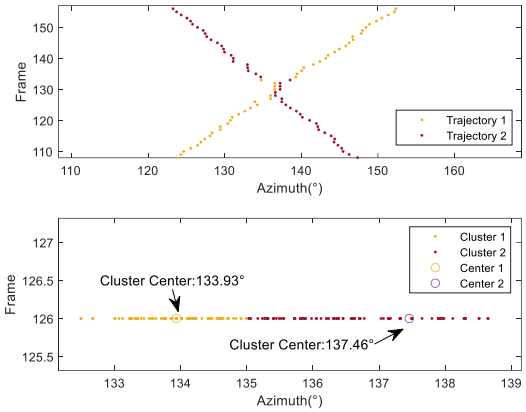


Figure 3 (b) –Particle clusters at the point of association

Figure 3 –Track Result with AUKF-ALM-RBMCDA

Figure 3(a) presents the trajectory tracking results obtained by the AUKF-ALM-RBMCDA algorithm, while Figure 3(b) displays the particle cluster distribution at frame 126 corresponding to Figure 3(a). As shown in Figure 3(a), compared to the unmodified algorithm, the adapt label management algorithm can accurately associate trajectories with their corresponding targets even at the points of trajectory crossing, frame 128 and frame 343. Combining Figures 3(a) and 3(b), it can be observed that the improved algorithm is still able to effectively distinguish between two clusters in the case of trajectory proximity, preserving the complete trajectories of different targets. At frame 133, due to the proximity of the two targets, their probability density distributions no longer satisfy the independence assumption. The overlap of the distributions leads to covariance divergence, resulting in a broader sigma point sampling range and greater deviation in the estimated target states. However, the proposed proximity-based label management method associates measurements based on historical labels, which may result in a “crossing” phenomenon.

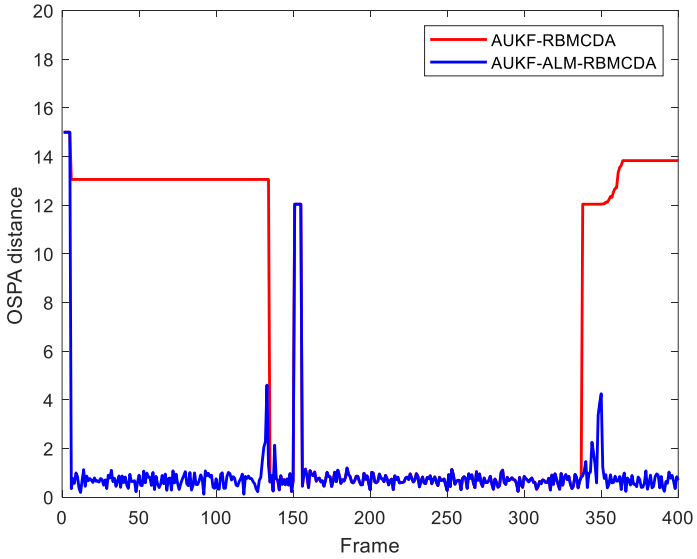


Figure 4 – Trajectory Tracking Error

The Optimal Subpattern Assignment (OSPA) distance is used as a metric to evaluate the association accuracy of the two algorithm here. OSPA is a performance evaluation metric specifically designed for multi-target tracking, which jointly accounts for localization errors between estimated and true targets, as well as penalties arising from mismatches in the number of targets. The OSPA distance is defined as follows:

$$\mathrm{OSPA}_p^c(A,B)=\left(\frac{1}{n}\left(\min_{\pi\in\Pi_n}\sum_{i=1}^{m}d_c(a_i,b_{\pi(i)})^p+c^p(n-m)\right)\right)^{1/p} \tag{23}$$

where $A=\{a_1,...,a_m\}$ denotes the set of estimated targets, $B=\{b_1,...,b_n\}$ denotes the set of true targets, $d(\cdot,\cdot)$ represents the Euclidean distance, p is typically chosen as 1 or 2, c denotes the cutoff distance, $\prod_n$ denotes all possible target assignments, $m=\min(|A|,|B|)$, $n=\max(|A|,|B|)$. When calculating trajectory errors, frame 134 to 338 are considered successfully associated. The resulting error is shown in Figure 4. Figure 4 shows that the AUKF-RBMCDA algorithm exhibits significant errors during frame 1 to 134 and frame 338 to 400. In contrast, the AUKF-ALM-RBMCDA algorithm shows correct associations except for normal errors caused by track initiation delays during frame 1 to 5 and 151 to 155. The trajectory crossings at frame 133 and 350 lead to expected errors.

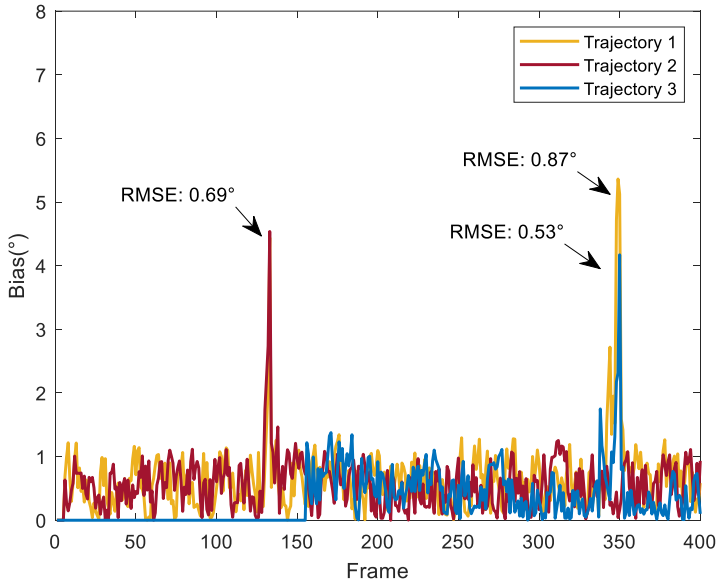


Figure 5 – Estiamte Error of Three Trajectory

Figure 5 presents the tracking errors of the three trajectories, with the root mean square errors (RMSE) calculated as 0.87°, 0.69°, and 0.53°, respectively. In areas without trajectory crossings, the errors of the three trajectories generally range between 0° and 1°, At trajectory crossing points, the errors for Trajectory 1 at frame 133 and 349 are 4.44° and 5.36°, the error for Trajectory 2 at the crossing is 4.46° and the error for Trajectory 3 is 4.17°.

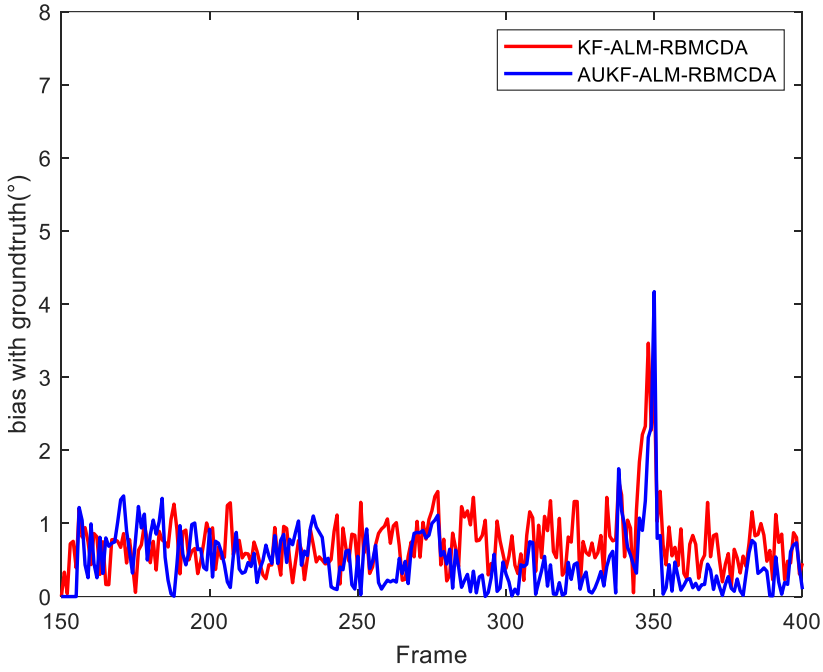


Figure 6 – Tracking Error of Trajectory 3

To verify the accuracy of the AUKF-ALM-RBMCDA algorithm during target maneuvers, a comparative analysis between the KF-ALM-RBMCDA algorithm and AUKF-ALM-RBMCDA algorithm for tracking Trajectory 3 is presented in Figure 6. As shown in Figure 1, Trajectory 3 appears only after frame 150 and is absent beforehand; thus, the tracking error of the algorithm is zero prior to frame 150. Therefore, only the tracking error after frame 150 is analyzed here. Between frames 150 and 184, the errors of both algorithms are nearly identical. After frame 184, except at trajectory crossing points, the AUKF-ALM-RBMCDA algorithm maintains an error angle below 1°, outperforming the KF-ALM-RBMCDA algorithm, At trajectory crossing points, when two targets are in close proximity, a similar “crossing” phenomenon as previously described occurs, resulting in a slight increase in error.

4. Conclusion

This paper focuses on pure bearing multi-target tracking using RBMCDA algorithm, incorporating an adaptive unscented Kalman filter method and a target proximity-based label management strategy. By employing the label management strategy, particles corresponding to different targets in close proximity are separately clustered and matched to trajectories. The adaptive unscented Kalman filter is employed to estimate tracking performance under nonlinear conditions. Simulation results demonstrate that the improved target label management strategy achieves higher association accuracy during target crossings, while the AUKF-based enhancement yields lower tracking errors for maneuvering targets.

ACKNOWLEDGEMENT

This work is supported by the National Natural Science Foundation of China under Grant Nos. 62271162.

REFERENCES

1. Howe BM, Miksis-Olds J, Rehm E, Sagen H, Worcester PF, Haralabus G. Observing the oceans acoustically. Front Mar Sci.2019;6:426.
2. Zhao W, Li X, Pang Z, Hao C. A novel distributed bearing-only target tracking algorithm for underwater sensor networks with resource constraints. IET Radar Sonar Navig. 2024;18(7):1161-1177.
3. Kumar M, Mondal S. Recent developments on target tracking problems: A review. Ocean Eng. 2021;236:109558.
4. Chitre M, Shahabudeen S, Stojanovic M. Underwater acoustic communications and networking: recent advances and future challenges. Mar Technol Soc J. 2008;42(1):103-116.
5. Wang Y, Wang H, Li Q, Xiao Y, Ban X. Passive sonar target tracking based on deep learning. J Mar Sci Eng. 2022;10:181.

6. Li X, Lu B, Ali W, Jin H. Passive tracking of multiple underwater target in incomplete detection and clutter environment. Entropy. 2021;23:1082.
7. Zhang P, Mu Z. Real-time multi-target tracking method based on GMM-RBMCDA. Computer Appl Res. 2012;29(11):4345-4349.
8. Luo J, Wang Z, Chen Y, Wu M, Yang Y. An improved unscented particle filter approach for multi-sensor fusion target tracking. Sensors. 2020;20(23):6842.
9. Qi B, Hui C, Hao Y, Li X, Mathiopoulos PT. Improved density-based data association method under azimuth rate constraint. In: Proceedings of 2024 OES China Ocean Acoustics (COA); 2024; Harbin, CHINA. p. 1-7.
10. Madhukar PS, Prasad LB. State estimation using extended Kalman filter and unscented Kalman filter. In: Proceedings of the 2020 International Conference on Emerging Trends in Communication, Control and Computing (ICONC3); 2020; Lakshmangarh, India. p. 1-4.

DAILY PROGRAM

AUGUST 21 (THU)

09:00 – 10:20	Receiver Design and Robust Signal Processing for Underwater Acoustic Communications 324 (3F)
Chair	Jungpyo Hong (Changwon National University)
09:00	A Message Passing Decision Feedback Receiver with IDI-Progressive Framework for Underwater acoustic OTFS Communications
	Yang Yang, Lu Ma, Boon-Chong Seet
09:20	An Adaptive RAKE Receiver Based on Bionic Underwater Acoustic Communication by Mimicking Dolphin
	Tiany Liu, Songzuo Liu, Yipeng Xing, Jiaxuan Li, Gang Qiao
09:40	Iterative Underwater Acoustic DSSS Processing Algorithm for Integrated Communication and Localization System in Non-Uniform Doppler and Strong Noise Interference Channels
	Jinhao Deng, Jie Wu, Hongyu Cui
10:00	Nonlinear Composite Chirp Preamble for Enhanced Synchronization in Narrowband Underwater Acoustic Communications
	Bogeun Seo, Sangman Han, Haklim Ko, Hojun Lee

Factor graph-based IDI progressive receiver architecture for UWA-OTFS

Yang YANG[123]; Lu MA[1234]; Boon-chong SEET[5]
[1] National Key Laboratory of Underwater Acoustic Technology, Harbin Engineering University, Harbin 15001，China
[2] Key Laboratory of Marine Information Acquisition and Security (Harbin Engineering University), Ministry of Industry and Information Technology, Harbin 150001，China
[3] College of Underwater Acoustic Engineering, Harbin Engineering University, Harbin 150001，China
[4] Sanya Nanhai Innovation and Development Base of Harbin Engineering University，Sanya 572024，China
[5] The Department of Electrical and Electronic Engineering, Auckland University of Technology, Auckland 1010, New Zealand

ABSTRACT
Orthogonal time frequency space (OTFS) modulation presents a promising solution for underwater acoustic (UWA) mobile communication systems. However, the wideband Doppler effect in UWA channels induces non-uniform Doppler shifts, significantly complicating signal recovery processes. This study proposes a factor graph-based inter-Doppler interrupt (IDI) progressive receiver architecture for UWA-OTFS systems. The message-passing equalizer and channel decoder are integrated into a progressive receiver framework, where the decoder's extrinsic information is utilized for subsequent equalizer iterations. Following multiple decoder iterations, the channel IDI depth D is updated to incorporate additional inter-carrier interference (ICI) into the system model, thereby enhancing overall performance. Simulations validate the robust performance of the proposed architecture.

Keywords: Underwater acoustic communication, OTFS, Doppler effect, mobile communication

1. INTRODUCTION

As an emerging modulation technique, Orthogonal Time Frequency Space (OTFS) modulation demonstrates superior performance in time-varying channels through delay-Doppler (DD) domain signal processing, offering a novel solution for underwater acoustic (UWA) mobile communications[1]. Factor graph (FG)-based message passing (MP) interference cancellation techniques have been widely adopted in UWA-OTFS communications [2], [3]. Yan Li et al. [4] proposed specific improvements to the conventional MP algorithm framework and evaluated system performance under various modulation schemes, path indices, and Doppler factors through comprehensive simulations. Zhekai Xue et al. [5] introduced a unitary approximate MP detector for OTFS systems, demonstrating superior bit error rate (BER) performance compared to conventional AMP and MP approaches. To address the high computational complexity of matrix inversion in OTFS detection, Yang Zhang et al. [6] developed a belief propagation detector. The detector leverages equivalent channel sparsity and approximate probability updates to reduce computational complexity, while incorporating damping factors to enhance algorithmic convergence.

However, the wideband Doppler effect in UWA channels results in non-uniform Doppler frequency shifts in received signals [7], which not only degrades system performance but also significantly increases the complexity of signal recovery . To address this challenge, inspired by the progressive framework considering ICI in OFDM demodulation [3], this paper proposes a factor graph-based Inter-Doppler Interrupt (IDI) progressive receiver architecture. This architecture effectively resolves the signal detection challenges in UWA-OTFS systems by outputting soft information through message-passing equalization algorithms in conjunction with Low-Density Parity-Check (LDPC) decoders.

The rest of the paper is organized as follows. In Section II, we briefly describe the UWA-OTFS system model and the equivalent channel matrix is analyzed in the delay-Doppler domain. Section III proposed an IDI progressive receiver architecture for UWA-OTFS system, the proposed scheme is then evaluated by numerical simulations and test in Section IV and Section V. Finally, conclusions are drawn in Section VI.

[1] heu_yangyang@hotmail.com
[2] malu@hrbeu.edu.cn

2. UWA-OTFS COMMUNICATION SYSTEM

2.1 Transmitted signal

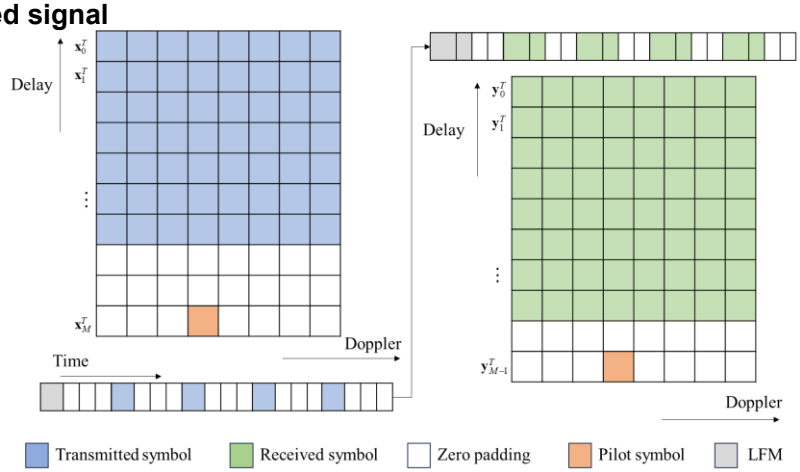


Fig. 1 OTFS modulation scheme

Within the DD domain framework, we define $\mathbf{X} \in \mathbb{C}^{M \times N}$ as the modulated signal sequence. By applying the Zak transform with a rectangular pulse shaping function $\mathbf{G}_{tx} = \mathbf{I}_M$, we can establish the mathematical correlation between the time-domain baseband signal $\mathbf{s} \in \mathbb{C}^{MN \times 1}$ and $\mathbf{X}$ in DD domain. This relationship is mathematically formulated as

$$\mathbf{s} = vec\left(\mathbf{X} \cdot \mathbf{F}_N^{\dagger}\right). \tag{1}$$

The passband signal $\tilde{s}(t)$ can be obtained by up-conversion. The UWA channel model, characterized by multiple propagation paths with distinct Doppler scaling factors, can be mathematically represented as [10]

$$h(t) = \sum_{p=1}^{P} h_p \delta((1+\alpha_p)t - \tau_p) \tag{2}$$

where, h_p, α_p and τ_p represent the amplitude gain, Doppler factor and delay of the p^{th} path. Then the passband received signal can be expressed as

$$y(t) = \sum_{p=1}^{P} h_p \tilde{s}((1+\alpha_p)t - \tau_p) + \tilde{n}(t) \tag{3}$$

where $\tilde{n}(t)$ represents the Gaussian noise component within the passband. Subsequently, by applying resampling with the estimated parameter $\hat{\alpha}$, we obtain

$$\tilde{y}_r(t) = \sum_{p=1}^{P} h_p \tilde{s}[(1+b_p)(t-\tau_p')] + \tilde{n}(t/(1+\hat{\alpha})) \tag{4}$$

where b_p is the residual Doppler factor after resampling, and equivalent delay $\tau_p' = \tau_p / (1+b_p)$. After down-conversion operation, the received baseband signal $y_r[i'] = y_r(t)\big|_{t=i'/B}$. The discrete-time sequence of the received signal is $i' = 0, 1, ..., L-1$, where $L = [MN + B\tau_{\max}]$ is defined the maximum delay scale. Subsequently, the received baseband vector signal can be represented as $\mathbf{r} \in \mathbb{C}^{MN \times 1}$. Following synchronization, MN samples are extracted from signal $y_r[i']$ to derive the baseband received vector signal $\mathbf{r} \in \mathbb{C}^{MN \times 1}$. The time domain relationship is characterized by

$$\mathbf{r} = \mathbf{G} \cdot \mathbf{s} + \mathbf{n} \tag{5}$$

where $\mathbf{G} \in \mathbb{C}^{MN \times MN}$ is the equivalent baseband channel matrix in time domain, the element of the matrix is that

$$G(i', i) = \sum_{p=1}^{P} h_p e^{-j2\pi f_c (1+b_p)\tau_p'} e^{j2\pi b_p f_c i'/B} \times \mathrm{sinc}\left((1+b_p)i' - i - (1+b_p)\tau_p' B\right) \tag{6}$$

where f_c is the center frequency. Due to the effect of b_p, the band-sparsity character of channel matrix $\mathbf{G}$

is presented.

2.2 The relationship of input and output in DD domain

Defined $\mathbf{x} \in \mathbb{C}^{NM\times 1}$, $\mathbf{y} \in \mathbb{C}^{MN\times 1}$ is the transmitted and received vector signal in DD domain, respectively,

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_0 \\ \vdots \\ \mathbf{x}_{M-1} \end{bmatrix} = vec(\mathbf{X}^T), \mathbf{y} = \begin{bmatrix} \mathbf{y}_0 \\ \vdots \\ \mathbf{y}_{M-1} \end{bmatrix} = vec(\mathbf{Y}^T) \tag{7}$$

where $\mathbf{x}_m, \mathbf{y}_m \in \mathbb{C}^{N\times 1}, m = 0,\ldots,M-1$ are the subcarrier-level vectors along the Doppler axis, as shown as in Fig. 1. The input-output relationship in DD domain can be represent

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w} \tag{8}$$

$\mathbf{H} \in \mathbb{C}^{MN\times MN}$ is the equivalent channel matrix in DD domain, and Fig 2 shows the magnitude heatmap of $\mathbf{H}$ under different Doppler scale b_p . The element of H is

$$H[kN+l, k'N+l'] = \sum_{n_1=0}^{N-1}\sum_{n_2=0}^{N-1} e^{-j2\pi(n_1 l - n_2 l')/N} G\left(n_1 M + k, n_2 M + k'\right). \tag{9}$$

The relationship between matrices $\mathbf{H}$ and $\mathbf{G}$ can be found in ref. [11]. The prominent main diagonal line exhibits the subcarrier gains, corresponding to the desired signal component. The prominent main diagonal line exhibits the subcarrier gains, corresponding to the desired signal component. Two distinct bright lines parallel to the main diagonal indicate the presence of multipath components. The channel exhibits three propagation paths, characterized by three bright lines centered around distinct band structures. The residual Doppler effect predominantly influences adjacent subcarriers, with the interference magnitude diminishing proportionally to the inter-subcarrier distance. Additionally, increased Doppler spread manifests as enhanced energy dispersion in non-diagonal elements, consequently intensifying the Inter-Carrier Interference (ICI). The channel matrix $\mathbf{H}$ typically demonstrates energy concentration along its main diagonal and adjacent off-diagonal elements. We adopt an assumption that $H[kN+l, k'N+l'] \approx 0,$ if $|l-l'| > D$, by modeling the IDI exclusively for a single subcarrier with respect to its D adjacent subcarriers, where D denotes the IDI-depth parameter.

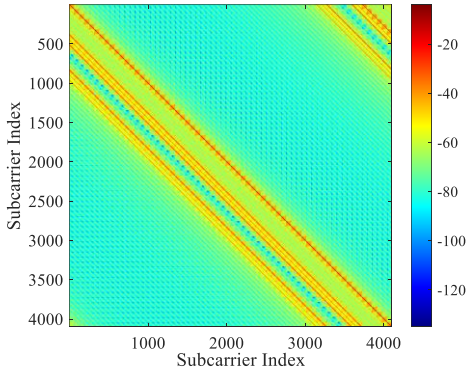


Fig 2. Equivalent channel matrix $\mathbf{H}$ in the DD domain

3. IDI PROGRESSIVE RECEIVER ARCHITECTURE

Unlike narrowband communication OTFS, the Doppler effect in UWA communication leads to Doppler index shifts and phase rotations of DD domain symbols, rather than simple frequency shifts [2]. Additionally, time scaling also affects interference in the delay domain, with the interference range determined by the delay index l_p and Doppler index k_p of the p^{th} path. In this section, considering the impact of residual Doppler factors b_p , we will explore the equalization algorithms for UWA-OTFS.

3.1 Based on FG equalization algorithm

In this section, we introduce the IDI-depth D to quantify the degree of IDI. Assuming each symbol is affected by interference from only S_D adjacent symbols, where D represents the IDI-depth parameter. After

the resampling preprocessing operation, S_{IDI} becomes significantly smaller than NM , consequently rendering matrix $\mathbf{H}$ sparse. Let $\mathcal{U}(b)$ and $\mathcal{V}(a)$ denote the sets of indexes with non-zero elements of in the b-*th* row and a-*th* column, respectively, Under the constraint of D , then $|\mathcal{V}(c)| = |\mathcal{U}(b)| = S_D$ for all rows and columns. Following Eq.(8) ,we model the system as a sparsely-connected FG with NM variable nodes corresponding to $\mathbf{x}$ and NM observation nodes corresponding to $\mathbf{y}$. In this FG, each observation node $y[b]$ is connected to the set of S_D variable nodes $\{x[c], c \in \mathcal{U}(b)\}$. Similarly, each variable node $x[c]$ is connected to the set of S_D observation nodes $\{y[b], b \in \mathcal{V}(c)\}$.

Eq. (8) can be modeled as

$$y[b] = x[c]H[b,c] + \underbrace{\sum_{d \in \mathcal{U}(b), d \neq b} x[d]H[b,d] + n[b]}_{\xi_{b,c}}. \tag{10}$$

We assumed that 1) all the transmitted symbols $a_j \in \mathbb{A}$ are equally likely; 2) the components of $\mathbf{y}$ are approximately independent for a given $x[c]$, due to the sparsity of $\mathbf{H}$. 3) the interference terms $\xi_{b,c}$ defined in (10) are independent for a given c . 4) for each $y[b]$, a variable $x[c]$ is isolated from the other interference terms, which are then approximated as Gaussian noise with an easily computable mean and variance.

The joint maximum a posterior probability detection rule for estimating the transmitted signals is given by

$$\mathbf{x} = \arg\max_{\mathbf{x} \in \mathbb{A}^{NM \times 1}} \Pr\left(\mathbf{x} | \mathbf{y}, \mathbf{H}\right) \tag{11}$$

which has a complexity exponential in NM . Since the joint maximum a posteriori (MAP) detection can be intractable for practical values of N and M , based on above assumption, we consider the symbol-by-symbol MAP detection rule for $c = 1, \ldots, MN$

$$\hat{x}[c] \approx \arg\max_{a_j \in \mathbb{A}} \prod_{b \in \mathcal{V}(a)} \Pr\left(y[b] | x[c] = a_j, \mathbf{H}\right). \tag{12}$$

In the MP algorithm, the connections and the messages passed between the observation and variable nodes. The mean and variance of the interference terms are used as messages from observation nodes to variable nodes. On the other hand, the message passed from a variable node $x[c]$ to the observation nodes $y[b]$, $b \in \mathcal{V}(c)$, is the probability mass function (pmf) of the alphabet $\mathbf{p}_{c,b} = \{p_{c,b}(a_j) | a_j \in \mathbb{A}\}$.

The details of the steps in iteration i in the MP algorithm are detailed below. The mean and variance of interference in (10) are computed based on the extrinsic information from a channel decoder. Define $p_{c,b}$ as the extrinsic probability fed back from the channel decoder. We have

$$\mu_{\xi}^{(i)} = \sum_{d \in \mathcal{U}(b), d \neq b} \sum_{j=1}^{Q} p_{b,d}^{(i-1)}(a_j) a_j H[b,d] \tag{13}$$

and

$$\left(\sigma_{\xi}^{(i)}\right)^2 = \sum_{d \in \mathcal{U}(b), d \neq b} \left(\sum_{j=1}^{Q} p_{b,d}^{(i-1)}(a_j) |a_j|^2 |H[b,d]|^2 - \left| \sum_{j=1}^{Q} p_{b,d}^{(i-1)}(a_j) a_j H[b,d] \right|^2 \right) + \sigma^2 \tag{14}$$

$$\bar{x}[d] = \sum_{j=1}^{Q} p_{c,b}^{(i-1)}(a_j) a_j .$$

Message passings from variable nodes $x[c]$ to observation nodes $y[b]$: The pmf vector $\mathbf{p}_{c,b}$ (variable $x[c]$ is equal to a_j) can be updated as

$$p_{c,b}^{(i)}(a_j) = \rho \cdot \tilde{p}_{c,b}^{(i)}(a_j) + (1-\rho) \cdot p_{c,b}^{(i-1)}(a_j) \tag{15}$$

where the damping factor ρ performs weighted averaging between the currently computed message and the message from the previous iteration, yielding the final updated message $p_{c,b}^{(i)}(a_j)$. When ρ approaches 1,

the algorithm exhibits aggressive updates, potentially achieving faster convergence but risking instability or oscillatory behavior. When ρ approaches 0, the algorithm demonstrates conservative updates, resulting in slower but more stable convergence. Furthermore,

$$\begin{aligned}\tilde{p}_{c,b}^{(i)}(a_j) &\propto \prod_{d\in\mathcal{V}(c),d\neq b} \Pr\left(y[d]\middle|x[c]=a_j,\mathbf{H}\right) \\ &= \prod_{d\in\mathcal{V}(c),d\neq b} \frac{f^{(i)}(d,c,j)}{\sum_{k=1}^{Q} f^{(i)}(d,c,k)}\end{aligned} . \tag{16}$$

The likelihood function $f^{(i)}(d,c,k)$ quantifies the degree of support that the observation $y[d]$ provides for different possible values of variable $x[c]$, expressed as

$$f^{(i)}(c,a,k)=\exp\left(\frac{-\left|y[c]-\mu_{c,a}^{(i)}-H_{c,a}a_k\right|^2}{\left(\sigma_{a,c}^{(i)}\right)^2}\right). \tag{17}$$

Compute the convergence indicator η as

$$\eta^{(i)}=\frac{1}{NM}\sum_{a_j=\mathbb{A}}^{NM}\Diamond\left(\max_{a_j=\mathbb{A}} p_c^{(i)}(a_j)\geq 1-\gamma\right) \tag{18}$$

for some small $\gamma>0$ and where

$$p_c^{(i)}(a_j)=\prod_{d\in\mathcal{V}(c)}\frac{f^{(i)}(d,c,j)}{\sum_{k=1}^{Q} f^{(i)}(d,c,k)} \tag{19}$$

where $\Diamond(\cdot)$ is an indicator function which gives a value of 1, if the expression in the argument is true, and 0 otherwise. If $\eta^{(i)}>\eta^{(i-1)}$, then we update

$$\hat{x}[c]=\arg\max_{a_j=\mathbb{A}} p_c^{(i)}(a_j), c=1,\ldots,NM. \tag{20}$$

Symbol decision updates are executed only when the current iteration yields superior estimates compared to the previous iteration. The MP algorithm terminates upon satisfying at least one of the following conditions: 1) $\eta^{(i)}=1$. 2) Maximum number niter of iterations is reached.

3.2 Framework

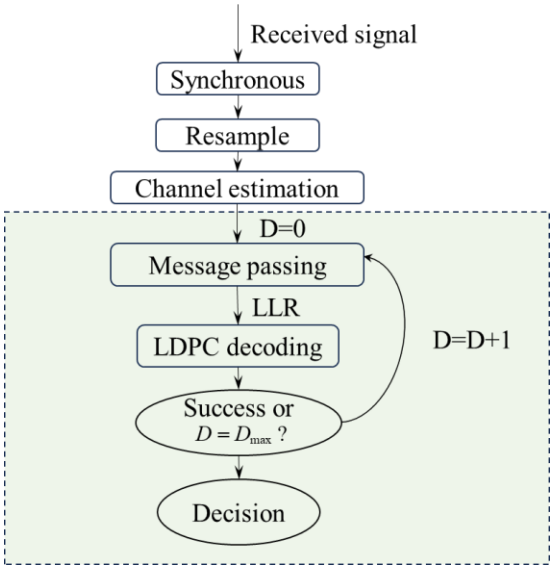


Fig. 3 OTFS progressive receiver structure

Preprocessing resampling enhances the sparsity of channel matrix $\mathbf{H}$. Nevertheless, due to residual Doppler factors, the energy of multipath components at the receiver disperses across multiple DD units, resulting in significant IDI. This phenomenon complicates the detection and equalization processes. The progressive IDI receiver aims to mitigate IDI while maintaining a balance between system complexity and

BER performance.

The DD channel response is estimated using pilot symbols. Based on the estimated channel parameters, a sparse matrix $\mathbf{H}$ is constructed, determining the interference sets $\mathcal{U}(b)$ and $\mathcal{V}(c)$ for each symbol. A detailed description of the OMP estimator can be found in [12]. The MP equalizer and channel decoder are integrated into a progressive receiver framework, where the extrinsic information from the decoder is utilized in subsequent equalizer iterations. After multiple decoder iterations, the IDI depth D is updated to incorporate additional interference components into the system model, enhancing performance.

4. SIMULATION

In this section, a UWA-OTFS system with a center frequency of 7 kHz and a bandwidth of 4 kHz is considered. The system parameters are summarized in Table 1. Employing 4-QAM modulation, the system incorporates a zero-padded guard interval comprising half of the total delay symbols $M_{zp} = M/2$, yielding an effective data rate of 1.5 kbps. The cyclic property of the channel matrix block is maintained for channel delay spreads up to 8 ms.

Table 1 – Simulation parameters

Parameter	Value	Parameter	Value
Bandwidth	4 kHz	Number of delay bins	16
Center frequency	7 kHz	Modulation scheme	4-QAM
Number of Doppler bins	64	Communication rate	1500bps

Performance evaluation utilizes a stochastic UWA multipath channel model as described in [10]. The channel parameters are characterized by: The time intervals between two consecutive paths are exponentially distributed with an average value of 2 ms. Rayleigh-distributed path amplitudes with exponentially decaying power profiles along the delay axis, Doppler scaling factors uniformly distributed in the range $\left[-1\times10^{-3}, 1\times10^{-3}\right]$ with standard deviation σ_v m/s.

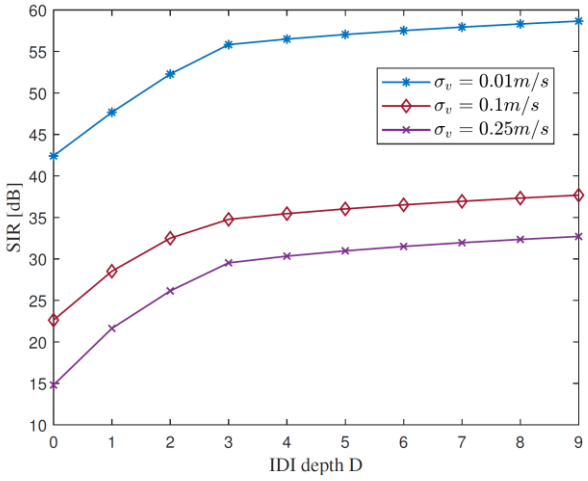


Fig. 4 SIR *vs* IDI depth D

By evaluate signal to IDI ratio (SIR), matrix $\mathbf{H}$ under IDI is quantified

$$\chi_{SIR} = \mathop{\mathbb{E}}_{l}\left[\frac{\sum_{l':|l-l'|\le D}\left|H[kN+l,k'N+l']\right|^2}{\sum_{l':|l-l'|>D}\left|H[kN+l,k'N+l']\right|^2}\right]. \tag{21}$$

Fig. 5 illustrates that the Signal-to-Interference Ratio (SIR) stabilizes and plateaus with increasing IDI depth *D*, due to IDI primarily originates from interference between adjacent Doppler pulses. The system's outage performance demonstrates progressive enhancement with increasing IDI depth D. In channels exhibiting minimal path variations, outage performance convergence is achieved rapidly at low D values, whereas channels with significant path variations require larger D values to attain optimal performance.

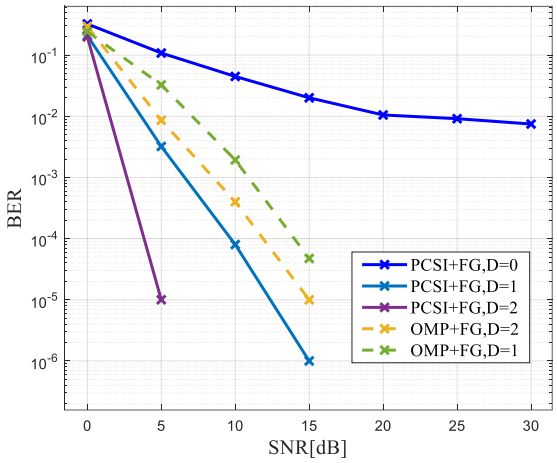


Fig. 5 The BER performance comparison of different receivers.

Fig. 5 presents a comparative analysis of the FG-based receiver performance with varying IDI depth D under both perfect CSI (PCSI) and imperfect CSI obtained through OMP estimation. When $D = 0$ (i.e., without considering IDI), the FG algorithm exhibits limited symbol detection capability. As the IDI depth D increases, a progressive enhancement in BER performance is observed. These results further substantiate the significance of incorporating IDI considerations in UWA-OTFS communications.

5. SEA FIELD

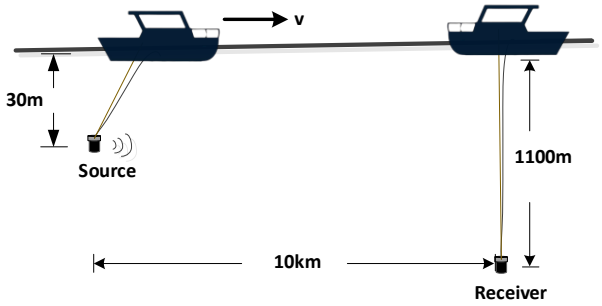


Fig. 6 Scenario diagram

Experimental sea trials were implemented in the South China Sea in May 2024, where the water column depth measured approximately 5.5 km. The receiving array was deployed at a depth of 1100 m, while the transmitting source was positioned at 30 m below the surface, maintaining a consistent outbound velocity of 2 m/s, as shown in Fig. 6. The experimental protocol involved the transmission of OTFS signal packets, each with a duration of 6 seconds and a transmission interval of 6 minutes, culminating in a total of 31 packet transmissions. The initial separation between the transmitting and receiving vessels was 10 km, followed by a convergence-divergence trajectory. The OTFS signals were transmitted at a data rate of 2000 bps, employing a carrier frequency of 3 kHz.

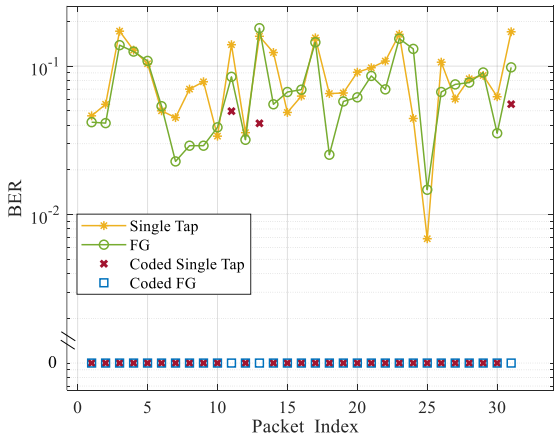


Fig. 7 BER performance comparison

Fig. 7 shows the BER results for 31 transmitted data packets during the communication experiment. It compares the BER performance of four receiver algorithms: Single tap equalizer, the proposed FG equalizer, Single tap equalizer after LDPC and FG equalizer after LDPC. Performance evaluation of hard decision outcomes reveals that MP equalization yields lower uncoded BER relative to single tap equalization. The proposed IDI progressive receiver achieves successful demodulation for all transmitted packets.

6. CONCLUSIONS

This paper proposes a novel IDI-aware progressive receiver framework. Taking into account residual Doppler effects in resampled UWA channels, we first establish the input-output relationship in the DD domain, characterizing the sparse representation of UWA channels. Subsequently, considering the trade-off between receiver complexity and BER performance, we develop a factor graph-based progressive receiver framework incorporating IDI effects. Simulation results validate both the significance of IDI depth consideration and the achievable BER performance of the proposed receiver.

ACKNOWLEDGEMENTS

This research was supported in part by the National Key Research and Development Program of China under Grant 492 2023YFC3010800, in part by the National Natural Science Foundation of China under Grant 62271161, in part by the Key 493 Research and Development Program of Shandong Province under Grant 2022CXGC020409, and in part by the National Key 494 Laboratory of Underwater Acoustic Technology under Grant 2023-JCJQ-LB-072-08.

REFERENCES

1. Xue Y, Zhu X, Zheng Y R. Orthogonal time-frequency space modulation for underwater mobile acoustic communications[J]. The Journal of the Acoustical Society of America, 2025, 157(2): 1378-1390.
2. Raviteja P, Phan K T, Hong Y, et al. Interference cancellation and iterative detection for orthogonal time frequency space modulation[J]. IEEE transactions on wireless communications, 2018, 17(10): 6501-6515.
3. Zhang H, Zhang T. A low-complexity message passing detector for OTFS modulation with probability clipping[J]. IEEE wireless communications letters, 2021, 10(6): 1271-1275.
4. Li Y, Chen S, Yang W. Simulation study of OTFS underwater acoustic communication technology[C]//Third International Symposium on Computer Applications and Information Systems (ISCAIS 2024). SPIE, 2024, 13210: 99-105.
5. Xue Z, Wang Q, Jing L, et al. Unitary approximate message passing detection method for OTFS underwater acoustic communication system[C]//2023 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC). IEEE, 2023: 1-5.
6. Zhang Y, Zhang Q, He C, et al. Low-Complexity Belief Propagation detector for OTFS System[C]//2021 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC). IEEE, 2021: 1-5.
7. Wang X, Shi X, Wang J, et al. Iterative lmmse-sic detector for dse-aware underwater acoustic otfs systems[J]. IEEE Transactions on Vehicular Technology, 2024, 73(7): 9895-9910.
8. Wang Z, Zhou S, Catipovic J, et al. Factor-graph-based joint IBI/ICI mitigation for OFDM in underwater acoustic multipath channels with long-separated clusters[J]. IEEE Journal of Oceanic Engineering, 2012, 37(4): 680-694.
9. Zhang Y, Wang Y, Liu Y, et al. A Deep Learning Receiver for Underwater Acoustic OTFS Communications With Doppler Squint Effect[J]. IEEE Wireless Communications Letters, 2025.
10. Berger C R, Zhou S, Preisig J C, et al. Sparse channel estimation for multicarrier underwater acoustic communication: From subspace methods to compressed sensing[J]. IEEE transactions on signal processing, 2009, 58(3): 1708-1721.
11. Thaj T, Viterbo E. Low complexity iterative rake decision feedback equalizer for zero-padded OTFS systems[J]. IEEE transactions on vehicular technology, 2020, 69(12): 15606-15622.
12. Shi C, Zhao L, Cui Y, et al. Joint detection and channel estimation for MIMO-OTFS systems[J]. IEEE Transactions on Vehicular Technology, 2024.

An adaptive RAKE receiver based on bionic underwater acoustic communication by mimicking dolphin

Tianyi LIU[1]; Songzuo LIU[2]; Yipeng XING[3]; Jiaxuan LI[4]; Gang QIAO[5]
[1] National Key Laboratory of Underwater Acoustic Technology, Harbin Engineering University, Harbin 15001, China
[2] Key Laboratory of Marine Information Acquisition and Security (Harbin Engineering University), Ministry of Industry and Information Technology, Harbin 150001, China
[3] College of Underwater Acoustic Engineering, Harbin Engineering University, Harbin 150001, China
[4] Sanya Nanhai Innovation and Development Base of Harbin Engineering University, Sanya 572024, China

ABSTRACT
To mitigate multipath effects in bionic underwater acoustic communication using the spectral profile of dolphin whistle for frequency modulation, aiming at the problem of high computation required for accurate channel estimation methods under long multipath delay spread in the deep sea, this paper proposes a low-complexity adaptive linear RAKE receiver based on the orthogonality of the modulation frequency. First, an adaptive threshold setting method based on the noise segment after cross-correlation is utilized for effective multipath selection to reduce the interference of different noise levels. Subsequently, the index of "average distance of energy distribution" is introduced to measure the demodulation of each path to further reduce the impact of the noise-paths on RAKE reception. Finally, the demodulation results of each path are combined to obtain the multipath diversity gain and improve the decoding performance. Simulation and sea trials verified its feasibility. Using this method for sea trial data, the average BER was reduced by an order of magnitude and the data packet transmission success rate has increased by 13.7%. The low complexity makes the proposed method is easier to be realized in engineering, and provides a new idea for the practical application of deep-sea robust bionic underwater acoustic communication.

Keywords: Bionic underwater acoustic, Dolphin whistle, adaptive RAKE Receiver

1. INTRODUCTION
In recent years, underwater acoustic communication (UWAC) has become increasingly crucial for marine information acquisition and underwater wearable technologies. (1). However, the widespread deployment of traditional UWAC devices, particularly in high-density applications such as offshore oil field monitoring systems is causing significant impact on marine ecosystems. Studies have shown that long-term or acute exposure to artificial acoustic signals can severely affect marine wildlife, potentially damaging their auditory systems, navigation abilities (2, 3, 4). In response to these challenges, bionic UWAC has emerged as an innovative solution, offering an environmentally friendly approach to UWAC, by precisely mimicking marine biological acoustic characteristics. This creates a new paradigm for achieving harmonious coexistence with marine ecosystems (5).

Since dolphin whistles are frequency-modulated signals, their frequency continuously changes over time, making them ideal for transmitting information (6, 7). The duration of the whistle typically ranges from several hundred milliseconds to several seconds, with frequency changes spanning from several hundred hertz to several tens of kilohertz (8), which aligns with the operating frequency band of underwater acoustic transducers. Therefore, embedding information into the time-frequency spectrum contour of dolphin whistles via frequency modulation is the primary method for bionic UWAC (9, 10). Reference (11) proposes a bionic UWAC method based on frequency-shift keying (FSK). This method generates a baseband signal by

[1] liutianyi@hrbeu.edu.cn
[2] liusongzuo@hotmail.com
[3] xingyipeng@hrbeu.edu.cn
[4] leejx_janette@163.com
[5] qiaogang@hrbeu.edu.cn

modulating information, which is then added to the time-frequency spectrum contour curve of the dolphin whistle signal at a certain ratio to obtain the synthesized whistle time-frequency spectrum. The synthesized whistle is then regenerated as the biomimetic communication signal. At the receiver, the received synthetic whistle is coherent-multiplied with locally generated dolphin whistle signals with a fixed frequency difference, followed by low-pass filtering to obtain the FSK signal for information demodulation, thereby achieving bionic communication. This method has a simple modulation and demodulation principle, consumes fewer system resources, and is easier to implement in engineering, providing technical support for the practical application of bionic UWAC algorithms.

Multipath propagation is a significant issue in UWAC applications. When multiple propagation paths exist between the sound source and receiver, it leads to signal amplitude attenuation and inter-symbol interference (ISI), thereby affecting the performance of UWAC systems. Therefore, appropriate channel estimation and equalization techniques are often required to counteract the effects of multipath propagation. Currently, common methods used in bionic UWAC include the RAKE receiver (12, 13) and the virtual time reversal mirror (14, 15, 16). The time reversal (TR) method has a significant reducing effect on multipath expansion in array receiver, but for single-element receiver, the reduction in multipath expansion is not prominent. Reference (17) proposes a channel shortening technique based on the minimum mean square error (MMSE) algorithm, which can significantly reduce channel multipath expansion for both single-element and array reception, and overcome the error flattening phenomenon of TR technique. Due to the sparsity of the underwater acoustic channel, compressive sensing (CS) theory is commonly used in UWAC for channel estimation (18, 19, 20). However, the deep sea exhibits significant ocean stratification and different sound propagation patterns, resulting in noticeable changes in sound velocity gradients during sound propagation. This leads to significant multipath expansion in the received acoustic signals, typically ranging from several hundred milliseconds to several seconds (21). Under such conditions, the computational complexity of using CS methods for channel estimation is high, making it impractical for engineering implementation.

Therefore, to address the issue of excessive computational complexity required for precise channel estimation methods under deep-sea long multipath delay expansion, this paper proposes a low-complexity adaptive linear RAKE receiver. First, an adaptive threshold setting method based on the correlated noise floor after correlation is employed to select effective multipaths, thereby reducing the interference of different noise levels on the correlation peaks. Then, the index of “average distance of energy distribution” is introduced to evaluate the demodulation performance of each path, further mitigating the impact of selected noisy paths on the RAKE receiver. Finally, the demodulation results of each path are merged to obtain multipath diversity gain, thereby improving decoding performance. Simulation and sea trial experiments have validated the effectiveness of this method.

The remainder of this paper is organized as follows: In Section 2, we introduce the bionic UWAC algorithm by mimicking dolphin whistle based on frequency shift keying. Section 3 introduces the adaptive RAKE receiver proposed in this paper. The analysis of simulation and the data processing results of sea trial are shown in Section 4. Finally Section 5 summarizes all the work and gives the conclusion.

2. BIONIC UWAC ALGORITHM BY MIMICKING DOLPHIN WHISTLE

2.1 Dolphin Whistle Signal Model

The time-frequency characteristics of dolphin whistle signals can be determined by the time-frequency spectrum contour curve of the whistle signals. Based on the fundamental time-frequency spectrum contour curve of dolphin whistle signals, the whistle signals can be modeled and synthesized (22), with the formula being

$$s_w(t) = \sum_{r=1}^{R} A_r(t)\cos\left(\int_0^t 2\pi f_{w,r}(\tau)\,\mathrm{d}\tau + \varphi_{0,r}\right),\ 0 < t \leqslant T_\omega \tag{1}$$

where T_w denotes the duration of the dolphin's whistle. R denotes the maximum harmonic order. $f_{w,r}(t) = r \times f_{w,1}(t)$ denotes the time-frequency spectrum contour curve of the rth harmonic, which can be obtained by doubling the fundamental frequency time-frequency spectrum contour $f_{w,1}(t)$. $A_r(t)$ denotes the amplitude of the rth harmonic as a function of time and $\varphi_{0,r}$ denotes the initial phase of the rth harmonic.

2.2 Modulation Algorithm

Assume that each whistle signal can modulate binary information represented by $\boldsymbol{a}=[a_1,a_2,\cdots,a_N]$. This is converted into a bipolar code $\boldsymbol{c}=[c_1,c_2,\cdots,c_N]$, where $c_i \in \{-1,1\}$. N represents the number of binary information bits that each whistle signal can modulate. If the code symbol width is T, then $N = \lfloor T_w/T \rfloor$ ($\lfloor\cdot\rfloor$ denoting floor function). The rectangular wave baseband signal obtained after modulation of $\boldsymbol{c}$ is

$$m_c(t) = \sum_{i=1}^{N} c_i \times u(t - iT), 0 < t \leq NT, u(t) = \begin{cases} 1, 0 < t \leq T \\ 0, otherwise \end{cases} \quad (2)$$

Based on the principles outlined in Reference (12), a sine wave signal of half is selected to smooth the transition at the transition point $c_i \neq c_{i+1}$, and the end of $m_c(t)$ is padded with zeros to ensure that the duration of $m_c(t)$ matches that of the whistle spectral contour $f_{w,1}(t)$, thereby obtaining the final baseband signal $m_w(t)$. Multiplying $m_w(t)$ by the spectral shift Δf and adding it to the dolphin whistle signal's time-frequency spectral contour $f_{w,1}(t)$ yields the synthesized whistle time-frequency spectral contour $f_s(t)$, with the formula being

$$f_s(t) = m_w(t) \times \Delta f + f_{w,1}(t), 0 < t \leq T_w \quad (3)$$

where Δf represents the absolute value of the frequency difference between the signal frequency of each symbol and the carrier frequency f_c in frequency-shift keying (FSK). This value determines the magnitude of the change in the synthesized whistle signal spectrum. The smaller Δf is, the smaller the change in the synthesized whistle signal spectrum, and the more similar it becomes to the original signal. The overall frequency modulation process is illustrated in Figure 1. Finally, based on the whistle signal parameter model in Equation (1), the synthesized whistle signal $s_s(t)$ is generated.

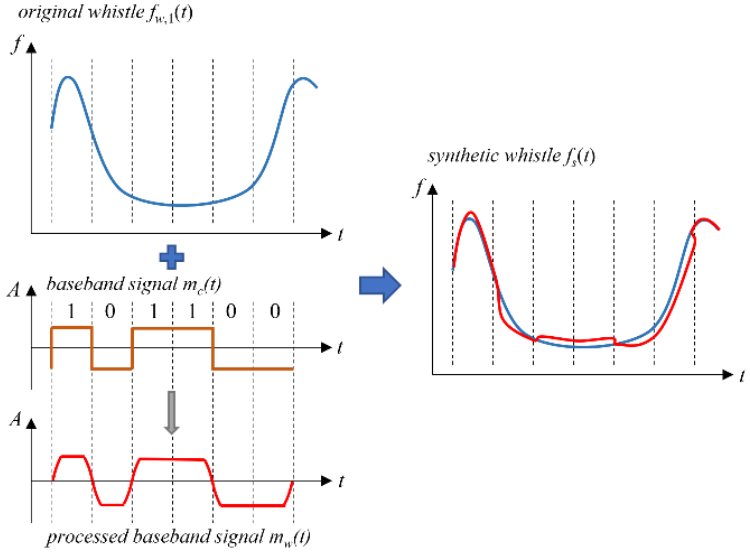


Figure 1 – Schematic diagram of FSK communication spectrum modulation by mimicking dolphin

The frame structure of the FSK communication signal by mimicking whistle is shown in Figure 2. Each frame contains one original dolphin whistle signal $s_w(t)$ and N_F synthesized whistles $s_s(t)$. There is a guard interval between adjacent whistles to reduce the impact of the underwater multipath channel.

original whistle $s_w(t)$	guard interval	synthetic whistle $s_{s,1}(t)$	guard interval	synthetic whistle $s_{s,2}(t)$	.....	synthetic whistle $s_{s,N}(t)$

Figure 2 – Frame structure of FSK communication spectrum modulation by mimicking dolphin

2.3 Demodulation Algorithm

At the receiver, the original dolphin whistle signal $s_w(t)$ is first used to synchronize the received signal. Then, the doppler factor is estimated using a correlator array (23), and doppler compensation is performed on the signal through resampling. After extracting the received synthetic whistle signal $r_s(t)$, a local whistle signal $s_1(t)$ is generated at the receiver based on the time-frequency spectrum contour $f_1(t)$.

$$s_1(t) = \cos\left(\int_0^t 2\pi f_1(\tau)\mathrm{d}\tau + \varphi_{0,r}\right), 0 < t \leq T_w \quad (4)$$

where the local time-frequency spectrum contour $f_1(t)$ differs from the dolphin whistle spectrum contour $f_{w,1}(t)$ by a fixed frequency f_c, which is the carrier frequency of the final FSK signal.

After coherently multiplying the received synthetic whistle $r_s(t)$ with the local whistle signal $s_1(t)$, we obtain

$$
\begin{aligned}
r(t) &= r_s(t)s_1(t) \\
&= A'(t)\cos\left(\int_0^t 2\pi f_s(\tau)\mathrm{d}\tau + \varphi_{0,r}\right)\cdot\cos\left(\int_0^t 2\pi f_1(\tau)\mathrm{d}(\tau) + \varphi_{0,r}\right) \\
&= \frac{A'(t)}{2}\left[\cos\left(\int_0^t 2\pi(f_c + m_w(\tau)\times\Delta f)\mathrm{d}\tau\right)\right. \\
&\quad \left.+\cos\left(\int_0^t 2\pi(2f_{w,1}(\tau) - f_c + m_w(\tau)\mathrm{d}\tau + 2\varphi_{0,r})\right)\right]
\end{aligned}
\tag{5}
$$

where the first term represents the low-frequency component, which is the desired FSK signal $r_{\mathrm{FSK}}(t)$ used for information demodulation, while the second term represents the high-frequency component, which is undesired interference and can be filtered out using a low-pass filter.

For 2FSK signals, non-coherent orthogonal energy demodulation (24) can be used. The specific processing flow is shown in Figure 3. Extract the ith symbol $r_i(t)$ of $r_{\mathrm{FSK}}(t)$, and calculate the energy $E_{i,1}(t)$ and $E_{i,2}(t)$ of $r_i(t)$ at its two possible carrier frequencies $f_1=f_c+\Delta f$ and $f_2=f_c-\Delta f$.

$$
\begin{aligned}
E_{i,1}(t) &= \left[\int_0^T r_i(t)\cos(2\pi f_1 t)\mathrm{d}t\right]^2 + \left[\int_0^T r_i(t)\sin(2\pi f_1 t)\mathrm{d}t\right]^2 \\
E_{i,2}(t) &= \left[\int_0^T r_i(t)\cos(2\pi f_2 t)\mathrm{d}t\right]^2 + \left[\int_0^T r_i(t)\sin(2\pi f_2 t)\mathrm{d}t\right]^2
\end{aligned}
\tag{6}
$$

The decision is made by comparing the magnitudes of $E_{i,1}(t)$ and $E_{i,2}(t)$. If $E_{i,1}(t)>E_{i,2}(t)$, then $a_i^*=1$. If $E_{i,1}(t)<E_{i,2}(t)$, then $a_i^*=0$. Repeating this process yields all demodulation information $\boldsymbol{a}^*=[a_1^*,a_2^*,\cdots,a_N^*]$.

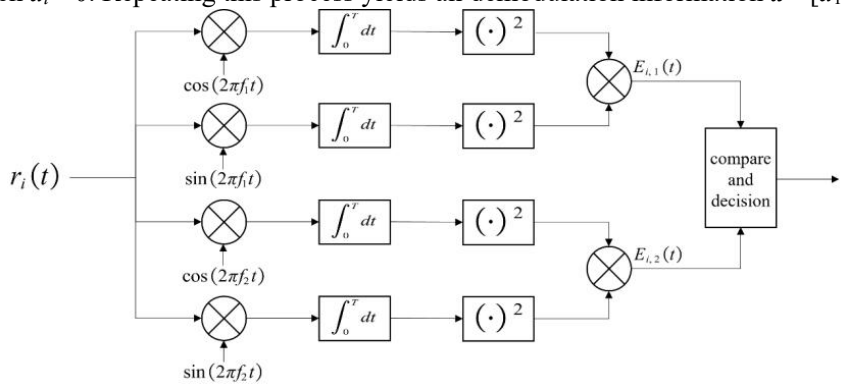


Figure 3 – Orthogonal demodulation method for 2FSK

3. ADAPTIVE RAKE RECEIVER

3.1 Basic Principles of RAKE Receiver

The frequency of dolphin whistle signals changes continuously over time. The time-frequency spectrum contour of adjacent symbols in the synthesized whistle signals are not identical. Due to the multipath effects of the acoustic channel, the interference from the previous symbol with the current symbol in the spectrum is either non-overlapping or has a very small overlap range. Therefore, the dolphin whistle FSK communication method can exhibit a certain degree of resistance to ISI even without channel equalization. However, the underwater acoustic channel is complex, and multipath effects still impact the performance of bionic communication algorithms. RAKE receiver technology is a multipath diversity reception technique that adjusts the weights of time-resolved multipath signals individually, combining them into an enhanced signal to boost the energy of the useful signal and improve decoding performance.

By detecting the delay of multipath signals through the correlation peak between the local synchronization whistle signal and the received signal, the signals at each delay are demodulated using orthogonal energy demodulation as shown in Section 2.3, and the demodulation results from each path are then combined to obtain multipath diversity gain. Figure 4 shows the structure of the RAKE receiver. The selection of multipath delay estimation and merging methods are the two major factors affecting the performance of the RAKE receiver. Merging the demodulation results of noisy paths may lead to performance degradation. This paper proposes an adaptive RAKE receiver to address this problem, which selects effective multipaths by setting adaptive thresholds and introduces the index of "average distance of energy distribution" to adaptively adjust the multipath merging method. The following sections will provide a detailed description of these two

improvements.

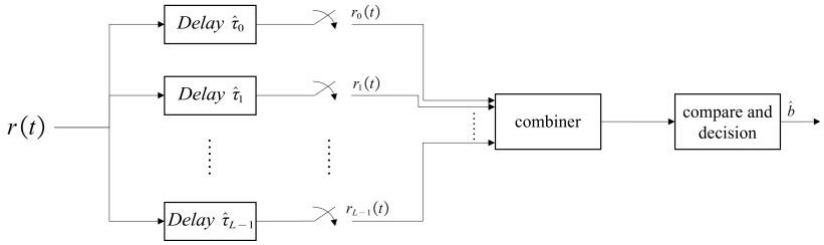


Figure 4 – Structure of RAKE receiver

3.2 Adaptive Threshold Setting Method for Cross-correlation Peak Detection

Due to the influence of noise, the cross-correlation peaks may not necessarily be valid multipaths, so a threshold needs to be set to minimize the interference caused by noise on multipath selection. If the threshold is set too low, it is easy to select noise paths, which affects the performance of subsequent RAKE combiner. However, if the threshold is set too high, it may cause valid paths to be missed, thereby reducing the multipath diversity gain. Therefore, a fixed threshold often cannot adapt to different channel structures and noise levels, so an adaptive threshold is often required. A method proposed in (25) for setting an adaptive threshold based on the correlated noise floor after cross-correlation involves calculating the mean and variance of the correlated peak noise segment to set the threshold as

$$\Psi = \lambda\sigma_{nc} + \mu_{nc} \tag{7}$$

Where λ is a constant. Based on experience, λ can be set to 5.

3.3 Adaptive Multipath Merging Scheme

There are three common RAKE combining methods: selective combining (SC), equal gain combining (EGC) and maximum ratio combining (MRC). SC selects one signal from all branches as the final result. EGC combines all branches with the same weight. MRC assigns different weights to each signal based on its SNR, thereby maximizing the SNR of the combined signal. Theoretically, MRC is the most optimal of the three methods.

After selecting effective multipaths through adaptive threshold settings, the signals from each path are demodulated using orthogonal energy demodulation. To further avoid the influence of merged noise paths on the final demodulation results, the index of "average distance of energy distribution" is introduced to evaluate the quality of the demodulation results, thereby assisting RAKE merging. Since binary frequency-shift keying is used, two energy decision values, f_0 and f_1, are ultimately generated. When $f_0 > f_1$, it is judged as "0", when $f_0 < f_1$, it is judged as "1". The definition of the "average distance of energy distribution" is

$$energy_distance = \sum_{i=1}^{N} \left| \frac{f_{0i} - f_{1i}}{f_{0i} + f_{1i}} \right| \Big/ N \tag{8}$$

where N is the total number of modulated bits. The closer *energy_distance* is to 1, the greater the difference between the two decision quantities, indicating that the demodulation result for that path is more reliable. If the maximum *energy_distance* exceeds 0.95 or one path has an *energy_distance* significantly greater than the others, the result of that path is directly selected (i.e., SC). Otherwise, the path with an *energy_distance* greater than the average *energy_distance* of all paths is selected for MRC. Figure 5 illustrates the adaptive selection process of the multipath merging scheme.

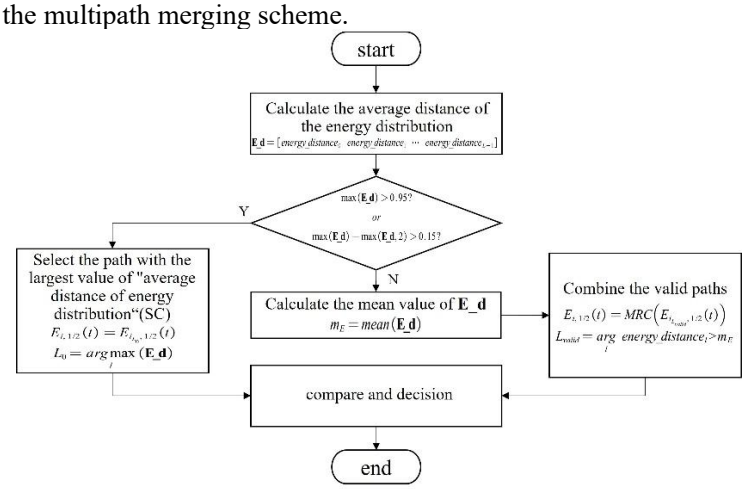
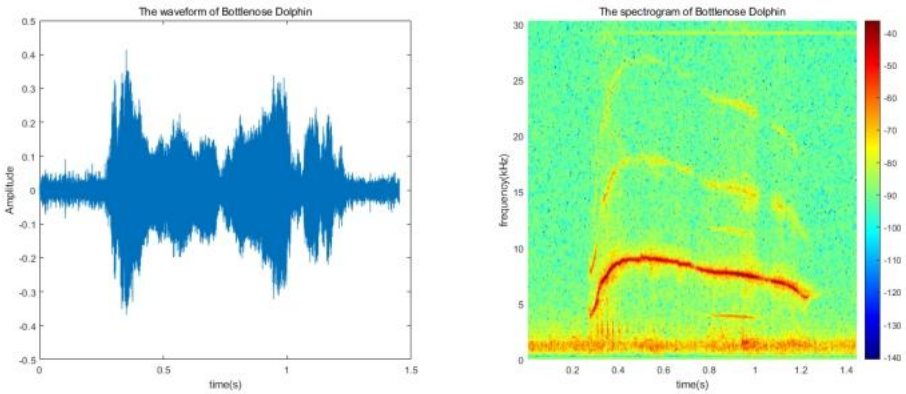


Figure 5 – The adaptive selection process of the multipath merging scheme

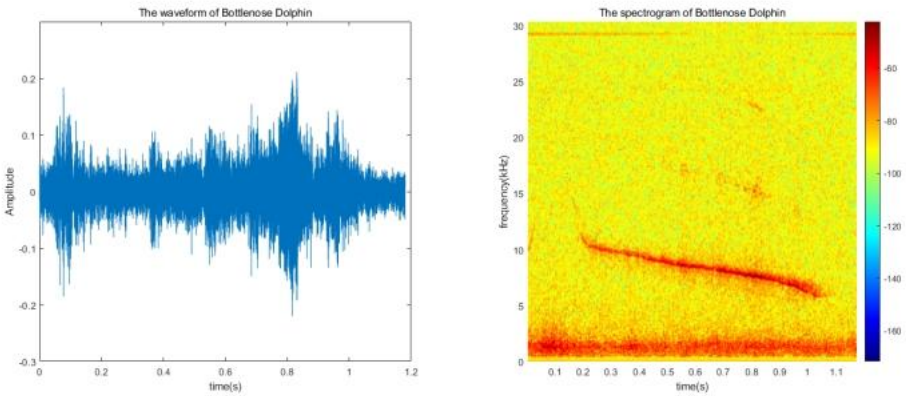
4. SIMULATION AND SEA TRIALS

4.1 Simulation Analysis

A bottlenose dolphin whistle signal was selected as the simulated whistle communication object. This whistle signal is available in the Watkins Marine Mammal Sound Database (26), with its waveform and time-frequency spectrum shown in Figure 6(a). The whistle signal has a duration of approximately 1 s, with a clear time-frequency spectrum contour and the frequency continuously varies over time. The whistle shown in Figure 6(b) exhibits good correlation characteristics, making it suitable for use as a synchronization signal.



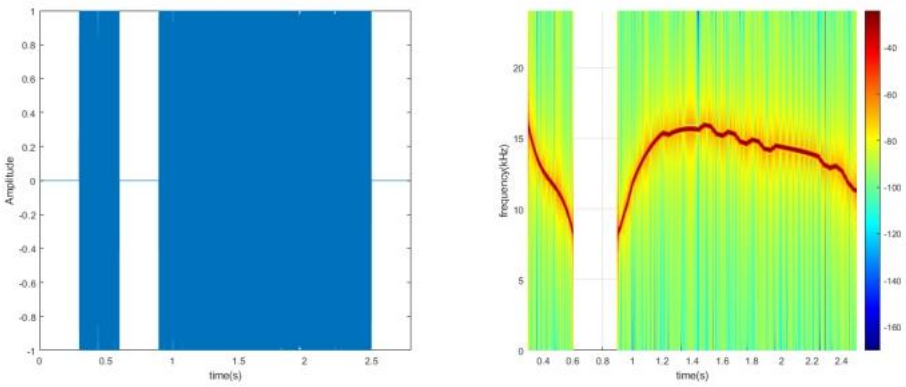
(a)whistle 1



(b)whistle 2

Figure 6 – Bottlenose Dolphin whistles

Based on the time-frequency spectrum contour extraction method described in Reference (13), the fundamental time-frequency spectrum contour of the whistle was extracted, and its frequency range was adjusted to 8–16 kHz for subsequent UWAC experiments by mimicking dolphin. Each frame of communication signals contains a segment of raw whistle for synchronization, doppler and channel estimation, and synthesized whistles for information transmission. The transmitted signal waveform is shown in Figure 7.



(a) time-domain waveform (b) time-frequency spectrum

Figure 7 – Transmission signal waveform

The simulated channels are based on 20 sets of channel impulse responses obtained from channel estimation using actual measurement data collected in the South China Sea in the summer of 2023. Figure 8 shows examples of some of the actual measured underwater acoustic channels, which can be seen to have a delay extension of 1–2 s and a complex and variable multipath structure.

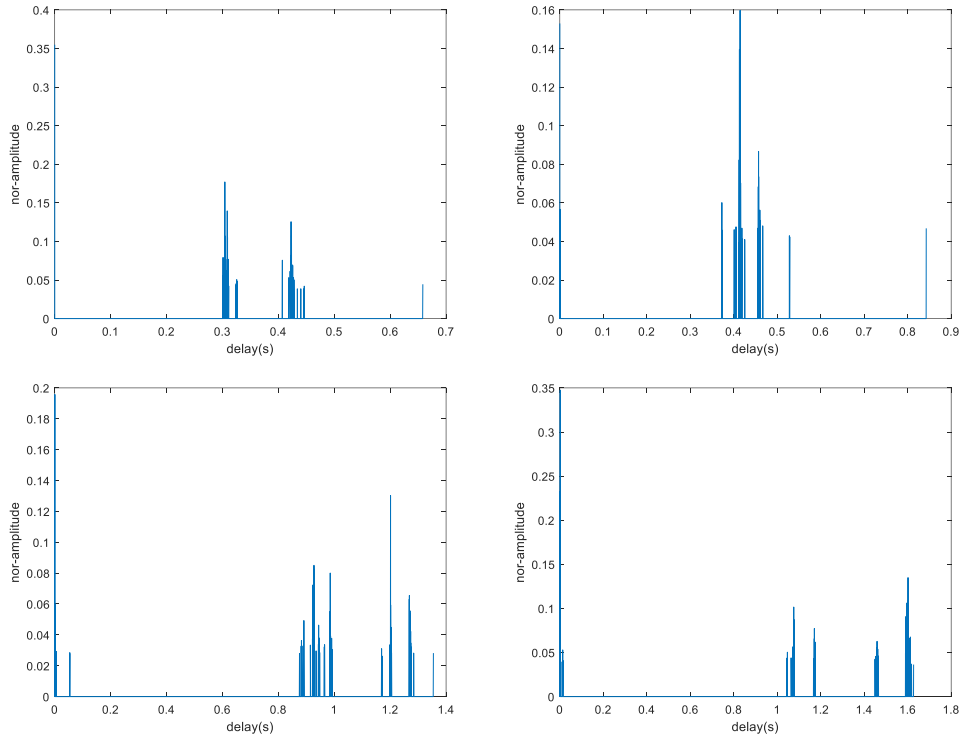


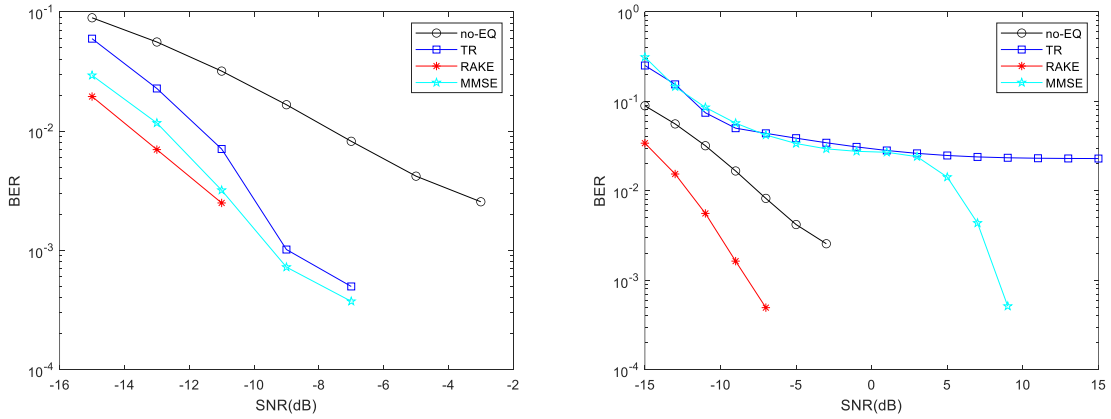
Figure 8 – The actual measured underwater acoustic channels

Based on these measured channels, we simulated and compared the bit error rate(BER) performance of bionic UWAC by mimicking dolphin under different channel estimation and equalization schemes. The specific simulation parameters are shown in Table 1.

Table 1 – The settings of simulation parameter

Parameter Name	Parameter Value
Symbol Duration (T)	0.08 s
Modulated Frequency Difference (Δf)	200 Hz
Carrier Frequency of FSK (f_c)	200 Hz
Sampling Rate (f_s)	48 kHz
Communication Band (B)	8-16 kHz
Maximum Number of the Estimating Multipaths (N)	10

Monte Carlo simulations were conducted using multiple sets of deep-sea measured acoustic channels to obtain the average BER curves for the RAKE receiver, TR and MMSE equalization under the orthogonal matching pursuit (OMP) channel estimation and corss-correlation peak selection multipath scenarios, as shown in Figure 9. As can be seen from the figure, when using OMP channel estimation, all three methods achieve certain performance gains. However, when using cross-correlation channel estimation, due to increased channel estimation errors, the performance of TR and frequency-domain MMSE equalization drops sharply, while the RAKE receiver still maintains good performance. In deep-sea channels, where the multipath delay of the channel is often very long, the RAKE receiver based on cross-correlation channel estimation offers a better trade-off between complexity and performance.



(a) OMP (b)cross-correlation

Figure 9 – SNR - BER curves for different equalization schemes

Figure 10 compares the average BER performance of the traditional RAKE receiver (with a fixed threshold of 0.3 for cross-correlation channel estimation and MRC of all selected paths), the RAKE receiver with adaptive threshold for cross-correlation channel estimation, and the improved adaptive RAKE receiver with "average distance of energy distribution" assistance under multiple measured channels. As shown in the figure, the RAKE receiver with an adaptive threshold achieves improved performance because the adaptive threshold can adapt to different noise levels, thereby reducing the impact of noise on multipath selection. Meanwhile, the RAKE receiver that uses the "average distance of energy distribution" to assist in merging decisions further reduces the probability of merging into noisy paths, leading to a further decrease in BER.

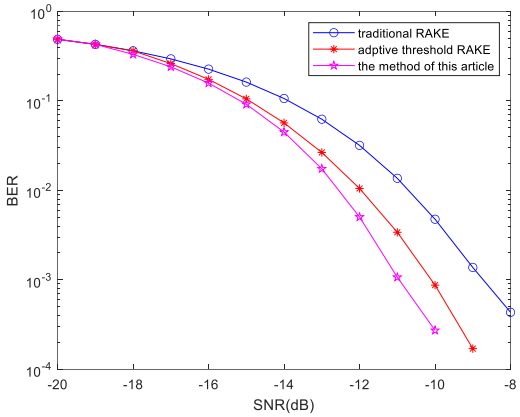


Figure 10 – SNR - BER curves for different RAKE receiver schemes

4.2 Sea Trial Verification

In September 2024, an UWAC experiment using communication signals by mimicking dolphin whistle was conducted in the South China Sea. The signal parameters used in the sea trial were consistent with those in the previous simulation. A total of 328 valid data sets were received and processed. The performance of the no-equalization, traditional RAKE receiver (with a fixed threshold of 0.3 for cross-correlation channel estimation and MRC of all selected paths) and the improved adaptive RAKE receiver proposed in this paper were compared. Their BERs and data packet transmission success rate are shown in Table 2. As a result, the proposed adaptive threshold RAKE receiver with "average distance of energy distribution" assistance outperforms the traditional RAKE receiver. Compared to the results before equalization, the average BER is reduced by one order of magnitude, and the data packet transmission success rate is improved by 13.7%, validating the effectiveness of the proposed method.

Table 2 – Results of sea trial data processing under different schemes

Demodulation method	no-equalization	traditional RAKE receiver	the improved adaptive RAKE receiver proposed in this paper
BER	2.37×10^{-2}	8.55×10^{-3}	4.81×10^{-3}
data packet transmission success rate	78.31%	91.10%	92.01%

5. CONCLUSIONS

Embedding information into the time-frequency spectrum contour of dolphin whistle via frequency modulation is the primary method for bionic UWAC. However, the multipath effects of underwater acoustic channels can degrade communication performance, so channel estimation and equalization are typically required at the receiver. To address the issue of excessive computational complexity required for precise channel estimation methods under deep-sea long multipath delay expansion conditions, this paper proposes a lower-complexity adaptive RAKE receiver to mitigate multipath effects. An adaptive threshold setting method based on the correlated noise floor after cross-correlation is employed to select effective multipaths, reducing interference from different noise levels on correlation peaks. The index of "average distance of energy distribution" is introduced to evaluate the demodulation performance of each path, further mitigating the impact of selecting noisy paths on the RAKE receiver. Finally, the demodulation results of all valid paths are merged to obtain multipath diversity gain, enhancing system performance. Simulation and sea trial validated the effectiveness of this method. By applying this method to process sea trial data, the average BER was reduced by an order of magnitude, and the data packet transmission success rate improved by 13.7%. This method is simple in principle, has low computational complexity, and is easier to implement in engineering, providing new insights for the practical application of robust bionic UWAC in deep-sea environments.

ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China under Grant No. 62231011.

REFERENCES

1. Qian YJ, Su H, Han GJ, et al. Ecologically friendly full-duplex data transmission scheme for underwater acoustic sensor networks. IEEE Internet Things J., 2020, 8(9): 7676–7690.
2. Holt MM, Noren DP, Dunkin RC, et al. Vocal performance affects metabolic rate in dolphins: implications for animals communicating in noisy environments. J. Exp. Biol., 2015, 218(11): 1647–1654.
3. Liu SC, Gao LJ, and Su DP. Deep learning based underwater acoustic channel estimation exploiting physical knowledge on channel sparsity. ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp) / ACM International Symposium on Wearable Computers (ISWC). ELECTR NETWORK, SEP 21-26, 2021, 655–659.
4. Richardson WJ, Greene CR, Malme CI, et al. Marine mammals and noise. Academic Press, 1995.
5. Songzuo Liu, Mohsin Abrar Khan, Muhammad Bilal, and Habib Hussain Zuberi. Low Probability Detection Constrained Underwater Acoustic Communication: A Comprehensive Review. IEEE Commun. Mag., 2025, 63(2): 21–30.
6. Qiao G, Bilal M, Liu SZ, et al. Biologically Inspired covert underwater acoustic communication—A review. Phys. Commun., 2018, 30: 107–114.
7. Ma TL. Bionic underwater acoustic communication by mimicking cetaceans whistles. [Master dissertation], Harbin Engineering University, 2017.
8. Niu FQ, Xue RC, Zhou ZM, et al. Analysis of differences on whistle classification and characteristics of Indo-Pacific bottlenose dolphins(Tursiops aduncus) in different environment. Acta Acustica, 2020; 45(2):189-195.
9. Liu SZ, Ma TL, Qiao G, et al. Biologically inspired covert underwater acoustic communication by mimicking dolphin whistles. Appl. Acoust., 2017; 120: 120-128
10. Qiao G, Ma TL, Liu SZ, et al. A frequency hopping pattern inspired bionic underwater acoustic communication. Phys. Commun., 2021, 46: 101288.
11. Ma TL, Liu SZ, Qiao G, et al. Bionic Underwater Acoustic Communication by Mimicking Dolphin Whistle Based on Frequency Shift Keying. J. Electron. Inf. Techn., 2022, 44(06): 2045-2053
12. Liu SZ, Liu BJ, Yin YL, et al. M-ray covert underwater acoustic communication by mimicking dolphin sounds. J. Harbin Eng. Univ. 35 (1) (2014) 119–125.
13. Liu SZ, Qiao G, and Ismail A. Covert underwater acoustic communication using dolphin sounds. J. Acoust. Soc. Am., 2013, 133(4): EL300–EL306.
14. Liu SZ, Ma TL, Qiao G, et al. Bionic communication by dolphin whistle with continuous-phase based on MSK modulation. IEEE International Conference on Signal Processing. IEEE, 2016.
15. Li Z, Jiang J, Chen X D F. Bio-inspired secure underwater acoustic communication strategy by shifting the natural whale whistle phase. Appl. Acoust, 2023, 214(Nov.):1.1-1.9.
16. Li Z, Jiang J, Chen G, et al. Bio-inspired secure underwater acoustic communication by quasi-orthogonal

keying. Appl. Acoust, 2024, 218(000).
17. Fan WW. Research on the key technology of AUV acoustic Frequency-Hopping FSK communication. [Doctoral dissertation], Harbin Engineering University, 2015.
18. Stojanovic M, Preisig J. Underwater acoustic communication channels: Propagation models and statistical characterization. IEEE Commun. Mag., 2009; 47(1): 84−89.
19. Yang T C. Properties of underwater acoustic communication channels in shallow water. J. Acoust. Soc. Am., 2012; 131(1): 129−145.
20. Jiang WH, Tong F, Zhang HT, et al. Dynamic discriminative compressed sensing estimation of hybrid sparse underwater acoustic channel. Acta Acustica , 2021; 46(6): 825−834.
21. Zhao YF, Tong F, Ma BL, et al. Multipath-cluster-wise joint equalization method for long-range deep-sea single-element underwater acoustic communication. Acta Acustica , 2024, 49(04): 764-773.
22. Buck JR and Morgenbesser HB. Synthesis and modification of the whistles of the bottlenose dolphin, Tursiops truncatus. J. Acoust. Soc. Am., 2000, 108(1): 407–416.
23. Yu JT. Research on Mobile Underwater Acoustic Communication Based on Hyperbolic Frequency Modulation Signal. [Master dissertation], Zhejiang University, 2017.
24. Sun HS. Mobile Underwater Acoustic Communication based on FH-FSK. [Master dissertation], Harbin Engineering University, 2019.
25. Sun DJ, Wu J, Zheng HC .The line-of-sight peak detection and tracking of underwater acoustic DSSS communications in the doubly spread channel. Appl. Acoust, 2024, 216(Jan.):109761.1-109761.10.
26. Watkins W. Watkins marine mammal sound database. https://cis.whoi.edu/science/B/whalesounds/about.cfm, 2021.

PROCEEDINGS of the
16th International Conference on Theoretical and Computational Acoustics

August 18 to 22, 2025 BEXCO, Busan Korea

Iterative Underwater Acoustic DSSS Processing Algorithm for Integrated Communication and Localization System in Non-Uniform Doppler and Strong Noise Interference Channels

Jinhao Deng[1,2,3]; Jie Wu[4]; Hongyu Cui[1,2,3]
[1] 1. National Key Laboratory of underwater Acoustic Technology, Harbin Engineering University, Harbin 150001, China;
[2] Key Laboratory of Marine Information Acquisition and Security(Harbin Engineering University),Ministry of Industry and Information Technology; Harbin 150001,China;
[3] College of Underwater Acoustic Engineering, Harbin Engineering University, Harbin 150001, China
[4] The Southern Marine Science and Engineering Guangdong Laboratory Guangzhou), Guangzhou 511458, China

ABSTRACT

The integration of underwater communication and positioning technology using a compact ultrashort baseline array is analyzed. Due to the broadband characteristic of underwater acoustic signals, the received signal experiences not only severe multipath and Doppler interference during maneuvering but also strong noise interference, which significantly degrades the receiver's processing performance. Moreover, the modulation of communication bits disrupts the phase continuity of the signal, introducing additional phase jumps in the measurement results, which is unacceptable for high-precision measurement systems. To address these challenges, this paper proposes an integrated communication and positioning receiver processing method. First, a double differential encoding scheme is employed to mitigate constant Doppler effects, while the bit-interleaved coded modulation with iterative decoding (BICM-ID) technique is utilized to compensate for the signal-to-noise ratio (SNR) loss introduced by differential encoding. Then, a joint estimation method for data bits and the channel is proposed to achieve robust communication and high-precision measurements. Furthermore, a beamforming approach based on the maximum output signal-to-interference-plus-noise ratio (SINR) criterion is proposed to suppress strong interference signals.

Keywords: Non-Uniform Doppler, Strong Noise Interference, DSSS, Underwater Acoustic Communication;

1. INTRODUCTION

Recent advances have demonstrated the potential of integrated underwater acoustic (UWA) communication and positioning to enhance the efficiency of underwater operations by enabling simultaneous data exchange and localization system through shared hardware and signal processing frameworks [1]. Direct-sequence spread spectrum (DSSS) technology has emerged as a promising enabler for such integration, owing to its robustness against interference and high multipath resolution capability. Specifically, DSSS suppresses narrowband interference and ambient noise through wideband pseudorandom code modulation, ensuring reliable communication in complex UWA channels [2]. Meanwhile, its inherent bandwidth expansion facilitates precise multipath component identification, supporting high-accuracy positioning [3]. Meanwhile, the ultra-short baseline (USBL) system has become a commonly used positioning device for AUVs due to its simple structure and easy installation [4]. While DSSS and USBL systems provide foundational capabilities, their performance is fundamentally constrained by three challenges. First, the limited sound speed in water results in severe Doppler effects within UWA channels, which not only induce frequency shifts in the received signals but also cause dynamic compression or expansion of the signal pulse width [5]. Second, integrated systems are often installed on mobile platforms (e.g., AUV or ships), and the platform motion generates directional strong noise, significantly degrading the performance of the receiver [6]. Finally, the phase discontinuities introduced by the bits modulation directly reduce positioning accuracy, as they disrupt the phase continuity required for high-precision ranging algorithms.

Doppler mitigation in UWA communications primarily follows two approaches. The first employs estimation/compensation using Doppler-insensitive reference signals like LFM/HFM [7], but these only estimate frame-averaged Doppler shifts, failing in dynamic scenarios. The second uses differential modulation (e.g., DDPSK [8]), which handles constant Doppler at SNR costs but degrades under accelerated

motion. Recent advances combine these approaches: PSP architectures [9] integrate channel estimation with decoding for time-varying cases, while bit-interleaved coded modulation with iterative decoding (BICM-ID) systems [10] employ iterative demodulation-decoding to recover SNR losses. However, these solutions remain vulnerable to environmental interference.

While time/frequency-domain approaches (e.g., adaptive notch filtering, periodic interference cancellation) are limited by the non-stationary and wideband nature of UWA channels [11], spatial array processing remains the dominant solution for interference suppression. Among these, subspace-based methods demonstrate strong interference rejection capabilities but require extensive prior information and typically operate at high SNR conditions [12]. Alternatively, beamforming techniques enhance the output signal-to-interference-plus-noise ratio (SINR) by steering the main lobe toward the desired signal while nulling interference directions. Conventional approaches are widely used, yet their effectiveness is constrained in USBL systems due to the limited array aperture and sparse element configuration [13].

Existing research on interference cancellation for integrated communication positioning systems remains limited. In RF systems, phase-locked loops (PLLs) track phase variations effectively, but their performance degrades in underwater acoustic channels due to severe multipath effects [14]. An alternative approach involves reconstructing reference signals from successfully decoded DSSS bits to compensate for symbol-transition-induced delays in time-delay measurements.

In this paper, we propose an iterative UWA DSSS processing algorithm for integrated communication and localization system. The main contributions are as follows

1) To combat constant Doppler effects in UWA channels, the proposed method employs a DDPSK to mitigate residual Doppler after resampling, while utilizing a BICM-ID framework to recover the inherent SNR loss from differential encoding.

2) For enhanced interference resistance against strong directional noise, a beamforming approach based on the maximum SINR criterion is developed.

3) A joint data-and-channel estimation scheme is introduced to simultaneously achieve robust communication performance and high-precision positioning capability

The rest of this paper is organized as follows. Section 2 describes the structure and algorithm of the proposed receiver. Numerical simulation results are presented in Section 3. The data analysis of sea experiments is shown in Section 4. Finally, Section 5 summarizes the whole paper.

2. THE PROPOSED RECEIVER

The overall structure of the receiver can be divided into three main parts: the preprocessing for multi-channel array signals, the BICM-ID and the joint data and channel estimation. The block diagram of the proposed receiver is shown in Fig.1.

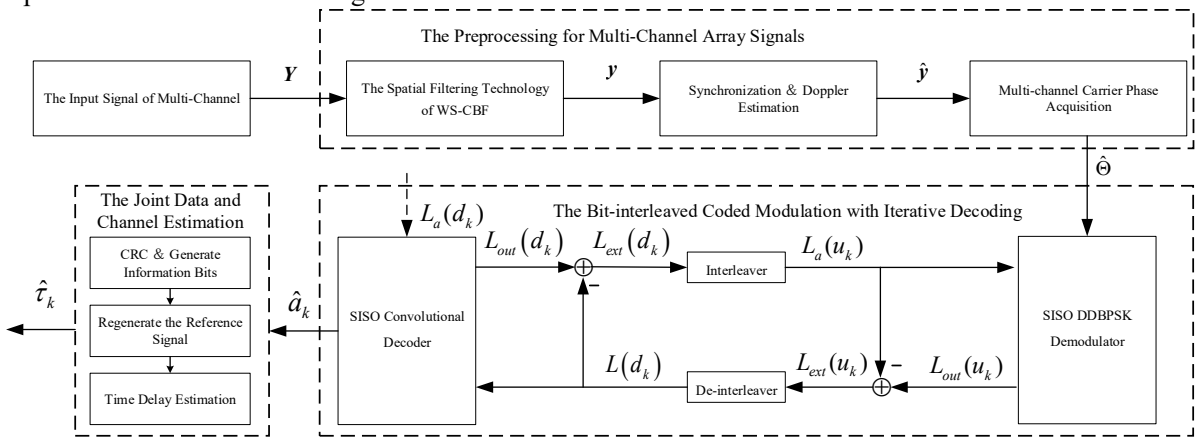


FIG.1 Block diagram of the proposed receiver

Upon receiving multi-channel signals $\boldsymbol{Y}$ by the USBL array, the receiver first suppresses strong noise interference from specific directions using array signal processing techniques and the processed signal is $\boldsymbol{y}$. Subsequently, synchronization and Doppler estimation methods are employed to estimate the initial time delay and velocity of a frame of DSSS signal. Following this, the carrier phase of the received signal $\hat{\Theta}$ is estimated, and the information bits are decoded utilizing a BICM-ID structure. After decoding, the Cyclic Redundancy Check (CRC) is applied to verify the decoded information bits, and the transmitted information bits are regenerated. Finally, the regenerated reference signal is used to re-estimate the time delay $\hat{\tau}_k$ of the received signal.

2.1 The spatial filtering technology of WS-CBF

For the conventional beamformers (CBF) used in the UWA wideband communication systems, it only needs to satisfy the distortionless constraint in the direction of signal arrival. This constraint ensures a 0 dB main lobe response in the desired direction while achieving maximum spatially uniform white noise gain equal to the number of array elements. However, when strong directional interference exists at the receiver, simply steering the beam toward the signal direction may not yield optimal performance. If strong interference happens to align with these deep nulls, the conventional beamformer can still demonstrate significant interference suppression capability.

Based on this concept, we propose an improved spatial-domain processing method derived from CBF. By deliberately steering the beam away from its nominal direction, the proposed approach maximizes the difference between the beam response toward the desired signal and that toward the interference direction, thereby achieving optimal noise and interference suppression performance and maximizing the output signal-to-interference-plus-noise ratio (SINR). This method is feasible for broadband communication systems and exhibits promising application potential. Although the beam-steering direction may no longer have physical significance (i.e., it does not strictly align with the actual signal direction), it ensures the highest signal quality and maximum output SINR for communication signal processing. The SINR output by the beam is defined as:

$$SINR_{out} = \frac{\sigma_{ys}^2}{\sigma_{yi}^2 + \sigma_{yn}^2} = \frac{\sigma_s^2 \boldsymbol{w}^H \boldsymbol{p}_s \boldsymbol{p}_s^H \boldsymbol{w}}{\sum_{d=1}^{D} \sigma_d^2 \boldsymbol{w}^H \boldsymbol{p}_d \boldsymbol{p}_d^H \boldsymbol{w} + \boldsymbol{w}^H \sigma_n^2 \boldsymbol{\rho}_n \boldsymbol{w}} \tag{1}$$

The power of the beamformed output signal, interference, and noise are denoted by σ_{ys}^2, σ_{yi}^2 and σ_{yn}^2, respectively. The weighting vector of the conventional beamformer is $\boldsymbol{w}^{\mathrm{H}} = \left[w_1^*, w_2^*, \cdots, w_M^* \right]$, $\boldsymbol{p}_s$ represents the direction response vector corresponding to the expected signal, and $\boldsymbol{p}_d$ represents the direction response vector corresponding to the interference signal. By traversing different directions in space, the variable $SINR_{out}$ reaches its maximum. Thus, the CBF method based on the optimal SINR criterion is called (SINR-CBF, SCBF), and its formula is:

$$\begin{aligned} \hat{\Omega}_0 &= \arg\max_{\Omega_0 \in \Theta} \left(\frac{\sigma_s^2 \boldsymbol{w}^H \left(\Omega_0 \right) \boldsymbol{p}_s \left(\Omega_s \right) \boldsymbol{p}_s^H \left(\Omega_s \right) \boldsymbol{w} \left(\Omega_0 \right)}{\sum_{d=1}^{D} \sigma_d^2 \boldsymbol{w}^H \left(\Omega_0 \right) \boldsymbol{p}_d \left(\Omega_d \right) \boldsymbol{p}_d^H \left(\Omega_d \right) \boldsymbol{w} \left(\Omega_0 \right) + \boldsymbol{w}^H \left(\Omega_0 \right) \sigma_n^2 \boldsymbol{\rho}_n \boldsymbol{w} \left(\Omega_0 \right)} \right) \\ &= \arg\max_{\Omega_0 \in \Theta} \left(SINR_{out} \left(\Omega_0 \right) \right) \end{aligned} \tag{2}$$

The Ω_s and Ω_d represent the propagation directions of the desired signal and the interference signal respectively. While CBF requires a priori knowledge of the desired signal's direction of arrival (DOA) to compute the array weight vector, the proposed SCBF method eliminates this requirement. It operates effectively without knowing the DOAs of either the desired signal or interfering signals. It is observed from empirical data that the levels of noise and interference vary across different array elements. If equal-gain combining is directly applied to all channels, those with higher noise and interference power may degrade the overall combined performance. Inspired by the principle of multichannel diversity combining, channels with higher signal-to-noise ratios (SNRs) should be assigned larger weighting factors, thereby contributing more significantly to the combined signal, while channels with lower SNRs should receive smaller weights. In the special case where all channels have equal noise power, the weighting factors are determined by the amplitude of the received signals. These characteristics enable maximal ratio combining (MRC) to achieve the maximum output SINR under any fading condition. Accordingly, by estimating the variance of the noise-plus-interference for each channel which is denoted as $\sigma_{m,i+n}^2$ for the mth channel, the weighting vector is written as:

$$w_{am} = \frac{\sigma_{1,i+n}^2}{\sigma_{m,i+n}^2} \tag{3}$$

By incorporating the variances of interference and noise into the weighting process, the conventional steering vector-based beamforming (SCBF) can be better adapted to practical environments. This enhanced

method is referred to as the Weighted Signal-to-Interference-plus-Noise Ratio Conventional Beamforming (Weight-SINR-CBF, WSCBF). The corresponding weight vector is given by:

$$\boldsymbol{w}_{ws}^{H} = \left[w_{a1} w_1^*, w_{a2} w_2^*, \cdots, w_{aM} w_M^* \right] \tag{4}$$

Accordingly, the signal after multichannel processing can be expressed as:

$$\boldsymbol{y} = \boldsymbol{w}_{ws}^{H} \boldsymbol{Y} \tag{5}$$

2.2 The BICM-ID method

We should calculate a posterior probability (APP) of the input bit u_k, and then APP log likelihood ratio (LLR) information displayed in eq (6) is exported by DDPSK demodulator outputs.

$$L(u_k) = \ln \frac{P(u_k = +1 | \hat{\Theta})}{P(u_k = -1 | \hat{\Theta})} \tag{6}$$

Finally, system makes the decision of bit u_k on the basis of eq (7):

$$\hat{u}_k = \begin{cases} +1 & L(u_k) > 0 \\ -1 & L(u_k) < 0 \end{cases} \quad k = 0,1,\cdots,M \tag{7}$$

Hence, the eq (7) can be calculated as:

$$L(u_k) = \ln \frac{\sum_{(s',s) \in \sum_k^+} \alpha_k(s') \cdot \gamma_k(s',s) \cdot \beta_{k+1}(s)}{\sum_{(s',s) \in \sum_k^-} \alpha_k(s') \cdot \gamma_k(s',s) \cdot \beta_{k+1}(s)} \tag{8}$$

Where s′,s are a couple of states that denote the state transition which is caused by the kth moment input bit u_k, and $\sum_k^+$ and $\sum_k^-$ represent the transitions caused by $u_k = +1$ and $u_k = -1$. By analyzing related researches, we find that α,β metric can be calculated by γ metric recursively. Thus, the kernel of this problem becomes how to calculate the γ metric. The transmitter binary sequence of u_k or v_k in each receiving channel are consistent. Therefore, we can calculate γ_k metric in eq (9):

$$\gamma_k(s',s) = P(u_k) \cdot e^{-\frac{E_s}{N_0} \left\| \hat{\Theta}_k - h_k \cdot v_k \right\|^2} \tag{9}$$

Where the h_k is the channel impulse response and n_k follow the complex Gaussian distribution. In this paper, we apply linear prediction method to estimate channel coefficient. When neglecting the noise term, the temporary channel estimation $\tilde{h}_k$ can be calculated as:

$$\tilde{h}_k = \hat{\Theta}_k \cdot v_k^* \tag{10}$$

When substituting eq (10) into eq (9), the extended γ_k metric is:

$$\gamma_k(s',s) = P(u_k) \cdot e^{-\frac{E_s}{N_0} \left\| \hat{\Theta}_k - v_k \cdot \sum_{n=1}^{N} P_n \cdot \hat{\Theta}_{k-n} \cdot \tilde{v}_{k-n}^* \right\|^2} \tag{11}$$

In eq (11), N stands for the symbol number evolved currently decoding process. We let $P_n = 1/N$ and the first term of γ_k by eliminating the public terms can be calculated as:

$$P(u_k) = e^{\frac{u_k \cdot L_a(u_k)}{2}}, \; u_k = -1, +1 \tag{12}$$

$L_a(u_k)$ is the input logarithmic likelihood ratio of a priori information. Eventually, the final expression of γ_k metric in DDPSK algorithm is shown in the following:

$$\gamma_k(s',s) = \begin{cases} \dfrac{L_a(u_k)}{2} + \dfrac{E_s}{N_0} \cdot \left| \dfrac{1}{N} \cdot \hat{\Theta}_k \cdot v_k \cdot \sum_{n=1}^{N} \hat{\Theta}_{k-n} \cdot \tilde{v}_{k-n}^* \right| & if \; u_k = +1 \\ -\dfrac{L_a(u_k)}{2} + \dfrac{E_s}{N_0} \cdot \left| \dfrac{1}{N} \cdot \hat{\Theta}_k \cdot v_k \cdot \sum_{n=1}^{N} \hat{\Theta}_{k-n} \cdot \tilde{v}_{k-n}^* \right| & if \; u_k = -1 \end{cases} \tag{13}$$

$\hat{\Theta}$ is passed through the SISO multi-channel DDPSK demodulator along with the prior information $L_a(u_k)$ and demodulator outputs $L_{out}(d_k)$. The extrinsic information can be calculated as:

$$L_{ext}(u_k) = L_{out}(u_k) - L_a(u_k) \tag{14}$$

$L_a(u_k)$ is the input of the prior log likelihood ratio (LLR) information which is an all zero-vector in the first scan and it can be refreshed by the output of channel decoder. We can finally obtain the estimated result of information sequence $\hat{a}_k$ while the iterations reach the maximum. If not, the extrinsic information from convolutional decoder can be calculated as:

$$L_{ext}(u_k) = L_{out}(d_k) - L(d_k) \tag{15}$$

$L_{ext}(u_k)$ feeds back to DDPSK demodulator for a new iteration. After several times, a more accurate decision would be made.

2.3 The joint data and channel estimation

After estimating the original information bit sequence $\hat{a}_k$, the CRC check is performed to verify its correctness. Upon successful verification, the encoded information sequence v_k is regenerated using the encoder. Based on the initial velocity $\hat{v}_0$ obtained from the synchronization preamble, the new reference signal is obtained by re-utilizing the velocity and the information sequence v_k, which is expressed as:

$$\hat{c}_1(t) = c_1\left(t - \frac{\hat{v}_0 t}{c}\right)e^{iv_k}, v_k \in \left\{0, \frac{2\pi}{M}, \cdots, \frac{(M-1)2\pi}{M}\right\} \tag{16}$$

The correlation of the received signal $\hat{\boldsymbol{y}}$ with the reference signal $\hat{c}_1(t)$ defined in Eq.16 yields the output $r_k(\tau)$. By selecting the peak with the highest energy in $r_k(\tau)$, the time delay $\hat{\tau}_k$ of the detected peak is given by:

$$\hat{\tau}_k = \arg\max_{\tau} |r_k(\tau)| \tag{17}$$

3. NUMERICAL SIMULATIONS

In this section, we illustrate the performance of the proposed system over the additive white Gaussian noise (AWGN) channel and time-varying channel. In our system, we set the outer code as R = 1/2 convolutional code with the [133,171] generator, and the length of information bits is 136 and tailed with 6 bits in this simulation. Furthermore, the initial phase of DDBPSK modulation is 0. the parameters of DSSS communication system are that bandwidth is 2.5kHz, the center frequency is 12.5kHz, and the length of spreading code is 127. Without loss of generality, an m-sequence is employed as the spreading code due to its nearly ideal autocorrelation characteristics.

3.1 the performance of BICM-ID method

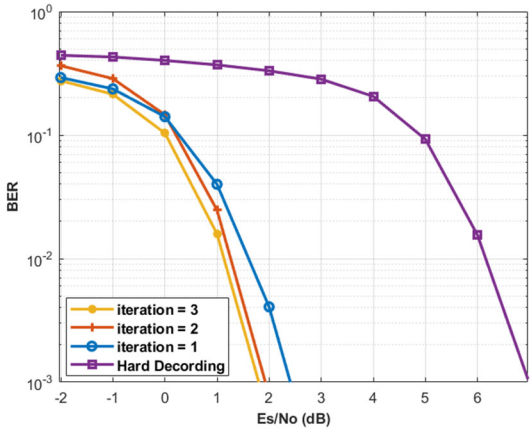


FIG.2. Comparison of decoding performance between BICM-ID method and double-differential hard decoding method under AWGN channel

Under the AWGN channel, the performance of the BICM-ID method is analyzed and compared with the conventional double-differential hard-decision method through simulation. With identical system parameters, the bit error rate (BER) performance of the two receiver schemes is evaluated, as shown in the Fig.2. It can be observed that the BICM-ID approach achieves approximately a 6dB performance gain over the double-

differential hard decoding method. This demonstrates that BICM-ID effectively compensates for the SNR loss typically associated with double-differential schemes.

Through simulation analysis of the performance of the BICM - ID method and the conventional differential hard-decision decoding method in the AWGN channel, it is found that the BICM-ID method shows significant advantages. As can be seen from the BER performance curve in Fig. 2, under the same Es/N0 condition, the BER of the BICM-ID method is approximately 6 dB lower than that of the conventional method. At the same time, as the number of iterations increases, the BER curve of the BICM-ID further decreases, which indicates that iterative decoding can effectively improve the system performance.

3.2 the performance of WSCBF method

The interference suppression performance of the proposed WS-CBF method is analyzed and compared with that of the conventional CBF method. A six-element ultrashort baseline array is selected as the receiving array and the array configuration is illustrated in Figure 2. The array geometry and the azimuth simulation parameters are listed in Table 1.

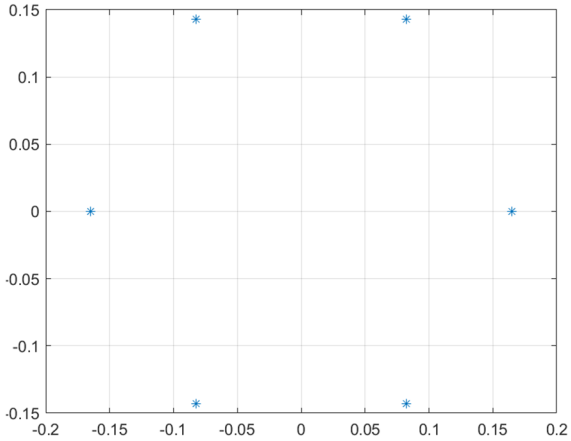
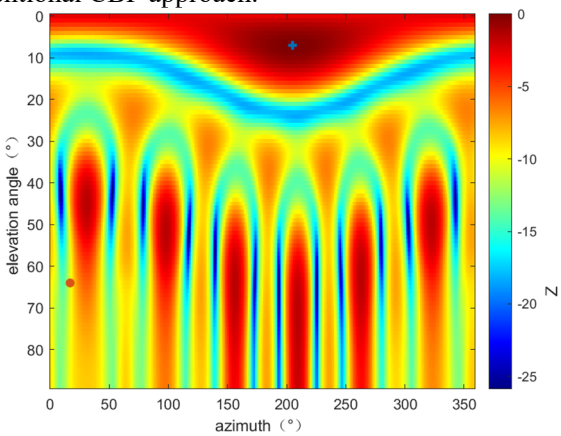


FIG.3. USBL Configuration Diagram of a uniform 6-element array.

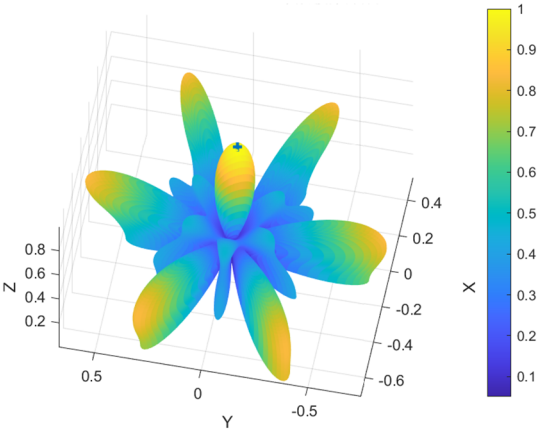
Table.1 Difference in beamforming gain between the two methods at the target and interference angles

Number of elements	6
Inter-element spacing	0.165m（circle radius）
Direction of the target	elevation angle: 7°， azimuth: 205°
Direction of the interferer	elevation angle: 64°， azimuth: 17°

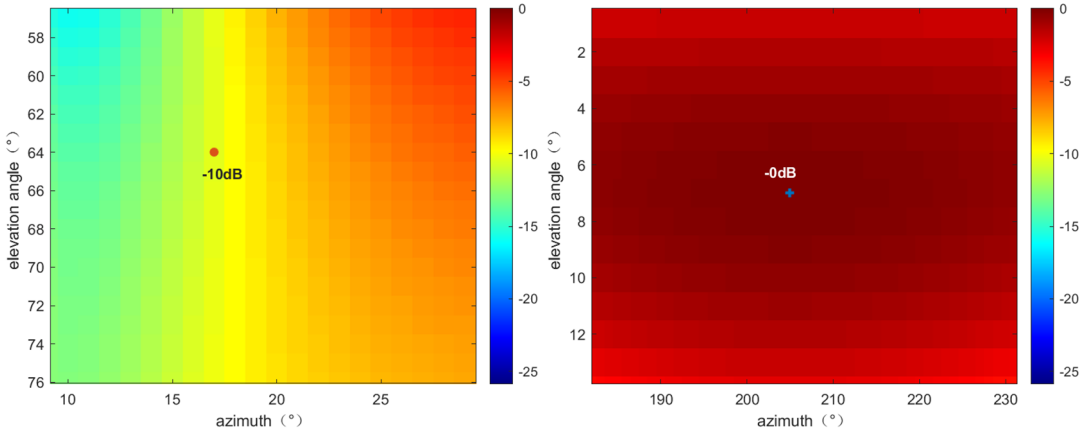
The beam patterns of the two methods, CBF and WS-CBF, are simulated and shown in Fig.4 and Fig.5, respectively. The gains at the target and interference directions are highlighted. A comparison of the results is provided in Table 2. The gain difference between the target-aligned direction and the interference direction for the CBF method is 10 dB, while that for the WS-CBF method is 23.33dB. This indicates that the proposed method achieves an approximate 13dB improvement in interference suppression compared to the conventional CBF approach.



(a) 2D beam pattern

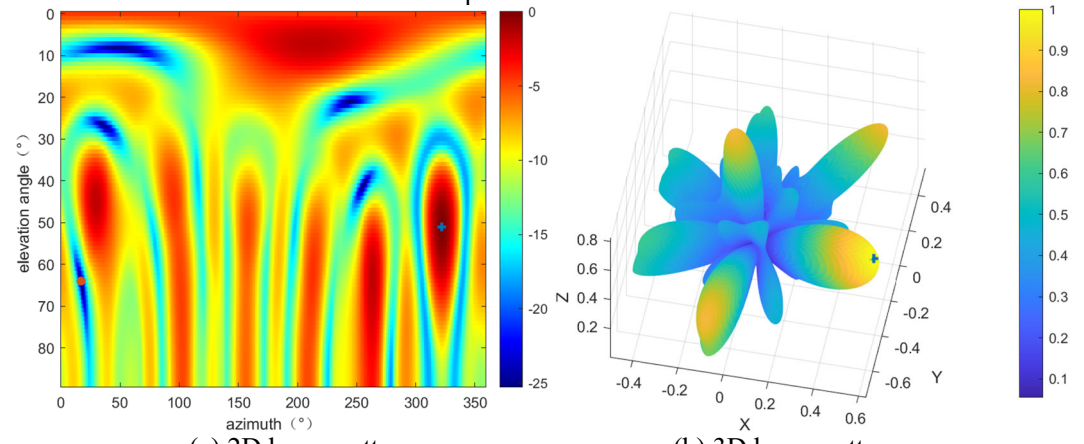


(b) 3D beam pattern

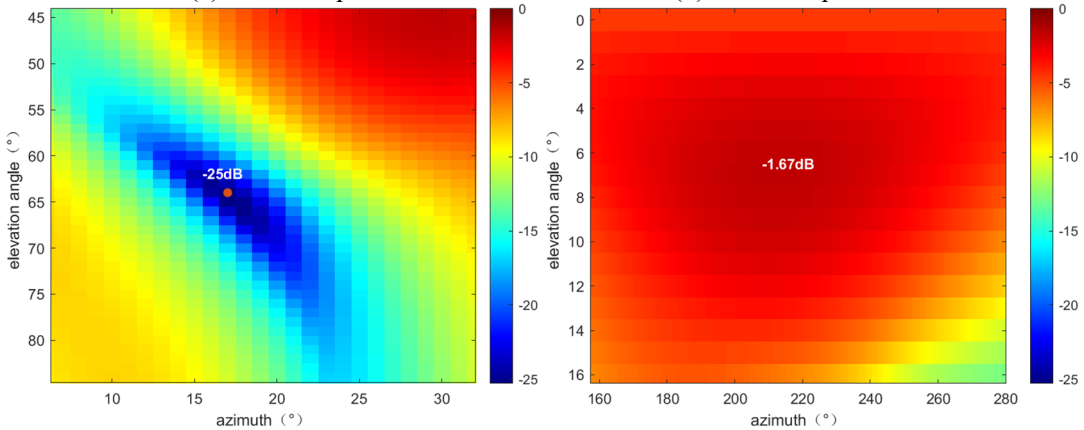


(c) The interference directional gain (d) The target directional gain

FIG.4. Beam pattern of the CBF method.



(a) 2D beam pattern (b) 3D beam pattern



(c) The interference directional gain (d) The target directional gain

FIG.5. Beam pattern of the WS-CBF method.

Table.2 Difference in beamforming gain between the two methods at the target and interference angles

The processing method	Target directional gain (dB)	Interference Directional Gain (dB)	Relative gain (dB)
CBF	0	-10	10
WS-CBF	-1.67	-25	23.33

3.3 the performance of the joint data-and-channel estimation method

A simulation analysis of delay estimation is conducted. The duration of the signal frame is 5.08 seconds. In the AWGN channel and an input SNR of 10 dB, the delay estimation result for one frame is shown in FIG.6. In the conventional method, which does not incorporate joint bit-modulation estimation, the estimated delay is consistently affected by modulation-induced fluctuations. In contrast, the proposed joint data-and-

channel estimation method yields a much smoother and more stable delay estimation.

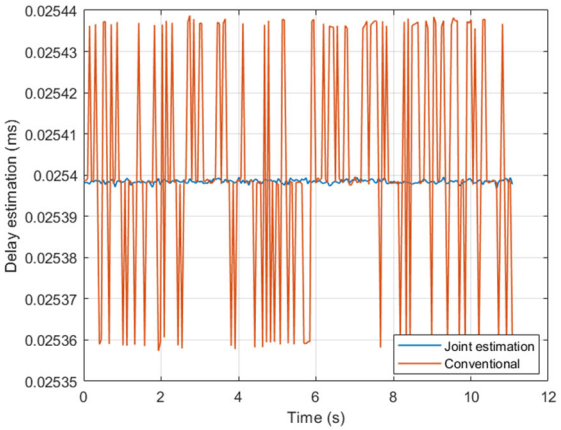


FIG.6. Comparison of results from two Delay estimation methods.

4. EXPERIMENTAL ANALYSIS

In the sea trial data, the signal has a bandwidth of 5 kHz and a center frequency of 12.5 kHz. The spreading code length is 255 chips, and the sampling rate is 200 kHz. For BPSK modulation, the duration of one signal frame is 14.84 seconds. During the experiment, the vessel was moving at a speed of 3 knots, and a total of eight communication frames were collected under continuous sailing conditions. The output SINR results from array processing are shown in FIG.7. It is shown that the WS-CBF method achieves an overall processing gain of 10 dB compared to the CBF method.

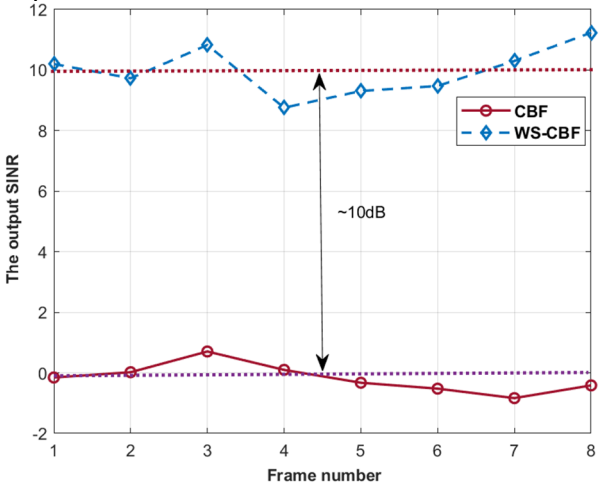
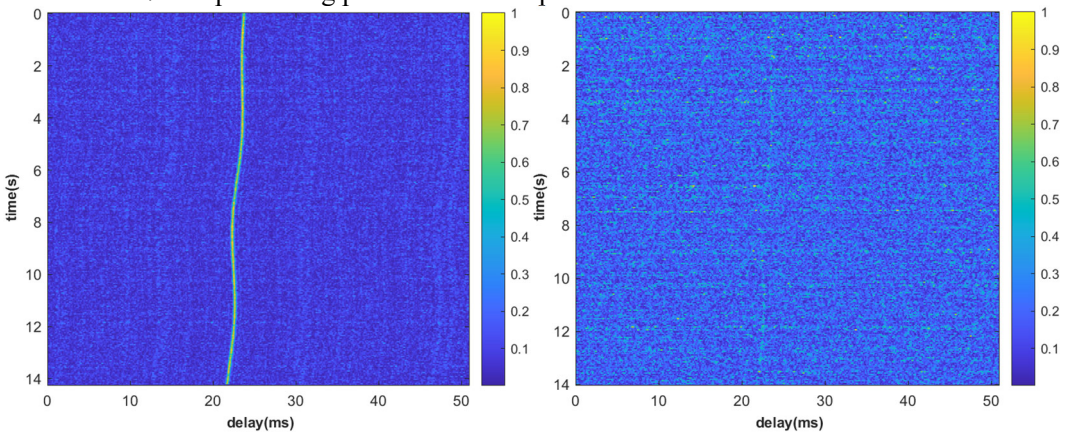


FIG.7 The processing performance comparison of WS-CBF and CBF methods



(a) The CIR by WS-CBF method (b) The CIR by CBF method

FIG.8 The CIR comparison of WS-CBF and CBF methods

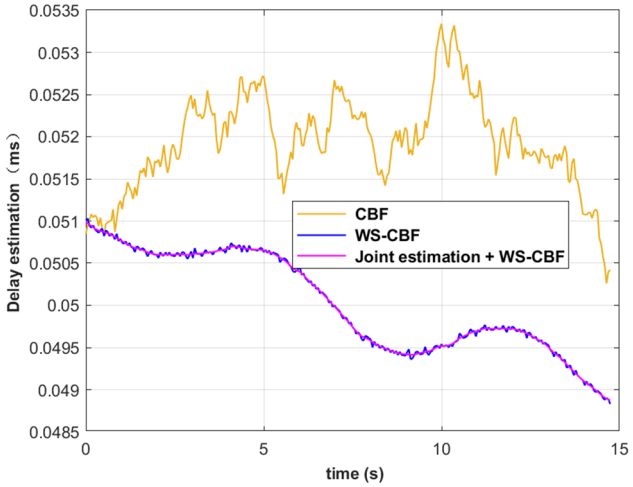


FIG.9 The processing performance comparison of WS-CBF with Joint estimation and CBF methods

Randomly select the signal result of one frame processed by either the WS-CBF or CBF method, and draw the corresponding channel-impulse-response (CIR) diagram as shown in FIG.8. It shows that the channel path of the target processed by the WS-CBF method is more distinct. One frame from the 8 collected frames is selected to analyze the performance of the joint delay estimation. The delay processing result is shown in FIG.9. As observed from the figure, the proposed receiver achieves the best delay estimation performance, followed by the method without joint estimation. The CBF method yields the poorest performance.

5. CONCLUSIONS

In this paper, we propose an iterative UWA DSSS processing algorithm for integrated communication and localization system. First, utilizing a BICM-ID framework to recover the inherent SNR loss from differential encoding. The simulation result shows that the BICM-ID approach achieves approximately a 6dB performance gain over the double-differential hard decoding method. Second, the WS-CBF method is proposed to enhanced interference resistance against strong directional noise. In the case of a uniform six-element array and a specific target direction, the proposed method demonstrates a 13 dB improvement in processing gain compared to the conventional CBF method. Third, A joint data-and-channel estimation scheme is introduced to simultaneously achieve robust communication performance and high-precision positioning capability. Finally, the robustness and effectiveness of the proposed receiver are validated through sea trial communication data collected under strong interference.

ACKNOWLEDGEMENTS

This paper is supported by the National Key R&D Program of China (2023YFC2810300),

REFERENCES

[1] U. Draz et al., “Hybrid Underwater Localization Communication Framework for Blockchain-Enabled IoT Underwater Acoustic Sensor Network, ”*IEEE Internet Things J.*, vol. 12, no. 11, pp. 16858-16885, 1 June1, 2025.

[2] D. Sun, J. Wu, H. Cui, Y. Fu, X. Hong, and C. Zheng, “The line-of-sight peak detection and tracking of underwater acoustic DSSS communications in the doubly spread channel,” Appl. Acoust., vol. 216, p. 109761, 2024.

[3] Y. Han, C. Zheng, and D. Sun, “Signal design for underwater acoustic positioning systems based on orthogonal waveforms,” Ocean Eng., vol. 117, pp. 15–21, 2016.

[4] D. Sun, J. Ding, C. Zheng, and W. Huang, “An underwater acoustic positioning algorithm for compact arrays with arbitrary configuration,” *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 1, pp. 120–130, Mar. 2019.

[5] ROSS M, VAN UFFELEN L. “Characterizing the self noise of a seaglider AUV using a passive acoustic monitor,” in Proc. *OCEANS MTS/IEEE Kobe Techno-Oceans (OTO)*, 2022, pp. 1-5.

[6] M. Wang, C. Zheng, D. Sun, and J. Xu, “Sonar micro-doppler effect induced by ship oscillatory motion in waves: Phenomenon, model, and experimentation,” *JASA Express Letters*, vol. 5, no. 2, p. 026003, Feb. 2025.

[7] B. Sharif, J. Neasham, O. Hinton, and A. Adams, “A computationally efficient doppler compensation system for underwater acoustic communications,” *IEEE J. Ocean. Eng*., vol. 25, no. 1, pp. 52–61, Jan. 2000.

[8] Z. Liu, K. Yoo, T. C. Yang, S. E. Cho, H. C. Song, and D. E. Ensberg,“Long-range double-differentially coded spread-spectrum acoustic communications with a towed array,” *IEEE J. Oceanic Eng.*, vol. 39, no. 3, pp. 482–490, Jul. 2014.

[9] X. Xu, S. Zhou, A. K. Morozov, and J. C. Preisig, “Per-survivor processing for underwater acoustic communications with direct-sequence spread spectrum,” *J. Acoust. Soc. Amer*., vol. 133, no. 5, pp. 2746–2754, May 2013.

[10] D. Sun, J. Wu, X. Hong, C. Liu, H. Cui, and B. Si, “Iterative double differential direct-sequence spread spectrum reception in underwater acoustic channel with time-varying doppler shifts,” *J. Acoust. Soc. Amer.*, vol. 153, no. 2, pp. 1027–1041, Feb. 2023.

[11] X. Zhao, H. Niu, S. Wang and Y. Shi, "A Phase Difference Method for Wind Speed and Direction Measurements Based on An Optimized Adaptive Cross-correlation Algorithm," in Proc.*2024 IEEE/CIC International Conference on Communications in China (ICCC)*, Hangzhou, China, 2024, pp. 1952-1957.

[12] R. Zeng and Y. Wang, "Orthogonal Angle Domain Subspace Projection-Based Receiver Algorithm for Underwater Acoustic Communication," in *IEEE Commun. Lett.*, vol. 22, no. 5, pp. 1102-1105

[13] T. C. Yang, “Temporal resolutions of time-reversal and passive-phase conjugation for underwater acoustic communications,” IEEE J. Ocean. Eng., vol. 28, no. 2, pp. 229–245, Apr. 2003.

[14] H.-u.-R. Khan and M. A. Landolsi, “Modeling and analysis of delay locked loop tracking with quadrature-spread DSSS signals over fading channels,” in Proc. *Int. Conf. Control Decis. Inf. Technol.*, 2016, pp. 360–365.

Nonlinear Composite Chirp Preamble for Enhanced Synchronization in Narrowband Underwater Acoustic Communications

Bogeun SEO[1]; Sangman HAN[1]; Haklim KO[1]; Hojun LEE[1*]
[1] Hoseo University

ABSTRACT
In narrowband underwater communications, conventional synchronization signals employing simple linear chirps suffer from degraded synchronization performance due to low chirp rates that inadequately cope with underwater channel conditions. In this paper, we propose a novel preamble design method that combines various nonlinear chirps to enhance synchronization performance in narrowband underwater communication environments. Through numerical simulations, we compared the synchronization performance of the proposed preamble structure with existing preamble methods, demonstrating that the proposed approach achieves superior synchronization performance.
Keywords: Narrowband underwater communications, Nonlinear chirp, Preamble, Synchronization

1. INTRODUCTION

Narrowband underwater communication is an essential approach due to the highly restricted frequency resources available in underwater environments [1-2]. The acoustic characteristics of underwater channels result in severe signal attenuation at higher frequencies, making communication primarily reliant on very low-frequency, narrowband systems [1-2]. Additionally, efficient utilization of time and frequency resources is crucial in underwater environments to allow multiple users to communicate simultaneously within the same spatial region. However, time-division approaches are challenging to implement reliably due to variations in underwater sound speed and dynamic environmental conditions; thus, frequency-division schemes are predominantly employed [3-4]. Due to technical limitations of underwater communication equipment, the overall available bandwidth is highly constrained, necessitating extremely narrowband communication techniques to support multiple users sharing the limited frequency resources.

Consequently, designing robust synchronization signals becomes critically important in narrowband underwater communication environments characterized by restricted frequency resources and complex channel conditions. Particularly, conventional synchronization signals using simple linear chirps suffer from poor synchronization performance, as their inherently low chirp rates are insufficient to overcome the Doppler effects and multipath propagation commonly encountered in practical underwater scenarios [3-4].

In this paper, we propose a robust synchronization signal design approach that combines various nonlinear chirp signals. To increase the effective chirp rate, the total duration of the synchronization signal is divided into multiple time slots, each allocated chirp waveforms determined by pseudo-noise (PN) sequences, which dictate both chirp directions (up/down) and shapes (convex/concave). Numerical simulations comparing the proposed nonlinear composite chirp preamble with existing methods demonstrate that the proposed technique achieves superior synchronization performance even under Doppler effects and multipath propagation conditions.

2. PROPOSED METHOD

The synchronization signal proposed in this paper for narrowband underwater communications divides the total synchronization signal duration T into N_c equal-length time slots, and each slot is constructed by combining various nonlinear chirp signals based on two pseudo-noise (PN) sequences.

Initially, two independent PN sequences, $\boldsymbol{P} = [p_1, \ \ldots, p_n, \ldots, p_{N_c}]$ and $\boldsymbol{Q} = [q_1, \ \ldots, q_n, \ldots, q_{N_c}]$, each of length N_c and consisting of binary values (0 or 1), are generated. These PN sequences determine the chirp waveform allocated to each time slot. The instantaneous frequency $f_n(t)$ of the chirp signal assigned to the

* corresponding author: hjlee@hoseo.edu

n-th slot is defined as follows:

$$f_n(t) = f_{start}^n + (f_{end}^n - f_{start}^n)\left(\frac{t}{T_s}\right)^{\beta_n}, \quad \text{for } 0 \le t \le T_s \tag{1}$$

where the duration of each time slot is $T_s = T/N_c$, and the chirp direction (related to f_{start}^n and f_{end}^n) is determined by the PN sequence $\boldsymbol{P}$ as follows:

$$f_{start}^n = \begin{cases} f_0, & \text{for } p_n = 1 \\ f_1, & \text{for } p_n = 0, \end{cases} \tag{2}$$

$$f_{end}^n = \begin{cases} f_1, & \text{for } p_n = 1 \\ f_0, & \text{for } p_n = 0, \end{cases} \tag{3}$$

In equations (2) and (3), $f_0 = f_c - B/2$ and $f_1 = f_c + B/2$ where f_c and B denote the center frequency and bandwidth of the narrowband preamble signal. Additionally, the chirp curvature (β_n) is determined by the PN sequence $\boldsymbol{Q}$, resulting in either a convex ($\beta_n < 1$ where $q_n = 0$) or concave ($\beta_n > 1$ where $q_n = 1$) waveform shape.

Finally, the chirp signal generated in the n-th slot, $s_n(t)$, is expressed as follows:

$$s_n(t) = \sqrt{2}\cos\left(2\pi \int_0^t f_n(\tau)\, d\tau\right), \quad \text{for } 0 \le t \le T_s, \tag{4}$$

Using the above formulation, chirp signals corresponding to the Nc time slots are generated and concatenated to form the complete preamble. At the receiver, the preamble is detected and its synchronization timing is estimated through cross-correlation with the received signal.

3. COMPUTER SIMULATIONS

To evaluate the synchronization performance of the proposed preamble structure, computer simulations were conducted. Three comparison methods were used: (1) a conventional linear chirp with duration T and bandwidth B; (2) a nonlinear chirp with the same duration and bandwidth, using a curvature factor β; and (3) a structure dividing the duration T into N_c time slots, with each slot assigned either an up- or down-linear chirp according to a PN sequence. The spectrograms of these three comparison methods and proposed method are shown in Figure 1, and the detailed simulation parameters are summarized in Table 1.

Table 1 – Simulation parameters

Parameters	Values
Bandwidth, B	200 Hz
Curvature factor, β	2
The number of slots, N_c	6
Center Frequency, f_c	32 kHz
Sampling rate, f_s	200 kHz
Doppler Shift	10 Hz
Doppler Spread	5 Hz

The underwater acoustic channel used in the simulation was generated using the Bellhop ray tracing model based on a measured vertical sound speed profile from the East Sea of Korea. The corresponding underwater acoustic channel model is illustrated in Figure 2.

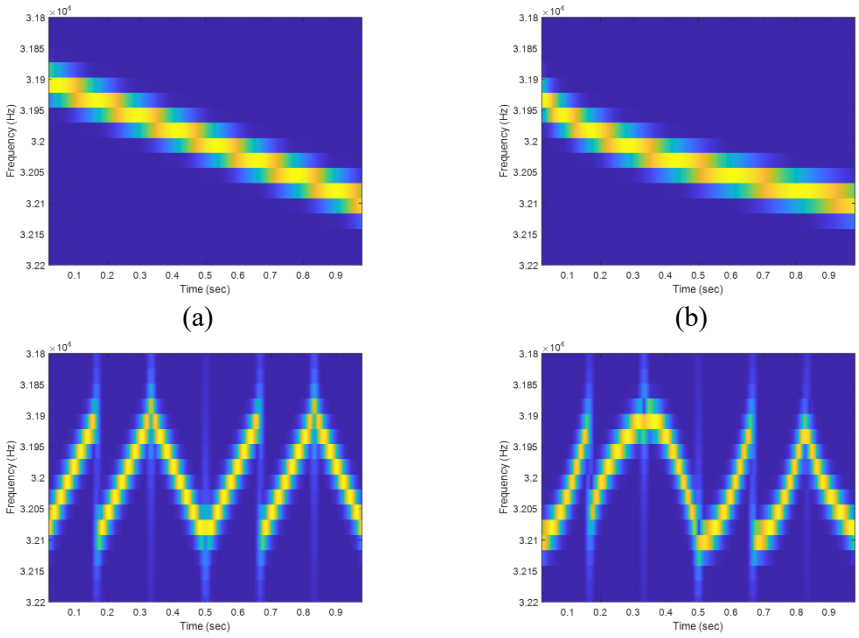


Figure 1 – The spectrograms of comparision and proposed preambles: (a) linear chirp; (b) nonlinear chirip; (c) up/down chirp sequence; (c) proposed method

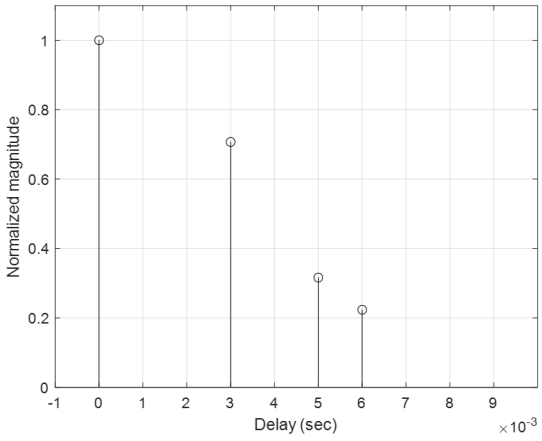


Figure 2 – Channel impulse response.

The channel was modeled as a Rician fading channel with a K-factor of 10 dB. Doppler shift and spread were set to 10 Hz and 5 Hz, respectively.

The synchronization performance of the proposed method and the comparison methods was evaluated in terms of the root mean square error (RMSE) of the estimated synchronization timing, as shown in Figure 3. In this figure, the black circle marker represents the conventional linear chirp, the blue triangle marker denotes the nonlinear chirp with curvature, the purple square marker corresponds to the multi-slot linear up/down chirp method, and the red hexagon marker indicates the proposed method. As shown in the Figure 3, the proposed method exhibits larger RMSE than the other methods when the signal-to-noise power ratio (SNR) is below 0 dB. However, as the SNR increases (above 3 db), the proposed method outperforms all other techniques.

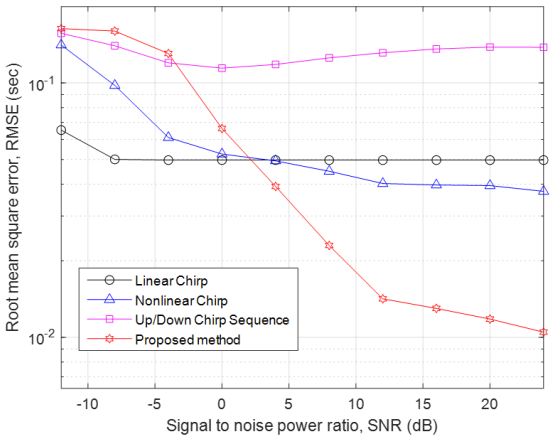


Figure 3 – RMSE comparisons.

4. CONCLUSIONS

This paper proposed the PN-sequence-based nonlinear composite chirp preamble for synchronization in narrowband underwater acoustic communications. By combining various chirp directions and curvatures across time slots, the proposed method improves synchronization robustness under multipath and Doppler conditions. Simulation results confirmed that the proposed approach outperforms conventional linear and nonlinear chirp-based methods, demonstrating its effectiveness for bandwidth-constrained underwater environments.

ACKNOWLEDGEMENTS

This work was supported by the Korea Research Institute for defense Technology planning and advancement (KRIT) through the Domestic Development Support Program for Weapon System Components.

REFERENCES

1. F. J. Cañete, J. López-Fernández, C. García-Corrales, A. Sánchez, E. Robles, F. J. Rodrigo, and J. F. Paris, "Measurement and modeling of narrowband channels for ultrasonic underwater communications," Sensors, vol. 16, no. 2, pp. 256, Feb. 2016.
2. Z. Ying, L. Huanyu, L. Bingyu, L. Li, W. Hao, and R. Peng, “An underwater acoustic semantic communication approach to underwater image transmission,” Intell. Mar. Technol. Syst, vol. 3, no. 5, Feb. 2025.
3. W. Xie, E. Zhang, L. You, D. Wang, Z. Wang and F. Liqun, "Combating Multi-Path Interference to Improve Chirp-Based Underwater Acoustic Communication," ICC 2024 - IEEE International Conference on Communications, Denver, CO, USA, 2024, pp. 5377-5382.
4. C.V. María, D.T. Dídac, R.R. Ignacio, U.R. Jorge, and A.C. Rafael, "Signal Processing to Characterize and Evaluate Nonlinear Acoustic Signals Applied to Underwater Communications," Electronics, vol. 13, no. 21, pp. 4192, Oct. 2024

DAILY PROGRAM

AUGUST 21 (THU)

09:00 – 10:40	Ultrasound and Acoustic Materials for Performance Enhancement and Structural Monitoring 325 (3F)
Chair	Ivan Smirnov (Oxford Suzhou Centre for Advanced Research)
09:00	Tesla Valve-Inspired Metamaterial Design for Enhanced Acoustic Performance in Load-Bearing Mortar
	Kebede Alemayehu Moges, Youngbeom Beak, Sungwoo Park, Sukhoon Pyo
09:20	Complementary metamaterial for ultrasound transmission through elastic barrier
	Ki Yong Lee, Wonju Jeon
09:40	Simulation of ultrasonic reactors with different frequencies for processing flowing liquid media
	Ivan Smirnov, Mohan Zhang, James Kwan
10:00	Investigation of the Impact of Surface Holes on the Sound Absorption Performance of Rubber Coatings
	Bo Hu, Xingchao Yu, Jintong Gao, Sijia Wang, Bowen Shi, Haoyang Zhang
10:20	Characterization and Early Detection of Hydrogen Embrittlement in Offshore Bolts using Acoustic Emission
	Nokhaiz Sabir

Tesla Valve-Inspired Metamaterial Design for Enhanced Acoustic Performance in Load-Bearing Mortar

Alemayehu Moges KEBEDE[1]; Nazhiefah DALILA[1], Youngbeom BEAK[2], Sungwoo PARK[2]*, Sukhoon PYO[1*]

[1]Department of Civil, Urban, Earth, and Environmental Engineering, Ulsan National Institute of Science and Technology (UNIST), Ulsan 44919, South Korea

[2]School of Architecture and Building Science, Chung-Ang University, Seoul, South Korea

ABSTRACT

Achieving efficient low-frequency sound absorption in compact, load-bearing structures remains a critical challenge in architectural acoustics. This work introduces a hybrid acoustic metamaterial, inspired by Tesla valve geometry and embedded Helmholtz resonators, 3D-printed and integrated within high-strength mortar. The design leverages space-coiling pathways for subwavelength absorption (250–450 Hz) and resonator-tuned attenuation for higher frequencies (1200–2500 Hz), achieving broadband performance in a structurally robust form factor. Fabricated using PLA and cast into mortar blocks, the system attained a compressive strength of 72 MPa while achieving absorption coefficients >0.9 and up to 129% improvement over conventional perforated mortar. Experimental measurements and COMSOL simulations exhibited excellent agreement, confirming the effectiveness of geometric tuning. This integrated approach overcomes the thickness–performance trade-off and enables next-generation cementitious composites for noise mitigation in buildings, transportation infrastructure, and urban environments.

Keywords: Acoustic metamaterials, Sound absorption, Tesla valve-inspired design, high-strength mortar, wave scattering, energy dissipation, coiled structure

1. INTRODUCTION

Low-frequency environmental noise (below 500 Hz), often caused by urban traffic, industrial machines, and HVAC systems, poses a significant challenge in building acoustics. Traditional solutions like porous foams, mineral wool, and micro-perforated panels offer limited effectiveness due to their bulk, fragility, or incompatibility with structural roles [1][2][3]. In contrast, cement-based materials provide load-bearing capacity but suffer from poor acoustic absorption, particularly in the low-frequency range. Surface-perforated mortar (SPM) represents one advance by introducing geometrically tuned voids that function like Helmholtz resonators [4][5]. However, the cylindrical uniformity of SPM cavities limits their ability to achieve broadband absorption or subwavelength performance. This gap has motivated recent research into acoustic metamaterials, engineered geometries that induce wave manipulation through internal structure rather than bulk material properties. Inspired by the Tesla valve, a unidirectional fluidic diode, we propose a Tesla valve-inspired acoustic metamaterial cast directly into high-strength mortar. Its asymmetric, tortuous geometry induces complex wave paths, increasing effective acoustic length without altering thickness. By tuning internal angles and integrating resonators, we demonstrate a scalable system capable of absorbing sound across low to mid frequencies while maintaining compressive strength of 72 MPa. Unlike conventional resonators limited by geometry or frequency bandwidth, our approach integrates multiple dissipation and resonance mechanisms into a single, load-bearing structure.

2. METHODOLOGY

Tesla valve-inspired units with internal angles of 30°, 45°, 60°, and 90° were designed using CAD software.

1* shpyo@unist.ac.kr
2* spark22@cau.ac.kr

These structures feature multiple internal baffles that create elongated, space-coiled channels, thereby increasing the acoustic path length.

Each unit was 3D printed with PLA (polylactic acid) at 100% infill to ensure structural rigidity. For the DET-90 design, Helmholtz resonators were integrated, utilizing neck diameters of 2–4 mm and cavity depths of 20 mm and 40 mm. The geometries were tailored for subwavelength interactions, achieving high tortuosity while maintaining a compact footprint. Minor surface roughness from the printing process was considered acceptable and may have enhanced viscous dissipation.

The mortar matrix followed ASTM C305 [6], using a water-to-binder ratio of 0.35 and a sand-to-cement ratio of 1. A polycarboxylate superplasticizer (1% by binder weight) was included to ensure sufficient workability. The mix consisted of cement, silica sand, and silica fume. Compressive strength was tested on 50 mm cubes after steam curing at 90°C for 48 hours to ensure the matrix could support the embedded metamaterials without compromising mechanical integrity.

PLA metamaterial inserts were embedded into 300 mm × 300 mm × 150 mm mortar panels during casting, with their front faces left exposed to replicate surface acoustic treatments. Only the front-facing apertures were accessible to incident sound waves, while the remaining portions were encased in structural mortar. Samples underwent 24 hours of ambient curing, followed by accelerated curing. Alignment and bonding were confirmed visually and dimensionally.

Acoustic performance was evaluated using the in-situ impedance method (ISO 13472-1) [7] with a PU probe over a frequency range of 200–2500 Hz, as shown in Figure 1. The surface absorption area ratio (SAAR) was calculated to assess broadband performance. Numerical simulations were conducted in COMSOL Multiphysics using the Thermoviscous Acoustics and Poroacoustics modules. The mortar was modeled using the Johnson-Champoux-Allard (JCA) framework with parameters: porosity (φ = 37%), flow resistivity (σ = 8300 Pa·s/m²), and tortuosity ($\alpha\infty$ = 2.13).

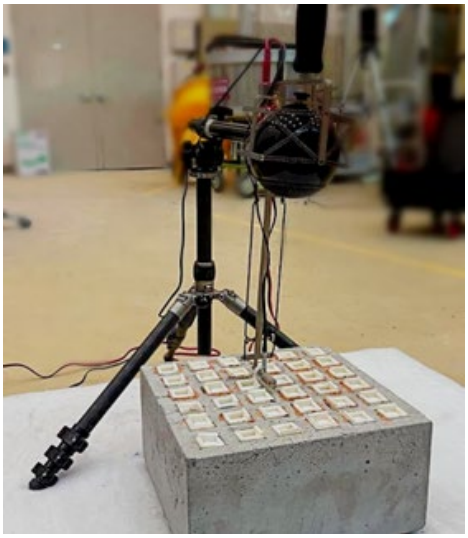


Figure 1 – In situ sound absorption

Table 1 – Summary of acoustic and mechanical performance for selected configurations

Configuration	Angle (°)	Resonator	Peak SAC (Hz)	SAAR
DET-REF	–	No	1100	0.17
DET-30	30	No	250, 832	0.35
DET-45	45	No	250, 650	0.37
DET-60	60	No	250, 515	0.39
DET-90	90	No	250, 450	0.34

DET-90-AG20	90	Yes	250, 450, and 1250	0.30
DET-90-AG40	90	Yes	250, 450, and 1750	0.32

3. RESULTS AND DISCUSSION

3.1 Low-Frequency Absorption

As the valve angle increased, the acoustic path lengthened without adding thickness, resulting in a shift of the resonance frequency toward lower values (see Figure 2). Specifically, DET-30 exhibited a peak absorption at 832 Hz, while DET-90 showed a resonance peak at 450 Hz. The surface absorption area ratio (SAAR) also improved with increasing angle, reaching a maximum of 0.39 for DET-60, which achieved an optimal balance between acoustic tortuosity and impedance matching. Larger valve angles increased tortuosity, slowing wave propagation and extending the effective path length, key factors in shifting resonance to lower frequencies [8]. However, in highly tortuous designs like DET-90, excessive bending can cause internal reflections and diminish overall absorption.

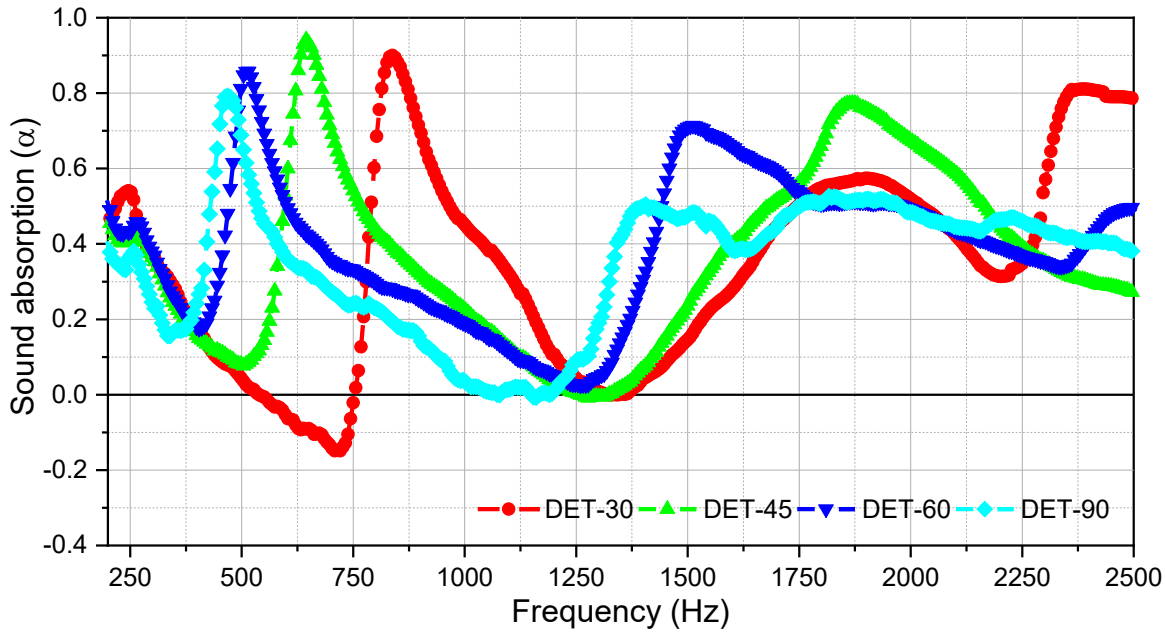


Figure 2 – Sound absorption coefficient (SAC) spectra of Tesla valve-inspired geometries. Increasing internal valve angle shifts resonance to lower frequencies, demonstrating tunable subwavelength absorption via geometrical design.

Thermoviscous boundary effects became more pronounced in the narrower and more complex channels. Viscous damping rose from 3419.6 Np/m at 30° to 3822.3 Np/m at 90° [9][10], while flow resistivity increased from 1.81×10^6 to 3.68×10^6 Pa·s/m [11], enhancing impedance continuity with air (see Figure 3, Table 2). These increases in damping and resistance indicate stronger wave-surface interactions and greater energy loss due to friction. Nevertheless, overly high flow resistance at larger angles may hinder wave penetration, highlighting a trade-off between maximizing absorption and maintaining adequate wave transmission.

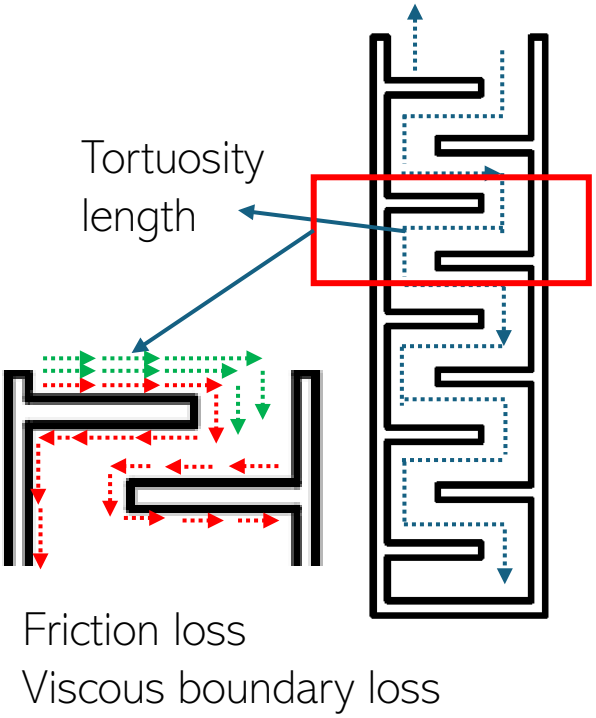


Figure 3 – Visualization of Tesla valve-inspired channels, higher angles increase acoustic path tortuosity, enhancing low-frequency absorption through resonance shifting and viscothermal loss mechanisms.

Table 2 – Governing mechanisms and geometric trends in Tesla valve-inspired acoustic metamaterials.

Mechanism	Key Formula	Affected by Geometry	Trend with Angle ↑
Resonance Shift	$f_1 \approx \frac{c}{4L_{eff}}$	L_{eff} ↑	f ↓
Viscous Loss	$\alpha_v = \frac{1}{a}\sqrt{\frac{\eta\omega^2}{\rho_0}}$	$a = D/2$ ↓	α_v ↑
Thermal Loss	$\alpha_t = \alpha_v \cdot (\gamma - 1)\sqrt{Pr}$	a ↓	α_t ↑
Flow Resistance	$R = \frac{\rho_0 cktDL}{2S^2}\left[1 + (\gamma - 1)\sqrt{Pr}\right]$	D ↓, L ↑	R ↑
Tortuosity	$\tau = \left(\frac{L_{pore}}{L_{straight}}\right)^2$	Path complexity ↑	τ ↑

3.3. Helmholtz Resonators for Mid-Frequency Enhancement

DET-90-AG20 and AG40 introduced resonant peaks at 1250 Hz and 1750 Hz, respectively. While these additions slightly lowered overall SAAR, they broadened the absorption bandwidth. The resonators worked as expected, producing localized pressure buildup and oscillatory airflow. The Q-factor of these resonators was controlled via neck diameter and cavity depth [12]. Integration into the Tesla valve frame enabled dual-resonance functionality without additional thickness.

3.4. Hybrid Combinations and Synergistic Effects

Combining geometries (e.g., DET-30/45) yielded constructive interference across broader ranges. DET-30/45 achieved SAAR = 0.412. Over-complex combinations like DET-30/45/60 showed diminished

performance, likely due to destructive phase interference. This indicates an optimal balance between diversity of path lengths and coherence of resonant modes. Distributed resonators can cover more frequencies, but spatial and phase interactions must be managed.

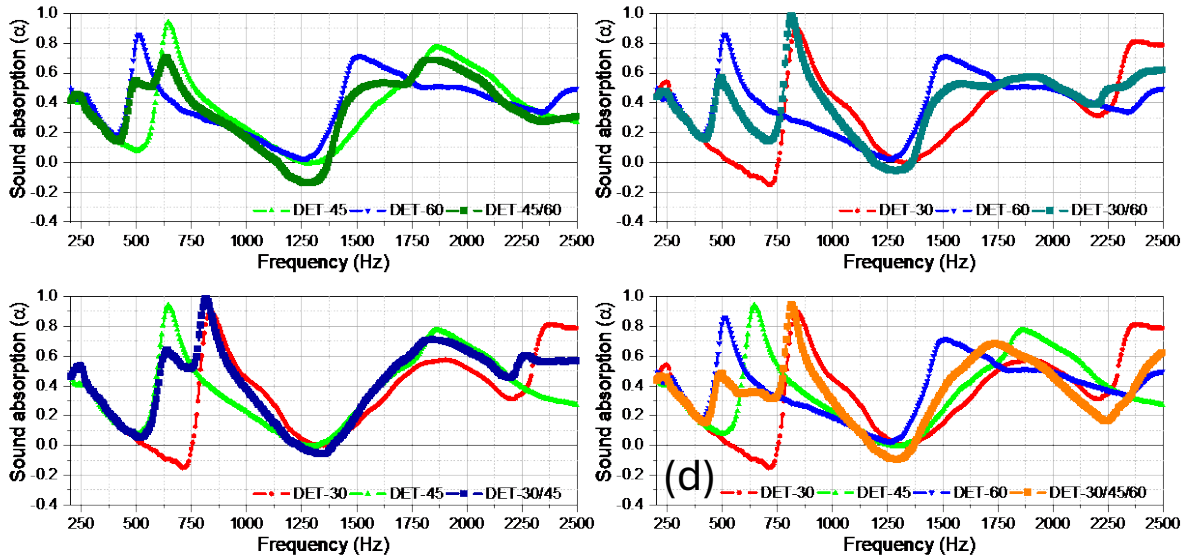


Figure 4 – Sound absorption performance of combined Tesla valves (a) DET-45/60, (b) DET-30/60, (c) DET-30/45, and (d) DET-30/45/60

3.5. COMSOL Simulation Validation

Simulated SAC curves closely matched experimental results (±5%). Pressure field plots confirmed standing waves in Tesla geometries and local cavity resonance in Helmholtz modules. The simulations reinforced the validity of geometric tuning and dissipative mechanisms. Modeling also showed strong agreement in phase and amplitude patterns, validating the use of JCA parameters for cement-based composites.

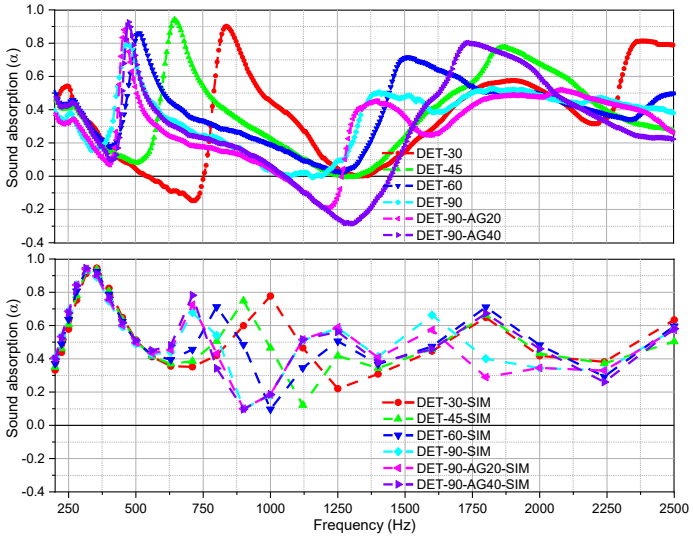


Figure 5 – Experimental and numerical analysis of acoustic response in Tesla valve and Helmholtz resonator-based metamaterials: (a) Experimental sound absorption coefficient (SAC) for Tesla valve geometries, (b) Simulated SAC results showing corresponding resonance behavior.

4. Conclusion

This study demonstrates a novel, structurally integrated acoustic metamaterial design using Tesla valve geometries. By adjusting internal angles and integrating resonators, subwavelength low-frequency absorption was achieved without compromising strength. Hybrid designs extended functionality into mid-frequency ranges. The mortar-based system achieved SAC >0.8 at 450 Hz and 72 MPa strength, suitable for load-bearing applications. This approach decouples structural and acoustic design constraints, enabling sound-absorbing infrastructure with tunable design.

Future work will explore multi-material printing, machine learning-based geometry optimization, and integration into precast panels and façade elements. The ability to embed tunable resonators within strong, castable matrices offers potential for sustainable, high-performance building components across architectural, industrial, and environmental domains.

Acknowledgment

Supported by the National Research Foundation of Korea (No. RS-2023-00212366).

References

[1] Y.-S. Wong, V. Sekar, S. Y. Eh Noum, and S. Sivanesan, "Effect of Thickness and Perforation Size on the Acoustic Absorption Performance of a Micro-Perforated Panel," *MATEC Web Conf.*, vol. 335, p. 03016, 2021, doi: 10.1051/matecconf/202133503016.

[2] E. Gliścińska and M. Michalak, "Composite acoustic system for low-frequency sound absorption," *Sci. Rep.*, vol. 15, no. 1, pp. 1–15, 2025, doi: 10.1038/s41598-025-89151-5.

[3] Z. Li, X. Li, Z. Wang, and W. Zhai, "Multifunctional sound-absorbing and mechanical metamaterials via a decoupled mechanism design approach," *Mater. Horizons*, vol. 65, no. 5, pp. 75–87, 2022, doi: 10.1039/d2mh00977c.

[4] K. A. Moges, S. Park, and S. Pyo, "Sound absorption characteristics of surface perforated mortar with micro-sized pores," *Appl. Acoust.*, vol. 195, p. 108811, Jun. 2022, doi: 10.1016/J.APACOUST.2022.108811.

[5] S. Park, K. A. Moges, and S. Pyo, "Experimental study on the sound absorption performance of surface-perforated mortar," *Constr. Build. Mater.*, vol. 307, p. 124824, Nov. 2021, doi: 10.1016/J.CONBUILDMAT.2021.124824.

[6] ASTM C305 and ASTM C305-20, "Standard Practice for Mechanical Mixing of Hydraulic Cement Pastes and Mortars of Plastic Consistency," *Annu. B. ASTM Stand.*, vol. 14, no. 2, p. 147, 1998, doi: 10.1177/089033449801400218.

[7] ISO and 13472-1, "Acoustics — Measurement of sound absorption properties of road surfaces in situ — Part 1: Extended surface method," vol. 2004, 2004.

[8] Y. jun Xin *et al.*, "Labyrinthine acoustic metamaterials with triangular self-similarity for low-frequency sound insulation at deep subwavelength dimensions," *Results Phys.*, vol. 54, Nov. 2023, doi: 10.1016/j.rinp.2023.107151.

[9] Y. Champoux and J. F. Allard, "Dynamic tortuosity and bulk modulus in air-saturated porous media," *J. Appl. Phys.*, vol. 70, no. 4, pp. 1975–1979, 1991, doi: 10.1063/1.349482.

[10] M. R. Stinson, "The propagation of plane sound waves in narrow and wide circular tubes, and generalization to uniform tubes of arbitrary cross-sectional shape," *J. Acoust. Soc. Am.*, vol. 89, no. 2, pp. 550–558, 1991, doi: 10.1121/1.400379.

[11] J.F.Allard and N.Atalla, *Propagation of Soundin Porous Media:Modelling Sound Absorbing Materials*. Andover. England: A John Wiley and Sons, Ltd, 2009.

[12] H. Duan *et al.*, "Acoustic Metamaterials for Low-Frequency Noise Reduction Based on Parallel Connection of Multiple Spiral Chambers," *Materials (Basel).*, vol. 15, no. 11, 2022, doi: 10.3390/ma15113882.

Complementary metamaterial for ultrasonic transmission through elastic barrier

Ki Yong Lee[1]; Wonju Jeon[1];

Korea Advanced Institute of Science and Technology, Republic of Korea

ABSTRACT

We introduce a complementary metamaterial characterized by triple negative effective properties of mass density, bulk modulus, and shear modulus, designed to facilitate wave transmission through elastic barriers. The transformation method with affine mapping is adopted for deriving theoretical condition for elastic complementary media, revealing that triple negativity is essential for canceling out the physical space of blocking medium. To attain triple negativity, the metamaterial features multiple inclusions systematically arranged in a base matrix, inducing dipolar, monopolar, and quadrupolar resonances through the collective motion of inclusions and matrix. Time-domain simulations in an in-house digital twin demonstrate that the metamaterial, when attached to the incident side of an elastic barrier, remarkably enhances energy transmission of Gaussian pulses. The proposed metamaterial holds broad potential for diverse applications, including antireflection layers, non-destructive testing, underwater communication, and transcranial ultrasonic neuroimaging.

Keywords: Complementary metamaterial; Multiple resonances; Triple-negative effective properties; Elastic barrier; Ultrasonic energy transmission;

[1] gf@aaa.com

PROCEEDINGS of the
16th International Conference on Theoretical and Computational Acoustics

August 18 to 22, 2025 BEXCO, Busan Korea

Simulation of ultrasonic reactors with different frequencies for processing flowing liquid media

Ivan Smirnov[1], Mohan Zhang[2], James Kwan[3]

[1]Oxford Suzhou Centre for Advanced Research
[2]Oxford Suzhou Centre for Advanced Research
[3]Department of Engineering Science, University of Oxford

ABSTRACT

Given the importance of protecting the environment and human health, the treatment of industrial waste, including metalworking fluids (MWF) or wastewater, is being given high priority. Ultrasound technology offers an innovative means of overcoming the limitations of traditional methods for water treatment: improved process efficiency, reduced energy consumption, and lower generation of secondary pollutants make it a promising alternative in the field of water treatment. In this paper, we consider a type of universal reactor for processing flowing liquid using focused ultrasound. With the effectiveness in generating cavitation as the central principle, we perform computer simulations of the reactor, focusing on the frequency and power of ultrasound, in order to screen out the optimal structural parameters of the reactor as well as the matching ultrasound parameters. Based on the simulation results, we formulated the basic principles for operation of a fluid treatment reactor and designed a model on the basis of which a device for experiments was manufactured. This prototype was used for experimental validation of the reactor installation with various input condition parameters in order to determine the optimal parameters of the treatment mode.

[1] ivan.smirnov@oxford-oscar.cn
[2] mohan.zhang@oxford-oscar.cn
[3] james.kwan@balliol.ox.ac.uk

Investigation of the Impact of Surface Holes on the Sound Absorption Performance of Rubber Coatings

Bo Hu [1]; Xingchao Yu [2]; Jintong Gao [3]; Sijia Wang [4]; Bowen Shi [5]; Haoyang Zhang [6]

[1,2,3,4,5,6] Harbin Engineering University , China

ABSTRACT

In the design and application of underwater sound-absorbing coatings, introducing surface holes and embedding sensors is an important technical approach to enhancing their performance and expanding their applications. Therefore, understanding the impact of different hole parameters on the acoustic performance of these coatings is of great significance. To address this issue, this paper establishes a finite element model of a sound-absorbing coating with surface holes. The effects of various hole parameters on the acoustic performance of large rubber-coated samples are investigated by measuring the sound absorption coefficient and the average sound absorption coefficient. The results indicate that, compared to hole depth, increasing the hole area has a smaller impact on the sound absorption coefficient. Furthermore, the negative influence on the coating's absorption ability can be effectively minimized by positioning the larger end of the conical holes closer to the incident wave.

PROCEEDINGS of the
16th International Conference on Theoretical and Computational Acoustics

August 18 to 22, 2025 BEXCO, Busan Korea

Characterization and Early Detection of Hydrogen Embrittlement in Offshore Bolts using Acoustic Emission

Nokhaiz Sabir[1]
[1]University of Warwick

ABSTRACT

Hydrogen can cause corrosion and cracking in steel, which leads to costly failures in industries like offshore energy and oil and gas. Large bolts and connectors are especially vulnerable, and their failure can result in severe accidents, including the 2012 Gulf of Mexico incident that leaked 432 barrels of drilling fluid. Hydrogen-Induced Cracking (HIC) in bolts is hard to detect and can lead to significant dangers.

Acoustic Emission (AE) technology detects the release of strain energy during damage events like corrosion or cracking, helping identify and locate damage. By monitoring AE over time, it is possible to assess the condition of a structure and estimate its remaining useful life. There is a strong interest in using AE technology to understand how hydrogen embrittlement affects critical bolts and connectors.

The project aims to create a method for using AE to monitor bolts for hydrogen embrittlement in real service conditions. This includes designing an experimental setup to determine the best parameters for detecting AE from hydrogen embrittlement and selecting suitable sensors for effective monitoring.

DAILY PROGRAM

AUGUST 21 (THU)

14:20 – 15:40	AI and Acoustic Sensing Applications in Industry, Ecology, and Medicine 323 (3F)
Chair	Jae Hak Jeong (Korea National University of Transportation)
14:20	Artificial Intelligence-Powered Acoustic Fingerprint-Based DeepScan for Product Quality Control Testing
	Lingguang Chen, Nicholas Kalvaitis, Sean F. Wu
14:40	Real-Time Smart Buoy System for Cetacean Monitoring
	Ching-Tang Hung, Wei-Lun Li, Meng-Fan Tsai, Wei-Chun Hu, Chi-Fang Chen
15:00	Reproduction and Analysis of Pulse Wave Propagation and Velocity (PWV) in the
	Jae-Hak Jeong
15:20	Model refinements to better estimate ship-generated noise impact on endangered
	Maximilian Lauch

Artificial Intelligence-Powered Acoustic Fingerprint-Based DeepScan for Product Quality Control Testing

Lingguang Chen[1]; Nicholas Kalvaitis[2]; and Sean F. Wu[3]

ABSTRACT

This paper introduces a novel quality control (QC) methodology: the Artificial Intelligence-Powered Acoustic Fingerprint-Based DeepScan (AIAF DeepScan) system for real-time, non-contact, and non-invasive product inspection in manufacturing environments. Every mechanical or electromechanical system emits sound during operation, carrying unique acoustic fingerprints that reflect its internal structure, dynamic behavior, and material properties. AIAF DeepScan leverages these signatures by extracting acoustic features from measured signals and comparing them against pre-established databases to automatically, accurately, and reliably differentiate acceptable (GO) from defective (NG) products. Furthermore, when trained with known fault modes, the system can identify root causes of defects. The AIAF DeepScan workflow consists of three core stages: (1) Reference Database Construction using labeled GO and NG product samples; (2) Signal Processing and AI Model Training involving noise cleansing, dimensionality reduction, and acoustic feature extraction; and (3) Real-Time Deployment for automated QC decision-making on the production line. Case studies in electric motor testing demonstrate the system's ability to detect subtle anomalies under challenging factory conditions, achieving a high classification accuracy with sub-second inference times and actionable root-cause diagnostics. These results showcase AIAF DeepScan's potential to transform quality assurance processes across high-throughput manufacturing sectors.

Keywords: AIAF DeepScan, Quality Control, Root Cause Analysis

1. INTRODUCTION

In modern manufacturing, product quality control (QC) is essential to ensure operational reliability, reduce defect-related warranty costs, and maintain brand integrity. Two common strategies—in-line and end-of-line (EOL) QC testing—play a critical role in identifying defects during and after production. In-line QC enables early-stage defect detection and process control, while EOL QC ensures final compliance before shipment. Both are indispensable in high-speed, automated production environments.

Traditional QC methodologies rely on a combination of visual inspections, mechanical probing, electrical measurements, and manual functional testing. While effective in specific contexts, these techniques face considerable limitations:

- Contact-based methods can damage delicate components or introduce wear.
- Manual inspections are slow, subjective, and inconsistent.
- Conventional sensors and probes are constrained by mechanical complexity and accessibility.
- Automated vision systems are restricted to surface-level assessments and often lack the sensitivity to detect internal defects or dynamic failures.

[1] Signal-Wise LLC, U.S.A., lingguang.chen@signal-wise.com
[2] Wayne State University, U.S.A., nicholas.kalvaitis@ signal-wise.co.kr
[3] Wayne State University, U.S.A., sean_wu@wayne.edu

To overcome these shortcomings, significant attention has turned to non-contact, non-invasive QC techniques, such as:

- Machine vision systems for surface inspection,
- Infrared thermography for detecting thermal anomalies,
- Laser Doppler vibrometry for measuring mechanical resonance, and
- Ultrasound and passive acoustic methods for probing internal structure and defects.

However, these methods are often limited by environmental variability (e.g., lighting, temperature, ambient noise), line-of-sight constraints, and a lack of diagnostic depth. Acoustic techniques, while promising, are particularly sensitive to background noise and require sophisticated signal interpretation tools to extract meaningful insights under real-world conditions.

Given these limitations, there is a growing demand for more robust, intelligent, and adaptive QC technologies that can operate in real-world manufacturing settings—offering high accuracy, repeatability, and diagnostic insight without slowing down production.

This unmet need is the impetus of developing advanced solutions such as the Artificial Intelligence-Powered Acoustic Fingerprint-Based (AIAF) DeepScan system,[1] which leverages the inherent acoustic signatures of products and AI-driven analytics to enable real-time, non-contact, and non-invasive quality inspection with high fidelity and diagnostic capability.

The underlying principle of AIAF DeepScan lies in the fact that all acoustic signals have unique characteristics or fingerprints. Conversely, specific acoustic characteristics or fingerprints determine how the acoustic signals sound.

Using acoustics fingerprints, it is possible to identify the sound generation mechanisms and acquire a better understanding of why a specific sound is generated. Potential applications of acoustics fingerprints are boundless, ranging from product QC testing, machine failure analyses to intelligence gathering, etc.

This paper presents applications of using the AIAF DeepScan system to analyze not only electric motors QC testing, but also their defects.

2. AIAF DeepScan for In-Line and End-of-Line Product QC Testing

2.1 Theoretical Foundation

All sounds emitted by a mechanical or an electromechanical product in production contains a unique set of acoustic characteristics—acoustic fingerprint—a multidimensional signal that reflects the internal mechanical structure, material properties, assembly quality, and dynamic condition of the product. These acoustic emissions, or **acoustic fingerprints**, are encoded in the time-varying pressure fluctuations in the surrounding air and can be captured using microphone arrays or vibration sensors.

Let $s(t)$ represent the raw time-domain acoustic signal captured from a product under test. AIAF DeepScan takes $s(t)$ as the input data and processes it through a composite transformation function $\mathcal{T}$ to produce a diagnostic feature vector $\mathbf{f} \in \mathbb{R}^{n}$:

$$\mathbf{f} = \mathcal{T}[s(t)] \equiv \mathcal{F} \circ \mathcal{D} \circ \mathcal{P}[s(t)], \tag{1}$$

where $\mathcal{P}[s(t)]$ indicates preprocessing the raw data $s(t)$ including cleansing the input data $s(t)$, normalization, windowing, etc. $\mathcal{D}$ implies a dimensionality reduction and signal separation using, e.g., Principal Component Analysis (PCA),[2] Autoencoders.[3] $\mathcal{F}$ represents feature extraction by using, e.g., spectral entropy, harmonics, Mel-frequency Cepstral Coefficients (MFCCs), kurtosis, skewness, etc.[4, 5] The symbol $\circ$ signifies function composition, which means that the output of one function becomes the input to the next.

Accordingly, Eq. (1) indicates that the raw data $s(t)$ is preprocessed by function $\mathcal{P}[s(t)]$, then processed by dimensionality reduction function $\mathcal{D}$, and finally targeted acoustic features are extracted by function $\mathcal{F}$.

The extracted features **f** are input to a supervised machine learning model $\mathcal{C}$, which may

be a convolutional neural network (CNN), support vector machine (SVM), or ensemble classifier, trained to distinguish between known product conditions:

$$y = \mathcal{C}(\mathbf{f}), \quad y \in \{\mathrm{GO}, \mathrm{NO}_1, \mathrm{NO}_2, \cdots\}, \tag{2}$$

where y is the predicted quality class (e.g., GO = acceptable product, NG_1 = loose component, NG_2 = motor misalignment, etc.).

This acoustic-based inference is resilient to random noise and mechanical variation, thanks to both physical richness of acoustic signatures and the statistical power of modern AI classifiers.

2.2 Architecture and Operation

The AIAF DeepScan system operates in three primary stages:

Stage 1: Baseline Database Construction

- High-fidelity acoustic data are collected from known good and defective units.
- Each defective class represents a known defect mode (e.g., rotor imbalance, etc.).
- The acoustic fingerprints are annotated, structured, and indexed by time-frequency domain representations (e.g., spectrograms, wavelet transforms).

Stage 2: AI Model Training and Signal Cleansing

- Background noise and irrelevant transients are suppressed using adaptive filters, spectral subtraction, and ICA (Independent Component Analysis).[6]
- Dimensionality is reduced using PCA or deep autoencoders to retain dominant diagnostic features while eliminating redundancy and noise.
- Classifiers are trained on labeled datasets using k-fold cross-validation, optimizing for accuracy, recall, and inference speed.

Stage 3: Real-Time Quality Control Testing

- During production, acoustic signals are captured in-line or at the EOL stage in a fully non-contact, passive manner.
- Features are extracted and classified, and decisions are made in real time.
- Anomalous units are flagged, and the corresponding defect types are reported using root-cause inference.

2.3 Advantages Over Traditional Methods

Comparing to the traditional QC methodologies, AIAF DeepScan system possesses many advantages. Table 1 summarizes the key advantages of AIAF DeepScan system, which includes but not limited to be fully non-contact, non-invasive, automated, objective, accurate, reliable, and reliable. Moreover, it enables a real-time inline inspection without disrupting the workflow.

Table 1 compares pros and cons of traditional QC methodologies versus AIAF DeepScan

Traditional QC Methods	AIAF DeepScan
Contact-based or intrusive testing	Fully non-contact, non-invasive
Highly dependent on human operators	Automated, objective, and repeatable
Sensitive to ambient noise and line variability	Robust to noise via AI-driven preprocessing and denoising
Limited to surface or interface-level assessment	Captures deep structural and dynamic behavioral characteristics
Binary classification (GO/NG) only	Multiclass classification with root-cause attributions
Requires halting production for some inspections	Real-time inline inspection whtout disrupting workflow

2.4 Illustrative System Diagram

Figure 1 displays the conceptual block diagram of AIAF DeepScan, whose key components are as follows:

- Testing objects (e.g., electric motors, bearings, etc.) emit acoustic waves on the assembly lines.
- Microphones measure the acoustic signals emitted by testing objects and from background environments.

AIAF DeepScan system takes these acoustic signals as input and perform data acquisition, including denoising, conducting various types of mathematical transformations, extracting targeted acoustic features, using pre-established AI models to analyze and classify the extracted acoustic feature, making GO/NG decisions, and displaying the root causes of NG products accurately, reliably, continuously, and automatically.

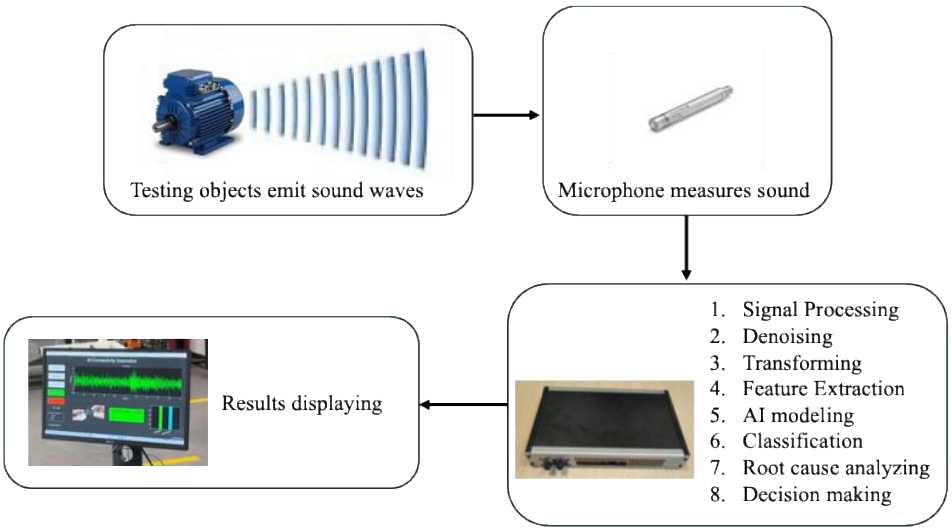


Figure 1. A conceptual diagram of the AIAF DeepScan system for carrying out non-contact, non-invasive product QC testing on an assembly line.

3. Real-World Validations

AIAF DeepScan has been successfully validated experimentally in different manufacturing settings for different test object. Despite significant ambient noise from conveyors and pneumatic tools, the system consistently achieved:

- Classification Accuracy: >99.7% for GO/NG determination.[7]
- Inference Time: <200 ms per unit, enabling in-line integration.
- Defect Identification: Detection of otherwise undiagnosed issues (e.g., rotor misalignment masked by compliant housing).

These validation results underscore AIAF DeepScan's potential to revolutionize QC testing by combining acoustic science, AI, and systems engineering to deliver actionable diagnostics in high-throughput, high-noise manufacturing environments.

3.1 PASS/FAIL Test Under Adverse Acoustic Condition

Figure 2 illustrates the experimental setup for in-line testing of electric motors using AIAF DeepScan. A unidirectional microphone was positioned approximately 15 cm from each motor, which operated at either constant or variable speed settings. The background sound pressure level (SPL) on the factory floor exceeded 90 dB(A), while the motor emissions were around 80 dB(A), yielding a Signal-to-Noise Ratio (SNR) < –10 dB(A)—a particularly challenging condition for traditional QC systems.

Despite this low SNR, AIAF DeepScan successfully extracted diagnostic acoustic fingerprints and made real-time PASS/FAIL determinations based on feature matching confidence. Figures 3 and 4 present sample results, where the "Match %" column represents the confidence level of a match with known GO or NG acoustic fingerprints. A threshold of 60% was used; products scoring above this threshold were deemed acceptable (PASS), while those below failed the QC test (FAIL).

Figures 5 and 6 summarize testing results for electric motors running at 6,000 RPM under two different load conditions—15 Nm and 45 Nm. The system achieved continuous, automated, and highly reliable QC testing across both test scenarios, with results transmitted instantly to the central production monitoring system for traceability and quality tracking.

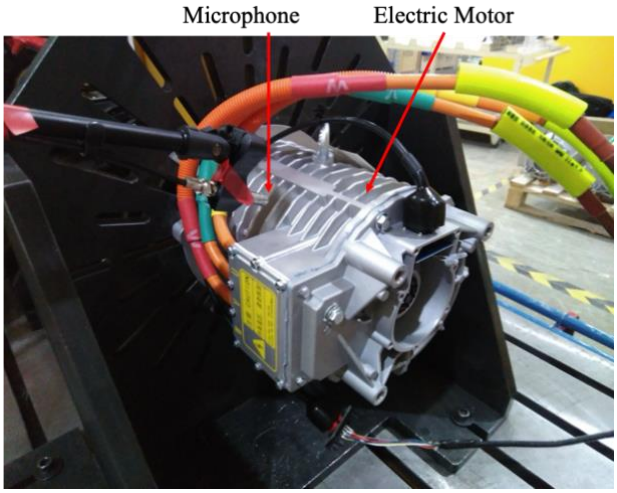


Figure 2. Test setup for AIAF DeepScan to perform inline electric motors QC testing and defects analyses on a production line.

Motor #1

Speed (rpm)	Detect times	Match (%)
2400	0	6
2500	50	94
2600	0	19
3000	0	14

Motor #2

Speed (rpm)	Detect times	Match (%)
2400	0	6
2500	0	4
2600	0	13
3000	50	99

Motor #3

Speed (rpm)	Detect times	Match (%)
2400	50	100
2500	0	9
2600	0	23
3000	0	36

Motor #4

Speed (rpm)	Detect times	Match (%)
2400	0	2
2500	0	3
2600	50	100
3000	0	16

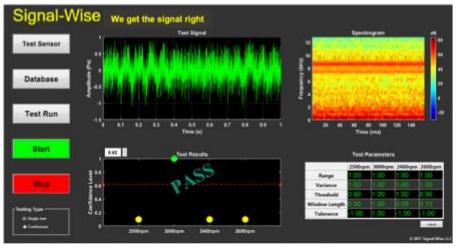


Figure 3. Sample results of electric motors QC testing at different constant speeds with an average confidence level higher than 98% in the presence of high random background noise.

Motor #5

Speed (rpm)	Detect times	Match (%)
2400	0	17
2500	0	13
2600	0	12
3000	21	47

Motor #6

Speed (rpm)	Detect times	Match (%)
2400	0	15
2500	0	13
2600	48	85
3000	0	19

Motor #7

Speed (rpm)	Detect times	Match (%)
2400	0	18
2500	50	93
2600	0	15
3000	0	23

Motor #8

Speed (rpm)	Detect times	Match (%)
2400	49	86
2500	0	18
2600	0	17
3000	0	21

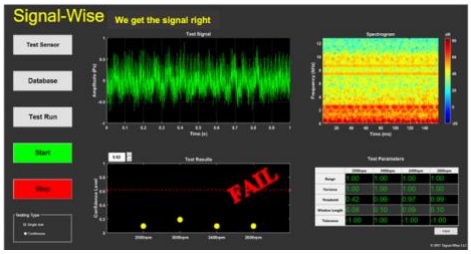


Figure 4. The higher the percentage is, the higher the confidence level of PASS/FAIL decision is. The threshold level is 60%, below which the confidence level of detecting an acceptable product (e.g., an electric motor) is low and therefore, the product fails the QC testing.

Motor Power Supply: 15 NM

Product ID	Initial Speed (RPM)	End Speed (RPM)	Run-up Time (sec)	Speed Div. (RPM)	Failing Speed (RPM)	Result	Comments
15M01T01	500	6000	11	275	N/A	Pass	Motor 01 "PASS"
15M01T02	500	6000	11	275	N/A	Pass	
15M01T03	500	6000	11	275	N/A	Pass	
15M01T04	500	6000	11	275	N/A	Pass	
15M02T01	500	6000	11	275	N/A	Pass	Motor 02 "PASS"
15M03T01	500	6000	11	275	N/A	Pass	Motor 03 "PASS"
15M04T01	500	6000	11	275	N/A	Pass	Motor 04 "PASS"
15M05T01	500	6000	11	275	N/A	Pass	Motor 05 "PASS"
15M06T01	500	6000	11	275	N/A	Pass	Motor 06 "PASS"
15M06T02	500	6000	11	275	N/A	Pass	
15M06T03	500	6000	11	275	N/A	Pass	
15M06T04	500	6000	11	275	N/A	Pass	
15M07T01	500	6000	11	275	N/A	Pass	Motor 07 "PASS"
15M07T02	500	6000	11	275	N/A	Pass	
15M07T03	500	6000	11	275	N/A	Pass	
15M07T04	500	6000	11	275	N/A	Pass	
15M08T01	500	6000	11	275	N/A	Pass	Motor 08 "PASS"
15M08T02	500	6000	11	275	N/A	Pass	
15M08T03	500	6000	11	275	N/A	Pass	
15M08T04	500	6000	11	275	N/A	Pass	
15M09T01	500	6000	11	275	N/A	Pass	Motor 09 "PASS"
15M09T02	500	6000	11	275	N/A	Pass	
15M09T03	500	6000	11	275	2837.5	Fail	
15M09T04	500	6000	11	275	N/A	Pass	
15M10T01	500	6000	11	275	3387.5	Fail	Motor 10 "FAIL"
15M10T02	500	6000	11	275	3387.5	Fail	
15M10T03	500	6000	11	275	3112.5 3387.5	Fail	
15M10T04	500	6000	11	275	3387.5	Fail	
15M11T01	500	6000	11	275	N/A	Pass	Motor 11 "FAIL"
15M11T02	500	6000	11	275	3662.5 5587.5	Fail	
15M11T03	500	6000	11	275	5587.5	Fail	

Signal-Wise Technologies

Figure 5. Summary of QC testing results for different electric motors running at 6,000 RPM under a power supply of 15 NM on a production line.

Motor Power Supply: 45 NM

Product ID	Initial Speed (RPM)	End Speed (RPM)	Run-up Time (sec)	Speed Div. (RPM)	Failing Speed (RPM)	Result	Comments
45M01T01	500	6000	11	275	N/A	Pass	Motor 01 "PASS"
45M01T02	500	6000	11	275	N/A	Pass	
45M02T01	500	6000	11	275	N/A	Pass	Motor 02 "PASS"
45M03T01	500	6000	11	275	N/A	Pass	Motor 03 "PASS"
45M04T01	500	6000	11	275	N/A	Pass	Motor 04 "PASS"
45M05T01	500	6000	11	275	N/A	Pass	Motor 05 "PASS"
45M06T01	500	6000	11	275	2837.5	Fail	Motor 06 "FAIL"
45M06T02	500	6000	11	275	2837.5	Fail	
45M06T03	500	6000	11	275	2837.5	Fail	
45M06T04	500	6000	11	275	2837.5	Fail	
45M07T01	500	6000	11	275	N/A	Pass	Motor 07 "PASS"
45M07T02	500	6000	11	275	N/A	Pass	
45M07T03	500	6000	11	275	N/A	Pass	
45M07T04	500	6000	11	275	N/A	Pass	
45M08T01	500	6000	11	275	N/A	Pass	Motor 08 "PASS"
45M08T02	500	6000	11	275	N/A	Pass	
45M08T03	500	6000	11	275	N/A	Pass	
45M08T04	500	6000	11	275	N/A	Pass	
45M08T05	500	6000	11	275	N/A	Pass	
45M09T01	500	6000	11	275	N/A	Pass	Motor 09 "PASS"
45M09T02	500	6000	11	275	N/A	Pass	
45M09T03	500	6000	11	275	N/A	Pass	
45M09T04	500	6000	11	275	N/A	Pass	
45M10T01	500	6000	11	275	3387.5	Fail	Motor 10 "FAIL"
45M10T02	500	6000	11	275	N/A	Pass	
45M10T03	500	6000	11	275	3387.5	Fail	
45M10T04	500	6000	11	275	3387.5	Fail	
45M11T01	500	6000	11	275	5587.5 5862.5	Fail	Motor 11 "FAIL"
45M11T02	500	6000	11	275	5587.5 5862.5	Fail	
45M11T03	500	6000	11	275	5587.5 5862.5	Fail	

Signal-Wise Technologies

Figure 6. Summary of QC testing results for different electric motors running at 6,000 RPM under a power supply of 45 NM on a production line.

3.2 Defects Analysis and Root Cause Inference

A key differentiator of AIAF DeepScan is its ability to not only classify GO vs. NG products but also determine the root causes of failure. This is enabled by the fact that each failure mode—such as bearing wear, unbalanced shafts, or misalignment—generates distinct acoustic patterns embedded in the signal's frequency content, harmonics, and order features.

In this study, motors were tested during acceleration (run-up) from 500 RPM to 6,000 RPM over 11 seconds, followed by deceleration back to 500 RPM. Conventional QC relies heavily on auditory inspection during this phase, which is subjective, inconsistent, and vulnerable to ambient interference.

Using AIAF DeepScan, acoustic signals are processed objectively and compared against pre-labeled defect fingerprints. Results are repeatable, standardized, and diagnostically informative.

Figure 7 presents results for an acceptable motor (#15M07T01) during run-up. The time-domain waveform, order spectrum, and defect confidence level all remained below the 60% threshold, confirming the absence of abnormal sound signatures. This motor passed the QC test.

Real-time constraints were also met: data acquisition, signal processing, and decision-making were completed within the 11-second run-up window, and decision feedback was generated immediately.

In contrast, Figure 8 shows results for motor #15M11T01. High defect confidence levels—approaching 100%—were observed throughout the acceleration phase, clearly indicating abnormal sound generation. This motor was flagged as NG and failed the QC test.

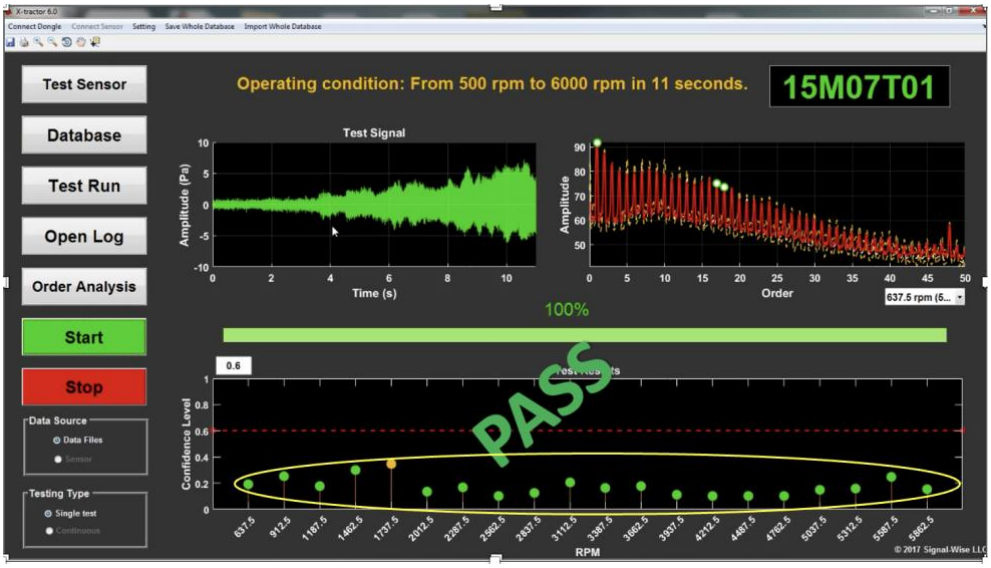


Figure 7. QC testing for electric motor #15M07T01 during runup from 500 RPM to 6,000 RPM. Since no abnormal sounds were detected throughout runup, this electric motor did not produce any abnormal sounds and therefore, it passed the QC test.

Figure 8. QC testing for electric motor #15M11T01 during runup from 500 RPM to 6,000 RPM. Since in this case the confidence level of detecting the presence of abnormal sounds was very high throughout the runup, therefore, this electric motor has abnormal sounds and failed the QC test.

3.3 Root Cause Attribution via Acoustic Fingerprint Comparison

To pinpoint the source of failure in motor #15M11T01, AIAF DeepScan compared its acoustic fingerprint against that of the acceptable motor #15M07T01 (see Figure 9). Key differences emerged in the **order spectrum**, where #15M11T01 displayed prominent half-order harmonics—highlighted by black ellipses—which were absent or weak in the GO reference. The amplitudes of these half orders exceeded acceptable bounds, indicating dynamic imbalance, possibly due to rotor misalignment.

Such spectral features are difficult or impossible to detect via human auditory inspection, particularly when masked by random background noise. However, AIAF DeepScan reliably isolates these signals and identifies their structural meaning.

Finally, Figure 10 compares the deceleration phase signals of both motors. Again, differences in spectral structure were detected automatically, providing further confirmation of defect presence and the system's diagnostic precision.

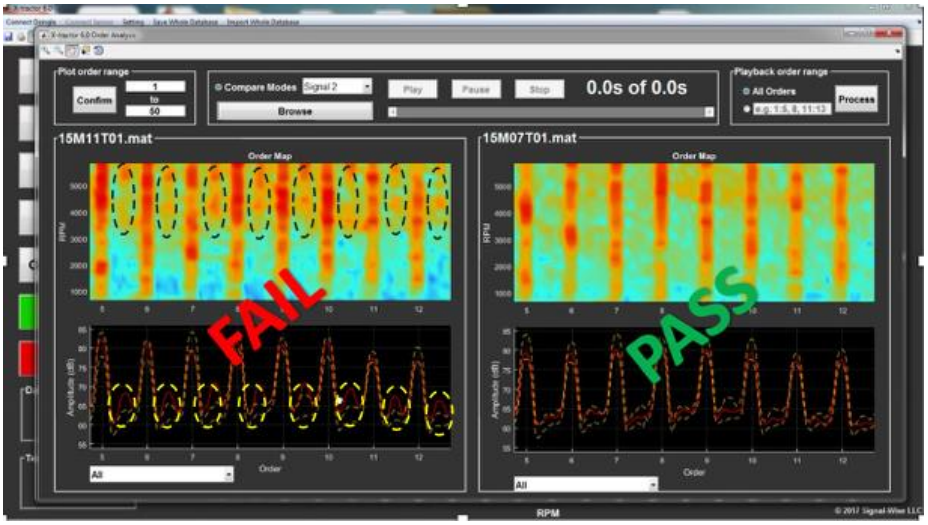


Figure 9. Comparison of the acoustics fingerprints of electric motor #15M11T01 versus those of electric motor #15M07T01. Results show that the former has many half orders, as indicated by the black elliptic circles on the left top of Figure 9, whose amplitudes fell outside the boundary of an acceptable electric motor, as indicated by the yellow elliptic circles on the left bottom of Figure 9. In comparison, these half orders were quite weak, and their amplitudes all fell inside the boundaries of an acceptable electric motor. Therefore, the root cause of electric motor #15M11T01 emitting abnormal sounds was due to the presence of half orders, which could indicate a rotor misalignment.

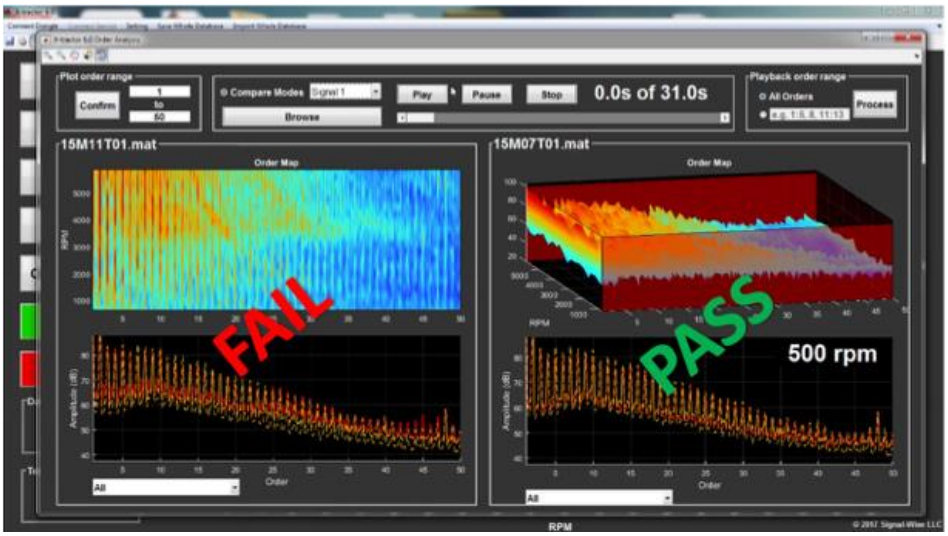


Figure 10. Comparison of the acoustic behaviors of electric motor #15M11T01 and those of #15M07T01 in deceleration from 6,000 RPM to 500 RPM are compared. It is impossible to discern an acceptable electric motor from an unacceptable one based on auditory inspection only. Using AIAF DeepScan, however, it not only enables one to differentiate them but also show the root cause of electric motor defects.

4. Conclusion

The AIAF DeepScan system represents a paradigm shift in product quality control, offering a non-contact, non-invasive, and intelligent alternative to traditional inspection methods. Its core strengths lie in:

- High diagnostic accuracy under adverse noise conditions (SNR < –10 dB),
- Real-time inference suitable for in-line and EOL integration (<200 ms),
- Defect classification and root-cause analysis, enabling predictive maintenance and process optimization,
- Standardization and repeatability, eliminating operator subjectivity.

Validated in industrial settings for electric motor inspection, AIAF DeepScan has demonstrated its robustness, scalability, and transformative potential. Future work includes deploying this system for other components—such as pumps, gearboxes, and assemblies—and integrating it into cloud-based manufacturing intelligence platforms for full traceability and closed-loop quality assurance.

REFERENCES

1. Lingguang Chen, “Acoustic Inspection System and Method,” U.S. Patent Application (#19/181,754), April 2025.
2. Jolliffe, I. T., & Cadima, J. (2016). *Principal component analysis: A review and recent developments*. Philosophical Transactions of the Royal Society A, 374(2065), 20150202.
3. Hinton, G. E., & Salakhutdinov, R. R. (2006). *Reducing the dimensionality of data with neural networks*. Science, 313(5786), 504–507.
4. Davis, S. B., & Mermelstein, P. (1980). *Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences*. IEEE Transactions on Acoustics, Speech, and Signal Processing, 28(4), 357–366.
5. Tzanetakis, G., & Cook, P. (2002). *Musical genre classification of audio signals*. IEEE Transactions on Speech and Audio Processing, 10(5), 293–302.
6. Hyvärinen, A., & Oja, E. (2000). *Independent component analysis: Algorithms and applications*. Neural Networks, 13(4–5), 411–430.
7. Nicholas Kalvaitis and Lingguang Chen, “3-Sigma Test Report,” AIAF Test Report, April 2025.

PROCEEDINGS of the
16th International Conference on Theoretical and Computational Acoustics

August 18 to 22, 2025 BEXCO, Busan Korea

Real-Time Smart Buoy System for Cetacean Monitoring

Tsai Meng-Fan[1], Hung Ching-Tang[2], Li Wei-Lun[3], Chen Chi-Fang[4], Hu Wei-Chun[5]

[1]Ocean Sound Company
[2]Ocean Sound Company
[3]Ocean Sound Company
[4]National Taiwan University
[5]National Taiwan University

ABSTRACT

This research focuses on developing a cost-effective and high-performance smart buoy system designed for the real-time acoustic and visual monitoring of cetaceans. By integrating advanced underwater acoustic sensors, multi-channel hydrophone arrays, and image processing techniques, the proposed system can rapidly detect, localize, and estimate cetacean populations. Data and imagery are processed through real-time communication and edge computing, facilitating immediate analysis and visualization. These robust and reliable smart buoys, deployed in marine environments, support the development of an extensive acoustic-visual dataset. Additionally, a user-friendly web-based analytical platform significantly enhances user accessibility and data interpretability. This innovative monitoring approach makes a substantial contribution to marine biodiversity conservation and has extensive practical applications across various marine-related industries.

[1]mftsai@oceansound-tw.com
[2] cthung@oceansound-tw.com
[3] pywl03040911@gmail.com
[4] chifang@ntu.edu.tw
[5] william_hu@outlook.com

Reproduction and Analysis of Pulse Wave Propagation and Velocity (PWV) in the Human Aorta

Jae-Hak Jeong[1]
[1] Korea National University of Transportation, Chungju, Republic of Korea

Pulse waves are one of the most fundamental physiological acoustic signals, and their propagation velocity (Pulse Wave Velocity, PWV) is a significant biomarker for vascular health. PWV reflects arterial properties such as diameter, wall thickness, and stiffness, making it useful in clinical diagnostics and modeling. This study presents a mock circulatory robot platform (MCRP) that reproduces pulse waveforms propagating through a human aorta. The MCRP mimics the human aorta's physiological structure, mechanical properties, and function. It enables direct measurement of carotid-femoral PWV (cfPWV) under controlled conditions, and the measured values match theoretical predictions. Using the MCRP, we compare conventional reflection wave separation with an AI-based method for PWV estimation. The AI model is trained on time-domain waveform features, and its estimates are evaluated against the ground-truth cfPWV obtained from the MCRP. This study demonstrates that the MCRP can provide both a reliable simulation environment and valid reference data for PWV analysis. Combining physical modeling with AI interpretation improves the accuracy and applicability of non-invasive vascular assessment.

Keywords: Blood Pressure, Circulatory, Healthcare AI, Vascular Age, Pulse Wave Velocity

1. INTRODUCTION

Pulse wave velocity (PWV) is a critical biomarker reflecting arterial stiffness and is widely used in the assessment of cardiovascular health [1,2]. It is influenced by vascular parameters such as diameter, elasticity, and wall thickness [3]. While various clinical studies have attempted to extract PWV from in vivo blood pressure waveforms, their progress has been limited due to significant inter-patient variability and the lack of high-resolution reference datasets, especially in young adults [4]. To overcome these challenges, this study introduces a mock circulatory robot platform (MCRP) capable of reproducing pulse waveforms that propagate through age-specific aortic models. The MCRP offers a controlled experimental environment to investigate the physics of pulse wave propagation under known vascular conditions.

2. Mock Circulatory Robot Platform (MCRP)

The MCRP is composed of a modular system that includes a left ventricular pumping unit, compliance chamber, tunable peripheral resistance, and interchangeable artificial aortic models [5]. Six artificial aortas were fabricated using biomimetic silicone materials to replicate mechanical stiffness profiles representing vascular ages ranging from 20 to 70 years [6]. Each segment underwent tensile testing to validate its stiffness and was calibrated to generate target PWV values consistent with physiological data. The system operates under programmable hemodynamic conditions—typically 70 mL stroke volume at 60 BPM—to simulate physiologically realistic systolic and diastolic pressures of 120/80 mmHg. Real-time pressure waveform acquisition at multiple sites along the aorta was achieved using high-fidelity pressure sensors. This reproducible setup enables systematic analysis of pulse wave propagation and morphological changes across vascular conditions [7].

[1] jaehak.jeong@ut.ac.kr

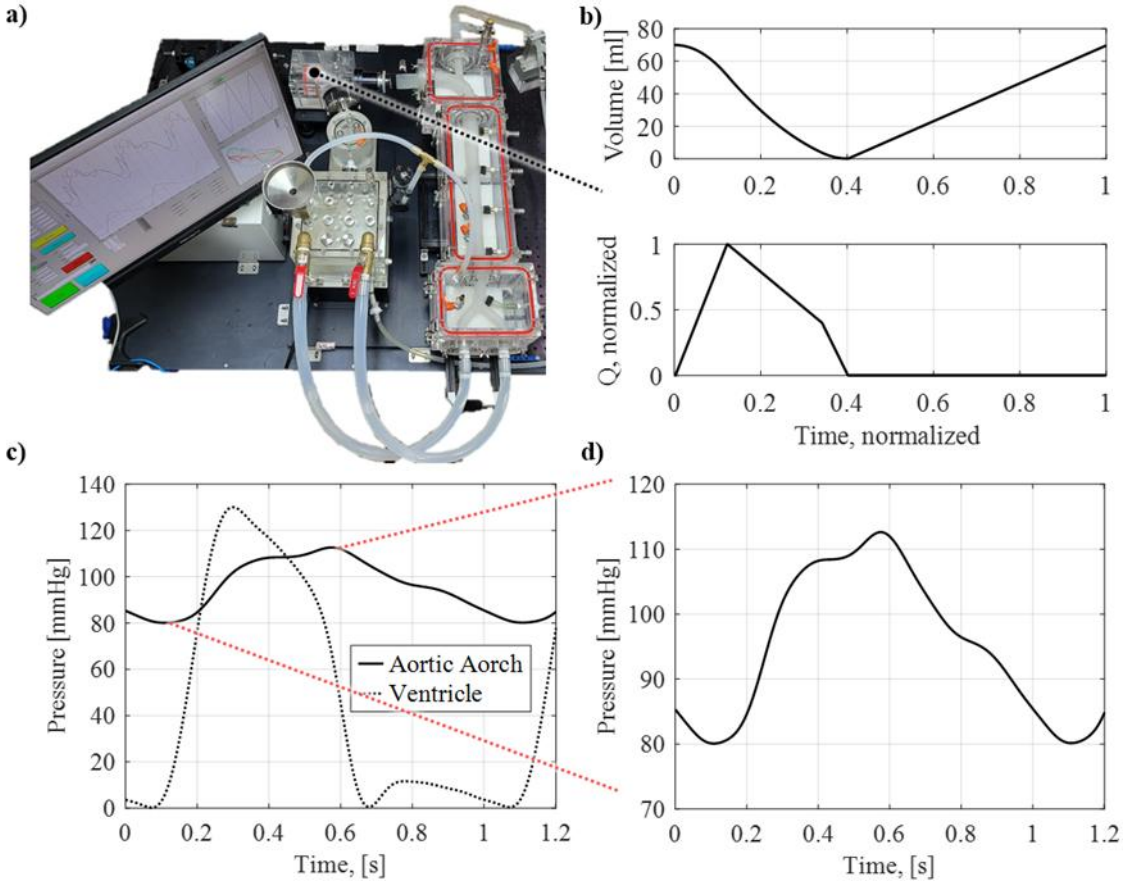


Figure 1: Reproduction of BP waves using the MCRP; a) fabricated MCRP, b) volume and flow waveforms input to the ventricular pump, c) BP waves measured in the ventricle and aortic arch, d) reproduced central BP wave with similar magnitude and waveform to humans

3. Methodology

3.1 Physics-Based PWV Estimation

PWV was estimated using a wave separation method based on theoretical principles of forward and reflected wave components [8,9]. By assuming a triangular flow waveform, central aortic pressure signals were decomposed via frequency-domain Fourier analysis into forward and backward components. The temporal delay (pulse transit time, PTT) between these two waveforms was computed via cross-correlation, from which the PWV was derived. This method offers interpretability and aligns with established arterial wave propagation theory.

3.2 AI-Based Estimation via CNN

To validate and further explore waveform patterns, a one-dimensional convolutional neural network (1D-CNN) was trained on a synthetic dataset of 330 waveforms generated from the MCRP [10]. The model architecture included two convolutional layers and fully connected classifiers for age group prediction. Each waveform was resampled to 500 Hz and zero-padded to 1024 length for uniform input. Data augmentation through temporal shifting generated approximately 10,000 waveforms to improve robustness. Although the focus of this paper is on physical PWV analysis, the AI model served to evaluate the distinctiveness and stability of waveform morphology across age-matched conditions.

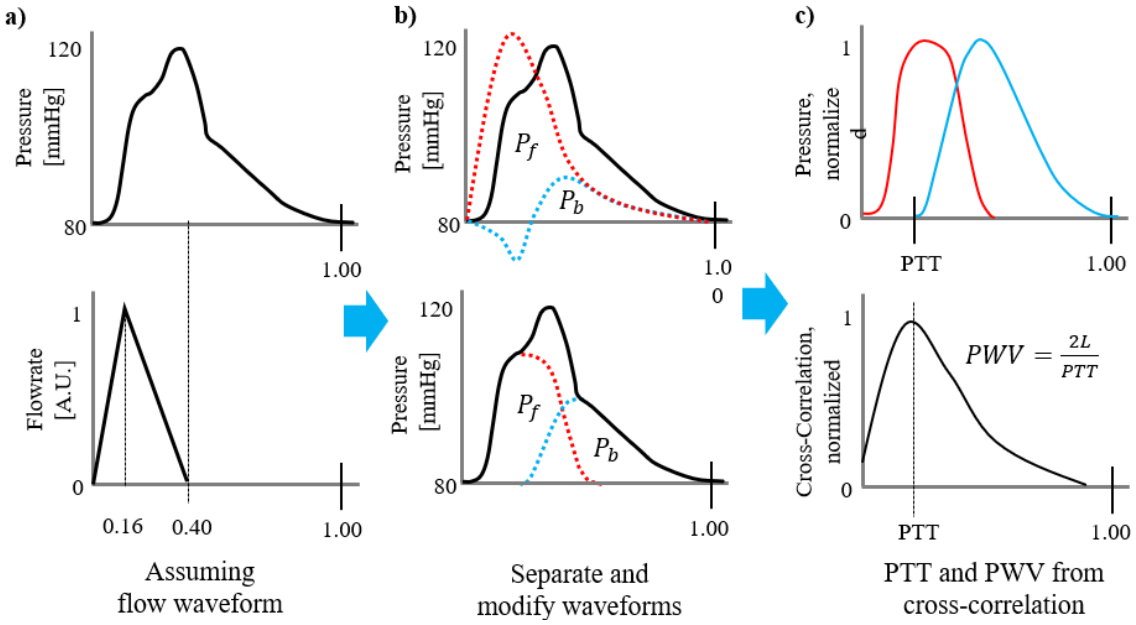


Figure 2: Estimation of PWV using the separation of the forward and backward waves of the central BP wave; a) Assumption of triangular flow wave from BP wave, b) separation of the wave in the frequency domain and waveform correction, c) calculation of PTT and PWV using cross-correlation.

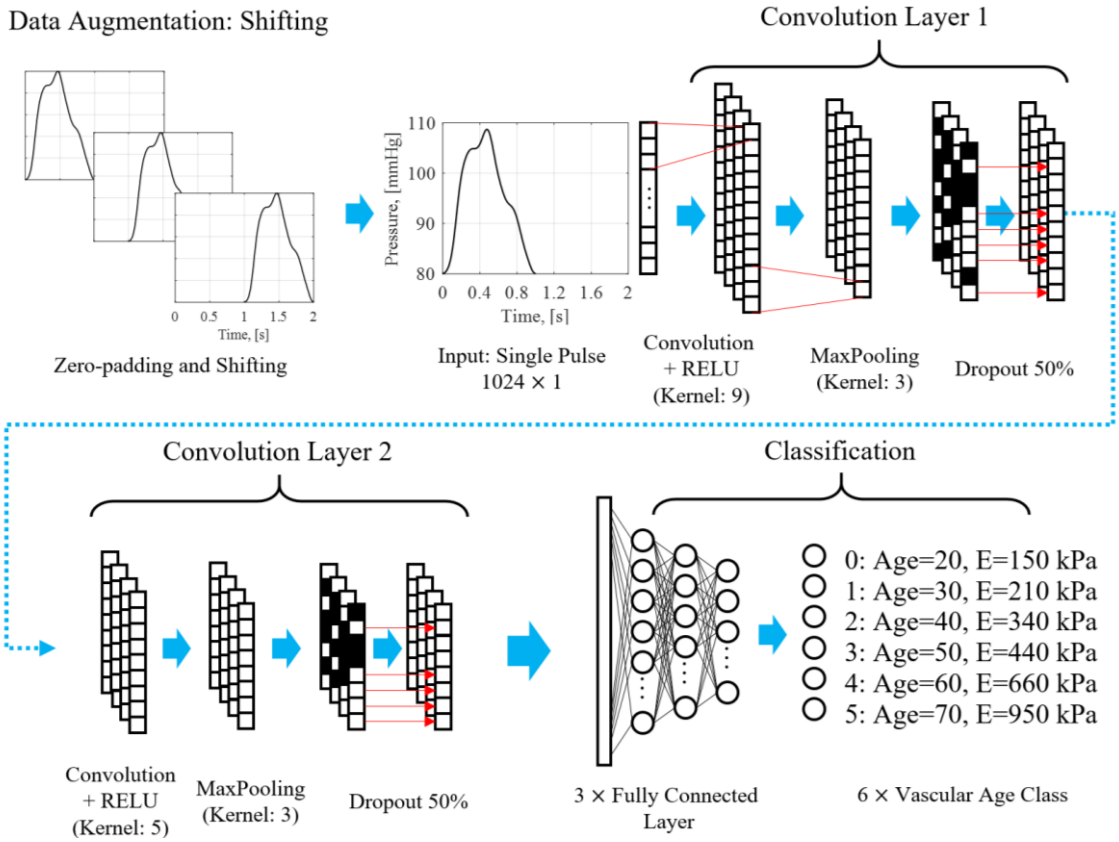


Figure 3: Data-driven estimation of vascular age: preprocessing of input data, and 1D-CNN model consisted of 2 convolution layers for classification of 6 age group.

4. Results and Analysis

The physics-based PWV estimation exhibited high accuracy overall, with root-mean-square error (RMSE) of 0.48 m/s in velocity and approximately 4.3 years in estimated vascular age. However, it showed a consistent tendency to overestimate vascular stiffness in younger aortic models due to low waveform augmentation. In contrast, the AI-based method achieved a classification accuracy of 90.15% and reduced vascular age RMSE to 3.04 years. Transfer learning applied to clinical datasets particularly enhanced performance in the 20–30-year age group, effectively addressing overestimation biases. Both approaches confirmed the capacity of the MCRP to produce physiologically valid waveforms suitable for PWV assessment and waveform classification.

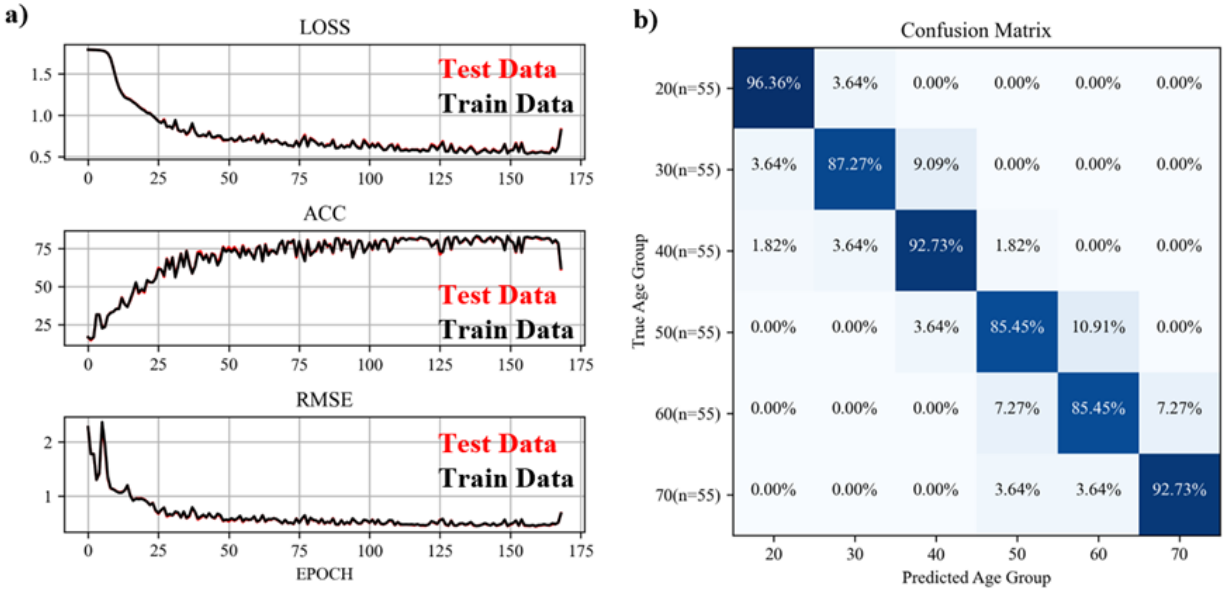


Figure 4: Learning curve and confusion matrix of 1D-CNN trained by simulator data.

5. Conclusion

This study demonstrates that the mock circulatory robot platform provides a robust and repeatable environment for reproducing and analyzing pulse wave propagation in the human aorta. The controlled simulation of age-specific aortic stiffness allows for accurate benchmarking of PWV estimation methods under known conditions. Furthermore, combining physical modeling with AI interpretation offers a multi-perspective approach to analyzing vascular dynamics. Future work will expand this framework to include waveform propagation in the brachial and radial arteries, enabling a more comprehensive understanding of multi-site vascular wave transmission and serving as a reference platform for biosignal-based health monitoring technologies.

ACKNOWLEDGEMENTS

This research was supported by the 2025 Glocal University 30 Project at Korea National University of Transportation. (본 연구는 2025 년 국립한국교통대학교 글로컬대학 30 사업의 지원을 받아 수행되었음.)

."

REFERENCES

1 Westerhof N., et al. (2008). Arterial stiffness and wave reflection: determinants of cardiovascular risk. Hypertension, 52(4), 586–593.

2 O'Rourke M. F., et al. (2002). Clinical applications of arterial stiffness; definitions and reference values. Am J Hypertens, 15(5), 426–444.

3 Nichols W. W., O'Rourke M. F. (2005). McDonald's Blood Flow in Arteries. Hodder Arnold.

4 Laurent S., et al. (2006). Expert consensus document on arterial stiffness: methodological issues and clinical applications. Eur Heart J, 27(21), 2588–2605.

5 Jeong J. H., et al. (2023). Fabrication of cardiovascular simulator for reproducing human blood pressure wave and inverse estimation method of cardiovascular parameters. Ph.D. Dissertation, KAIST.

6 Jeong J. H., et al. (2023). Age-specific artificial aorta fabrication with matched stiffness. Unpublished data.

7 Mitchell G. F. (2009). Arterial stiffness and wave reflection: biomarkers of cardiovascular risk. Hypertension, 53(5), 754–755.

8 Qasem A., et al. (2007). Estimation of aortic PWV using frequency-domain analysis of pressure waveform. J Biomech Eng, 129(5), 930–934.

9 Westerhof B. E., et al. (2006). Methodology for the determination of wave reflection and arterial stiffness. Hypertension, 48(4), 603–608.

10 Chen X., et al. (2021). Deep learning-based estimation of arterial stiffness from PPG. IEEE Trans Biomed Eng, 68(5), 1483–1493.

PROCEEDINGS of the
16th International Conference on Theoretical and Computational Acoustics

August 18 to 22, 2025 BEXCO, Busan Korea

Model refinements to better estimate ship-generated noise impact on endangered species

Maximilian Lauch[1]

[1]Fisheries and Oceans Canada

ABSTRACT

Robust modelling of underwater ship-generated noise can add much to our understanding of the impact to marine species. Here we present modeling efforts for the Salish Sea, demonstrating refinement of a vessel noise model through sensitivity analyses and validation via comparison to in-situ passive acoustic recordings. The model supports initiatives to minimise the acoustic disturbance of endangered southern resident killer whales.

The Salish Sea has varied coastlines, bathymetry, geoacoustic conditions and oceanographic properties, while hosting international shipping lanes. Model inputs tested included: vessel source-levels; ray-trace angles; geoacoustic profiles; numerical ray-tracing tolerances; and horizontal resolution of the 3-D sound-speed grid. The sensitivity analysis quantifies how key modelling and input data choices influence both the root-mean-square error (RMSE) between predicted and observed sound pressure levels (SPL), and computational cost. For each parameter set, we computed the RMSE of the model output against the passive acoustic recordings time-series to identify the optimal input parameters, ray-tracing methods, and site-specific adaptations.

The analysis framework presented highlights a comprehensive yet efficient approach to estimate underwater noise emissions and therefore allow for the design of mitigation actions to address ship-generated noise that we hope can be applied beyond our present example.

[1]maximilian.lauch@dfo-mpo.gc.ca

DAILY PROGRAM

AUGUST 21 (THU)

14:20-15:40	Modeling and Detection of Acoustic Signatures from Complex Underwater Structures 324 (3F)
Chair	Kyungmin Baik (Korea Research Institute of Standards and Science)
14:20	Comparative analysis of array characteristics of different planar arrays
	Jiaxin Du, Nan Zou, Guangpu Zhang, Jingze Huang
14:40	Effects of appendages on the turbulence and flow noise of a submarine model using high-order scheme
	Peng Jiang, Shijun Liao, Bin Xie
15:00	Modelling and simulation of Ship-Excited Micro-Doppler Effect
	Minhao Wang, Cuie Zheng, Tianye Na
15:20	Research on nonlinear Acoustic Detection of Submarine Pipelines Based on 3D Height Field Modeling
	Peng-Yue Wang, Jian-Jun Zhu, Pei-Hong Wang, Guo-Qing Zhao

Comparative analysis of array characteristics of different planar arrays

Zou Nan[1] Du Jiaxin[2] Zhang Guangpu[3]Huang Jingze[4]
[1] National Key Laboratory of Underwater Acoustic Technology,Harbin Engineering University; Key Laboratory of Marine Information Acquisition and Security(Harbin Engineering University),Ministry of Industry and Information Technology; College of Underwater Acoustic Engineering, Harbin Engineering University,China
[2] National Key Laboratory of Underwater Acoustic Technology,Harbin Engineering University; Key Laboratory of Marine Information Acquisition and Security(Harbin Engineering University),Ministry of Industry and Information Technology; College of Underwater Acoustic Engineering, Harbin Engineering University,China
[3] National Key Laboratory of Underwater Acoustic Technology,Harbin Engineering University; Key Laboratory of Marine Information Acquisition and Security(Harbin Engineering University),Ministry of Industry and Information Technology; College of Underwater Acoustic Engineering, Harbin Engineering University,China
[4] National Key Laboratory of Underwater Acoustic Technology,Harbin Engineering University; Key Laboratory of Marine Information Acquisition and Security(Harbin Engineering University),Ministry of Industry and Information Technology; College of Underwater Acoustic Engineering, Harbin Engineering University,China

ABSTRACT

Different array designs make the performance of sonar different. Array design and array performance analysis have positive guiding significance for the development of sonar technology and equipment. In this paper, a new hexagonal array is designed and compared with the most commonly used circular array to analyze the performance. The analysis shows that in the horizontal dimension, the side lobe of the semi-hexagonal array is lower than that of the semi-circular array in most areas, and the side lobe of the semi-hexagonal array is higher in very few areas. This is due to the fact that the hexagonal array is composed of a linear array, which still retains the ' left and right port ambiguity ' problem of the linear array, but the azimuth of the higher side lobe is predictable . In the pitch dimension, there is no significant difference between the two formations. According to the azimuth spectra of the five methods of CBF, MVDR, MUSIC, Eigenspace and EAAIS-CBF, it is concluded that in the corresponding interest interval, the resolution of the two arrays is approximately the same regardless of the algorithm, but the side lobe of the semi-hexagonal array is lower than that of the semi-circular array.

Keywords: designs make, hexagonal array, array characteristics, array signal processing technology.

1. INTRODUCTION

In recent years, with the rapid development of array signal processing methods, they are widely used in communication, radar, sonar, astronomy, medicine and other fields. While optimizing the signal processing algorithm, array design and array performance analysis are indispensable parts before signal processing. Shen designed a 32-element spherical crown array antenna based on the characteristics of low-orbit satellite broadcast signal receiving terminals. The antenna with this formation structure solves the problem that the satellite broadcast signal is difficult to meet the coverage requirements of high-gain and multi-beams[1]. Huang designed a sparse array design algorithm for adaptive beamforming using the alternating direction method of multipliers (ADMM), which significantly outperforms several other sparse array design strategies in terms of output signal-to-noise ratio [2].Supakit Kawdungta designed an antenna array using the modified fruit fly optimization algorithm (MFOA), and the proposed algorithm has the advantages of easy implementation, global search space and fast convergence rate [3]. In the field of hydroacoustics, the

[1] djx1763980@163.com

tremendous development of small underwater unmanned motion platforms has driven the need for array improvement. Different array designs make the performance of sonar different, and the array design and array performance analysis, etc. have positive guiding significance for the development of sonar technology and equipment. Limited by the physical space or carrying capacity of unmanned platforms, traditional line arrays will not be suitable for carrying on the bow of unmanned platforms.Tesei, A proposed to solve the problem of detecting marine traffic in a specific sea area by hosting a small volumetric acoustic array on a submerged glider [4].X Sun designed a conical helical array with a small aperture, which was mounted on the head of an Autonomous Underwater Vehicle (AUV). aimed at localizing hydroacoustic sources [5]. At present, there are three main types of bow sonar array adopted by sonar designers, including cylindrical array, spherical array and conformal array. The array formation needs to be determined according to the mission tasks undertaken by the bow array and the space provided by the platform. In general, an important feature of the development of bow formation is the diversity of formation types.

Hydroacoustic arrays utilize spatially distributed sensor arrays to collect data from the acoustic field and then process the collected data. This technology is known as array signal processing technology, which has the advantages of high gain, flexible beam control, strong anti-jamming ability and high spatial resolution. This is also an important reason why it is widely used in the field of hydroacoustic detection. However, in the actual hydroacoustic detection process, the marine environment is complex and variable, and the acoustic field data collected by the array includes both the target signals of interest and a variety of interferences, such as marine environmental noise, platform self-noise and so on. The problem of weak target detection under the background of strong interference has been a key concern of scholars. The more commonly used methods are beam forming method [6-9] and subspace reconstruction method [10-12]. The beam forming method is to weight the received signals of different array elements in the array to achieve the purpose of suppressing interference and enhancing the target, so as to display the weak target in the beam forming output results. Commonly used beam forming methods are CBF, MVDR, LCMV, etc. The subspace reconstruction method utilizes orthogonal projection to project the array received signal into the subspace domain, and uses the differences between targets to reconstruct the signal subspace and weaken the effect of strong interference on weak targets. In addition, the sparse method uses the sparsity of the target in the whole spatial range to distinguish the strong target and the weak target. [13-14].

In this paper, a new array is designed : hexagonal array, which is compared with the most commonly used planar array and circular array. Based on the analysis of the beam pattern, the advantages and disadvantages of the two arrays are compared and summarized. Based on the two arrays, the applicability of hexagonal array in underwater acoustic signal processing is analyzed by simulating several classical algorithms.

2. ARRAY MODEL AND CHARACTERISTIC ANALYSIS

A hexagonal array is an array that is a combination of six line arrays. Each array element is uniformly distributed on each subarray, each subarray has 16 array elements, the spacing of the array elements is half a wavelength, the vertex position is not deployed sensors, a total of M = 96 array elements. The circular array is an array of 96 array elements uniformly distributed on a circle with radius R. R is taken as the length of the hexagonal array edge. The geometric models of the two arrays are shown in Figure 1.

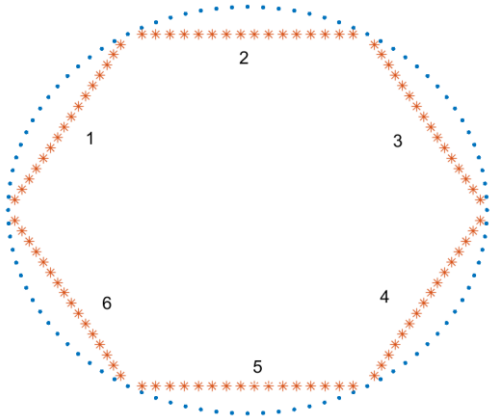


Figure 1 – Geometric models of circular array and hexagonal array

The orientation map of the array is an important tool for evaluating the performance of the array, which can reflect the key characteristics of the array such as gain characteristics, interference suppression ability, spatial resolution ability and so on. A coordinate system is established with the center of the hexagonal array as the reference origin, the angleθ with the positive direction of the x-axis is the horizontal azimuth angle and

the counterclockwise is positive, and the angleϕ with the positive direction of the z-axis is the vertical pitch angle. The corresponding unit vector of propagation direction is

$$\mathbf{v}(\theta,\phi) = -[\sin\phi cos\theta \quad \sin\phi sin\theta \quad \cos\phi]^T \tag{1}$$

where $(\cdot)^T$ denotes the transpose.

The orientation vector of the hexagonal array can be expressed as

$$\mathbf{a}(\theta,\phi) = \left[e^{-j2\pi f\mathbf{v}(\theta,\phi)\boldsymbol{p}_1/c}, e^{-j2\pi f\mathbf{v}(\theta,\phi)\boldsymbol{p}_2/c}, \ldots, e^{-j2\pi f\mathbf{v}(\theta,\phi)\boldsymbol{p}_M/c} \right]^T \tag{2}$$

where, $\boldsymbol{p}_m = [p_{xm}, p_{ym}, p_{zm}]^T$ is the position coordinate of the hexagonal array element.

In the same coordinate system, the position of the array element of the circular array can be expressed as

$$\boldsymbol{p}_m = [R\sin\phi\cos\frac{2\pi m}{M}, R\sin\phi\sin\frac{2\pi m}{M}, R\cos\phi] \tag{3}$$

Then the orientation vector can be expressed

$$\mathbf{a}(\theta,\phi) = \left[e^{-j2\pi fR\sin(\phi)\cos(\theta-\frac{2\pi}{M})/c}, e^{-j2\pi fR\sin(\phi)\cos(\theta-\frac{2\pi*2}{M})/c}, \ldots, e^{-j2\pi fR\sin(\phi)\cos(\theta-\frac{2\pi*M}{M})/c} \right]^T \tag{4}$$

According to the definition of the orientation map of an array: the response of an array of a certain orientation to a signal of a given direction has

$$P(\theta,\phi) = w^H \mathbf{a}(\theta,\phi) \tag{5}$$

where w is the weighted vector corresponding to a desired direction (θ_0,ϕ_0), which is taken as the conventional beamforming vector. $(\cdot)^H$ denotes the covariance transpose.And $\mathbf{a}(\theta,\phi)$ is the guiding vector of the array.

The orientation map can reflect the spatial response characteristics of the beamformer, indicating the gain of the array for signals incident from different angles. Considering that the circular array is usually used as a semicircular array in practical engineering applications due to the masking effect, only the semi-hexagonal array composed of subarrays 1, 2 and 3 and the semicircular array in the corresponding region are compared and analyzed in the following simulations. The horizontal azimuth and vertical pitch angle ranges are set to [1°, 360°] and [1°,90°], respectively, and are uniformly sampled at 0.2° intervals. The orientation maps of the semi-hexagonal and semi-circular arrays in the (100°, 40°) direction are obtained as shown in Figure 2 and 3.

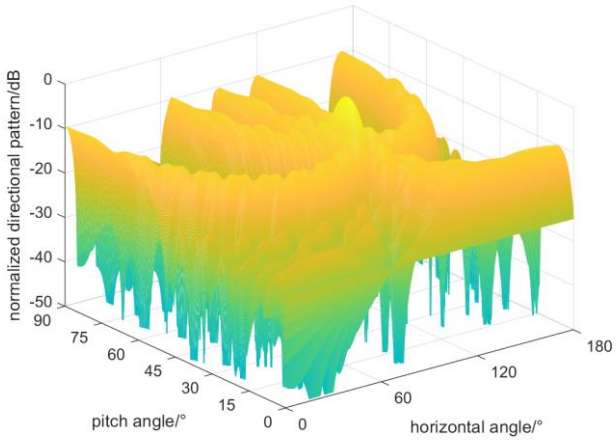


Figure 2 – Two-dimensional pattern of semi-hexagonal array

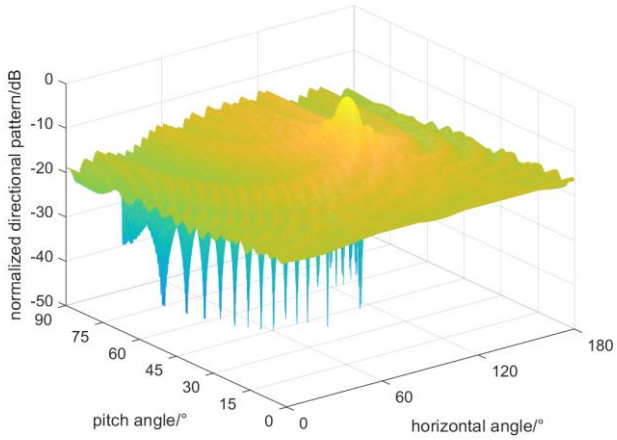


Figure 3 – Two-dimensional pattern of semi-circular array

It can be seen that the main peak value appears in the direction of (100°, 40°). The beamformer has the largest response in this direction, corresponding to the direction of the desired signal. The response in other directions is lower than that in this direction, indicating that it is suppressed to a certain extent. In order to visualize the main peak value, the sections of the directional maps passing through (100°, 40°) parallel to the *xoz* plane and *yoz* plane are made respectively, and the results of the semicircular and semi-hexagonal arrays are displayed in Figure 4.

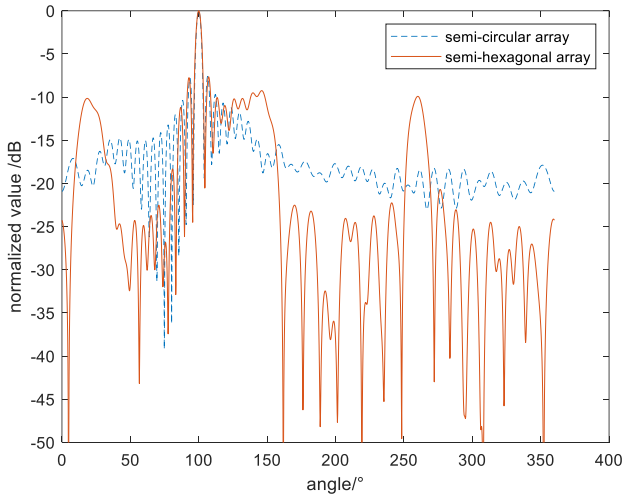


Figure 4 – Comparison of the pattern of different sections

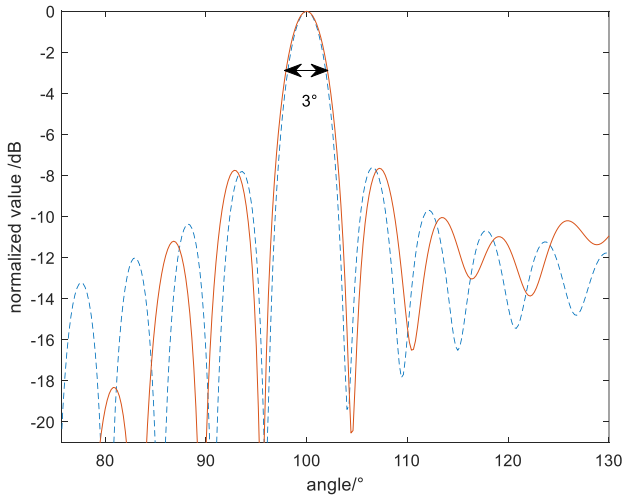


Figure 5 – Comparison of local enlarged images

From Figure 4, it can be seen that the two types of arrays have only one predefined main peak value in the 100° direction in the full 360° range, and the side lobes are lower than the main peak value. In addition, the semicircular array has a higher side lobe than the semi-hexagonal array in most of the area ranges.And the side lobe of the semi-hexagonal array is higher in very few areas, specifically near the 20°, 140°, and 260° directions. This is due to the fact that the semi-hexagonal array is essentially a combination of three line array subarrays, which still retains some of the characteristics of the line array, and the above three angles correspond to the symmetry direction of 100° about the different subarrays, which is the higher side lobes caused by the "left and right side fuzzy" problem of the line array. At the same time, each subarray has a certain suppression effect on the side lobes generated by the other subarrays due to the "port and starboard problem", so that the highest side lobes are still lower than the main peak value. It is also worth noting that the highest side lobe in the 260° direction in the semi-hexagonal array is due to the fact that the signal is incident from the 100° direction, and the main receiving subarray is the No.2 array, which has a symmetrical side lobe in the 260° direction for signals incident from the 100° direction. This shows that the hexagonal array has different signal incidence directions, and the main role of the array is also different. Figure 5 is a partial enlargement of Figure 4, which is labeled with the -3dB beamwidth of the two arrays. Half-hexagonal array and semi-circular array of the beamwidth are about 3° around. The beamwidth reflects the signal resolution ability of the array. Then the resolution of the two arrays to the signal is similar.

Changing the incident horizontal direction angle to 90°, the vertical pitch angle is unchanged.The results are shown in Figure 6. It can be seen that the semicircular array in the full 360 ° has the same directivity, there is no line array end of the projected direction of the problem of decreasing resolution, with the spatial angle shift invariance. On the other hand, the half-hexagonal array is analyzed as before. In the direction of 150°, 270°, and 30° appear higher than the semicircular array, and the 270° direction is the highest. Although the half-hexagonal array is not as good as the semicircular array with the same pointing in the full 360 °, its pointing is a periodical change to 60 ° as a cycle. Comparing the orientation plots in the pitch dimension in Figures 4 and 6 again.There is no significant difference between the two formations in the pitch dimension.

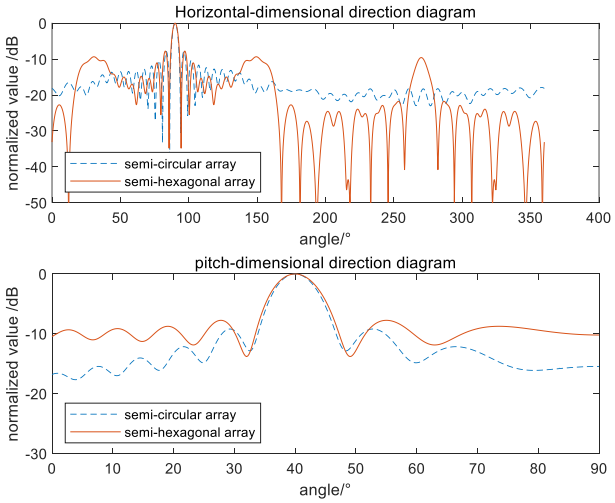


Figure 6 – Comparison of patterns of different sections changing the incident direction

3. ALGORITHM APPLICATION

3.1 Array Received Signal Model

In order to compare the application effect of the two types of arrays and apply them in specific algorithms, the array received signal model is constructed first.To simplify the model, only the horizontal azimuth is considered.According to the theory of spatial spectrum estimation, assuming that there are D sources incident from the $\theta_1,\theta_2,\cdots,\theta_D$ direction. The signal matrix corresponding to the array received signal can be expressed as

$$X(t) = A(\Theta)S(t) + N(t) \tag{6}$$

where, $S(t)$ is the source data, $N(t)$ is the received noise signal of the array, which is set as Gaussian white noise, and the noise signals between array elements are uncorrelated.

$$A(\Theta) = [a(\theta_1) \quad a(\theta_2) \quad ,\cdots, \quad a(\theta_D)] \tag{7}$$

is the orientation vector matrix of the corresponding individual D source. Then the covariance matrix of the array received signal is

$$R = \mathrm{E}\left[XX^{\mathrm{H}} \right] = A(\Theta) R_s A^H(\Theta) + \sigma_n^2 I_M \tag{8}$$

In engineering implementation, the covariance matrix is often estimated by discrete sampling of the received signal.

$$\hat{R} = \frac{1}{K} \Sigma_{k=1}^{K} X_k X_k^{\mathrm{H}} \tag{9}$$

3.2Conventional Algorithm Application

In the spatial spectrum theory, the representative algorithms are CBF algorithm, MVDR algorithm and MUSIC algorithm. The CBF algorithm is a representative of the non-data-dependent fixed beamforming algorithm[15], and its azimuth spectrum estimation result is

$$P_{cbf}(\theta) = \mathbf{a}^H(\theta) R \mathbf{a}(\theta) \tag{10}$$

Some scholars have proposed data-dependent adaptive beamformers. The basic principle is based on the array output noise plus interference power minimization constraint, referred to as the MVDR algorithm[16]. The azimuth spectrum estimation result is

$$P_{\mathrm{MVDR}}(\theta) = 1 / \mathbf{a}^H(\theta) R^{-1} \mathbf{a}(\theta) \tag{11}$$

Later, with the development of high resolution algorithms, American scholars Schmist R O and others proposed the MUSIC algorithm, which divides the received data into signal and noise subspaces[17]. The azimuthal spectrum of this algorithm is estimated as

$$P_{\mathrm{MUSIC}}(\theta) = 1 / \mathbf{a}^H(\theta) \mathbf{U}_N \mathbf{U}_N^H \mathbf{a}(\theta) \tag{12}$$

where, $\mathbf{U}_N$ represents the noise subspace.

Hoffman used the feature structure of the correlation matrix to project the scanning vectors into the signal subspace for the azimuthal estimation, and put forward the Eigenspace algorithm. space algorithm[18]. The azimuth spectrum estimation result is

$$P_{\mathrm{Eigen}}(\theta) = (\mathbf{a}^H(\theta) \mathbf{U}_{\mathbf{S}} \mathbf{\Lambda}_{\mathbf{S}}^{-1} \mathbf{U}_{\mathbf{S}} \mathbf{a}(\theta)) / (\mathbf{a}^H(\theta) R^{-1} \mathbf{a}(\theta) \mathbf{a}^H(\theta) R^{-1} \mathbf{a}(\theta)) \tag{13}$$

The EAAIS method proposed by Ren reconstructs the subspace by using the difference between the interference and the target to construct suitable judgment conditions[11]. After eliminating the effect of strong interference, the CBF power spectrum output is then performed.

The simulation conditions are set as follows: the receiving array is unchanged, two broadband signals are received and each signal is uncorrelated, the signal-to-noise ratios are 20dB and -30dB, respectively, the incidence angles are 70° and 100°, respectively, and the noise is Gaussian white noise that is uncorrelated with the signals. The results are shown in Figure 7.

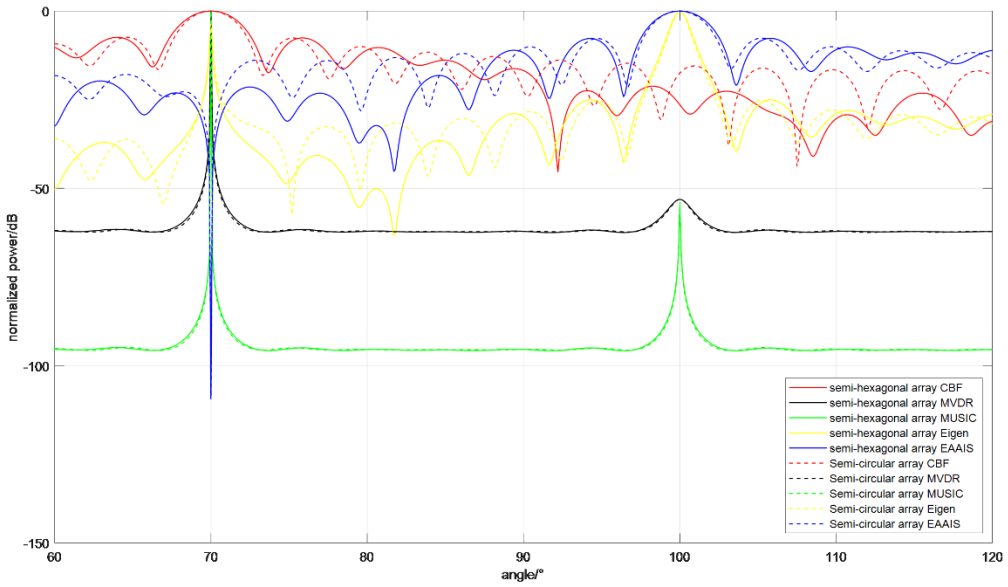


Figure 7 – Comparison of multiple azimuth spectrum estimation algorithms for two arrays

As can be seen from Figure 7, the resolution of the two arrays is approximately the same for both algorithms in the corresponding intervals of interest, which verifies that the resolution of the two arrays is almost the same. Due to the superiority of the MVDR and MUSIC algorithms, there is no significant difference in the sidelobes of the two arrays.The CBF algorithm itself has a higher sidelobe level. At this point, in the region of interest, the sidelobe of the semi-hexagonal array is lower than the sidelobe of the

semi-circular array, which is consistent with the characteristics shown in the orientation diagram.The strong interference in the 70° direction is significantly suppressed in the EAAIS-CBF algorithm. Since the CBF algorithm is still used subsequently, the resolution is not high. The eigenspace algorithm allows for a higher resolution with a lower sidelobe. In the range of interest, the semi-hexagonal array also has better results with lower side flaps than the semi-circular array. The above simulation results show that the reasonable design of array type, combined with excellent algorithms, can lead to better array signal processing. The above simulation results show that the reasonable design of the array type, combined with the excellent algorithm, can make the array signal processing better.

4. CONCLUSIONS

From the 2D orientation map, it can be concluded that both types of arrays can effectively realize 360° omni-directional detection without blurring. In the horizontal dimension, the sidelobe of the hexagonal array is lower than the circular array in most of the area range, and the sidelob of the hexagonal array is higher in a few areas, but the orientation can be predicted. In the pitch dimension, there is no obvious difference between the two types of arrays. Several classical algorithms are simulated. From the normalized power spectrum, in the corresponding interest interval, the resolution of the two arrays is approximately the same regardless of which algorithm. Using MVDR and MUSIC algorithms, there is no significant difference in the sidelobes of the two arrays. In the EAAIS-CBF algorithm, the strong interference in the 70 ° direction is obviously suppressed. The Eigenspace algorithm can obtain higher resolution with lower sidelobes. In addition to MVDR and MUSIC, in the range of interest, the side lobe of the semi-hexagonal array is also lower than that of the semi-circular array, which has better effect. Therefore, this study shows that the reasonable design of the array type, combined with the excellent algorithm, can make the array signal processing better.

ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China （62301182）.

REFERENCES

1. H. Shen, J. Li, K. Hao, Z. Wang, X. Li and D. Yang, "Design and simulation of low-orbit satellite broadcastsignal receiving and processing terminal spherical crown antenna array," 2024 IEEE International Conference on Signal, Information and Data Processing (ICSIDP), Zhuhai, China, 2024, pp. 1-6.
2. H. Huang, H. C. So and A. M. Zoubir, "Sparse Array Beamformer Design via ADMM," 2022 IEEE 12th Sensor Array and Multichannel Signal Processing Workshop (SAM), Trondheim, Norway, 2022, pp. 336-340.
3. S. Kawdungta, D. Torrungrueng, A. Lang and C. Phongcharoenpanich, "Design of Antenna Arrays by Using Modified Fruit Fly Optimization Algorithm," 2021 18th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), Chiang Mai, Thailand, 2021, pp. 69-72.
4. Tesei, A., “Passive acoustic surveillance of surface vessels using tridimensional array on an underwater glider.” Oceans IEEE, 2015.
5. X. Sun and H. Zhao, "A Conical Helix Array For Small Autonomous Underwater Platforms," 2019 IEEE International Conference on Signal, Information and Data Processing (ICSIDP), Chongqing, China, 2019, pp. 1-5.
6. HUANG Jianli, WANG Yu, GONG Zaixiao, WANG Jun, WANG Haibin. Beam-domain deconvolution beamforming algorithm enhanced by compressive sensing[J]. ACTA ACUSTICA, 2025, 50(1): 97-108.
7. YAO Lin, LIU Xiaodong, CAO Jinliang, WANG Shuwen, ZHANG Dongsheng, WANG Yanbin. Weighted beamspace direction-of-arrival estimation based on subarrays[J]. ACTA ACUSTICA, 2020, 45(4): 497-505.
8. LI Tianxing, MO Yaxiao, SU Lin, HU Yuan, CAO Jianguo, MA Li. The high-resolution underwater azimuth estimation algorithm for function-feature subspaces[J]. ACTA ACUSTICA, 2024, 49(5): 956-966.
9. LIANG Guolong, ZHANG Boyu, QI Bin, HAO Yu, DU Zhiyao, LI Xiang. Underwater multitarget fusion tracking method for passive sonar[J]. ACTA ACUSTICA, 2024, 49(3): 501-512.

10. Li C, Xie L, Liu Z, Zhou B, Ma Q. A subspace spatial smoothing-based sparse reconstruction passive direction-of-arrival estimation method under strong interference. J Acoust Soc Am. 2025 Apr 1;157(4):2376-2391.
11. Ren S, Ge F X, Guo X, et al. Eigenanalysis-Based adaptive interference suppression and its application in acoustic source range estimation. IEEE J. Oceanic Eng., 2015; 40(4): 903—916.
12. SHENG Xueli, LI Dewen, CAO Ran, YIN Jiarui, ZHOU Xuan. Strong interference suppression for subspace judgment analysis[J]. ACTA ACUSTICA, 2023, 48(6): 1119-1127.
13. Li C, Liang G, Qiu L, Shen T, Zhao L. An efficient sparse method for direction-of-arrival estimation in the presence of strong interference. J Acoust Soc Am. 2023 Feb;153(2):1257.
14. Du Z, Hao Y, Qiu L, Li C, Liang G. Sparsity-based direction-of-arrival estimation in the presence of near-field and far-field interferences for small-scale platform sonar arrays. J Acoust Soc Am. 2024 Nov 1;156(5):2989-3005.
15. YAN Shefeng, HOU Chaohuan, MA Xiaochuan. Matrix spatial prefiltering approach for direction-of-arrival estimation[J]. ACTA ACUSTICA, 2007, 32(2): 151-157.
16. JIANG Guangyu, SUN Chao, LIU Xionghou, XIE Lei, ZHUO Jie. Minimum variance distortionless response multi-source resolving using inhomogeneous diagonal unloading[J]. ACTA ACUSTICA, 2019, 44(4): 555-565.
17. R. Schmidt, "Multiple emitter location and signal parameter estimation," in IEEE Transactions on Antennas and Propagation, vol. 34, no. 3, pp. 276-280, March 1986.
18. Hossain M D,Mohan A S.Eigenspace time- reversal robust capon beamforming for target localization in continuous random media.IEEE Antennas Wireless Propagat. Lett.,2017;16:1605-1608.

PROCEEDINGS of the
16th International Conference on Theoretical and Computational Acoustics

August 18 to 22, 2025 BEXCO, Busan Korea

On the turbulence and flow noise of a submarine model using high-order scheme

Peng JIANG[1]; Shi Jun LIAO[1]; Bin XIE[1,*]
[1] School of Ocean and Civil Engineering, Shanghai Jiao Tong University, Shanghai, 200240, China

ABSTRACT

This study employs high-fidelity numerical simulations to analyze the turbulent flow and acoustic radiation of the SUBOFF submarine model at Re = 1.2×10^7. We integrate a third-order numerical scheme with Wall-Modeled Large Eddy Simulation (WMLES) and the Ffowcs Williams-Hawkings (FW-H) acoustic analogy to investigate these phenomena. Initial validation against a benchmark turbulent and acoustic airfoil flow demonstrated high consistency in pressure coefficient distributions across WMLES, wall-resolved large eddy simulation (WRLES), and experimental measurements. While WMLES captured instantaneous turbulent structures well, limitations in simulating boundary layer transition were observed. The model's reliability for noise prediction was confirmed by strong alignment between predicted wall pressure fluctuation spectral density and experimental trends. Subsequently, a 103-million-cell computational grid ensures detailed resolution of the flow field, while 648 hydrophones are deployed to capture the far-field acoustics. Our findings reveal that flow separations and vortex shedding significantly impact the hydrodynamic and acoustic behaviors. The interaction between the sail and hull, for instance, leads to intensified surface pressure fluctuations, increasing power spectral density (PSD) amplitudes by up to an order of magnitude. Far-field sound pressure levels exhibit distinct dipole directivity patterns, with peak noise levels reaching 85.10 dB on the central plane. Furthermore, the wake characteristics include complex vortical structures such as horseshoe and necklace vortices. These results provide critical insights into turbulence noise generation for complex underwater geometries, thereby advancing research and supporting noise reduction optimization for underwater vehicles.

Keywords: Turbulent noise, SUBOFF, High-order scheme

* Corresponding author: Dr. Bin Xie (Email: xie.b.aa@sjtu.edu.cn)

PROCEEDINGS of the
16th International Conference on Theoretical and Computational Acoustics

August 18 to 22, 2025 BEXCO, Busan Korea

Modelling and simulation of Ship-Excited Micro-Doppler Effect

Minhao Wang[1],Cuie Zheng[2],Tianye Na[3]

[1]College of Underwater Acoustic Engineering, Harbin Engineering University, Harbin, 150001, China
[2]College of Underwater Acoustic Engineering, Harbin Engineering University, Harbin, 150001, China
[3]College of Underwater Acoustic Engineering, Harbin Engineering University, Harbin, 150001, China

ABSTRACT

The Bragg scattering effect on the ocean surface and the resulting modulated sidebands have been widely studied. However, recent research indicates that the motion of ships under wave excitation also contributes to the modulated sidebands. In this study, the oscillatory motion of a sound source fixed to the ship under various conditions is simulated using the wave spectrum and the ship motion model. The projection of the motion along the line connecting the sound source and the seabed receiver approximates an oscillatory curve composed of sine functions, which varies with time. For a single-frequency tone, the instantaneous frequency (IF) curve of the oscillation is derived based on the Doppler effect, resulting in a distorted signal. Simulation results demonstrate that the wave motion, source motion, and IF may exhibit similar peak frequencies. The frequency-domain analysis reveals modulated sidebands resembling Bessel-like functions, consistent with the findings of M. V. Brown, while time-frequency-domain analysis reveals periodic oscillations in the IF of the received signal. As noted by V. C. Chen, these observations reflect different manifestations of the micro-Doppler effect in the frequency and time-frequency domains. The modulated sidebands may be jointly induced by sea surface scattering and ship motion.

[1]wangminhao@hrbeu.edu.cn
[2]zhengcuie@hrbeu.edu.cn
[3]natianye@163.com

Research on nonlinear Acoustic Detection of Submarine Pipelines Based on 3D Height Field Modeling

Peng-yue WANG[3]; Jian-jun ZHU[1,2,3]; Pei-hong WANG[3,4]; Guo-qing ZHAO[4]
[1] National Key Laboratory of Underwater Acoustic Technology, Harbin Engineering University, Harbin 150001,,China,
[2] Key Laboratory of Marine Information Acquisition and Security (Harbin Engineering University), Ministry of Industry and Information Technology, Harbin 150001,China
[3] College of Underwater Acoustic Engineering, Harbin Engineering University, Harbin 150001, China
[4] Sanya Nanhai Innovation and Development Base of Harbin Engineering University, Sanya, China

ABSTRACT
This paper focuses on the research of low-frequency and high directional parametric array sound source detection technology for high-resolution detection of underwater pipelines in different states (suspended, semi buried, fully buried). A three-dimensional height field model integrating ship speed, sonar ping rate, submarine topography, and pipeline geometric characteristics (radius, burial depth, and angle relative to survey lines) is established. Specifically, the seafloor roughness is characterized using Gaussian power spectral density functions parameterized by root mean square height and correlation length. Numerical simulations of nonlinear acoustic detection processes for submarine pipelines are conducted based on Kirchhoff approximation and perturbation theory, generating sub-bottom profile detection images. Systematic analysis is performed on the detection effectiveness of parametric array sonar for suspended, partially buried, and fully buried pipelines. Research results demonstrate that the quantitative relationship between terrain root mean square height and roughness is revealed, and the three-dimensional height field modeling effectively characterizes spatial structures of submarine pipelines under different states, numerical simulation achieves high-directional acoustic source detection for submarine pipelines. This study provides an effective technical approach for optimizing the design of nonlinear acoustic detection systems and developing detection methodologies for buried pipelines in practical engineering applications.

Keywords: 3D height field modeling, Submarine Pipelines, Nonlinear Acoustic Detection

1. INTRODUCTION

As critical infrastructure for offshore oil/gas transportation, monitoring submarine pipeline integrity represents a key challenge in marine engineering [1][2]. Accurate identification of burial states (exposed, semi-buried, fully buried) is essential for operational safety [3][4]. Complex seabed environments pose significant acoustic detection challenges: Conventional sonar faces inherent limitations in balancing low-frequency sediment penetration against high-resolution target identification, while multipath scattering and reverberation noise from rough seabeds further degrade signal-to-noise ratios, obscuring weak scattering signatures of buried pipelines. Existing studies predominantly rely on idealized terrain assumptions (flat seabeds, homogeneous media), lacking systematic investigation of realistic seabed roughness and multi-state pipelines (free-spanning, semi-buried, fully buried).

This study develops a detection modeling approach integrating nonlinear acoustic theory with 3D height-field construction. Leveraging parametric array sonar's nonlinear properties, high-frequency primary waves undergo self-demodulation to generate low-frequency, highly directional beams [5][6][7], overcoming conventional sonar's penetration-resolution tradeoffs. Coupled with Gaussian power spectrum parameterization, we establish a dynamically adjustable seabed roughness model (root mean square height δ, correlation length l) simulating key geomorphic features (sandy beds, sand waves, bedrock). Combining Kirchhoff approximation and perturbation theory yields a unified mathematical representation of interface scattering, sediment volume scattering, and pipeline target scattering. Through parametric optimization of pipeline geometry (radius, burial depth, orientation) and sonar settings, we investigate high-resolution sub-bottom profiling mechanisms using the parametric array's dual-frequency advantages. This parametric design

achieves quantitative characterization of complex terrains and multi-state pipelines, providing an adaptable theoretical framework for subsea pipeline detection sonar development and monitoring technology research.

2. CONSTRUCTION OF TARGET HEIGHT FIELD MODEL

2.1 Construction of Three-dimensional Underwater Height Field

The modeling of underwater terrain height field is the mathematical foundation of acoustic detection, and its core lies in parameterizing the characterization of seabed surface morphology through spectral models. In response to the challenge of scarce measured data, this paper chooses the Gaussian power spectrum model to quantitatively characterize the roughness characteristics of the seabed through two parameters: root mean square height (δ) and correlation length (l). Compared to power-law spectral models (which require manual truncation of wave numbers to avoid scattering integral divergence), Gaussian spectra have the following advantages:

(1) Physical interpretability: Single scale rough features can be fully described by two parameters: root mean square height and correlation length l;

(2) Numerical stability: The exponential decay characteristic naturally satisfies the high-frequency truncation condition, avoiding the divergence problem in scattering field calculations;

(3) Theoretical compatibility: Its smooth attenuation characteristics have inherent consistency with the gentle phase assumption of Kirchhoff approximation.

Although composite roughness models, such as the Kirchhoff perturbation hybrid method, can theoretically handle multi-scale terrain, their complex boundary matching conditions can significantly reduce computational efficiency. Thorsos proved by comparing the exact solutions of integral equations that when the surface curvature radius is greater than the wavelength of the sound wave, the Kirchhoff approximation supported by Gaussian spectrum can maintain a scattering intensity prediction accuracy of better than 90% [8]. This conclusion provides a theoretical basis for using Gaussian spectra to generate random underwater surfaces in this article - while ensuring computational feasibility, it can still effectively capture the dominant influence of the main scale terrain on the acoustic scattering field.

The height of a random rough surface can be regarded as the superposition of multiple harmonics of different frequencies, and therefore a random rough surface model can be generated through linear filtering. The specific method is to filter in the frequency domain and then obtain the height of the rough surface through inverse Fourier transform. The amplitude of each harmonic can be represented by Monte Carlo method using independent Gaussian random variables .

From the above analysis, the two-dimensional rough height field with length L can be expressed as [9]:

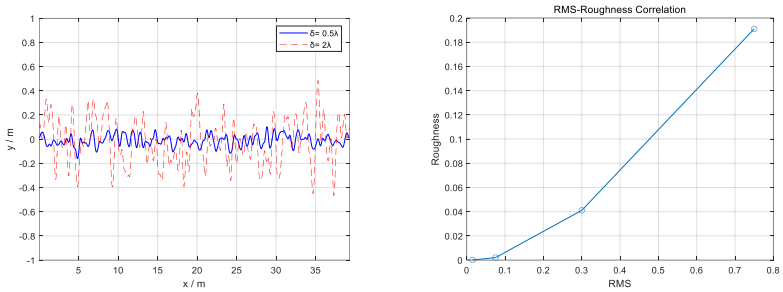
$$f\left(x_n\right)=1/L\times\sum_{j=-N/1+1}^{N/2}F\left(k_j\right)exp\left(ikx_n\right) \tag{1}$$

In the formula, $x_n=n\Delta x\left(n=-N/2+1,\ldots,N/2\right)$ is the nth sampling point, $F\left(k_j\right)$ and $f\left(x_n\right)$ is a Fourier transform pair, which can be expressed as:

$$F\left(k_j\right)=2\pi/\sqrt{2\Delta k}\times\sqrt{s\left(k_j\right)}\times\begin{cases}\left[N\left(0,1\right)+iN\left(0,1\right)\right], j=-N/2+1,\ldots,-1\\ N\left(0,1\right), j=0,..,N/2\end{cases} \tag{2}$$

When j>0, $F\left(k_j\right)=F\left(k_j\right)^*$,Among them, $k_j=2\pi j/L$, Δk is the wave number difference, $s\left(k_j\right)$ is the power spectral density of the rough surface, and the Gaussian distribution power spectral density is $s\left(k_j\right)=\left(\delta^2l/2\sqrt{\pi}\right)\times exp\left(-{k_j}^2l^2/4\right)$, $N\left(0,1\right)$ is a Gaussian random number. The above formula can be used to simulate a two-dimensional height field.

Based on a two-dimensional height field model, set sampling points (Nx=10240), correlate sonar parameters (sound velocity c=1500m/s, frequency f=10kHz, detection distance H=10m, beam opening angle θ=5 °) with terrain feature parameters (correlation length l, root mean square height δ), and establish a surface length L=100 × 2H × tan θ (θ is beam opening angle, 100 is scaling factor). By fixing the relevant length l=λ, the results of simulating rough seabed with root mean square heights of δ=0.5 λ and δ=2 λ are shown in Figure 1.



(a) 2D Seabed RMS Height Variations (b) RMS-Roughness Correlation

Figure 1 –Rough seabed simulation results

The root mean square height reflects the standard deviation of surface height distribution and is an important indicator for measuring the degree of surface roughness. Adjusting the root mean square height can effectively simulate seabed terrain with different roughness levels. A higher root mean square height will result in greater fluctuations, thereby increasing the scattering intensity and complexity of sound waves on the seabed surface. This not only affects the reflection and scattering characteristics of sound waves, but also significantly interferes with the echo characteristics of sonar signals.

first, The three-dimensional random rough surface model with a length of Lx in the X direction and Ly in the Y direction is represented as：

$$f(x_m, y_n) = \frac{1}{L_x L_y} \sum_{m_k=-M/2+1}^{M/2} \sum_{n_k=-N/2+1}^{N/2} F(K_{mk}, K_{nk}) \times exp[i(K_{mk} x_m + K_{nk} y_n)] \tag{3}$$

Among them, $x_m=m\Delta x$ and $y_n=n\Delta y$ respectively represent the m-th (n-th) sampling point in the x and y directions

$$F(K_{mk}, K_{nk}) = 2\pi[L_x L_y W(K_{mk}, K_{nk})]^{1/2} \times \begin{cases} 1/\sqrt{2} \times [N(0,1) + iN(0,1)], m_k \neq 0, \ldots, M/2; n_k \neq o, \ldots, N/2 \\ N(0,1), m_k = 0, \ldots, M/2; n_k = o, \ldots, N/2 \end{cases} \tag{4}$$

Among them, *W(Kmk, Knk)* is the power spectral density of a two-dimensional random rough surface, $K_{mk}=2\pi m_k/L_x$, $K_{nk}=2\pi n_k/L_y$ are wave numbers.

$$W(K_{mk}, K_{nk}) = \delta^2 \frac{l_x l_y}{4\pi} exp[-(k_x{}^2 l_x{}^2 + k_y{}^2 l_y{}^2)/4] \tag{5}$$

Among them, *lx* and *ly* are the relevant lengths in the X and Y directions, respectively, and *δ* is the root mean square height.

Based on a three-dimensional random rough surface model, simulate the seabed height field with a spatial grid number Nx=Ny=128, sonar parameters c=1500m/s, f=10kHz, λ=0.15m, H=10m, beam opening angle θ=5 ◦, and surface length Lx=Ly=100 × 2Htan θ (where 100 is the beam footprint ratio coefficient). Set the terrain parameters lx=ly=2λ and root mean square height δ=λ, and generate a three-dimensional seabed height field as shown in Figure 2.

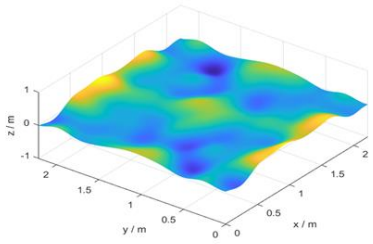


Figure 2 –3D Seabed Topography Model

The root mean square height (δ), as the core parameter for height field modeling, can effectively characterize the undulating characteristics of seabed topography. By comparing the measured data of typical seabed topography [10] [11], it was found that the δ values of different landforms showed significant differences in regularity (Figure 3):

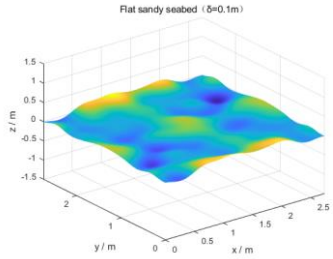
(1) Flat sandy seabed: δ is mostly distributed in the range of 0.05-0.15 m, and its height field exhibits low undulating characteristics. This type of terrain is commonly found in low-energy sedimentary environments

and is a typical background field for semi buried pipelines.

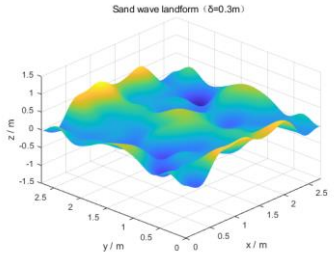
(2) Sand wave landform: The sand wave landform formed by hydrodynamic action has the characteristic of δ=0.2-0.5m, showing obvious periodic fluctuations. Research has shown[2] that such parameter combinations can effectively reproduce the forward enhanced scattering effect of sand wave terrain on sound waves.

(3) Bedrock outcrop area: δ can reach 0.3-1.2 m, and its height field includes steep local undulations, which need to be described using a composite roughness model.

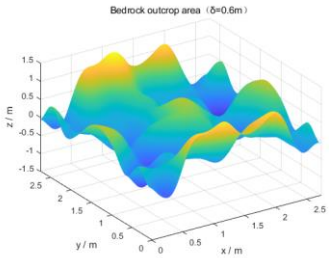
This article selects a parameter value of δ=0.15 m (corresponding to a sound wave wavelength of λ=0.15 m), which is consistent with the measured characteristics of nearshore sandy silt mixed seabeds [13]. By adjusting the δ value (0.05-0.6 m) through the system, typical scenarios such as flat seabed (δ=0.1 m) and moderate sand waves (δ=0.3 m) can be simulated separately. This parametric modeling method not only maintains the mathematical advantages of Gaussian spectrum, but also reproduces key terrain features through physical parameter adjustment.



(a) Flat sandy seabed



(b) Sand wave landform



(c) Bedrock outcrop area

Figure 3 –Underwater maps of different landforms

2.2 Construction Of Subsea Pipeline Height Field

The height field of the submarine pipeline is related to factors such as the distance *Hp* between the pipeline axis and the transducer, the installation angle θ (relative to the x-axis direction), the pipeline radius *rp*, the speed, and the detection ping rate. These factors determine the changes in depth and shape at the pipeline height location. The top view of constructing a three-dimensional underwater pipeline height field model is shown in Figure 4.

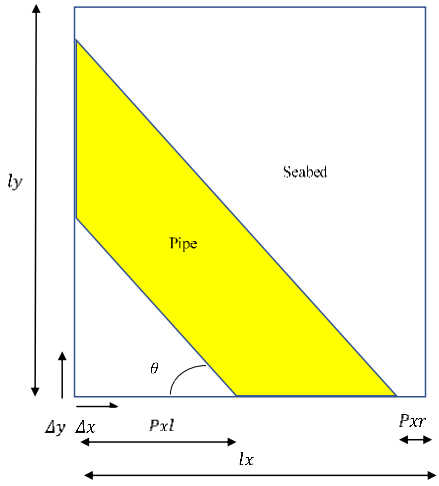


Figure 4 –Top view of 3D seabed height field model

Among them, the scale of the detection area (*lx*, *ly*), the number of grids (*Nx*, *Ny*), and the spatial step size

(Δx=lx/Nx, Δy=ly/ny) are defined. Combined with sonar parameters and scattering modeling requirements, the beam footprint scale *Lfp* and its multiple parameter multi are introduced, and the distance from the lower edge of the detection area to the left and right edges of the pipeline (*Pxl*, *Pxr*) is defined. Based on the geometric relationship of the model diagram, the expressions of the relevant parameters are derived as follows:

$$ly = lx = multi \times l_fp \tag{6}$$

$$\Delta x = lx/Nx \tag{7}$$

$$\Delta y = ly/Ny \tag{8}$$

$$P_{x_l} = lx/2 \tag{9}$$

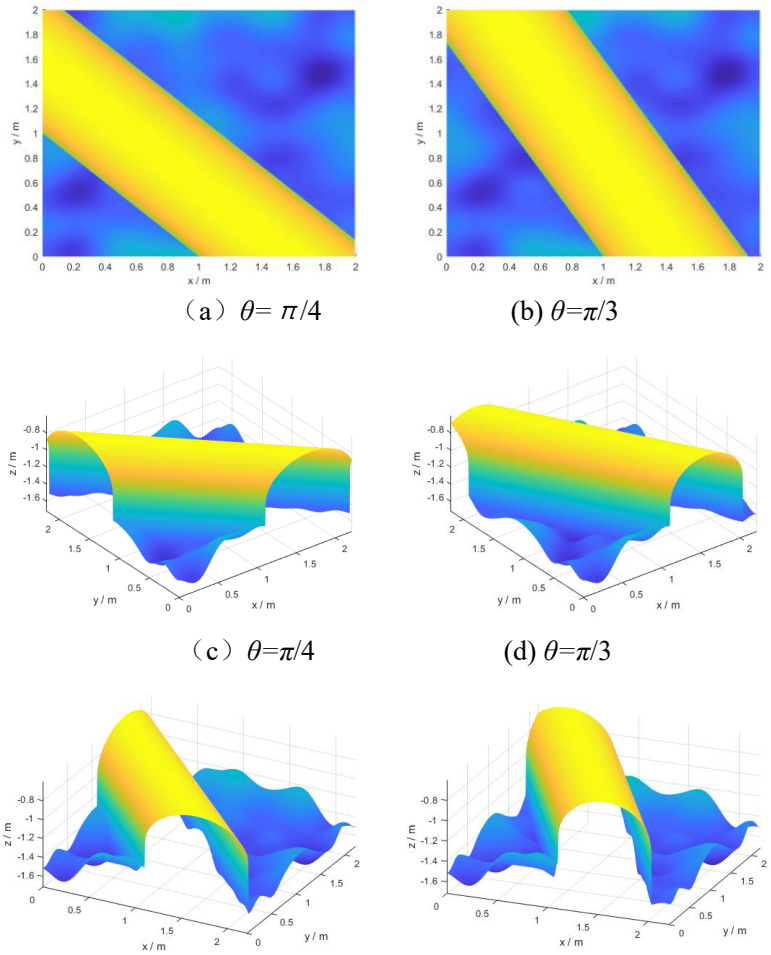
$$P_{x_r} = P_{x_l} + 2 \times r_p / \sin(\theta) \tag{10}$$

$$\Delta lp = r_p - (n-1) \times \Delta x \times \sin(\theta) \tag{11}$$

$$\Delta H = \sqrt{(r_p{}^2 - \Delta lp^2)} \tag{12}$$

Based on the three-dimensional seabed height field model, set the number of sample points *Ny=Nx*=384. Here, the surface lengths *lx* and *ly* in two directions are related to the number of sample points *Nx* and *Ny*, as well as the step sizes *Δx* and *Δy* in x and y directions, and to the pipeline radius (*rp*) and beam footprint scale. Obtain three views of the underwater pipeline height field (Figure 5).

The yellow area represents the pipeline section, and the blue area represents the seabed area. By changing the layout angle of the relationship, three-dimensional seabed pipeline height fields with different trends are generated, which facilitates the analysis of detection results for different pipeline trends in nonlinear acoustic detection of submarine pipelines.



（a）$\theta = \pi/4$ (b) $\theta=\pi/3$

（c）$\theta=\pi/4$ (d) $\theta=\pi/3$

（e）$\theta=\pi/4$ (f) $\theta=\pi/3$

Figure 5 –3D Pipeline Height Field (Tri-View)

Based on the three-dimensional height field model of submarine pipelines, further research is conducted

on the existence of different states of submarine pipelines, such as suspended, semi buried, and fully buried. Modeling and analysis are carried out for these three states of pipelines to explore their morphological characteristics under different burial conditions. As shown in Figure 6.

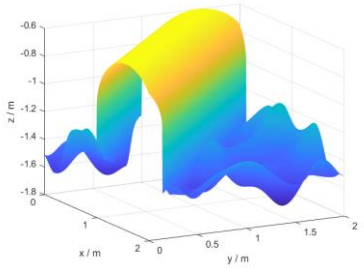


Figure 6 –3D Seabed Height Field of Suspended Subsea Pipeline

When detecting exposed underwater pipelines, due to the lack of cover and the influence of seabed sediments, the noise and seabed reverberation interference during sound wave propagation and scattering are relatively small, which can form clearer echo signals.

Semi buried submarine pipelines are common in practical marine engineering, and their surface cover and burial degree are influenced by various factors such as topography, seabed flow, and sediment accumulation. The height field of the semi buried submarine pipeline is shown in Figure 7.

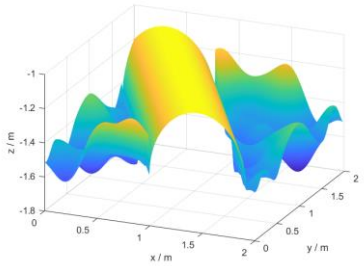


Figure 7 –3D Seabed Height Field of Partially Buried Subsea Pipeline

This type of pipeline not only has a certain buried protection effect, but also partially exposes itself. During the propagation and scattering of sound waves, noise and interference are more complex than in a suspended state.

In actual ocean engineering, to ensure the safety and reliability of submarine pipelines, burial measures such as sediment burial or pipeline covering are usually taken. Although these measures can effectively protect the pipeline, they also increase the difficulty of acoustic detection. Initialize the bottom height field of the seabed and the formation where the pipeline is located, and superimpose the constructed seabed pipeline height field to obtain the seabed height field containing buried pipelines. As shown in Figure 8.

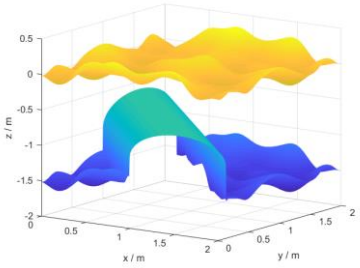


Figure 8 –3D Seabed Height Field of Fully Buried Subsea Pipeline

By adjusting the burial angle (θ), pipeline radius (rp), and the depth of the pipeline axis from the transducer (Hp), the shape of the underwater pipeline height field can be flexibly adjusted. This real-time configuration capability enables research to be more closely aligned with the actual pipeline status, providing data support that is more closely aligned with the actual detection situation.

3. NUMERICAL CALCULATION METHOD FOR SCATTERED SOUND FIELD

3.1 Calculation of Scattering Sound Field on the Underwater Surface

Kirchhoff approximation is generally considered to be applicable to surfaces with "slight fluctuations", more specifically to surfaces with curvature radii greater than wave length. Due to the fact that the Kirchhoff approximation is not limited to a small value of δ/λ (where δ is the root mean square height of the surface

and wavelength), it has been developed for applications in the medium to high frequency range. The main condition for the effectiveness of Kirchhoff approximation is that if the surface roughness is smooth enough in the horizontal dimension to avoid masking effects and interference from multiple reflected waves on the surface, it can be considered that the reflection of each point on the seabed surface is equivalent to the reflection of the cross-section of that point. This assumption was proposed by Brekhovskikh [14], which states that under a given incident angle θ inc and wave number k0 ($2\pi/\lambda$), the minimum interface curvature radius rc must satisfy the following conditions.

$$2k_0 r_c \cos(\theta_{inc}) \gg 1 \tag{13}$$

The reason for setting this restriction is to avoid the occurrence of a large number of sharp edges or points with small curvature radii, which can result in edge or masking effects. Even for small wave numbers, as long as the relevant length of the seabed surface is greater than the wavelength λ [15], the tangent plane approximation is effective. During the calculation phase, it is possible to check and exclude the influence of shielding effects or sound rays that may be incident on the seabed at angles exceeding the critical range Assuming that the sound source is located at a height H (P={0,0,H}) from the seabed surface, emits a sine wave with a frequency of f, has a sound pressure level of p0, a far-field directivity Di (R0), and a distance R0=| R-P | from the scattering point, the Kirchhoff sound scattering model is shown in Figure 9.

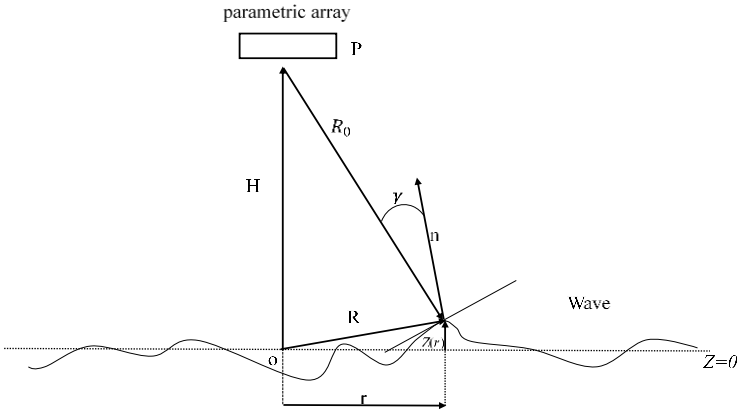


Figure 9 –Underwater sound scattering model

Representing the three-dimensional undulating seafloor surface as a seafloor terrain composed of surface elements, with each seafloor surface element tangent to the seafloor undulating surface; Within the range of the sound beam footprint, each seafloor element scatters the incident sound waves based on its physical characteristics such as distance from the sound source, beam width, angle between the seafloor element and the seafloor surface, density and sound velocity of the water body and seafloor sediment layer. The receiving transducer obtains all scattered sound waves from seafloor elements and sums them up in the time domain to obtain the time-domain observation field of seafloor scattered sound waves. The boundary conditions for point R on the seabed are as follows [16].

$$p_r(R) = \mathcal{R}_{01}(R_0, n) P_i(R) \tag{14}$$

$$\frac{\partial}{\partial n} P_r(\boldsymbol{R}) = -\mathcal{R}_{01}(\boldsymbol{R}_0, \boldsymbol{n}) \frac{\partial}{\partial n} P_i(\boldsymbol{R}) \tag{15}$$

Assuming that the water distribution above the seabed interface is uniform and the sound velocity distribution is constant at c0, in this case, Chotiros' proof [17] can be followed, and the incident sound pressure field received at point R can be expressed under approximate far-field conditions as:

$$p_i(\boldsymbol{R}) = p_0 G_i(\boldsymbol{R}_0) D_i(\boldsymbol{R}_0) \tag{16}$$

Among them, Gi (R0) is the Green's function of the incident sound pressure field.

$$G_i(R_0) = \frac{e^{-jk_0 R_0}}{R_0} \tag{17}$$

The expression for the radiated sound pressure at point P generated by the area element ds_R at point R is

$$dp_s(\boldsymbol{P}) = j \frac{\rho_0 c_0 k_0}{2\pi} v_r(R) G_r(\boldsymbol{R}_0) D_r(\boldsymbol{R}_0) d\boldsymbol{S}_R \tag{18}$$

The scattering observation sound field of time-domain area microelements is

$$dp_s(\boldsymbol{P},t)=\frac{\cos(\gamma(\boldsymbol{R}_0,\boldsymbol{n}))}{2\pi c_o R_0^{\,2}}\,p_0\left(D_i(\boldsymbol{R}_0)D_r(\boldsymbol{R}_0)\right)\times\mathcal{H}_{01}(\boldsymbol{R}_0,\boldsymbol{n})e^{'}\left(t-\left(\frac{2R_0}{C_0}\right)\right)d\boldsymbol{S}_{\boldsymbol{R}} \tag{19}$$

Sound field of backscattering from the seabed surface with an area of S

$$p_s(\boldsymbol{P},t)=\int_{(s)}dp_s(P,t) \tag{20}$$

3.2 Calculation of Subsea Volume Scattering Field

The analytical form of the acoustic pressure field received by point P (see Figure 10) from the seafloor volume element is obtained based on Ivakin's volume scattering method, and the expression of the local volume backscattering acoustic pressure field is derived based on Chernov's perturbation theory, which takes into account the changes in sound velocity and density in the seafloor sediment layer. At this time, the construction of the acoustic pressure field in the domain also follows the same principle as seabed modeling, but only applies the classical small perturbation theory within the volume range. The calculation model of underwater volume scattering sound field is shown in Figure 10.

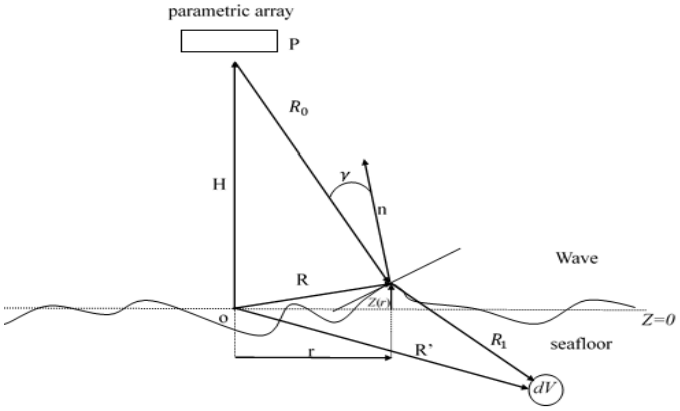


Figure 10 –Underwater volume scattering acoustic field model

Firstly, assuming an incident sound pressure field passes through the seabed and still follows the tangent plane approximation, it is expressed as:

$$p_t(R)=\left(1+\mathcal{H}_{01}(R_0,n)\right)p_i(R) \tag{21}$$

$$\frac{\partial}{\partial n}p_t(\boldsymbol{R})=\left(1+\mathcal{H}_{01}(\boldsymbol{R}_0,\boldsymbol{n})\right)\frac{\partial}{\partial n}p_i(\boldsymbol{R}) \tag{22}$$

The representation of the sound pressure field passing through the water sediment interface at point R is as follows:

$$p_t(\boldsymbol{R})=p_0G_i(\boldsymbol{R}_0)D_i(\boldsymbol{R}_0)T_{01}(\boldsymbol{R}_0,\boldsymbol{n}) \tag{23}$$

The incident sound pressure at point R is represented by the tangent plane approximation based on the Kirchhoff approximation principle, but the analytical form of the sound pressure field from the seabed volume element that passes through the seabed and scatters back to point P through the sedimentary layer is based on Ivakin's volume scattering method. The expression of the local volume backscattering sound pressure field is based on Chernov's perturbation theory, which takes into account the changes in sound velocity and density in the seabed sediment layer.

Scattering observation sound field of volume microelements:

$$\begin{aligned}dp_v(P,t)=&\frac{-n_1^2(\boldsymbol{R}')}{2\pi R_0^2c_0^2}\mu(\boldsymbol{R}')p_0D_i(\boldsymbol{R}_0)D_r(\boldsymbol{R}_0)T_{01}(\boldsymbol{R}_0,\boldsymbol{n})T_{10}(\boldsymbol{R}_1,-\boldsymbol{n})\\&\times\mathcal{F}^{-1}\left\{e^{-\alpha|f|R_1}\right\}*e^{''}(t-2((\overline{n_1}R_1+R_0)/c_0))dV_R\end{aligned} \tag{24}$$

The volume scattering sound field corresponding to the seabed surface with an area of S:

$$p_v(P,t)=\int_{(v)}dp_v(\boldsymbol{P},t) \tag{25}$$

3.3 Calculation of Scattering Sound Field in Submarine Pipeline

Combining the underwater scattering sound field model, the Kirchhoff approximation principle is applied to study the scattering field calculation of underwater pipelines in different states. The effectiveness of

Kirchhoff approximation depends on the fact that the relevant length l of the seabed surface is greater than the wavelength λ of the sound wave, that is, it satisfies the tangent plane approximation condition. In underwater pipeline modeling, the relevant length of the pipeline (i.e. pipeline radius rp) is set to be much greater than the wavelength λ, so Kirchhoff approximation is suitable for this application.

During the calculation process, it is necessary to check and eliminate the influence of shielding effects or sound rays that may be incident on the seabed at angles exceeding the critical range. The scattered pressure field on the surface of the pipeline is discretized into multiple scattering points, and the incident angle, reflection angle, and reflection coefficient of each scattering point are calculated and superimposed to obtain the total scattered sound field. Assuming the detection direction is perpendicular to the pipeline layout angle, the surface scattering model of the submarine pipeline is shown in Figure 11.

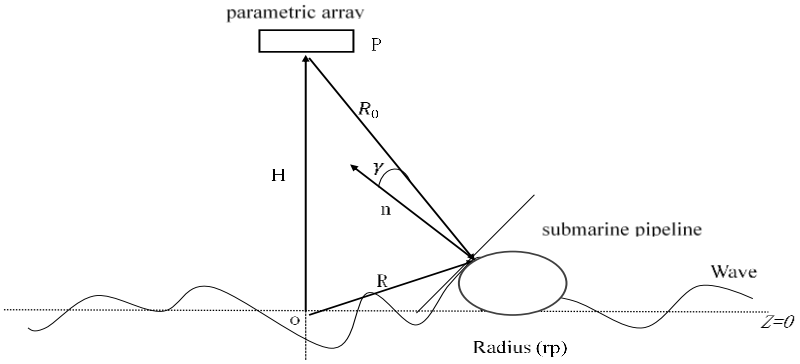


Figure 11 –Surface scattering model of submarine pipelines

Based on the seabed scattering model and combined with the geometric relationship of underwater pipelines, the calculation of the gamma angle has been simplified. By using Snell's law of refraction, the refraction angle of the scattering point is derived, and key parameters such as pipeline material density (*p*), pipeline radius (*rp*), burial depth (*Hp*), and sound velocity of the surrounding medium (*Cs*) are combined to calculate the reflection coefficient of the pipeline sediment layer interface $\mathcal{R}_{12}(\mathbf{R}_0, \mathbf{n})$.

On this basis, the calculation method for the backscattering pressure field on the surface of the submarine pipeline is the same as that for the scattering pressure field on the seabed surface, both based on Kirchhoff surface scattering theory. Therefore, the scattering pressure field of submarine pipelines is

$$p_{rp}(\mathbf{R}) = p_0 G_i(\mathbf{R}_0) D_i(\mathbf{R}_0) \mathcal{R}_{12}(\mathbf{R}_0, \mathbf{n}) \tag{26}$$

The expression for the radiated sound pressure at point P generated by the area element at point R is

$$dp_{\mathrm{sp}}(\boldsymbol{P}) = j \frac{\rho_0 c_0 k_0}{2\pi} v_r(R) G_r(\boldsymbol{R}_0) D_r(\boldsymbol{R}_0) d\boldsymbol{S}_R \tag{27}$$

Among them, Gr (R0) is the Green's function of the reflected sound pressure field, and Dr (R0) is the directivity of the receiver located at point P.

The scattering observation sound field of time-domain area microelements is

$$dp_{\mathrm{sp}}(\boldsymbol{P}, t) = \frac{\cos(\gamma(\boldsymbol{R}_0, \boldsymbol{n}))}{2\pi c_o {R_0}^2} p_0 \left(D_i(\boldsymbol{R}_0) D_r(\boldsymbol{R}_0) \right) \times \mathcal{R}_{12}(\boldsymbol{R}_0, \boldsymbol{n}) e' \left(t - \left(\frac{2R_0}{Cs} \right) \right) d\boldsymbol{S}_{\boldsymbol{R}} \tag{28}$$

Sound field of backscattering from a submarine pipeline with an area of S

$$p_{\mathrm{sp}}(\boldsymbol{P}, t) = \int_{(s)} dp_{\mathrm{sp}}(P, t) \tag{29}$$

This article studies the time-domain observation sound field model for underwater exploration, which is based on the condition of vertical detection of parametric array detection sound beams, and considers the influencing factors of highly directional detection sound beams in the calculation of underwater scattering sound fields. Figure 12 shows a schematic diagram of the construction principle of the time series model of the seabed surface, seabed sediment layer volume scattering, and seabed target pipeline under vertical incidence. At a given time (t), the seabed and target scattering field is composed of three parts: seabed surface scattering, seabed sediment layer volume scattering, and seabed pipeline scattering.

By constructing a scattering acoustic field model of the seabed and targets through scattering from the seabed surface, seabed volume, and seabed pipelines, it is easy to understand the detection parameters and physical processes of the interaction between sound waves and the seabed in the nonlinear acoustic detection process of the parametric array. Various parameter elements of the detection are considered as much as

possible, and the main factors affecting the time series of seabed scattering are explained and simulated more accurately and realistically.

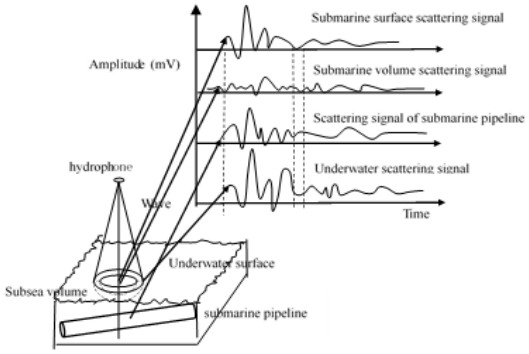


Figure 12 –Signal model of underwater and target scattering sound field

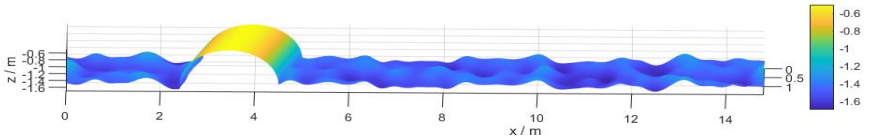
The total time-domain sound pressure field received at sound source P is the sum of the sound pressure field scattered from the seabed surface, the sound pressure field scattered from the seabed volume, and the sound pressure field scattered from the seabed pipeline. Specifically, it is represented as

$$p(P,t) = p_s(\boldsymbol{P},t) + p_v(P,t) + p_{sp}(\boldsymbol{P},t) = \int_{(s)} dp_s(\boldsymbol{P},t) + \int_{(v)} dp_v(\boldsymbol{P},t) + \int_{(s)} dp_{sp}(P,t) \quad (30)$$

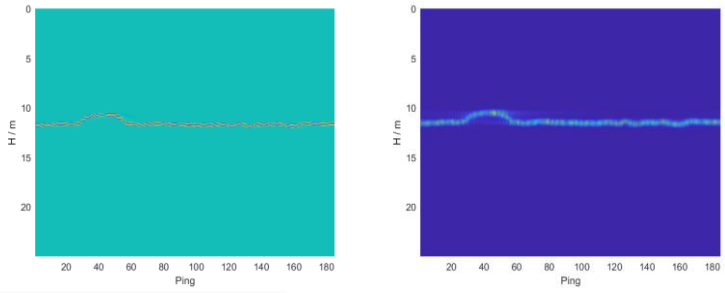
4. SIMULATION OF SUBSEA PIPELINE DETECTION

Generate shallow section detection images of submarine pipelines through numerical simulation, and verify the effectiveness of acoustic field calculation methods and three-dimensional height field models.

Set detection parameters: water depth H_W=10m, sediment layer thickness H_S=1.5m, adjust the distance between the pipeline axis and the transducer depth Hp=11.5m, pipeline radius rp=1m, and use continuous wave (CW) detection signal. The shallow section image of the exposed underwater pipeline is shown in Figure 13.



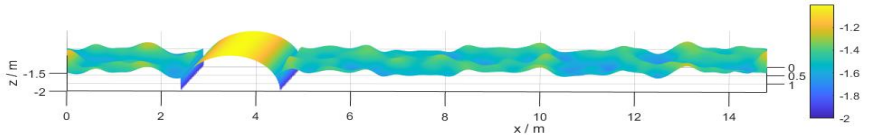
(a) 3D Seabed Height Field of Suspended Subsea Pipeline (H_p=11.5m,r_p=1m)



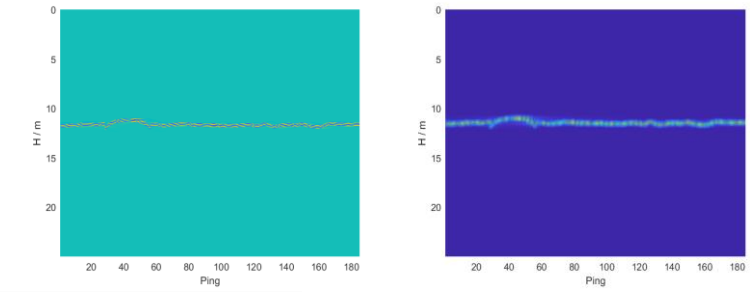
(b) Raw Echo Image (c) Coherently Processed Image

Figure 13 –Shallow geological profile of submarine pipeline in exposed state

Keeping other parameters constant, when changing the distance between the pipeline axis and the depth of the transducer, a shallow section detection profile image of a semi buried submarine pipeline is obtained (Figure 14).



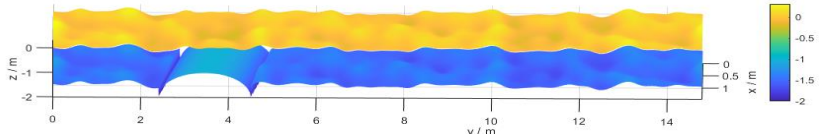
(a) 3D Seabed Height Field of Partially Buried Subsea Pipeline (H_p=12m,r_p=1m)



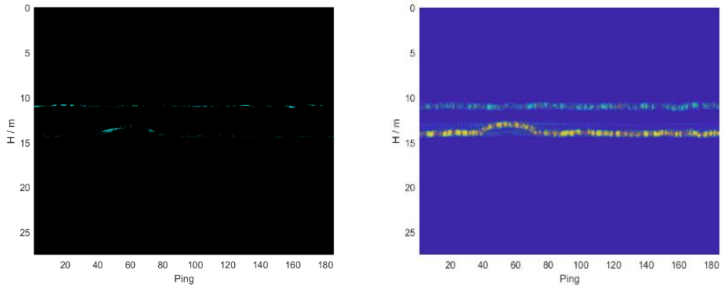
(b) Raw Echo Image (c) Coherently Processed Image

Figure 14 –Shallow geological profile of submarine pipeline in semi buried state

Keeping other parameters constant, adjust the distance between the pipeline axis and the depth of the transducer to obtain the shallow section detection profile image of the buried submarine pipeline as shown in Figure 15.



(a) 3D Seabed Height Field of Fully Buried Subsea Pipeline(H_p=15m,r_p=1m)



(b) Raw Echo Image (c) Coherently Processed Image

Figure 15 –Shallow geological profile of buried submarine pipelines

By adjusting the pipeline radius (*rp*), the distance between the axis and the transducer depth (*Hp*), the detection water depth (H_W), and the sediment layer thickness (H_S), shallow section images of submarine pipelines in different burial states (exposed, partially buried, fully buried) can be simulated. In field exploration, simulation conditions can be set according to the actual environmental parameters of the seabed to achieve multi scenario pipeline structure simulation. The simulation results show that:

Bare pipeline: completely exposed to the seabed interface, with high intensity of sound wave reflection signal. The edge contour of the pipeline is clear in the shallow section image, and the curvature characteristics are significant (Figure 13), forming a sharp contrast with the surrounding terrain.

Semi buried pipeline: partially embedded in sedimentary layers and partially exposed at the seafloor interface, its acoustic characteristics are easily confused with the surrounding terrain (such as sand waves and bedrock), especially in sandy or muddy sedimentary layers, where enhanced sound wave scattering leads to reduced target recognition (Figure 14).

Fully buried pipeline: The coverage of sedimentary layers causes attenuation of sound wave energy and multi-directional scattering, resulting in weak echo signals and low signal-to-noise ratio. The use of copy correlation processing technology can effectively suppress seabed roughness noise and multipath interference, improve imaging resolution and target contrast, and significantly present pipeline edge features and burial depth in shallow section maps (Figure 15).

The above results verify the effectiveness of the three-dimensional height field model and Kirchhoff approximation acoustic field calculation method, providing a reliable theoretical tool for the positioning, depth calculation, and structural damage assessment of exposed, semi buried, and fully buried pipelines.

5. CONCLUSIONS

By constructing a 3D height field model of the seafloor and pipelines, and applying the Kirchhoff approximation and perturbation theory, this study simulates the acoustic scattering field and shallow

subsurface imaging of pipelines in different burial states (exposed, semi-buried, fully buried). The Gaussian power spectrum parameterization (using RMS height δ and correlation length l) effectively represents typical seabed features like sand, sand waves, and bedrock outcrops, and the model's validity is confirmed through high-resolution imaging. The developed numerical method allows flexible adjustment of pipeline geometry (radius, burial depth, deployment angle) and sonar parameters (frequency, beam width), enabling adaptation to various scenarios. Specifically, changing the pipeline-to-transducer depth (Hp) simulates different burial states, while varying the deployment angle (θ) helps analyze scattering from different pipeline orientations.

This study provides an effective technical approach for the optimization design of nonlinear acoustic detection systems and the research of underwater pipeline detection methods in practical engineering applications. The framework of this method can also be extended to the field of multi-source acoustic data fusion, further enhancing the applicability of underwater target detection technology in dynamic marine environments and providing technical support for underwater target detection and observation in complex underwater environments.

ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China [grant numbers 42176188] and the Stable Supporting Fund of Acoustic Science and Technology Laboratory [numbers JCKYS2024604SSJS004].

REFERENCES

1. Schock S G, Tellier A, Wulf J, et al. Buried object scanning sonar[J]. IEEE journal of oceanic engineering, 2002, 26(4): 677-689.
2. Manik H M. Underwater acoustic detection and signal processing near the seabed[J]. Sonar Systems, 2011: 255-274.
3. Junejo N U R, Sattar M, Adnan S, et al. A survey on physical layer techniques and challenges in underwater communication systems[J]. Journal of Marine Science and Engineering, 2023, 11(4): 885.
4. Zhang H, Zhang S, Wang Y, et al. Subsea pipeline leak inspection by autonomous underwater vehicle[J]. Applied Ocean Research, 2021, 107: 102321.
5. Zhu J, Zhou T, Li T, et al. High-resolution sub-bottom profiling technology using parametric array and vector hydrophone[J]. Applied Acoustics, 2024, 223: 110077.
6. Lavia E F, Gonzalez J D, Blanc S. Modeling high-frequency backscattering from a mesh of curved surfaces using Kirchhoff approximation[J]. Journal of Theoretical and Computational Acoustics, 2019, 27(04): 1850057.
7. Olson D R. A series approximation to the Kirchhoff integral for Gaussian and exponential roughness covariance functions[J]. The Journal of the Acoustical Society of America, 2021, 149(6): 4239-4247.
8. Thorsos E I. The validity of the Kirchhoff approximation for rough surface scattering using a Gaussian roughness spectrum[J]. The Journal of the Acoustical Society of America, 1998, 83(1):78-92.
9. Briggs K B. Microtopographical roughness of shallow-water continental shelves[J]. IEEE Journal of Oceanic Engineering, 2002, 14(4):360-367.
10. Hellequin L, Boucher J M, Lurton X. Processing of high-frequency multibeam echo sounder data for seafloor characterization[J]. IEEE Journal of Oceanic Engineering, 2003, 28(1): 78-89.
11. Pouliquen E, Pautet L, Canepa G. Time domain modeling of seafloor scatter at low grazing angle using the small slope approximation[J]. The Journal of the Acoustical Society of America, 2003, 113(4_Supplement): 2298-2298.
12. Jackson D. The small-slope approximation for layered seabeds[C]//Proceedings of Meetings on Acoustics. AIP Publishing, 2013, 19(1).
13. Darmon M, Dorval V, Baqué F. Acoustic scattering models from rough surfaces: a brief review and recent advances[J]. Applied Sciences, 2020, 10(22): 8305.
14. Brekhovskikh L,Lysanov Y. Fundamentals of Ocean Acoustics (Modern Acoustics and Signal Processing). 2003.
15. Gragg R F, Wurmser D, Gauss R C. Small-slope scattering from rough elastic ocean floors: General theory and computational algorithm[J]. The Journal of the Acoustical Society of America, 2001, 110(6): 2878-2901.
16. Lee K, Seong W. Time-domain Kirchhoff model for acoustic scattering from an impedance polygon facet[J]. The Journal of the Acoustical Society of America, 2009, 126(1): EL14-EL21.
17. Chotiros N P . Reflection And Reverberation In Normal Incidence Echo-Sounding[J]. The Journal of the Acoustical Society of America, 1994, 96(5).

DAILY PROGRAM

AUGUST 21 (THU)

14:20 – 15:20	Computational and intelligent approaches in speech and acoustics 325 (3F)
Chair	Wan-Ho Cho (Korea Research Institute of Standards and Science)
14:20	Classifying American English Dialects Using Acoustic-Prosodic Features Extracted by openSMILE
	Rylen W. Garlitz
14:40	Development of accurate transmission line matrix method in acoustics and its application to underwater acoustics
	Wan-Gu Kim, Young Geul Yoon, Ho Youn Ji, Min-Seok Choi, Seong Hyeon Kim, Soonyeol Kwon, Bok Kyoung Choi, Byoung-Nam Kim, Keunhwa Lee, Sufyan Ali Memon
15:00	Evaluation of demodulated sound enhancement based on carrier waves boosting with distortion components of modulated wave for parametric array loudspeakers
	Mizuki IWAGAMI, Yoto IKEZAKI, Yuting GENG, Masato NAKAYAMA, Takanobu NISHIURA

PROCEEDINGS of the
16th International Conference on Theoretical and Computational Acoustics

August 18 to 22, 2025 BEXCO, Busan Korea

Classifying American English dialects using acoustic-prosodic features extracted by openSMILE

Rylen W. GARLITZ[1]
[1]University of Florida, United States of America

ABSTRACT

This study investigates the potential for classifying American English dialects using acoustic-prosodic speech features extracted with openSMILE (1). These features–such as pitch, formant frequencies, and energy contours–capture subtle phonetic cues linked to speaker-specific productions. Audio samples from the International Dialects of English Archive (IDEA) (2) were analyzed using the eGeMAPSv02 feature set (88-dimensional descriptors). Four broad U.S. regional dialects (West, South, Midwest, and Northeast) were examined using multiple classification models and evaluated using accuracy metrics.

Initial results showed poor performance in the four-region task (~23% accuracy), largely due to sample size limitations. However, a binary classification between the South and Midwest regions (with over 100 speakers each) achieved 76% accuracy. The top predictive features included harmonic and spectral measures, such as normalized H1-H2 variability and Hammarberg index.

These findings suggest that phonetic features alone can provide useful information for dialect classification, though larger datasets are needed for reliable performance. Future work will expand this approach to different corpora and explore alternative openSMILE features sets. These findings will inform future dissertation work on detecting pseudo-dialects–cases where speakers intentionally imitate another dialect–offering applications in forensic speech analysis and sociophonetics.

Keywords: Speech, openSMILE, Classification

1. INTRODUCTION

Dialect classification plays a crucial role in both theoretical and applied domains of linguistics. From a theoretical perspective, it enables large-scale investigations of regional variation and sound change (3). In applied applications, dialect-aware systems can improve the performance of automatic speech recognition (ASR), enhance forensic speaker profiling, and support voice-based technologies that interact more naturally with diverse users (4,5). English exhibits significant dialectal variation across regions, especially within the continental United States. These dialects differ not only in phonetic realizations and lexical choices but also larger prosodic features such as intonation, rhythm, and speech rate. Models should include both these segmental and suprasegmental features to produce the most accurate classification, but can acoustic-prosodic information alone accurately classify American English dialects?

This study investigates whether hand-crafted acoustic-prosodic features extracted using the openSMILE toolkit can classify American English dialects (1). openSMILE feature sets provide interpretable, linguistically motivated descriptors, including pitch, formant frequencies, voice quality measures, and energy contours–used in the classification of emotion, Autism Spectrum Disorder (ASD), Amyotrophic Lateral Sclerosis (ALS), and Parkinson's disease (6–8).

2. BACKGROUND

openSMILE (open-source Speech and Music Interpretation by Large-space Extraction) is a toolkit for audio analysis, processing, and classification. It has been used in several classification tasks including emotion (9), Amyotrophic Lateral Sclerosis (ALS) (7), and Autism Spectrum Disorder (ASD) (6). Jansson and Johnson (7) implemented a machine learning model using the eGeMAPSv02 feature set (10) to extract low-level descriptors of audio records of people with and without diagnosed ALS. They were able to create an 88% accurate model for classifying ALS speech against healthy controls. Beccaria et al. (6) also utilized

[1] rylengarlitz@ufl.edu

openSMILE with an eGeMAPS configuration to classify the speech of individuals with ASD. Using a Random Forest classifier, a Support Vector Machine, and a Decision Tree classifier, each model achieved approximately 83% classification accuracy. While openSMILE features have been widely used for clinical and affective speech classification, its application has received limited attention in dialect classification. Prior studies have primarily focused on embeddings or phonotactic cues for dialect discrimination (11). This study extends openSMILE's application into regional dialect identification, testing whether identifiable, low-level descriptors can capture phonetic and prosodic variation across U.S. English dialects.

3. METHODOLOGY

To evaluate this approach, this study used a subset of recordings from the International Dialects of English Archive (IDEA) (2), focusing on dialects within the United States. Recordings were preprocessed to isolate only the read speech portion, where informants read the standardized passage "Comma Gets a Cure" (12). "Comma Gets a Cure" was written to include many of J.C. Wells' standard lexical set words, allowing readers to produce a variety of English phonemes across different contexts. The recordings were also preprocessed for sampling rate and annotated with other metadata including state, region, and speaker sex. Other demographic metadata was available such as ethnicity, occupation, and age. These were not included after further data wrangling, due to the number of missing values.

3.1 Regions

At this point, a dataset was created containing the audio paths, speakers' states, and sexes. Due to the limited number of speakers per state, classification at the state level was not feasible. The speakers were instead grouped into four different regions for classification: West, Midwest, South, and Northeast. Regions were divided according to conventional U.S. geographic groupings. After dividing speakers into regions, there was skew in the number of speakers per region. To account for this, data was randomly sampled within each region to create equal groups. In total this resulted in 64 speakers from each region, with 256 speakers in the dataset. The dataset was split into training and testing subsets using *train_test_*split from *scikit-learn* (13). Specifically, 80% of the data was allocated for training and 20% for testing. To ensure reproducibility, a fixed random shuffle was applied by setting the *random_state* parameter to 5. This ensures that the data is randomly shuffled before splitting, but the split will always be the same each time the code is run, helping to ensure consistent results. After assigning each audio file to a region, openSMILE features were extracted.

3.2 openSMILE Features

openSMILE features were extracted using the eGeMAPSv02 configuration to pull 88-dimensional low-level and functional features (10). While this study used eGeMAPSv02, other feature sets, both pre-defined and custom, are available and discussed in Section 5. These features were used to train and evaluate three different classification models: Random Forest (RF), Support Vector Machine (SVM), and a K-Nearest Neighbors (KNN) model. All three models received the same training and testing data. These models were selected based on their prior use in classification tasks involving openSMILE features.

4. RESULTS AND DISCUSSION

4.1 Random Forest (RF)

Results from the baseline RF model are presented in Table 1 and visually in a confusion matrix in Figure 1. The model achieved an accuracy of 23%, which approximates chance level performance. This suggests that the model failed to learn from the openSMILE features and relied on chance-based predictions. As seen in the confusion matrix, most of the test cases are being classified as the Midwest region.

Table 1 – Random Forest classification report; baseline

RF-base	Precision	Recall	F1-Score	Support
Midwest	0.17	0.40	0.24	10
Northeast	0.22	0.12	0.15	17
South	0.80	0.27	0.40	15
West	0.13	0.20	0.16	10
Accuracy			0.23	52
Macro Avg	0.33	0.25	0.24	52
Weighted Avg	0.36	0.23	0.24	52

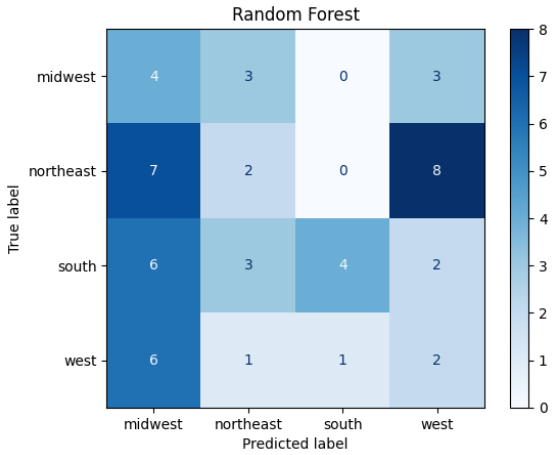


Figure 1 – Random Forest confusion matrix; baseline model

4.2 Support Vector Machine (SVM)

The SVM baseline model produced results shown in Table 2 and Figure 2. This model did not succeed at classifying dialectal regions. Two of the four dialects received no correct classification, given their zero precision, recall, and f1-scores. The accuracy is labelled as 23% and is marginally worse than chance. It seems that this model, like the RF model, did not extract meaningful patterns from the training data. Additionally, this model took the longest to train. Section 5 further discusses confounding factors.

Table 2 – SVM classification report; baseline

SVM-base	Precision	Recall	F1-Score	Support
Midwest	0.17	0.70	0.28	10
Northeast	0.00	0.00	0.00	17
South	0.00	0.00	0.00	15
West	0.42	0.50	0.45	10
Accuracy			0.23	52
Macro Avg	0.15	0.30.	0.18	52
Weighted Avg	0.11	0.23	0.14	52

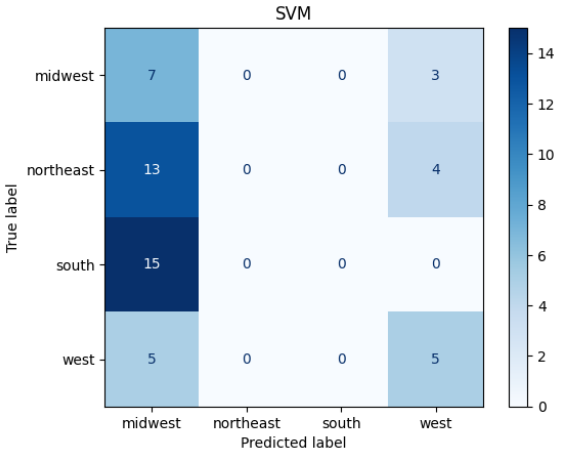


Figure 2 – SVM confusion matrix; baseline model

4.3 K-Nearest Neighbors (KNN)

Lastly, the KNN model results are also shown in Table 3 and Figure 3. This model also did not perform well with an accuracy of 19%. One region received no correct classification suggesting, much like the SVM above, the model failed to learn meaningful patterns. Further discussion on confounding factors for these models will be discussed in Section 5.

Table 3 – KNN classification report; baseline

KNN-base	Precision	Recall	F1-Score	Support
Midwest	0.21	0.50	0.29	10
Northeast	0.23	0.18	0.20	17
South	0.00	0.00	0.00	15
West	0.25	0.20	0.22	10
Accuracy			0.19	52
Macro Avg	0.17	0.22	0.18	52
Weighted Avg	0.16	0.19	0.16	52

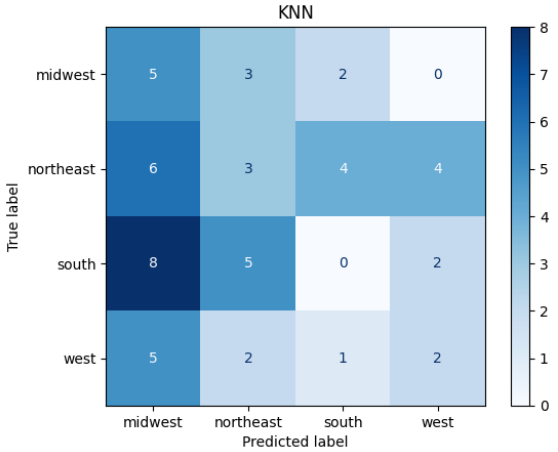


Figure 3 – KNN confusion matrix; baseline model

4.4 Binary Classification

Due to the poor performance of classifying four separate regions of the data, this study then focused on classifying only two regions, Midwest and South, with 114 and 161 speakers, respectively. The dataset was again split into training and testing subsets in the same manner as before using *train_test_*split from *scikit-learn* (13). Only a Random Forest model was employed for this binary task, as it was the only classifier that previously converged. Results from the model are presented below.

4.4.1 Random Forest

Random Forest model results for the binary region classification are presented in Table 4 and visually in Figure 4. This model demonstrated the predictive utility of the features for identifying the region. The accuracy is 76%, which is over chance. Given these results, it seems that the previous four-region classification was possibly confounded by the limited number of data points for each region. With a higher number of speakers per region the results may be improved.

Table 4 – Random Forest classification report; binary baseline

RF-bi-base	Precision	Recall	F1-Score	Support
Midwest	0.75	0.65	0.70	23
South	0.77	0.84	0.81	32
Accuracy			0.76	55
Macro Avg	0.76	0.75	0.75	55
Weighted Avg	0.76	0.76	0.76	55

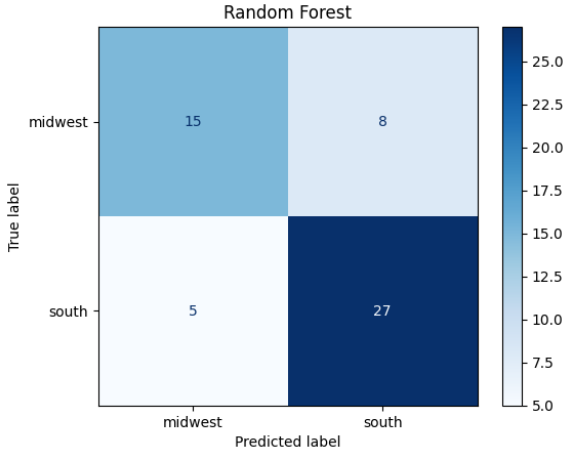


Figure 4 – Random Forest confusion matrix for binary region; baseline model

4.4.2 Important Features

To determine what hand-crafted features were most predictive in classifying the region I analyzed the "feature_importances_" attribute of the model. The top 5 most important features are presented in Table 7. The first feature "logRelF0-H1-H2_sma3nz_stddevNorm" captures variation in normalized spectral information. It is the standard deviation of the log ratio of the energy of the first F0 harmonic (H1) to the energy of the second F0 harmonic (H2). Several studies have looked at the effects of spectral tilt on speech intelligibility, voice quality, and the effect of emotion (14–16). Lower spectral tilt is associated with higher intelligibility in noise. "Anger" and "Joy" were associated with higher spectral tilt corresponding to louder speech with higher vocal effort. Both the second and the fourth feature, "mfcc2_sma3_stddevNorm" and "mfcc2V_sma3nz_amean" are derived measurements of the second Mel-Frequency Cepstral Coefficient (MFCC). The first being the normalized standard deviation and the other being the mean. MFCCs have been used as "gold-standard" features in several speaker classification/verification studies (17–19). The third feature, "hammarbergIndexUV_sma3nz_amean" is also a spectral feature related to the Hammarberg Index: the ratio of the strongest energy peak in the 0-2kHz range to the strongest peak in the 2-5kHz range (10,20). The Hammarberg Index, like MFCCs, have been used to successfully classify several speech features including vocal disorders and emotion (20,21). The final feature in Table 7 "slopeV0-500_sma3nz_amean" is another derived and averaged feature representing spectral slope: the linear regression slope of the log power spectrum between 0Hz and 500Hz. All the features in the top five for the RF model deal with spectral parameter information. It is likely that spectral information rather than frequency or energy related parameters has higher predicative power for speakers' dialect.

Table 7 – Random Forest Feature Importance

logRelF0-H1-H2_sma3nz_stddevNorm	0.029674
mfcc2_sma3_stddevNorm	0.026332
hammarbergIndexUV_sma3nz_amean	0.024938
mfcc2V_sma3nz_amean	0.020653
slopeV0-500_sma3nz_amean	0.018733

5. CONCLUSIONS

Overall, with the initial separation of speakers into four regions, all models performed poorly at classifying speakers' dialect with an accuracy approaching chance. When a model was trained on only the two largest regions of the dataset the performance improved. The RF model successfully predicted speakers' regional dialect based on their openSMILE features with an accuracy of 76%. Increasing the number of speakers could further enhance classification accuracy. Reviewing the top predicting features revealed that spectral parameters were the most informative including Hammarberg index, spectral tilt, spectral slope, and MFCCs.

Given that the results showed higher than chance accuracies, they seem to support that dialect can be distinguished using purely hand-crafted acoustic-prosodic features, without the inclusion of phonological information. The following paragraphs discuss key limitations and directions for future research.

There are several limitations with the current study which might have confounded some of the results. Namely, the quality of the audio from the dataset. Audio recording quality can have drastic effects on feature analysis. Poor quality audio would not capture as many details in the sound and could have significant distortion or noise in the audio stream. Controlling for this by using better quality audio recordings would possibly improve classification.

Another confounding factor is the division of speakers into broad regions. While I placed states into their typically respected regions, some of these might not be very similar to other dialects within the same region. For example, the New York dialect is quite a bit different than the Massachusetts dialect, but they are both placed into the same "Northeast" region. This may have affected classification accuracy. Additionally, I did not regulate the number of speakers from each state that contributed to the region. While it was not analyzed, there may exist a larger number of speakers from one state in a region that could be skewing the results. Future research should control for this potential imbalance.

Total number of audio samples is also a problem in many classification models. In the four-dialect classification there were only 64 speakers per region (256 speakers total) and 80% of that was used for training. The two-dialect model had an improvement in accuracy with double the number of speakers in each region so including more high-quality audio samples would likely enhance classification performance.

Future iterations of this research should investigate possibly training the models on a different set of features, for this paper I utilized eGeMAPSv02 because it was used most prevalently in previous research. There are, however, significantly more features and feature sets that can be used with openSMILE, as well as creating custom feature sets. Utilizing a different feature set with new/more features could introduce a key feature in this classification problem.

The current dataset that was used (IDEA) is probably not the most conducive for these types of audio analysis. Choosing a corpus with better quality audio would likely improve the quality of the features extracted and thus, improve the classification performance of those features in a model. Replicating this research on a corpus like TIMIT or a self-collected audio corpus with dialect grouping may yield interesting results.

REFERENCES

1. Eyben F, Weninger F, Gross F, Schuller B. Recent developments in openSMILE, the munich open-source multimedia feature extractor. 2013 Oct 21;835–8.
2. Meier P. International Dialects of English Archive [Internet]. 2025. Available from: https://www.dialectsarchive.com/
3. Labov W, Yaeger M, Steiner R. A Quantitative Study of Sound Change in Progress. In: Report on National Science Foundation Contract NSF-GS-3287. Philadelphia: University of Pennsylvania; 1972.
4. Quesenberry E, Morris K, Walker A. Shaping Interaction: The Impact of Voice Assistant Dialects on User

Experience. Conference Presentation presented at: SECOL 2025; 2025; Boca Raton, FL.
5. Paul D, Li J, Chang L, Petkov P, Su P, Padinjaru Veettil S, et al. Advancing Speech Accessibility with Personal Voice. Apple Machine Learning Research. 2023;
6. Beccaria F, Gagliardi G, Kokkinakis D. Extraction and Classification of Acoustic Features from Italian Speaking Children with Autism Spectrum Disorders. Proceedings of the RaPID-4 @ LREC 2022; 2022.
7. Jansson R, Johnson L. Using Machine Learning to Detect Speech Characteristics of People with ALS. 2024;
8. Bayestehtashk A, Asgari M, Shafran I, McNames J. Fully automated assessment of the severity of Parkinson's disease from speech. Computer Speech & Language. 2015 Jan;29(1):172–85.
9. Schuller B, Steidl S, Batliner A. The INTERSPEECH 2009 emotion challenge. In: Interspeech 2009 [Internet]. ISCA; 2009 [cited 2025 Mar 1]. p. 312–5. Available from: https://www.isca-archive.org/interspeech_2009/schuller09_interspeech.html
10. Eyben F, Scherer KR, Schuller BW, Sundberg J, Andre E, Busso C, et al. The Geneva Minimalistic Acoustic Parameter Set (GeMAPS) for Voice Research and Affective Computing. IEEE Trans Affective Comput. 2016 Apr 1;7(2):190–202.
11. Ghorbani S, Hansen JHL. Advanced accent/dialect identification and accentedness assessment with multi-embedding models and automatic speech recognition [Internet]. arXiv; 2023 [cited 2025 Jul 3]. Available from: https://arxiv.org/abs/2310.11004
12. Honorof DN, McCullogh J, Somerville B. Comma Gets a Cure. 2000.
13. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in python. J Mach Learn Res. 2011 Nov;12(null):2825–30.
14. Drugman T, Dutoit T, Bozkurt B. Excitation-based Voice Quality Analysis and Modification [Internet]. arXiv; 2020 [cited 2025 Jun 14]. Available from: http://arxiv.org/abs/2001.00582
15. Simantiraki O, Cooke M, Pantazis Y. Effects of Spectral Tilt on Listeners' Preferences And Intelligibility. In: ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) [Internet]. Barcelona, Spain: IEEE; 2020 [cited 2025 Jun 14]. p. 6254–8. Available from: https://ieeexplore.ieee.org/document/9054117/
16. Takeda S, Ueno Y, Nakasako N, Nakagawa H, Tsuru M, Isobe R, et al. Spectral-tilt features of emotional speech - Research on emotional-speech synthesis based on voice-quality conversion. In Paris, France; 2010.
17. Babu PA, Siva Nagaraju V, Vallabhuni RR. Speech Emotion Recognition System With Librosa. In: 2021 10th IEEE International Conference on Communication Systems and Network Technologies (CSNT) [Internet]. Bhopal, India: IEEE; 2021 [cited 2025 Apr 28]. p. 421–4. Available from: https://ieeexplore.ieee.org/document/9509714/
18. Usama Tanveer Gujjar M, Munir K, Amjad M, Rehman AU, Bermak A. Unmasking the Fake: Machine Learning Approach for Deepfake Voice Detection. IEEE Access. 2024;12:197442–53.
19. Staroniewicz P. Influence of Natural Voice Disguise Techniques on Automatic Speaker Recognition. Acoustics. 2018 Sep 1;
20. Hammarberg B, Fritzell B, Gaufin J, Sundberg J, Wedin L. Perceptual and Acoustic Correlates of Abnormal Voice Qualities. Acta Oto-Laryngologica. 1980 Jan;90(1–6):441–51.
21. Tamarit L, Goudbeek M, Scherer K. Spectral slope measurements in emotionally expressive speech. In: ISCA ITRW on Speech Analysis and Processing for Knowledge Discovery. 2008. p. paper 007.

Development of accurate transmission line matrix method in acoustics and its application to underwater acoustics

Wan-Gu KIM; Young Geul YOON; Ho Youn JI; Min-Seok CHOI; Seong Hyeon KIM;
Soonyeol KWON; Bok Kyoung CHOI; Byoung-Nam KIM[1]
Sea Power Reinforcement · Security Research Department, Korea Institute of Ocean Science & Technology, Busan 49111, Republic of Korea

Keunhwa LEE; Sufyan Ali MEMON
Department of Defense Systems Engineering, Sejong University, Seoul 05006, Republic of Korea

ABSTRACT
The transmission line matrix (TLM) method has been used extensively in the microwave field and is an important modeling method with many advantages, but its use in acoustics has been rare. We believe that one of the main reasons for this is that existing acoustic transmission line matrix method modelling studies have not presented the conditions for accurate modelling or a way to verify its accuracy. In this study, the relationship between mesh size and modelling accuracy was investigated for transmission line matrices composed of uniformly sized meshes in order to develop a practical acoustic transmission line matrix method. For this purpose, the acoustic field simulations for a rigid enclosure were performed with the acoustic transmission line matrix method and compared with the exact results from the image method, which can calculate the acoustic field without error. Additionally, we have used this method to create underwater acoustic simulators for cases where the medium is flowing or the sound source is moving. In particular, the calculation result of the simulator for the case where the sound source is moving were compared to the results of a validation experiment in a large water tank. The frequency components of the hydrophone received signal calculated for the case of a transmitter falling in a large water tank were in good quantitative agreement with the frequency components measured in the validation experiment.
Keywords: transmission line matrix method (TLM), underwater acoustics, numerical simulation, experimental validation, moving source, flow of medium, cavitation tunnel

ACKNOWLEDGEMENTS
This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (Ministry of Science and ICT) (No. RS-2022-00166977).

REFERENCES
1. Renato S. T. de Carvalho, and Júlio A. Cordioli, “Simulation of underwater acoustics and sediment wave propagation using the Discrete Huygens Modeling (DHM),” in Proceedings of Forum Acusticum. 2014 (European Acoustics Association, 2014).
2. Donard De Cogan, William J. O'Connor, and Susan Pulko, Transmission line matrix (TLM) in computational mechanics (CRC Press, New York, 2006).
3. Samir El-Masri, Xavier Pelorson, Pierre Saguet, and Pierre badin, “Development of the transmission line matrix method in acoustics applications to higher modes in the vocal tract and other complex ducts,” International Journal of Numerical Modelling: Electronic Networks, Devices and Fields **11**(3), 133-151 (1998).
4. Cheolsoo Park, Gun-Do Kim, Young-Ha Park, Keunhwa Lee, and Woojae Seong, “Noise Localization Method for Model Tests in a Large Cavitation Tunnel Using a Hydrophone Array,” Remote Sensing **8**(3), 195 (2016).

[1] bnkim@kiost.ac.kr

Evaluation of demodulated sound enhancement based on carrier waves boosting with distortion components of modulated wave for parametric array loudspeakers

Mizuki IWAGAMI[1]; Yoto IKEZAKI[1]; Yuting GENG[1]; Masato NAKAYAMA[1]; Takanobu NISHIURA[1]
[1] Ritsumeikan University, Japan

ABSTRACT
Parametric array loudspeakers (PALs) achieve sharp directivity utilizing the straightness of ultrasound. The ultrasound is modulated by the target sound, and emitted intensely from PAL. The target sound is demodulated as a difference frequency of the emitted ultrasonic components, due to the nonlinearity of air. However, PALs suffer from low demodulated sound pressure, limiting effectiveness in noisy environments. Higher-order components can contribute to the demodulated sound pressure when their frequencies align with the target frequency. To leverage this, we have proposed a method that combines two ultrasonic waves modulated with different carriers. The carrier frequencies are set to let higher-order components match the frequencies of the lower carrier and the target sound. These components, usually regarded as useless by-products, are utilized to boost the demodulated sound. Our previous work focused on sound pressure improvement, with insufficient evaluation of total harmonic distortion (THD). Therefore, in this paper, we evaluate the proposed method regarding both demodulated sound pressure and THD. The results show that the proposed method achieves higher sound pressure than conventional methods, due to the contribution of higher-order components. Furthermore, the proposed method achieves a level of THD comparable to that of conventional methods.
Keywords: Parametric array loudspeaker, Amplitude modulation, Demodulated sound enhancement, High-order components

1. INTRODUCTION

Parametric array loudspeakers (PALs) utilize intense ultrasonic signals to reproduce audible sound through nonlinear interactions in air. By exploiting the self-demodulation effect, PALs can generate difference-frequency components with highly directional characteristics [1], [2]. This unique property has enabled applications such as targeted audio transmission and acoustic spotlights [3].

When ultrasonic signals of differing frequencies propagate through air, nonlinear effects give rise to various secondary frequency components, including difference frequencies, sum frequencies, harmonics, and other high-order distortion products [4]. Typically, modulated signals composed of a carrier and its sidebands are transmitted [3], [5], and the audible signal is reconstructed via self-demodulation.

Despite their advantage in sharp directivity, PALs suffer from relatively low demodulated sound pressure levels compared to the emitted ultrasonic signals [1], [2]. This limitation reduces the intelligibility of the output in noisy environments. Previous studies have investigated methods to reduce harmonic distortion under over-boosted conditions by optimizing the spectral characteristics of the source signal [6]. These approaches have been shown to be effective, especially when the target signal has a harmonic structure. However, their applicability may be limited in situations where the signal lacks such regularity or when the focus is on utilizing, rather than suppressing, high-order components.

To overcome this limitation, we have proposed a method that enhances demodulated sound pressure by effectively utilizing high-order distortion components generated during nonlinear propagation [7]. These components, which are not typically utilized in conventional modulation schemes, reinforce the carrier frequencies through harmonic interactions and contribute to enhanced demodulated sound pressure. To facilitate this, two ultrasonic waves modulated with different carrier frequencies are combined, encouraging the generation of high-order components that align with either the lower carrier or the target frequency.

[1] {is0546ve@ed, is0492ss@ed, geng@fc, mnaka@fc, nishiura@is}.ritsumei.ac.jp

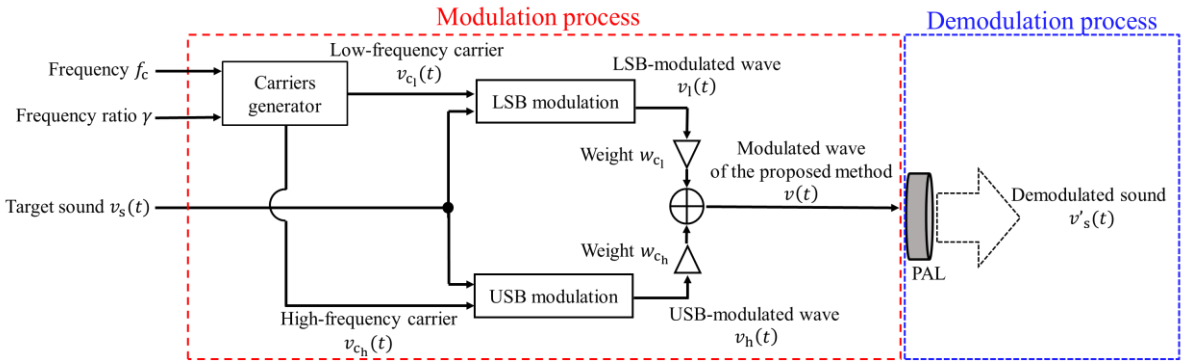


Figure 1 – Block diagram of the proposed method

In this paper, we evaluate the proposed method for enhancing the demodulated sound pressure of PALs by transmitting two modulated ultrasonic signals with different carrier frequencies. The proposed approach aims to leverage the constructive effects of high-order components to improve sound pressure without compromising sound quality. A quantitative evaluation is conducted in terms of the demodulated sound pressure and total harmonic distortion (THD), and the results demonstrate that the method improves demodulated sound pressure while maintaining sound quality.

2. PROPOSED METHOD

2.1 Overview of the Proposed Method

Figure 1 shows a block diagram of the proposed method.

The method uses two ultrasonic carrier waves, $v_{c_l}(t)$ and $v_{c_h}(t)$, which have different frequencies determined by a base carrier frequency f_c and a frequency ratio γ. The frequency ratio γ defines how many times higher the high-frequency carrier is compared to the low-frequency carrier. The low-frequency carrier $v_{c_l}(t)$ is modulated by the target signal $v_s(t)$ using lower sideband (LSB) modulation [8], and the high-frequency carrier $v_{c_h}(t)$ is modulated using upper sideband (USB) modulation [8]. The LSB-modulated wave $v_l(t)$ and the USB-modulated wave $v_h(t)$ are weighted by w_l and w_h, respectively, and summed to produce the final modulated wave $v(t)$, which is radiated from PALs to generate the audible sound $v'_s(t)$ through nonlinear demodulation. Details of the modulation process are described in Section 2.2.

The modulated wave is generated by combining two modulated signals. The low-frequency carrier is modulated by the target signal $v_s(t)$ using lower sideband (LSB) modulation [8], and the high-frequency carrier is modulated using upper sideband (USB) modulation [8]. The LSB-modulated wave $v_l(t)$ and the USB-modulated wave $v_h(t)$ are weighted by w_l and w_h, respectively, and summed to produce the final modulated wave $v(t)$. Details of the modulation process are described in Section 2.2.

When this modulated wave $v(t)$ is emitted at high intensity from the PALs, an audible sound $v'_s(t)$ is self-demodulated as in conventional PALs. In addition, the proposed method is characterized by the generation of frequency components matching the carrier frequencies through interactions between the two carriers. Since the carriers are boosted by specific high-order distortion components, the demodulated sound pressure is also enhanced as a result. The details of this demodulation process are presented in Section 2.3.

This enhancement arises from harmonic interactions between the modulated carriers, where high-order distortion components align at the carrier frequencies. These interactions regenerate energy at the carrier frequencies not through direct amplification, but through constructive interference among their harmonic components. The reinforced carriers intensify nonlinear interactions in air, resulting in stronger demodulated output. Some modulation components may also be indirectly enhanced through similar mechanisms.

Thus, the proposed method increases demodulated sound pressure by effectively combining modulation design with nonlinear propagation effects.

For practical implementation, the LSB-modulated signal $v_l(t)$ is emitted from the PAL-LC, which operates at a carrier frequency of 25 kHz, and the USB-modulated signal $v_h(t)$is emitted from the PAL-HC, operating at 50 kHz. These two PALs are positioned adjacent to each other to enable nonlinear interactions between their ultrasonic beams, which are essential for generating the high-order distortion components utilized in the proposed method.

2.2 Modulation Process in the Proposed Method

The proposed method employs two carrier signals with different frequencies, denoted as f_{c_l} and f_{c_h}, to

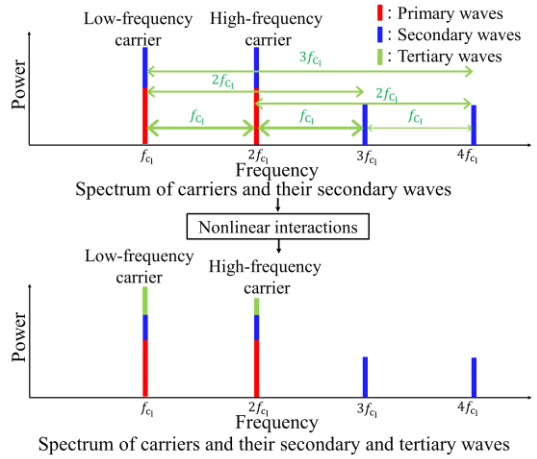


Figure 2 – Frequency relationship among the primary and secondary waves of the carriers, and their interactions

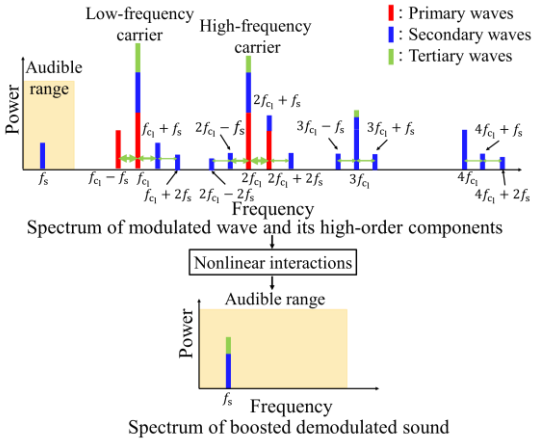


Figure 3 – Frequency relationship among the modulated waves and its high-order components, and their interactions

generate modulated ultrasonic waves. These frequencies are defined as:

$$f_{c_l} = f_c, \tag{1}$$

$$f_{c_h} = \gamma f_c, \tag{2}$$

where f_c is the base carrier frequency and γ is frequency ratio between the two carriers. In this study, $\gamma = 2$ is used, as described in Section 2.3.

Using these two carrier frequencies, we define the low-frequency and high-frequency carrier signals as follows:

$$v_{c_l}(t) = A_c \cos\left(2\pi f_{c_l} t\right), \tag{3}$$

$$v_{c_h}(t) = A_c \cos\left(2\pi f_{c_h} t\right), \tag{4}$$

where A_c denotes the amplitude of the carriers and t is the time index.

The target signal $v_s(t)$ is then modulated using lower sideband (LSB) modulation for the lower carrier and upper sideband (USB) modulation for the higher carrier. The resulting modulated waves are given by:

$$v_l(t) = \{1 + m_{ssb} v_s(t)\} A_c \cos\left(2\pi f_{c_l} t\right) + m_{ssb} \hat{v}_s(t) A_c \sin\left(2\pi f_{c_l} t\right), \tag{5}$$

$$v_h(t) = \{1 + m_{ssb} v_s(t)\} A_c \cos\left(2\pi f_{c_h} t\right) + m_{ssb} \hat{v}_s(t) A_c \sin\left(2\pi f_{c_h} t\right), \tag{6}$$

where, $\hat{v}_s(t)$ represents the Hilbert transform of $v_s(t)$, and m_{ssb} is the modulation factor, constrained within $0 < m_{ssb} \leq 1$.

Finally, the two modulated signals are weighted and summed to form the final ultrasonic signal:

$$v(t) = w_l v_l(t) + w_h v_h(t), \tag{6}$$

where, w_l and w_h are the weights applied to the LSB and USB components, respectively.

This configuration enables generation of specific high-order distortion components during nonlinear propagation. By carefully designing the carrier frequencies and modulation scheme, the method promotes constructive interactions among harmonic components, which reinforce the energy at the carrier frequencies. This design is crucial for the enhancement mechanism described in Section 2.3.

2.3 Demodulation Process in the Proposed Method

In the proposed method, the frequency ratio γ between the two carriers is set to 2, that is, $f_{c_h} = 2f_{c_l}$. This relationship is critical because it promotes the generation of specific high-order distortion components that align with the original carrier frequencies during nonlinear propagation. This configuration ensures that the interaction between the LSB-modulated wave $v_l(t)$ and the USB-modulated wave $v_h(t)$ generates high-order components that align with the carrier frequencies.

Figure 2 illustrates the interactions between the two carrier frequencies and their harmonic components. When the LSB- and USB-modulated signals are emitted simultaneously, nonlinear effects in air generate

second and third order components such as $f_{c_l}(= f_{c_h} - f_{c_l})$ and $3f_{c_l}(= f_{c_h} + f_{c_l})$. These components coincide with the carrier frequencies and lead to increased energy at f_{c_l} and f_{c_h}. This reinforcement occurs because harmonic components generated during nonlinear propagation constructively overlap at the carrier frequencies, without any increase in the original carrier amplitude.

Figure 3 illustrates how the modulated signals and the distortion components combine to reproduce the demodulated sound. For example, difference frequencies such as $f_{c_l} - (f_{c_l} - f_s)$ and $(f_{c_h} + f_s) - f_{c_h}$ match the target frequency f_s, as shown in Fig. 3. Various difference-frequency components emerge from the interaction of sidebands, many of which match the target audio frequency f_s. As the carrier components are reinforced, these interactions become stronger, resulting in an enhanced demodulated signal. Additionally, other secondary components separated by f_s may also contribute to the reproduction of the target sound through higher-order intermodulation effects. These results indicate that the proposed method improves demodulated sound pressure not by increasing the input amplitude, but by encouraging nonlinear interactions through intentional modulation design and carrier selection.

3. EVALUATION EXPERIMENTS

This section outlines the objective of the experiment conducted to evaluate the effectiveness of the proposed method. The proposed method aims to enhance demodulated sound pressure by utilizing high-order distortion components generated during nonlinear propagation. These components reinforce both the carrier and sideband frequencies through harmonic interactions, indirectly increasing the demodulated output. To evaluate the effectiveness of this approach, we compare the SPL of the demodulated signal and THD between the proposed method and a conventional single-carrier modulation approach. All experiments were conducted under identical conditions, with the modulation method being the only variable to ensure fair evaluation.

3.1 Experimental Conditions

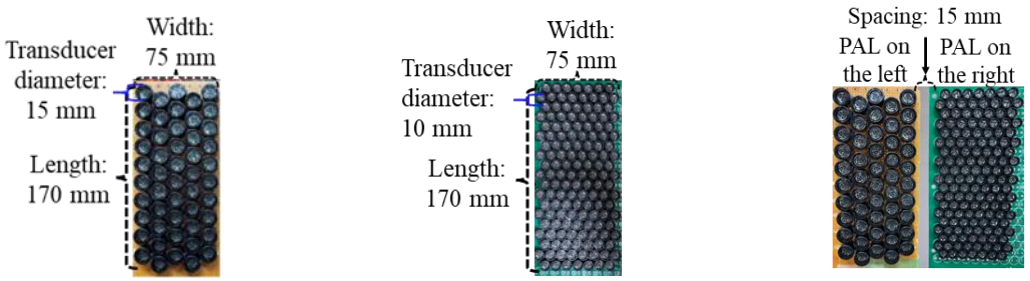
To implement the proposed method, it is ideal to use a single PAL capable of emitting both 25 kHz and 50 kHz carrier frequencies. However, such broadband ultrasonic transducers were not available for this study. Therefore, a PAL system was constructed using two separate PAL units, each equipped with ultrasonic transducers having different resonance frequencies. The structure of the PALs used in the experiment is shown on Fig. 4. Figure 4(a) shows a PAL using 25 kHz resonance frequency transducers (NICERA T2516A1), referred to as PAL-LC, for low-frequency carrier emission. Figure 4(b) shows a PAL using 50

Table 1 – Experimental equipment

Ultrasonic transducer for PAL-LC	NICERA, T2516A1
Ultrasonic transducer for PAL-HC	SPL, UT1007-Z325R
Power amplifier	VICTOR, PS-A2002
Microphone	SENNHEISER, MKH 8020
D/A, A/D converter	RME, Fireface UFX

Table 2 – Experimental conditions

Environment	Soundproof room ($T_{60} = 0.15$ s)
Ambient noise level	$L_A =$ 21.9 dB
Sampling frequency	192 kHz
Quantization	32 bits



(a) PAL-LC (b) PAL-HC (c) Combination of two PALs

Figure 4 – Photographs of the PALs employed in the experimental setu

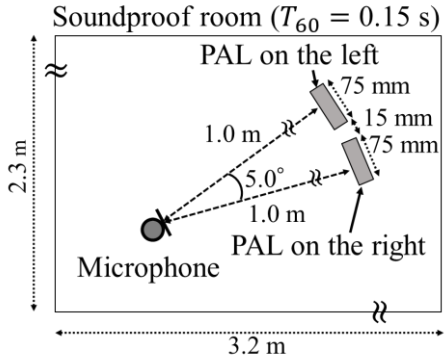


Figure 5 – Arrangement of the equipment

Table 3 – Practical measured conditions

Comparison patterns	The left PAL		The right PAL	
	Type of PAL	Modulation method	Type of PAL	Modulation method
Single-LSB	PAL-LC	LSB	-	-
Double-LSB	PAL-LC	LSB	PAL-LC	LSB
Single-USB	-	-	PAL-HC	USB
Double-USB	PAL-HC	USB	PAL-HC	USB
LSB+USB (w/o boost effect)	PAL-LC	LSB	PAL-HC	LSB
Prop. (w/ boost effect)	PAL-LC	LSB	PAL-HC	LSB

resonance frequency transducers (SPL UT1007-Z325R), referred to as PAL-HC, for high-frequency carrier emission. Figure 4(c) illustrates the configuration in which the PAL-LC and PAL-HC are arranged side by side. The spacing between the two PALs was set to 15 mm, which corresponds to the diameter of a NICERA transducer. The PAL-LC and PAL-HC were designed with equal radiation surface areas to maintain output balance between the two sources. The experimental equipment is summarized in Table 1. To enable a fair comparison among different modulation strategies, the experimental setup was designed to allow flexible assignment of modulation methods and carrier frequencies to the left and right PAL units. In the Double-LSB and Double-USB patterns, both sides use the same carrier frequency and modulation scheme, resulting in symmetric configurations. In contrast, the proposed method assigns a lower sideband signal with a 25 kHz carrier to the left PAL and an upper sideband signal with a 50 kHz carrier to the right PAL.

The spatial arrangement of the experimental setup is illustrated in Fig. 5. The two PALs were panned inward by 5 degrees so that their acoustic beams converge at a focal point 1.0 m in front of the array. A condenser microphone was placed 1.0 me in front of the PAL array and 1.2 m above the floor. The detailed experimental conditions for both the proposed and conventional methods are provided in Table 2.

This section describes the comparison patterns used in the evaluation experiment. The modulated signal $v(t)$ was constructed by combining two components: a lower sideband (LSB) modulated signal $v_{\mathrm{l}}(t)$ and an upper sideband (USB) modulated signal $v_{\mathrm{h}}(t)$, each modulated using different carrier frequencies. In this experiment, the carrier frequencies were set to $f_{\mathrm{c_l}} = 25$ kHz and $f_{\mathrm{c_h}} = 50$ kHz. Five comparison patterns were designed, as summarized in Table 3.

In the Single-LSB pattern, a single PAL-LC (parametric array loudspeaker designed for low-frequency carriers) was used to emit a 25 kHz LSB-modulated signal. In the Double-LSB pattern, two PAL-LC units were used on both sides to emit LSB signals simultaneously. The Single-USB pattern employed only one PAL-HC (for high-frequency carriers), emitting a 50 kHz USB-modulated signal. In the Double-USB pattern, two PAL-HC units were used to emit 50 kHz USB-modulated signals from both sides. In the proposed pattern (Prop. w/ boost effect), the PAL-LC and PAL-HC were positioned on the left and right, respectively. The PAL-LC emitted a 25 kHz LSB-modulated signal, and the PAL-HC emitted a 50 kHz USB-modulated signal. This configuration is designed to enable nonlinear interactions between the two carriers, generating distortion components that boost the carrier frequencies, thereby increasing demodulated sound pressure.

To further investigate the effect of carrier boosting, a simulated pattern called LSB+USB (w/o boost effect) was also included for comparison. In this pattern, the recorded signals from the Single-LSB and

Single-USB tests were combined post hoc, eliminating any physical interaction between the two PALs.

In all conditions, the modulation factor was set to $m_{\mathrm{ssb}} = 1.0$ to maximize the demodulated sound pressure without causing overmodulation. In the proposed method, the weights of the LSB and USB components were also adjusted so that the demodulated SPL from Single-LSB and Single-USB each reached 50 dB at a distance of 1.0 m.

To evaluate the effectiveness of the proposed method, we employed two quantitative metrics: demodulated sound pressure and THD.

First, the demodulated sound pressure was evaluated using a time-stretched pulse (TSP) signal covering the frequency range from 0 to 4.0 kHz [9]. For each comparison pattern, the sound was recorded, and the impulse response was calculated from the recorded signal via inverse filtering. From this impulse response, the direct sound component was extracted, and the sound pressure level L [dB] was calculated using the following:

$$L = L_{\mathrm{cali}} - 10\log_{10}\left(\frac{\sigma_{\mathrm{rec}}}{\sigma_{\mathrm{cali}}}\right)^2. \tag{7}$$

Here, L_{cali} is the sound pressure level measured using a sound level meter, σ_{cali} is the standard deviation of the recorded calibration signal, and σ_{rec} is the standard deviation of the recorded demodulated signal.

Next, to evaluate the impact of the proposed method on sound quality, we assessed the THD [10]. THD quantifies the degree of harmonic distortion; lower values indicate less spectral distortion. In this evaluation, logarithmic time-stretched pulse (log-TSP) signal [11] was used as the excitation signal. This signal enables temporal separation of the fundamental and harmonic components in the impulse response. Since out-of-band harmonics do not affect perceived sound quality, only audible-band harmonic components were used to compute THD. THD [%] was calculated using the following equation:

$$\mathrm{THD}[\%] = \frac{\sqrt{\sum_{n=2}^{n_{\max}} \sigma_{\mathrm{rec},n}^2}}{\sigma_{\mathrm{rec},1}} \times 100. \tag{8}$$

In Eq. (8), $\sigma_{\mathrm{rec},1}$ is the standard deviation of the amplitude of the fundamental component of the recorded signal, $\sigma_{\mathrm{rec},n}$ is that of the n-th harmonic component, and $n_{\max}$ is the maximum harmonic order considered.

3.2 Experimental Results and Discussions

This section presents the experimental results for the proposed and comparison methods and discusses the effectiveness of carrier boosting.

Figure 6 shows the sound pressure level of the demodulated sound for each comparison pattern. The horizontal axis represents the comparison patterns, and the vertical axis indicates the sound pressure levels of the demodulated sound. Error bars represent standard deviations. As shown in Fig. 6, the SPL of the proposed method (Prop. w/ boost effect) was 0.7 dB higher than that of Double-LSB, 0.6 dB higher than Double-USB, and 0.5 dB higher than LSB+USB (w/o boost effect). These results confirm that the proposed method increases the demodulated sound pressure through its boost effect.

However, from a practical perspective, further improvement in the demodulated SPL is necessary. Two main factors are considered to be responsible for the limited SPL improvement. The first is the insufficient emission level from the PALs, which limits the magnitude of the boost effect. In the present setup, the resonance frequencies of the PAL units do not match their respective carrier frequencies. To achieve higher emitted sound pressure, it is necessary to use PALs whose resonance frequencies match the carrier frequencies and to enlarge the aperture of the PALs.

The second factor is the limited contribution of the boost effect to the actual demodulated sound. Figures 7(a) and (b) show the power spectra under the LSB+USB (w/o boost effect) and the proposed method (w/ boost effect), respectively, highlighting the carrier boost effect. As shown, the power of the 50 kHz carrier in Prop. was 1.8 dB higher than in LSB+USB. Since the demodulated SPL is roughly proportional to the carrier level, as demonstrated in [12], this boost would ideally result in approximately a 1.0 dB increase in demodulated sound pressure. However, the observed gain was smaller than this theoretical value. To address this discrepancy, future work will incorporate phase control in the modulated wave to maximize the constructive interference of the high-order components. In summary, although the observed improvement was less than the theoretical maximum, the proposed method successfully increased the demodulated sound pressure through the boost effect.

Figure 8 shows the measured THD for each pattern, where the vertical axis represents THD in percentage and the horizontal axis indicates the comparison patterns. The THD of the proposed method (Prop.) was 2.7% lower than that of Single-LSB and 1.6% lower than that of Double-LSB. In contrast, it was 1.8% higher than

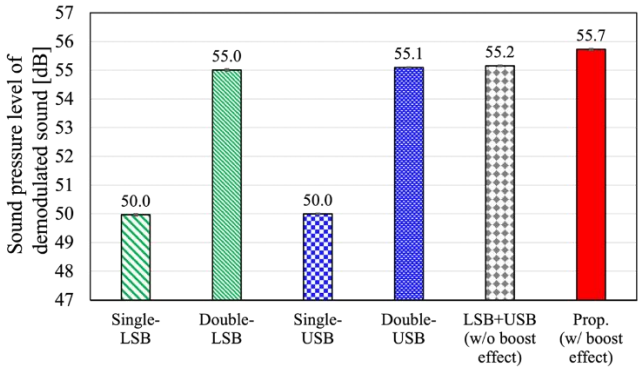
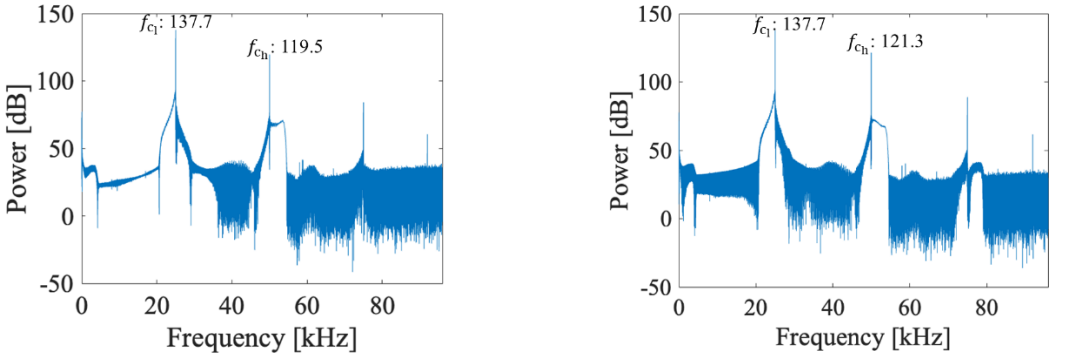


Figure 6 – Experimental results for demodulated sound pressure



(a) LSB+USB (w/o boost effect) (b) Prop. (w/ boost effect)

Figure 7 – Power spectra of the observed signal with microphone

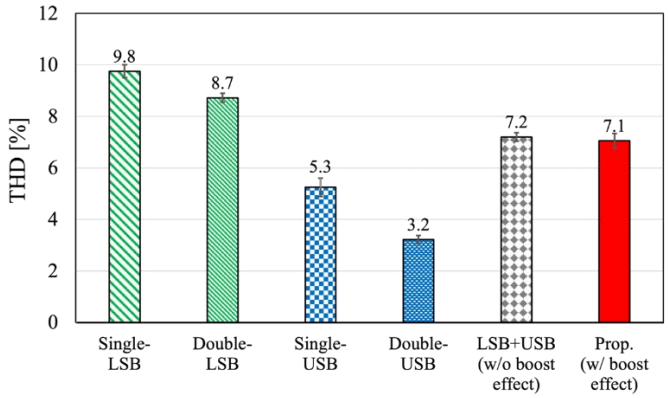


Figure 8 – Experimental results for THD

that of Single-USB and 3.9% higher than that of Double-USB. These results indicate that the THD of the proposed method is lower than that of LSB-modulated signals but higher than that of USB-modulated signals, which may be attributed to differences in the demodulation characteristics and the frequency responses of the loudspeakers used. Additionally, the THD of Prop. (w/ boost effect) was 0.1% lower than that of LSB+USB (w/o boost effect), suggesting that the boost effect does not cause significant degradation in sound quality. These results demonstrate that the boost effect of the proposed method enhances the demodulated sound pressure while maintaining spectral purity and auditory quality, similar to conventional single-carrier sideband modulation methods.

4. CONCLUSIONS

We proposed a demodulated sound enhancement method for PALs by combining two ultrasonic waves modulated with different carrier frequencies. The proposed method is designed to exploit high-order distortion components whose frequency combinations align with both the carrier and target frequencies,

thereby boosting the demodulated sound pressure. Experimental results demonstrated that the proposed method achieves higher demodulated sound pressure compared to conventional single-carrier modulation methods. This improvement is attributed to the effective contribution of high-order components. Furthermore, the THD remained comparable to that of conventional methods, confirming that sound quality is preserved. Although the sound pressure achieved did not reach the theoretical maximum, the findings indicate that the proposed method offers a promising direction for improving the performance of PALs in noisy environments. Future work will focus on optimizing carrier design and phase control to further increase the boost effect. These results contribute to the development of more efficient and high-quality directional audio systems.

ACKNOWLEDGEMENTS

This work was partly supported by Ritsumeikan University R-GIRO, ARC and RARA, and JSPS KAKENHI Grant Numbers JP23K28115, JP24K20803 and JP25H01158.

REFERENCES

1. P. J. Westervelt, “Parametric acoustic array,” The Journal of the Acoustical Society of America, vol. 35, no. 4, pp. 535–537, 1963.
2. C. Shi and W. Gan, “Development of parametric loudspeaker,” IEEE Potentials, vol. 29, no. 6, pp. 20–24, 2010.
3. M. Yoneyama, J. Fujimoto, Y. Kawamo, and S. Sasabe, “The audio spotlight: An application of nonlinear interaction of sound waves to a new type of loudspeaker design,” The Journal of the Acoustical Society of America, vol. 73, no. 5, pp. 1532–1536, 1983.
4. W. S. Gan, J. Yang, and T. Kamakura, "A review of parametric acoustic array in air," Applied Acoustics, vol. 73, no. 12, pp. 1211–1219, 2012.
5. T. Fujii, K. Ariyoshi, M. Nakayama, and T. Nishiura, "Control of maximum demodulation distance in parametric loudspeaker using variable-type acoustic lens with gas layer," IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, vol. E102-A, no. 12, pp. 299–309, 2019.
6. Y. Geng, M. Nakayama, and T. Nishiura, "Demodulated sound quality improvement for harmonic sounds in over-boosted parametric array loudspeaker," *Applied Acoustics*, vol. 186, Art. no. 108460, 2022.
7. Y. Ikezaki, Y. Yoto, M. Nakayama, and T. Nishiura, “Virtual multi-boosted amplitude modulation toward high-pressure audible sound with parametric array loudspeakers,” in *Proc. 2024 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, IEEE, 2024.
8. J. J. Croft, "Theory, history, and the advancement of parametric loudspeaker," White Paper, American Technology Corporation, 2001.
9. Y. Suzuki, F. Asano, H. Y. Kim, and T. Sone, "An optimum computer-generated pulse signal suitable for the measurement of very long impulse responses," The Journal of the Acoustical Society of America, vol. 97, no. 2, pp. 1119–1123, 1998.
10. T. Ge, and J. S. Chang, “Bang-Bang Control Class D Amplifiers: Total Harmonic Distortion and Supply Noise,” IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 56, no. 10, pp. 2353‑2361, 2009.
11. N. Moriya and Y. Kaneda, “Study of harmonic distortion on impulse response measurement with logarithmic time stretched pulse,” Acoustical Science and Technology, vol. 26, no. 5, pp. 462‑464, 2005.
12. Y. Ogami, T. Fukumori, M. Nakayama, and T. Nishiura, "Enhanced audio spot based on multiple carrier waves with extra parametric loudspeaker," in Proc. Inter-Noise 2017, pp. 3233–3240, 2017.

POSTER SESSION

AUGUST 19 (TUE)

14:20 – 16:00	Poster Session	326 (3F)
Chair(s)	Il Sung (Agency for Defense Development) Kyusik Chang (VRSound)	
1	A Study on the Measurement of Low-Altitude Small UAV Propulsion Noise	
	Il Sung, Seongil Kim, Wonjung Yoon, Johwan Yang	
2	A Hybrid Self-Attention and U-Net Architecture for three-dimensional Ocean Sound Speed Fields Prediction	
	Yingjie Li, Jixing Qin	
3	Deep Learning Approach to Identify Left Ventricular dysfunction from ECG in Patients with Left Bundle Block Patients	
	Chanjin Kwon, Hyeongjun Ha, Hyebin Gwag, Jongwon Seok	
4	Multilingual text-to-speech model based on StyleTTS 2	
	Hojune Lee, Minseo Kim, Jongwon Seok	
5	Acoustic and Visual Surveys to Analyze the Behavior and Signature Whistles of Indo-Pacific Humpback Dolphins (Sousa chinensis) along the Western Coast of Taiwan	
	Xiang-Hong Lin, Shih-Hsien Weng, Chi-Hung Lin, Wei-Chun Hu, Lien-Siang Chou, Chi-Fang Chen	
6	Investigation of the ultrasonic wave propagation in biological tissues with a blood channel filled with microparticles	
	Mohan Zhang, Ivan Smirnov, James Kwan	
7	Unconventional Dirac-Weyl semimetal and step edge states in acoustic realizations	
	Peng Wu, Yu-Gui Peng, Xue-Feng Zhu	
8	Brillouin Light Scattering Imaging of Undistorted GHz Surface Phonons and Acoustic-Exciton Transport Control	
	Jie Yang, Xiao-Ze Liu, Xue-Feng Zhu	
9	EEG-Audio Multimodal Emotion Recognition via Cross-Modal Attention	
	Hyoung-Gook Kim, Ye-Ji Yoo, Kyu-Hyeok Lee, Jin-Young Kim	

POSTER SESSION

AUGUST 19 (TUE)

14:20 – 16:00	Poster Session	326 (3F)
10	Fast Multi-Frequency DOA Estimation Using Frequency-Difference Sparse Bayesian Learning	
	Xiaoyan Liu, Haiqiang Niu, Haibin Wang	
11	An Improved Particle-Based Track-Before-Detect Algorithm for Underwater Weak Target Tracking	
	Yifei Li, Bin Qi, Guolong Liang	
12	Virtual sound source construction based on direct-to-reverberant ratio control using a line array loudspeaker consisted of parametric and electrodynamic loudspeakers	
	Takumi IRIGUCHI, Yuting GENG, Masato NAKAYAMA, Takanobu NISHIURA	
13	Estimating Reverberation Time in Enclosed Spaces by Analyzing Spectrogram Variations in Stereo Recordings	
	Juseong Kim, Joonwhi Kim, Jungyu Choi, Sungbin Im	
14	Robust TDOA Estimation via Spatial Coherence and CNN-Based Integration in Noisy and Reverberant Environments	
	Eojin Kim, Joonwhi Kim, Jungyu Choi, Sungbin Im	
15	Optimized Microphone Array Configurations for TDOA-Based Direction Finding	
	Jae-Hyoun Ha	
16	Low-Latency Android-Based Deployment of Neural Speech Enhancement for Hearing Aid Applications	
	Jin Tae Seok, Hyun Jae Shin, Seon Man Kim	
17	Low-Latency Deep Learning-Based Denoising of Self-Generated Footstep Noise in Quadruped Robots for Remote Acoustic Situational Awareness	
	Hyun Jae Shin, Jin Tae Seok, Seon Man Kim	
18	Study on railway rolling sound calculation model	
	Hyo-In Koh, Anders Nordborg	

POSTER SESSION

AUGUST 19 (TUE)

14:20 – 16:00	Poster Session	326 (3F)
19	An automatic target extraction algorithm for active acoustic detection based on image processing	
	Fan Yin, Yiming Gu, Haibin Wang, Chao Li, Leixin Nie, Hao Yin	
20	Passive source localization by two hydrophones in direct arrival zone and shadow zone	
	Kang Zheng, Jixing Qin, Shuanglin Wu	
21	Deep-sea active detection method based on vertical phased array	
	Naibin Chen, Dayong Peng, Li Ma, Tao Hu	
22	Analysis of Vertical Correlation Characteristics of Deep-Sea Acoustic Fields in Cold Eddy Environments Based on Remote Sensing Data	
	Pengxi Zhou, Tao Hu, Zhen Wang, Fujin Yang	
23	Analysis of spatial correlation of non complete deep-sea acoustic axis sound field	
	Zhifei Fang, Xinyi Guo	
24	Feasibility of seawater temperature observation using coastal acoustic tomography in Yeosu Bay of South Korea	
	Wan-Gu Kim, Young Geul Yoon, Ho Youn Ji, Min-Seok Choi, Seong Hyeon Kim, Bok Kyoung Choi, Byoung-Nam Kim, Siwoo Lee	
25	Implementation and Sea Trial Validation of an OFDM-based Broadband Underwater Acoustic Communication System for Seismic Data Transmission	
	Seung-Gyu Kim, Sae-Yong Park, Tae-Ho Im	
26	Smart Manufacturing-Oriented Dataset for Early Detection of Bearing Defects	
	Yunsu Kim, Seungwoo Lee, Yujin Bang, Jongwon Seok	
27	A Shallow-water Source Localization Neural Network Based on Source Depth Classification	
	Jing Guo, Juan Zeng	

POSTER SESSION

AUGUST 19 (TUE)

14:20 – 16:00	Poster Session	326 (3F)
28	Unambiguous Localization Method for Long-Baseline Acoustic Sources Based on Consistent Likelihood Decision	
	Dong Xu, Yan Wang, Bin Qi, Binchao Yang, Xiang Li	
29	High-Precision Underwater Source Localization Based on Motion Compensation and Ray Tracing	
	Binchao Yang, Yan Wang, Bin Qi, Dong Xu, Xiang Li	
30	Analysis of uncertainty aspects for multi-megawatt turbines in acoustic measurements	
	Eunkuk Son, Seungjin Kang, Jinjae Lee, Minsang Kang, Songjune Lee, Gwang-Se Lee	
31	Review of 3D acoustic intensimetry for the sound source localization based on a compact microphone array	
	In-Jee Jung, Wan-Ho Cho	
32	Pork Adulteration detection in Beef Using an Electronic Nose System	
	Tajmal Hussain, Jongwon Seok	
33	Lightweight Neural Codec for DIFAR Sonobuoy Signal Transmission Using Residual Quantization and Pruning	
	Yeonjin Park, Jungpyo Hong	
34	High-Speed Moving Object Tracking Using Deep Reinforcement Leaning	
	Jiinyong Ha, Jungpyo Hong, Jongwon Seok	
35	Web-Based Real-Time Speech Conversion System for Dysarthria	
	Kwanghyun Park, Jungpyo Hong	

PROCEEDINGS of the
16th International Conference on Theoretical and Computational Acoustics

August 18 to 22, 2025 BEXCO, Busan Korea

A Study on the Measurement of Low-Altitude Small UAV Propulsion Noise

Il Sung[1]; Seongil Kim[2]; Wonjoong Yoon[3]; Johwan Yang[4]
[1] Agency for Defense Development, Republic of Korea
[2] Agency for Defense Development (Retired), Republic of Korea
[3] Saevit-Tech, Republic of Korea
[4] Saevit-Tech, Republic of Korea

ABSTRACT

This study analyzes the acoustic characteristics of a small unmanned aerial vehicle (sUAV) operating at low altitude. Acoustic measurements were conducted using microphones capable of detecting audible frequencies and infrasound sensors for low-frequency noise detection. The microphone array consisted of four sensors arranged orthogonally to estimate the direction of arrival (DOA) of the sound source. The infrasound sensors were connected to porous hoses to reduce wind-induced noise.
The sUAV under investigation employs a fixed-wing airframe with a propeller-driven engine, rather than the more common quadcopter configuration. It is equipped with a reconnaissance camera that transmits real-time video and positional data.

This paper analyzes the sUAV's frequency and signal characteristics, and examines the feasibility of continuous detection and localization based on DOA estimation using the microphone array. The accuracy of acoustic-based localization was verified by comparing it with the actual position data of the sUAV. Tonal and harmonic components associated with the propeller's rotational speed were also identified through frequency analysis.

The results provide foundational data for the development of sUAV detection and tracking systems.

Keywords : Small Unmanned Aerial Vehicle (sUAV), Acoustic Measurement, Direction of Arrival (DOA) Estimation, Microphones and Infrasound Sensor, Propeller Noise Analysis

[1] isung92@gmail.com
[2] kim_seongil@yahoo.com
[3] ywj@saevit.co.kr
[4] yjh@saevit.co.kr

PROCEEDINGS of the
16th International Conference on Theoretical and Computational Acoustics

August 18 to 22, 2025 BEXCO, Busan Korea

A Hybrid Self-Attention and U-Net Architecture for three-dimensional Ocean Sound Speed Fields Prediction

Yingjie Li[1], Jixing Qin[2]
[1] The Institute of Acoustics of the Chinese Academy of Sciences
[2] The Institute of Acoustics of the Chinese Academy of Sciences

ABSTRACT

As a fundamental part of marine acoustics research, the prediction of ocean sound speed fields (SSFs) is important for underwater communications, marine resource exploration, and environmental monitoring. Accurate prediction of SSFs within specific three-dimensional geographic regions facilitates the characterization of acoustic propagation in complex environments, which is essential for supporting subsequent acoustic detection and communication. Current researches usually consider spatial and temporal characteristics of SSF independently. In this work, we propose a deep learning model combining a shifted-window self-attention mechanism with U-Net to predict 3D SSFs. The spatiotemporal characteristics of SSFs can be extracted by the multi-head self-attention mechanism, and their details can be enhanced through the localized spatial information recovery capability of U-Net. Based on experimental results using the SSF dataset from the South China Sea, our method outperforms traditional Convolutional Long Short-Term Memory networks in temporal extrapolation capability. Furthermore, our approach demonstrates superior adaptability to complex environmental regions such as thermocline zones. This research introduces a new approach for the prediction of 3D SSFs, enabling a transition from conventional static analysis to dynamic real-time forecasting. This advancement allows for dynamic communication path planning based on real-time SSF predictions, thereby enhancing the efficiency and reliability of underwater communication systems.

[1] liyingjie@mail.ioa.ac.cn
[2] qjx@mail.ioa.ac.cn

PROCEEDINGS of the
16th International Conference on Theoretical and Computational Acoustics

August 18 to 22, 2025 BEXCO, Busan Korea

Deep Learning Approach to Identify Left Ventricular dysfunction from ECG in Patients with Left Bundle Block Patients

Chanjin KWON[1], Hyeongjun HA[2], Hyebin GWAG[3], Jongwon SEOK[4]

[1] Changwon National University

[2] Changwon National University

[3]Samsung Changwon Hospital

[4] Changwon National University

ABSTRACT

Left ventricular systolic dysfunction (LVSD) is associated with increased mortality and is sometimes reversible when found early. While artificial intelligence (AI)-enabled electrocardiogram (ECG) has shown potential in screening for LVSD, it has not been validated in patients with left bundle branch block (LBBB). Given that LBBB may be both a cause and consequence of LVSD, this pilot study developed a deep learning model for LVSD detection in LBBB patients using a limited dataset. ECGs from 508 patients were analyzed, with LVSD defined as an ejection fraction ≤ 35%. An autoencoder-based anomaly detection model was combined with a convolutional neural network (CNN), followed by a lead-wise ensemble approach. The final model achieved an accuracy of 0.81, precision of 0.87, recall of 0.56, and AUC of 0.75. Despite dataset limitations, findings suggest potential for deep learning models in detecting LVSD among LBBB patients.

[1] kchanjini@gmail.com

[2] 500mltodtnqud@naver.com

[3] tgfbk@naver.com

[4] jwseok@changwon.ac.kr

PROCEEDINGS of the
16th International Conference on Theoretical and Computational Acoustics

August 18 to 22, 2025 BEXCO, Busan Korea

Multilingual text-to-speech model based on StyleTTS 2

Hojune LEE[1]; Minseo KIM[2]; Jongwon SEOK[3]
[1]Changwon National University, Republic of Korea
[2]Changwon National University, Republic of Korea
[3]Changwon National University, Republic of Korea

ABSTRACT

Recent advancements in text-to-speech (TTS) models have led to highly natural and expressive speech synthesis. However, generating consistent and high-quality speech across multiple languages remains a challenging task. In this study, we propose a multilingual TTS model based on StyleTTS 2, a diffusion-based model designed for human-level speech generation. We fine-tuned the base model using four language datasets—English, Spanish, French, and Korean—while maintaining a unified speaker identity. Key components of StyleTTS 2, including style diffusion, SLM-based adversarial discriminators, and differentiable duration modeling, are leveraged to enhance prosody, naturalness, and linguistic alignment. The proposed model maintains efficient inference speed and preserves synthesis quality across languages with minimal degradation in synthesis quality. This work highlights the adaptability of StyleTTS 2 for multilingual speech synthesis and presents a practical framework for extending high-quality TTS systems to diverse language environments.

Keywords: Text-to-speech, Multilingual, Cross-lingual

1. Introduction

Text-to-Speech (TTS) refers to the synthesis of natural human-like speech from text input(1). With recent advances in deep learning technologies, the performance of TTS models has significantly improved, enabling the generation of highly natural and expressive speech at a human level. Moreover, capabilities such as emotional expression, prosody control, multilingual synthesis, and zero-shot speaker adaptation have become possible. StyleTTS 2 is a TTS model that can synthesize speech in a zero-shot manner by extracting the speech style from a reference audio, using either single-speaker or multi-speaker speech data(2). However, the original StyleTTS 2 model is trained only on English, which limits its ability to synthesize speech in other languages. In this paper, we propose a multilingual TTS model based on StyleTTS 2, fine-tuned with multilingual datasets to enable zero-shot and cross-lingual speech synthesis for multilingual text. We also introduce preprocessing methods for speech data collected from various platforms to fine-tune the StyleTTS 2 model.

2. Related Work

2.1 StyleTTS 2

StyleTTS 2 is a text-to-speech (TTS) model that achieves human-level speech synthesis through a combination of a style diffusion model and adversarial training with a large speech language model (SLM). It models speaking style as a latent variable, which is sampled through a diffusion process. This approach enables the generation of diverse speaking styles even without reference audio. By sampling only a fixed-length style vector instead of the entire speech, the model significantly improves inference speed while preserving the flexibility and expressiveness of diffusion-based speech generation. This also allows for efficient end-to-end training.

2.2 Style Diffusion Model

StyleTTS 2 models speech style as a latent probabilistic variable and samples it using a diffusion model.

[1] lhj20183029@gmail.com
[2] kimminsae123@gmail.com
[3] jwseok@changwon.ac.kr

This approach enables the generation of an appropriate speaking style for a given text without relying on reference speech. In multi-speaker settings, only a fixed-length style vector is sampled, rather than the full speech, allowing faster inference while retaining the advantages of the diffusion model's rich generative capabilities. This is especially suitable for multilingual environments, where prosody and speaking style differ significantly across languages.

2.3 Speech Large Language Model (SLM)

Large-scale self-supervised speech language models (SLMs) have demonstrated the ability to extract rich linguistic and acoustic representations. StyleTTS 2 incorporates such models—pretrained on multilingual data, such as WavLM—as discriminators. These SLM-based discriminators enhance the naturalness of synthesized speech through end-to-end adversarial training. The training process transfers the knowledge embedded in the SLM encoder to the generative model, allowing it to learn a latent space optimized for speech synthesis. This approach not only improves the naturalness of generated speech but also provides strong generalization and robustness for out-of-distribution (OOD) text.

2.4 Differentiable Duration Modeling

Duration modeling determines how long each phoneme in the input text should be held during speech synthesis. Traditional TTS systems often rely on external aligners or non-differentiable methods, which can become bottlenecks. To address this, StyleTTS 2 introduces a differentiable duration modeling approach using a novel non-parametric upsampler that aligns frames to phonemes. Each phoneme's duration is modeled as a random variable, and the duration predictor is modified to output the probability that a phoneme lasts longer than a certain threshold. This enables end-to-end adversarial training and overcomes the limitations of conventional Gaussian upsampling methods, such as fixed width constraints. It also allows for effective modeling of variable-length alignments without additional training, improving the naturalness of synthesized speech in multilingual scenarios where phoneme durations vary between languages.

2.5 International Phonetic Alphabet (IPA)

The International Phonetic Alphabet (IPA) is a standardized system of phonetic notation commonly used in linguistics. It is designed to represent all the sounds of spoken languages in a consistent and precise manner. The IPA comprises 107 base characters, 52 diacritic marks, and 4 suprasegmental symbols. Since IPA provides a unified representation of pronunciation across languages, it enables more natural and accurate accent rendering in cross-lingual synthesis. Additionally, it significantly reduces the number of tokens required in the model compared to using individual phoneme systems for each language(3, 4).

2.6 EmergentTTS-Eval

EmergentTTS-Eval(5) is a recent evaluation framework designed to overcome the limitations of traditional TTS evaluation methods such as Mean Opinion Score (MOS) or Word Error Rate (WER), which often lack semantic coverage and are costly to scale. By leveraging a Large Audio Language Model (LALM) as a judge, EmergentTTS-Eval provides automated and fine-grained assessments of prosody, pronunciation, and contextual alignment in synthesized speech. It supports high correlation with human preferences and exhibits strong inter-judge consistency. We adopt this benchmark to ensure a comprehensive and reproducible evaluation of cross-lingual and expressive capabilities in TTS systems.

3. Method

3.1 Data Collection and Preprocessing

For model fine-tuning, we collected publicly available multi-speaker speech datasets in four languages: English (LibriTTS(6)), French (LibriVox(7)), Spanish (TEDx Spanish Corpus(8)), and Korean (AI Hub(9)). All audio files were resampled to 24 kHz and constrained to durations between 1.25 and 10 seconds. To prevent overlap with speaker identifiers used in the pretrained model, speakerIDs were relabeled with non-conflicting numerical labels. Table 1 summarizes the total duration of training and validation sets used for each language prior to fine-tuning.

Table 1 – Duration of datasets used for training.

Language	Train	Validation
English	2,281.91s	515.55s
French	2,532.89s	617.43s
Spanish	1,969.87s	466.78s
Korean	1,828.32s	428.12s

3.2 Phonetic Preprocessing

To train the model, transcriptions in each language were converted into the International Phonetic Alphabet (IPA). This was done using the Python library phonemizer in conjunction with eSpeak. Since the IPA symbols used vary by language, some characters may not be included in the pretrained model's input token set. Among the 178 original input tokens of StyleTTS 2, unused symbols were removed and replaced with new IPA characters required for the target languages, keeping the overall token size consistent.

3.3 Architecture Modification

To improve pronunciation accuracy and prosody in speech synthesis, we replaced the monolingual PL-BERT model used in StyleTTS 2 with multilingual PL-BERT model. This allows the model to better capture the contextual representation of phonemes across different languages, facilitating more natural and coherent multilingual synthesis.

4. Experiment

We trained our multilingual TTS model based on pretrained multispeaker StyleTTS 2 model. The base model was originally trained on VCTK—consisting of approximately 44,000 short clips from 109 native speakers with diverse accents—and the LibriTTS train-clean-460 subset, which includes about 245 hours of audio from 1,151 speakers. We used the multilingual datasets shown in Table 1 for fine-tuning. Each dataset contributed a limited amount of speech data (less than one hour), allowing us to evaluate the model's low-resource adaptation capabilities. The fine-tuning process was conducted using the preprocessed multilingual datasets on a single NVIDIA RTX 4090 GPU. The maximum input length was set to max_len = 200, equivalent to 2.5 seconds of audio, and training was performed over 30 epochs.

5. Results

To assess the effectiveness of the proposed multilingual TTS model, we conducted a series of experiments evaluating both monolingual and cross-lingual synthesis capabilities. We analyzed the quality of synthesized speech through visual inspection of mel-spectrograms, and automated evaluation using the EmergentTTS-Eval benchmark.

Figure 1 shows that the synthesized speech effectively captures the prosodic and spectral characteristics of the reference audio. The generated spectrograms exhibit clear structural alignment with the reference, including comparable energy contours, phoneme transitions, and overall rhythmic flow. In particular, the voiced and unvoiced segments are well separated and naturally distributed, indicating accurate phoneme-level duration modeling and sound articulation.

Figures 2 and 3 demonstrate that when synthesizing identical sentences (English and Korean) using four different reference speakers from English, Spanish, French, and Korean, the resulting spectrograms exhibit high consistency in phoneme sequencing and sentence-level structure. However, notable differences are observed in total utterance duration, phoneme-level segment lengths, and dominant frequency bands. These variations reflect the influence of each reference speaker's prosodic patterns and vocal characteristics, indicating that the model successfully performs cross-lingual style transfer. While maintaining identical textual content, the observed differences in rhythm and frequency distribution demonstrate the model's ability to adapt to speaker-specific delivery patterns while preserving linguistic accuracy.

Overall, these spectrogram-based evaluations confirm that the proposed multilingual TTS model not only maintains intelligibility across languages but also successfully preserves speaker-dependent expressive characteristics in both monolingual and cross-lingual synthesis scenarios.

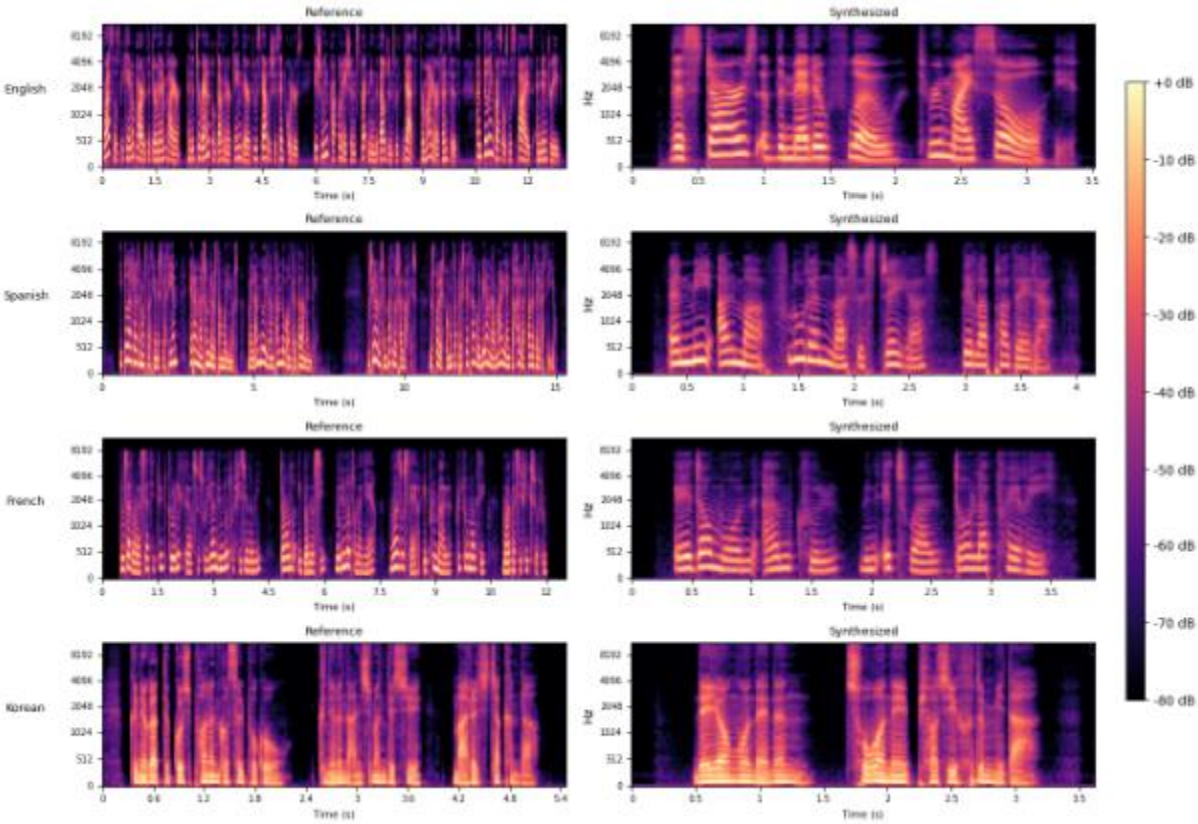


Figure – Mel-spectrogram comparison of reference and synthesized speech in four languages. Each row represents a language, with the reference audio on the left and the corresponding synthesized output on the right.

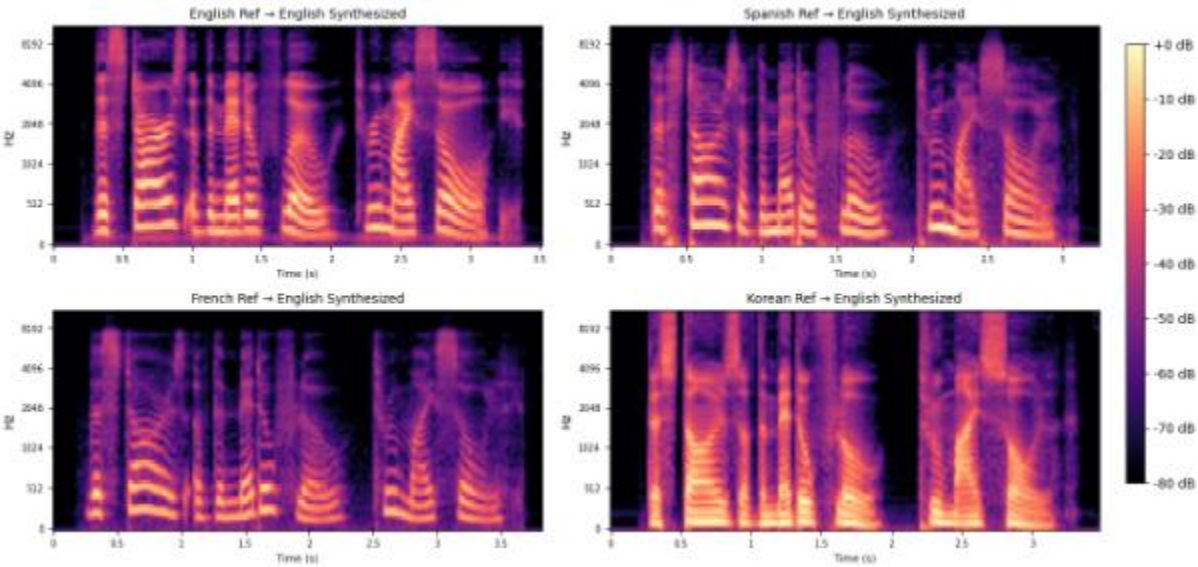


Figure – Mel-spectrograms of synthesized speech for the same English sentence, generated using reference audio from speakers of four different languages.

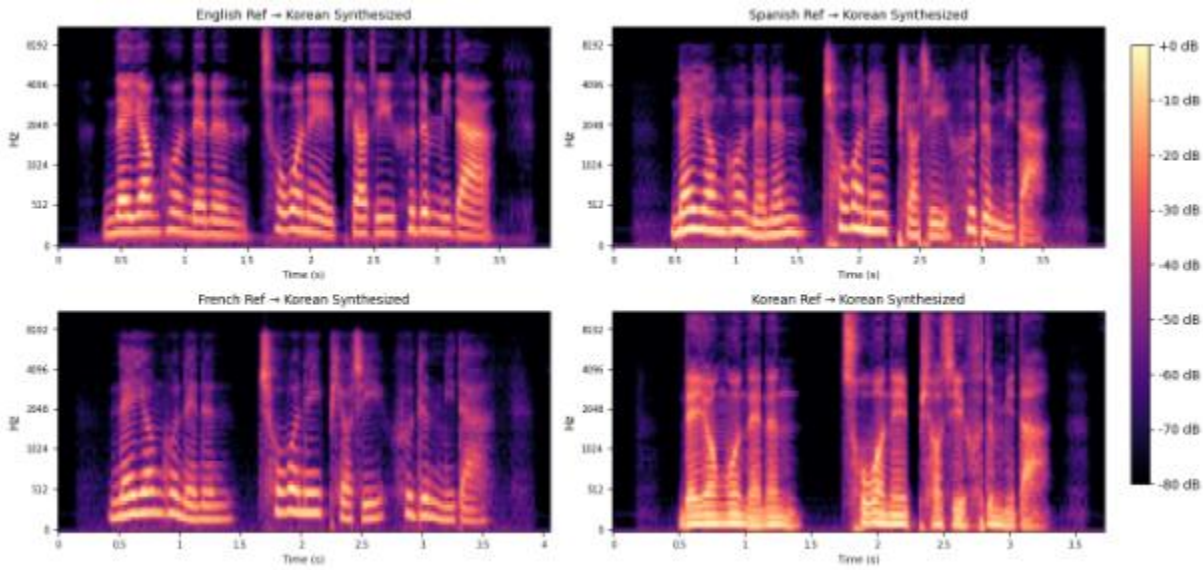


Figure – Mel-spectrograms of synthesized speech for the same Korean sentence, generated using reference audio from speakers of four different languages.

In addition, we evaluated both the base model and fine-tuned multilingual TTS models using the EmergentTTS-Eval benchmark, which includes 1,645 test prompts across six challenging categories: Emotions, Paralinguistics, Foreign Words, Syntactic Complexity, Complex Pronunciation, and Questions. A subset of 500 samples was used for evaluation using Gemini 2.0 Flash as a judger.

The fine-tuned model achieved a higher overall win rate (33.23%) compared to the base model (32.37%), with notable improvements in the Emotions (+6.14%) and Paralinguistics (+2.07%) categories. Word Error Rate (WER) and LLM-predicted MOS remained unchanged across both models. These results demonstrate that multilingual fine-tuning enhances expressive performance while preserving intelligibility.

Table 2 – Performance comparison between base model and fine-tuned multilingual models using EmergentTTS-Eval

Metric	Base model	Multilingual model
MOS	4.2066	4.2066
WER	16.62	16.62
Win Rate	32.37%	33.23%
Emotions Win Rate	26.79%	32.93%
Foreign Words Win Rate	29.38%	30.63%
Paralinguistics Win Rate	38.97%	41.04%
Pronunciation Win Rate	20.00%	18.52%
Syntactic Complexity Win Rate	32.39%	30.88%

6. Conclusion

In this study, we developed and evaluated a multilingual text-to-speech model based on StyleTTS 2, capable of zero-shot synthesis across multiple languages. Starting from an English-pretrained base model, we fine-tuned it using less than one hour of speech data per language, demonstrating the model's adaptability to multilingual data. The proposed system not only supports synthesis in multiple languages but also enables cross-lingual speech generation—synthesizing target language speech from reference audio in a different language.

These results demonstrate the potential of StyleTTS 2 for multilingual adaptation with limited data. In future work, expanding the training dataset may further improve synthesis quality and allow support for additional languages. Moreover, integration with ASR models could enable real-time multilingual translation and dubbing systems.

However, there is a risk of misuse, as the model can easily mimic an individual's voice. Future research should explore mechanisms—such as embedding digital watermarks in generated speech—to preserve the model's capabilities while preventing malicious use.

ACKNOWLEDGEMENTS

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea(RS-2022-NR075671).

This research was conducted using the "Emotional and Speaking Style Speech Synthesis Dataset" built with the support of the National Information Society Agency (NIA), funded by the Ministry of Science and ICT (MSIT), Republic of Korea.

REFERENCES

1. Tan X, Qin T, Soong F, Liu TY. A survey on neural speech synthesis. arXiv preprint arXiv:2106.15561. 2021 Jun 29.
2. Li YA, Han C, Raghavan VS, Mischler G, Mesgarani N. StyleTTS 2: Towards human-level text-to-speech through style diffusion and adversarial training with large speech language models. arXiv preprint arXiv:2306.07691. 2023 Jun 13.
3. Sanchez A, Falai A, Zhang Z, Angelini O, Yanagisawa K. Unify and conquer: How phonetic feature representation affects polyglot text-to-speech (TTS). arXiv preprint arXiv:2207.01547. 2022 Jul 4.
4. Lou H, Paik HY, Li S, Hu W, Yao L. Generalized multilingual text-to-speech generation with language-aware style adaptation. arXiv preprint arXiv:2504.08274. 2025 Apr 16.
5. Manku RR, Tang Y, Shi X, Li M, Smola A. EmergentTTS-Eval: Evaluating TTS models on complex prosodic, expressiveness, and linguistic challenges using model-as-a-judge. arXiv preprint arXiv:2505.23009. 2025 May 23.
6. Zen H, Dang V, Clark R, Zhang Y, Weiss RJ, Jia Y, Chen Z, Wu Y. LibriTTS: A corpus derived from LibriSpeech for text-to-speech. arXiv preprint arXiv:1904.02882. 2019 Apr 5.
7. Munich Artificial Intelligence Laboratories GmbH. M-AILABS Speech Dataset - French [Internet]. Munich: M-AILABS; 2017-2018 [cited 2025 Jul 12]. Available from: https://www.openslr.org/94/
8. Hernández-Mena CD. TEDx Spanish Corpus: Audio and transcripts in Spanish taken from the TEDx Talks [Internet]. Mexico City: Universidad Nacional Autónoma de México; 2019 [cited 2025 Jul 12]. Available from: https://openslr.org/67/
9. National Information Society Agency. Emotional and Speaking Style Speech Synthesis Dataset [Internet]. Seoul: AI Hub; 2020 [cited 2025 Jul 12]. Available from: https://aihub.or.kr/aihubdata/data/view.do?dataSetSn=466

Acoustic and visual surveys to analyze the behavior and signature whistles of Indo-Pacific humpback dolphins (*Sousa chinensis*) along the western coast of Taiwan

Xiang-Hong LIN[1]; Shih-Hsien WENG[2]; Chi-Hung LIN[3]; Wei-Chun HU[4]; Lien-Siang CHOU[5]; Chi-Fang CHEN[6]

Department of Engineering Science and Ocean Engineering, National Taiwan University, Taiwan
Department of Engineering Science and Ocean Engineering, National Taiwan University, Taiwan
Pvidence University, Taiwan
Ship and Ocean Technology Research Center, National Taiwan University, Taiwan
Institute of Ecology and Evolutionary Biology, National Taiwan University, Taiwan
Department of Engineering Science and Ocean Engineering, National Taiwan University, Taiwan

ABSTRACT

The Indo-Pacific humpback dolphin (*Sousa chinensis*), listed as Vulnerable by the IUCN, inhabits Taiwan's western coastal waters, where habitat degradation from coastal development poses serious threats. This study investigates the species at the Xin-Huwei River estuary through combined visual surveys and passive acoustic monitoring conducted in 2019–2021 and 2024. Using towed arrays and fixed acoustic stations, whistles and click-trains were recorded and analyzed with Raven Lite software. The focus was on identifying potential signature whistles—individually distinctive vocalizations used for social communication. By applying the SIGID method (Janik et al., 2013) and analyzing isolated whistle patterns, one candidate signature whistle was identified. The findings support the integration of visual and acoustic methods and provide a foundation for building a signature whistle catalog for *S. chinensis* in Taiwan, aiding future behavioral studies and conservation planning.

Keyword： Bioacoustics , Signature whistles , *Sousa chinensis*

[1] r12525095@ntu.edu.tw
[2] wengshihhsien@gmail.com
[3] chihunglin0505@gmail.com
[4] william_hu@outlook.com
[5] Chouls@ntu.edu.tw
[6] chifang@ntu.edu.tw

PROCEEDINGS of the
16th International Conference on Theoretical and Computational Acoustics

August 18 to 22, 2025 BEXCO, Busan Korea

Investigation of the ultrasonic wave propagation in biological tissues with a blood channel filled with microparticles

Mohan Zhang[1], Ivan Smirnov[2], James Kwan[3]

[1] Oxford Suzhou Centre for Advanced Research

[2] Oxford Suzhou Centre for Advanced Research

[3] Oxford University

ABSTRACT

Ultrasound, as a non-invasive and controllable method to achieve drug delivery in combination with microparticles, shows great promise in the accurate treatment of site-specific diseases such as revascularization or tumor. Here, we applied degradable porous PLGA microparticles loaded with drugs as cavitation enhancers ultimately can promote to achieve drug delivery. We experimentally measured the physical properties of a medium with these acoustically sensitive microparticles, such as the speed of sound, density, attenuation coefficient, and nonlinear coefficient, and applied them to simulate the propagation effect of various ultrasonic signals in a channel (simulating a vascular network) with different concentrations of microparticles. The work presents simulation results, in particular, it was found that the intensity of the ultrasound signal is inversely proportional to the concentration of microparticles, and the propagation of ultrasound to the focal point shows different signal variations for different microparticles. Also our approach makes it possible to study the use of porous microparticles in drug delivery and ultrasound imaging in a wider range of parameters than is usually available in experimental work.

[1] mohan.zhang@oxford-oscar.cn

[2] Ivan.Smirnov@oxford-oscar.cn

[3] james.kwan@balliol.ox.ac.uk

PROCEEDINGS of the
16th International Conference on Theoretical and Computational Acoustics

August 18 to 22, 2025 BEXCO, Busan Korea

Unconventional Dirac-Weyl semimetal and step edge states in acoustic realizations

Peng Wu; Yu-Gui Peng; Xue-feng Zhu

[1]*School of Physics and Innovation Institute, Huazhong University of Science and Technology, Wuhan 430074, China*

ABSTRACT

Topological materials have garnered significant attention for their topologically protected surface transport properties, demonstrating considerable application potential in novel waveguide devices and acoustic wave manipulation. Here, we report an unconventional Dirac-Weyl semimetal system that establishes a direct physical connection between Dirac points and Weyl points, transcending the conventional paradigm of their mere coexistence. Theoretical analysis and computational results reveal that the real meaningful coexistence of the two enhances the topological properties of the system, giving rise to a remarkable unidirectional chain-like ultra-long surface arc state. Furthermore, we propose a novel multi-channel transport mechanism mediated by geometrically engineered step edge states. Through the design of acoustic metamaterials with specifically tailored step structures, we achieve programmable dynamic control of multi-channel waveguides. This research provides both new design principles and an experimental platform for efficient manipulation and information transfer of acoustic waves.

Keywords: Acoustic metamaterials, topological surface states

1. INTRODUCTION

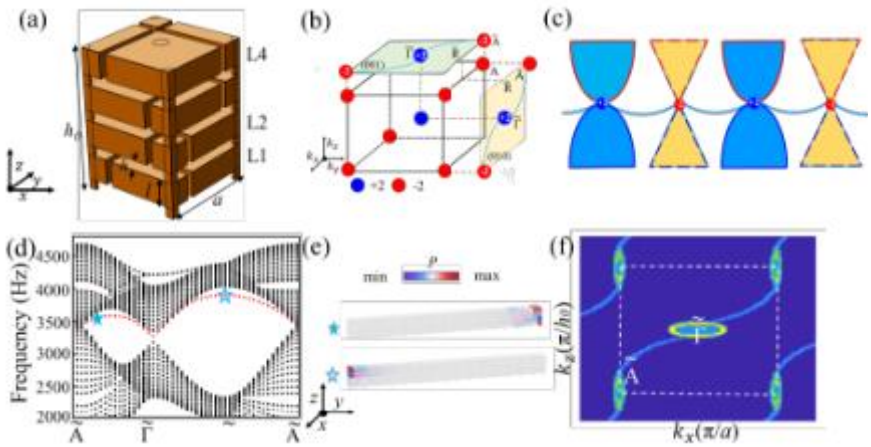


Figure 1 Ultra-long surface states in UDWS topological phases. (a) Front view of acoustic unit cell. (b) Schematic distribution of Weyl points (blue spheres) and Dirac points (red spheres) in the Brillouin zone, with their surface arc projections along the (010) and (001) directions. (c) Characteristic ultra-long surface arcs spanning the entire momentum space in the UDWS system. (d) Bulk dispersion along the (010) projection direction, where black solid lines represent bulk band projections and red solid lines indicate the existence of topological surface states. (e) Eigenmode distribution of surface arc states at the blue star marker in (d). (f) Isofrequency contour at 3335 Hz projected along the (010) direction.

By introducing glide symmetry into systems lacking spatial inversion symmetry, we realize an unconventional Dirac-Weyl semimetal (UDWS) that exhibits surface arc states traversing the entire

Brillouin zone. Through symmetry analysis, we confirm that ideal UDWS states can be achieved in chiral space groups 92 and 96. As illustrated in Fig. 1(a), Dirac points and Weyl points are located at the Brillouin zone center Γ and boundary A, respectively, each carrying topological charges of +2 and -2. In contrast to conventional systems where chirality-opposite Weyl points form internally connected short surface arcs, our UDWS demonstrates a distinct behavior in the (010) projected surface states. Remarkably, the surface arcs connect from the projection of Weyl points at Γ to Dirac points at A, and further extend toward the projection of Weyl points at Γ in adjacent Brillouin zones, as shown in Fig. 1(b). This configuration creates a chain-like surface arc pattern that originates at one terminus and traverses the entire Brillouin zone to connect with the other end. Through the construction of a phononic crystal model, we experimentally verify the characteristic long surface arcs in UDWS, as demonstrated in Figs. 1(c)-(e). This unique unidirectional chain-type surface arc establishes a nontrivial one-way phonon transport channel, suggesting significant potential for UDWS applications in phonon transport engineering.

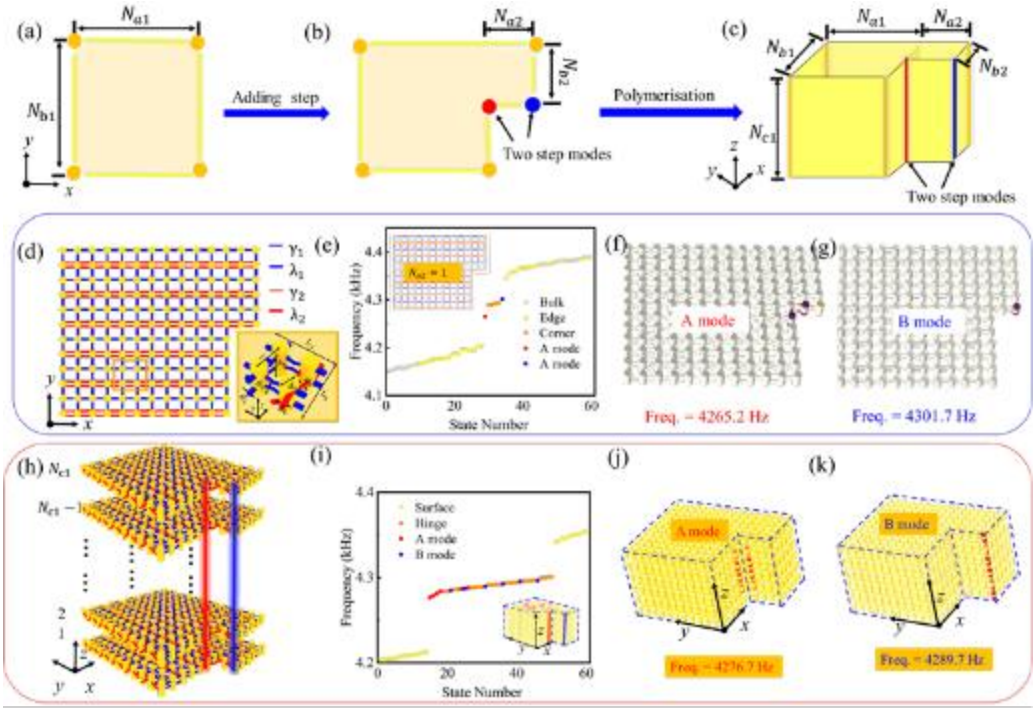


Figure 2. Construction schemes for zero-dimensional (0D) and one-dimensional (1D) step-edge states. (a-c) Schematic illustrations of step-edge state configurations. (d-g) Simulation design for realizing 0D step-edge states in a two-dimensional (2D) acoustic model. (h-k) Simulation design for implementing 1D step-edge states in a three-dimensional (3D) acoustic model.

Beyond utilizing edge states of topological insulators, an alternative effective approach to achieve high-performance edge transport is through step edge states (SESs). In our acoustic system, we introduce an additional geometric unit block to the finite array structure shown in Fig. 2(a), creating the stepped configuration illustrated in Fig. 2(b). This modification enables the observation of two distinct acoustic SESs, namely, mode A (red) at the step edge (270°) and mode B (blue) at the step position (90°). Specifically, we construct an acoustic model of single-layer step edge structure by incorporating three cell blocks into a finite configuration, as depicted in Fig. 2(d), where the unit cell is a quadrupole topological insulator, exhibited in the bottom-right inset. As shown in Fig. 2(e-g), the eigenvalue spectra and sound pressure field distributions reveal two step edge states emerging at frequencies 4265.5 Hz (red, mode A) and 4301.8 Hz (blue, mode B), in addition to conventional quadrupole corner states. These modes exhibit strong localization characteristics at the step edge (270°) and step position (90°), respectively.

The strongly localized modes in the 2D system suggest the possibility of achieving robust unidirectional energy transport along step-edge channels in 3D configurations. As shown in Figs. 2(b-c), we extend our design to three dimensions by stacking the single-layer step structure along the z-direction (Fig. 2h). The frequency spectrum and pressure field distributions (Figs. 2i-k) demonstrate that both modes A and B form stable 1D transport channels along the *z*-direction. Notably, mode B shows uniform distribution among four hinge states, indicating its consistency with hinge states and suggesting second-order topological characteristics in these step edge states. This spatially controlled, stable step edge configuration demonstrates significant potential for applications in acoustic energy harvesting, signal transmission, and vibration control, offering new possibilities for robust wave manipulation in engineered materials.

REFERENCES

1. H. Gao, Y. Kim, J. W. F. Venderbos, C. L. Kane, E. J. Mele, A. M. Rappe, and W. Ren, Dirac-Weyl semimetal: Coexistence of Dirac and Weyl fermions in polar hexagonal ABC crystals, Phys. Rev. Lett. 121, 106404 (2018)
2. H. L. He, C. Y. Qiu, X. X. Cai, M. Xiao, M. Z. Ke, F. Zhang, and Z. Y. Liu, Observation of quadratic Weyl points and double helicoid arcs, Nat. Commun. 11, 1820 (2020)
3. S. Long, J. Yang, H. Wang, Z. Yu, B. Yang, Q. Guo, Y. Xiang, L. Xia, and S. Zhang, Dirac–Weyl semimetal in photonic metacrystals, Opt. Lett. 48, 2349 (2023)
4. [26] R. Wu, J. Z. Ma, S. M. Nie, L. X. Zhao, X. Huang, J. X. Yin, B. B. Fu, P. Richard, G. F. Chen, Z. Fang, X. Dai, H. M. Weng, T. Qian, H. Ding, and S. H. Pan, Evidence for topological edge states in a large energy gap near the step edges on the surface of ZrTe5, Phys. Rev. X 6, 021017 (2016)
5. F. Schindler, Z. J. Wang, M. G. Vergniory, A. M. Cook, A. Murani, S. Sengupta, A. Y. Kasumov, R. Deblock, S. Jeon, I. Drozdov, H. Bouchiat, S. Guéron, A. Yazdani, B. A. Bernevig and T. Neupert, Higher-order topology in bismuth, Nature Phys 14, 918 (2018)
6. [28] M. S. Hossain, F. Schindler, R. Islam, Z. Muhammad, Y. X. Jiang, Z. J. Cheng, Q. Zhang, T. Hou, H. Y. Chen, M. Litskevich, B. Casas, J. X. Yin, T. A. Cochran, M. Yahyavi, X. P. Yang, L. Balicas, G. Q. Chang, W. S. Zhao, T. Neupert and M. Z. Hasan., A hybrid topological quantum state in an elemental solid, Nature. 628, 527 (2024)

PROCEEDINGS of the
16th International Conference on Theoretical and Computational Acoustics

August 18 to 22, 2025 BEXCO, Busan Korea

Brillouin Light Scattering Imaging of Undistorted GHz Surface Phonons and Acoustic-Exciton Transport Control

Jie Yang[1]; Xiao-ze Liu[2]; Xue-feng Zhu[1]

[1] *School of Physics and Innovation Institute, Huazhong University of Science and Technology, Wuhan 430074, China*

[2]*School of Physics and Technology, Wuhan University, Wuhan 430072, China*

ABSTRACT

Space-time-resolved imaging of undistorted surface phonon distribution and evolution is a crucial tool for studying the interactions between GHz lattice vibrations and matter, as well as structures. In this work, we propose to use the micro-focused Brillouin light scattering (μBLS) to achieve high-quality space-time-resolved imaging of one- or two-dimensional surface phonon distribution and evolution at a resonant frequency of 1.21 GHz. The μBLS method leverages its ultra-sensitivity to detect GHz surface phonons with an input radio frequency (RF) power of only -20 dBm. With excellent performance in spatial, temporal, and spectral resolution, the μBLS can effectively observe the tunability and transient spatial evolution of surface phonon distributions when the phase difference of RF signals or the launch pulses from the ports are varied. Furthermore, by integrating surface acoustic wave technology, we successfully demonstrate long-distance exciton transport in monolayer $MoSe_2$.Our proposed approach is significant for designing and characterizing on-chip phononic devices with versatile functionalities.

Keywords: Dynamic phononic lattice, Brillouin light scattering, $MoSe_2$, Exciton transport

INTRODUCTION

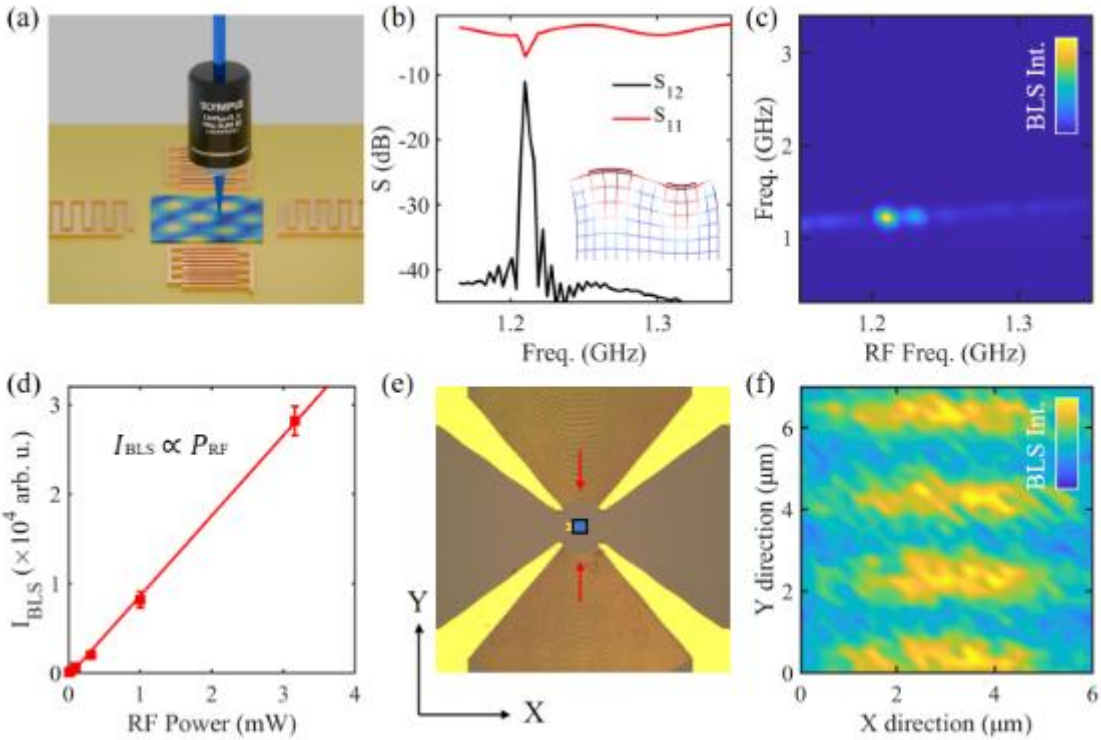


Fig. 1. (a) A schematic of the μBLS for SAW measurements. (b) The RF reflection/transmission spectra (S_{11} and S_{12}) in the frequency range of 1.15 to 1.35 GHz. The resonant frequency is located at 1.21 GHz. The inset shows the displacement field of SAW over one period. (c) A wide-frequency scanning of BLS spectra for the SAW (color bar: BLS intensity). The x-axis represents the RF frequency, while y-axis shows the blue-shifted frequency of photons. (d) The relationship between I_{BLS} and P_{RF}. (e) The asfabricated

double arc-shaped SAW device with two independent IDTs. The two IDTs are capable of applying separate RF inputs. (f) The BLS imaging of GHz SAW with the RF power is extremely low at -20 dBm.

Figure 1(a) shows the schematic of the μBLS for SAW measurements. In Fig. 1(b), we display both the reflection and transmission spectra of electrical RF signals for the as-fabricated SAW device, measured using a vector network analyzer. The results show that the resonant frequency (f_{SAW}) is around 1.21 GHz. The distinctive resonance peak indicates the effective excitation of the Rayleigh mode. In Fig. 1(c), we also present the μBLS measurement results at one location, compared with the electrical RF measurement in in Fig.1(b). Here, we present the relationship between the time-averaged BLS intensity and the RF excitation frequency with the scanning step of 2 MHz, where the x-axis and y-axis show the RF generator input and frequency shift of photons, respectively. Our results reveal that the measured BLS intensity is maximized at 1.21 GHz, where inelastic scattering between photons and phonons is strongest, in agreement with the resonant frequency shown in Fig. 1(b). In Fig. 1(d), we show the measured linear relationship between the intensities of surface phonons and BLS photons. When SAW is excited on the $LiNbO_3$ surface, it also generates a periodic refractive index variation, $\Delta n(x)$, in the piezoelectric crystal [1]

$$I_{BLS} \propto |\Delta n(x)|^2 = \frac{1}{4}|\Delta n_0|^2 \left(\frac{1+\cos(2k_{SAW}x - 2\omega_s t)}{2}\right), \#(1)$$

where Δn_0 denotes the amplitude of refractive index variation and is proportional to the amplitude of SAW (*viz.*, $\Delta n_0 \propto A$), A is amplitude, and $\omega = 2\pi f_{SAW}$ is angular frequency. It is worth noting that the μBLS exhibits superior detection sensitivity, as it can effectively detect surface phonon generated by an RF input as low as -20 dBm, as shown in Fig. 1(e) and 1(f), corresponding to medium deformations with amplitudes much smaller than a few picometers.

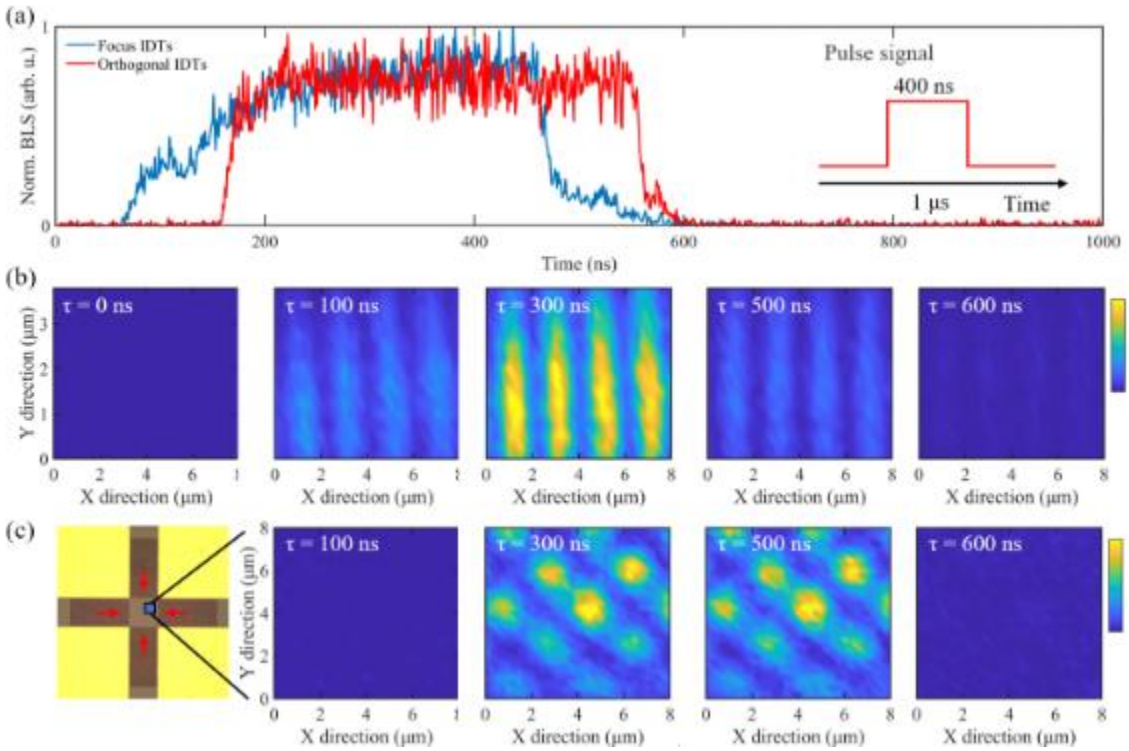


Fig. 2. The space-time-resolved evolutions of surface phonons for different devices. (a) The measured time-resolved BLS signals for the focus IDTs (double arc-shaped SAW device) and orthogonal IDTs, respectively. The inset schematically shows the input RF signal, where the period is 1 μs and the width of rectangular pulse is 400 ns. (b), (c) The BLS imaging of surface phonon evolutions at different times, which renders the formation processes of 1D (focus IDTs) and 2D (orthogonal IDTs) standing wave patterns, respectively.

The μBLS also exhibits superior performance in time-resolved imaging, which can be utilized to observe the space-time evolution of surface phonons, essential in the applications of signal-delay and quantum-control devices [2, 3]. Figure 2(a) presents the time-domain variations of normalized BLS

intensities for the focus (double arc-shaped device) and orthogonal IDTs, respectively, where the input RF signals have a period of 1 μs and rectangular pulse duration of 400 ns. Here, the delay times of rectangular pulses for the two cases are set to differ by 100 ns to demonstrate the superior time resolution. Figures 2(b) and 2(c) display snap-shots at different time moments, showing the emergence and disappearance of 1D and 2D standing-wave surface phonons with nanosecond time resolution. Note that the surface phonons are periodically localized in micrometer-scale regions with nanosecond modulations, which may have great potentials in the quantum-system control.

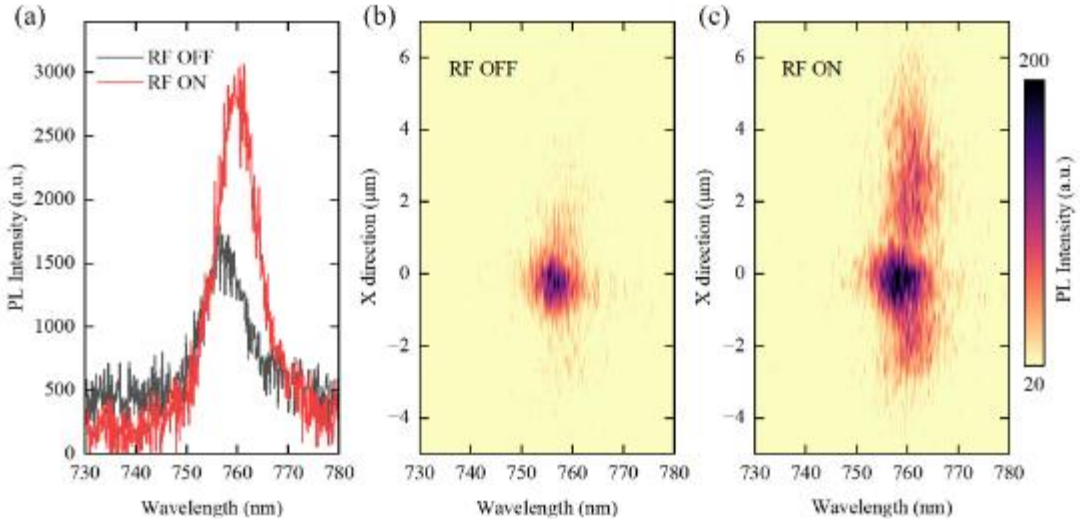


Fig. 3. SAW-driven transport of exciton at 150 K. (a) Comparisons of PL spectra of $MoSe_2$ before and after applying the surface acoustic waves by a RF excitation source, as indicated as "RF off" and "RF on". Here the "RF on" power is set at 100 mW, and the optical excitation power of $MoSe_2$ is set at 30 μW. Real-space PL mapping (b), when SAW is off, and (c) when SAW is on with 100 mW power.

Figure 3(a) presents the photoluminescence (PL) spectral response of monolayer $MoSe_2$ at 150 K. Upon SAW excitation (red curve), two distinct features are observed compared to the reference state without SAW (black curve): (1) PL intensity enhancement with approximately 2-fold increase in integrated intensity; (2) systematic redshift of exciton peak by about 5 meV. This acoustically-induced luminescence enhancement and band structure modulation demonstrate that SAW enables active manipulation of excitonic states through piezoelectric effects. Figures 3(b) and 3(c) further reveal the SAW-mediated control of exciton transport through spatially-resolved PL mapping. Under SAW OFF conditions (left panel), excitons exhibit typical diffusive transport characteristics with a diffusion length below 1 μm, consistent with previously reported exciton diffusion behavior [4, 5]. When applying SAW excitation with P_{SAW} = 100 mW (right panel), distinct exciton luminescence is observed at the sample edge approximately 6 μm away from the laser excitation spot along the wavevector direction. This phenomenon confirms that the periodic strain field generated by SAW can effectively facilitate directional exciton transport, surpassing the conventional diffusion limit. Notably, the transport efficiency shows nonlinear dependence on SAW power, suggesting the possible existence of a phonon-assisted exciton transport mechanism.

REFERENCES

1. J. Yang, et al. Brillouin light scattering imaging of undistorted field distribution and space-time evolution for GHz surface phonons. *Phys. Rev. Applied* **23**, L051001 (2025).
2. A. Noguchi, et al. Single-photon quantum regime of artificial radiation pressure on a surface acoustic wave resonator. *Nat. Commun*. **11**, 1183 (2020).
3. A. Iyer, et al. Coherent optical coupling to surface acoustic wave devices, *Nat. Commun.* **15**, 3993 (2024).
4. A. Javey, et al. Neutral exciton diffusion in monolayer MoS_2. *ACS Nano* **14**, 13433–13440 (2020).
5. M. Kulig, et al. Exciton diffusion and halo effects in monolayer semiconductors. *Phys. Rev. Lett.* **120**, 207401 (2018).

PROCEEDINGS of the
16th International Conference on Theoretical and Computational Acoustics

August 18 to 22, 2025 BEXCO, Busan Korea

EEG-Audio Mutimodal Emotion Recognition via Cross-Modal Attention

Ye-Ji Yoo[1]; Kyu-Hyeok Lee[1]; Hyoung-Gook Kim[1]; Jin-Young Kim[2]
[1]Kwangwoon University, Republic of Korea
[2]Chonnam National University, Republic of Korea

ABSTRACT

Recent advances in emotion recognition increasingly employ multi-modal approaches that integrate physiological and audio data for more accurate and robust emotional state inference. This paper proposes a novel EEG–audio multi-modal emotion recognition model that captures and fuses the spatio-temporal dynamics of both signals using a cross-modal attention mechanism. EEGNet and Temporal Convolutional Networks (TCN) extract spatio-temporal features from EEG data, while a pre-trained VGGish network combined with TCN obtains rich spatio-temporal embeddings from audio inputs. These EEG and audio features are aligned and fused via a cross-modal attention module, allowing each modality to emphasize complementary information and suppress irrelevant or noisy elements. The joint embeddings are then passed to a Multi-Layer Perceptron (MLP) for final emotion classification. Evaluated on DEAP dataset, the proposed approach outperforms EEG-only baselines and existing Conformer-based multi-modal methods, achieving higher classification accuracy, improved training stability, and enhanced interpretability of cross-modal feature interactions. This work demonstrates the effectiveness and efficiency of cross-modal attention for EEG–audio emotion recognition and suggests its potential to advance affective computing.

Keywords: EEG, Audio, Multimodla Emotion Recognition, Cross-Modal Attention

1. INTRODUCTION

Emotion classification is a key technology that enables more natural and emotionally rich interactions between humans and computers. It has been actively studied in various fields such as human-computer interaction, healthcare, and mental health monitoring. Early emotion recognition research was primarily based on unimodal data such as text, speech, images, and physiological signals, and achieved meaningful results within each modality [1, 2]. For instance, Electroencephalography (EEG)-based emotion recognition using hybrid Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) models has shown promising outcomes by effectively capturing brain activity patterns associated with emotional states [3].

However, due to the inherently complex and multidimensional nature of human emotions, unimodal approaches face fundamental limitations in accurately identifying emotional states. To address these challenges, recent studies have increasingly focused on multimodal emotion recognition techniques that integrate two or more modalities [4–6]. These approaches allow for more precise and detailed emotion detection by simultaneously analyzing various data types, such as physiological signals, visual behavior, and audio-visual information. In particular, the complementary integration of heterogeneous modalities has been shown to overcome the shortcomings of unimodal systems and to produce more robust and reliable emotion recognition models [7,8].

Among various multimodal approaches, the combination of visual and auditory informationis widely used due to its ease of data collection and the complementary nature of the information it provides. Notably, among multimedia content, videois considered one of the most effective and commonly consumed formats in modern society, as it includes both visual and auditory stimuli and has a strong capacity to evoke emotional responses. While watching a video, EEG signals can capture immediate physiological responses

to external stimuli, and the accompanying audio conveys emotional cues through features such as intonation, prosody, and timbre.

Nevertheless, EEG signals often suffer from a low Signal-to-Noise Ratio (SNR), which can degrade recognition accuracy, and audio signals alone may not be sufficient to represent complex emotional contexts. Therefore, integrating EEG and audio signals within a multimodal framework can serve as an effective strategy to compensate for the limitations of each individual modality and enhance overall emotion recognition performance.

Based on this perspective, this paer proposes a multimodal emotion recognition model that fuses EEG signals collected during video viewing with the corresponding audio signals from the video. By combining the complementary strengths of physiological and auditory signals, the model aims to achieve more accurate and robust emotion classification. This approach holds significant potential for real-world applications such as affective human-computer interfaces, personalized content recommendation, and mental health monitoring.

2. PROPOSED MULTIMODAL EMOTION RECOGNITION SYSTEM

Fig. 1 illustrates the overall architecture of the multimodal emotion recognition system that fuses spatio-temporal features extracted from two modalities: EEG and audio signals, to improve emotion recognition performance.

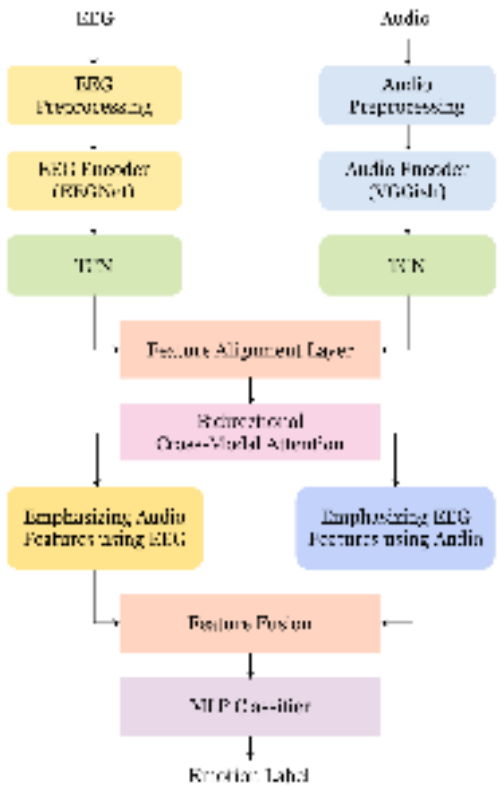


Figure 1 – Architecture of the proposed multimodal emotion recognition.

The system first employs EEGNet [9] and Temporal Convolutional Networks (TCN) [10] to effectively extract spatio-temporal features from EEG data. For the audio input, a pretrained VGGish network [11] combined with TCN is utilized to obtain rich spatio-temporal embeddings. These extracted EEG and audio features are then aligned and fused through a cross-modal attention module [12]. This module allows each modality to selectively emphasize complementary information while effectively suppressing irrelevant or noisy components. Finally, the fused embeddings are fed into a MLP classifier to accurately classify four emotion classes. This integrated framework effectively extracts and fuses high-level spatio-temporal features from the raw signals of each modality, maximizing the emotional information embedded in EEG and audio signals, thereby significantly improving the accuracy and stability of emotion recognition.

2.1 EEG-Based Spatio-Temporal Feature Encoding Using EEGNet and TCN

EEG captures human neural responses to stimuli with high temporal resolution. Notably, electrical

activity in specific brain regions and frequency bands is closely associated with various emotional states. Owing to these characteristics, EEG has been widely utilized in affective computing, and accurately extracting the spatio-temporal variations of EEG signals is crucial for improving emotion classification performance.

In this study, we propose an encoding framework for effectively extracting spatio-temporal features from EEG signals, comprising three stages: preprocessing, spatial feature extraction using an EEG-specific neural network, and temporal pattern learning through a temporal neural architecture. During preprocessing, artifacts such as eye blinks and muscle movements are removed from the raw EEG signals. The signals are then segmented into fixed time intervals and transformed into the frequency domain using Fourier transform. From the transformed data, the mean power within five major frequency bands—delta (0–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta (13–30 Hz), and gamma (30–45 Hz)—is calculated to obtain the power spectral density (PSD) for each channel.

To extract spatial features from these frequency-based inputs, we adopt EEGNet as the feature extractor. EEGNet processes time-series EEG data by applying temporal convolutions to learn frequency filters and then performs depthwise convolutions to capture the spatial distributions associated with each frequency band. This design effectively models spatial dependencies across EEG channels. A subsequent separable convolution layer summarizes temporal information from each channel independently and combines the resulting feature maps to generate compact spatial feature representations.

However, as the output of EEGNet still retains temporal structure, we further integrate a TCN to capture temporal dependencies more precisely. TCN utilizes a causal convolutional architecture with dilation to expand the receptive field, enabling the network to model long-range temporal relationships effectively. This structure is advantageous for capturing the sequential nature of emotional responses and provides better parallelization and training stability compared to traditional recurrent networks.

In summary, the proposed encoder—comprising EEGNet and TCN—jointly learns spatial and temporal characteristics of EEG signals, enabling the extraction of high-level spatio-temporal emotion-related features. This representation serves as a robust foundation for subsequent multimodal fusion and enhances the overall emotion recognition performance.

2.2 Audio Feature Encoding via VGGish and TCN

EEG-based emotion recognition models effectively reflect actual human emotional states but are vulnerable to noise and artifacts, limiting model robustness and generalization. Additionally, audio signals containing emotional information such as speech, rhythm, and timbre significantly influence the brain's emotional processing. Therefore, fusing EEG and audio modalities can complement each modality’s weaknesses and integrate complementary information, thereby enhancing accuracy and robustness in emotion recognition.

In this study, to utilize raw audio signals extracted from videos alongside EEG signals for multimodal emotion recognition, an audio encoder composed of preprocessing, VGGish-based spatial feature extraction, and TCN-based temporal feature encoding was designed. First, the audio signals undergo short-time Fourier transform with 50 ms Hamming windows and 10 ms hop length, converting the signal into the time-frequency domain. This process allows detailed analysis of frequency components in each short frame. Subsequently, a 20-band mel filter bank, reflecting human auditory characteristics, is applied to transform the frequency spectrum into the mel scale. The spectral components within each mel filter are summed, and logarithmic transformation is applied to generate the Log-Mel Spectrum (LMS)[13], extracting perceptually meaningful spectral features .

The resulting LMS serves as input to a pre-trained VGGish model, which is a deep convolutional neural network that captures diverse patterns along frequency and time axes within the audio spectrum, producing high-dimensional spatial feature maps. These spatial features effectively represent complex signal structures related to emotions, such as voice timbre, intonation, and rhythm.

Finally, the spatial feature maps extracted by VGGish are fed into a TCN to model temporal dynamics. Leveraging causal convolutions and dilations, TCN learns sequence dependencies over a wide receptive field, effectively capturing the dynamic characteristics of audio signals corresponding to emotional changes. This allows TCN to encode temporal variations of continuous spatial features, reflecting how emotional information evolves over time.

Consequently, the proposed audio encoder systematically extracts spatial and temporal features based on log-mel spectra, generating robust emotional representations suitable for fusion with EEG signals, thereby contributing to improved accuracy and reliability in multimodal emotion recognition models.

2.3 TCN-Based Joint Temporal Encoding of EEG and Audio Features

The TCN is a deep learning architecture designed for processing sequential data. Compared to Recurrent Neural Networks (RNNs), TCN allows for parallel computation and effectively mitigates issues related to vanishing or exploding gradients. Due to these architectural advantages, TCN ensures stable learning over long sequences and enables fast inference, making it particularly well-suited for tasks like emotion recognition, where the analysis of temporal signal patterns is critical.

TCN is composed of three key components. First, causal convolutionensures temporal causality by computing the output at each time step using only past information, without referencing future time steps. Second, dilated convolutionintroduces gaps between kernel elements (controlled by the dilation rate), allowing the network to expand its receptive field and capture long-range temporal dependencies efficiently. Third, residual connectionsfacilitate the flow of information across layers, improving both training stability and overall performance.

In this study, spatial features are first extracted from the raw inputs of two modalities—EEG and audio—using EEGNet and a VGGish-based encoder, respectively. Subsequently, a TCN is applied to both sets of spatial features to learn their temporal characteristics. In particular, we leverage dilated convolutions within the TCN to effectively capture changes in signal patterns over time for each segment. This process yields the following temporal feature vectors for the EEG and audio modalities:

$$\mathrm{F}^{\mathrm{EEG}}=\mathrm{TCN}\left(\mathrm{F}_{\mathrm{EEG}}^{spatial}\right),\ \mathrm{F}^{Audio}=\mathrm{TCN}\left(\mathrm{F}_{audio}^{spatial}\right), \tag{1}$$

where $\mathrm{F}_{\mathrm{EEG}}^{spatial}$ and $\mathrm{F}_{audio}^{spatial}$ refer to the spatial feature maps extracted by EEGNet and VGGish, respectively. The resulting $\mathrm{F}_{\mathrm{EEG}}^{temporal}$ and $\mathrm{F}_{\mathrm{Audio}}^{temporal}$ are the temporal embeddings generated by applying the TCN.

Next, the temporal feature vectors from both modalities are alignemnetd along the channel dimension to form a unified multimodal representation:

$$\mathrm{F}_{joint}=Alignment\left(\mathrm{F}^{\mathrm{EEG}},\mathrm{F}^{Audio}\right) \tag{2}$$

The alignmented feature vector F_{joint} contains comprehensive emotional information derived from both EEG and audio signals for a given video segment. It is subsequently fed into a Bidirectional Cross-Modal Attention (BCMA) module to model inter-modal interactions.

Through this approach, TCN effectively learns the temporal dynamics of affective signals, enhancing the fusion performance of EEG and audio inputs and playing a key role in improving the accuracy and robustness of the overall emotion recognition model.

2.4 Bidirectional Cross-Modal Attention-Based Fusion for EEG and Audio Features in Emotion Recognition

In this study, we introduce a BCMA mechanism to effectively capture the deep interactions between features extracted from two modalities: EEG and audio. The sequential features of each modality are defined as $\mathrm{F}^{\mathrm{EEG}}\in\mathbb{R}^{\mathrm{T}\times d}$ and $\mathrm{F}^{Audio}\in\mathbb{R}^{\mathrm{T}\times d}$, where T denotes the sequence length and d is the feature dimension.

The BCMA is designed such that the query vectors of one modality interact with the key-value pairs of the other modality, enabling each signal to dynamically select meaningful information from the counterpart modality and incorporate it into the fusion process.

First, the cross-modal attention with EEG as the query and audio as the key and value is computed as follows:

$$\mathrm{Q}^{\mathrm{EEG}}=\mathrm{F}^{\mathrm{EEG}}\mathrm{W}^{\mathrm{Q}},\ \mathrm{K}^{Audio}=\mathrm{F}^{Audio}\mathrm{W}^{k},\ \mathrm{V}^{Audio}=\mathrm{F}^{Audio}\mathrm{W}^{\mathrm{V}}$$
$$\mathrm{F}^{\mathrm{EEG}\rightarrow Audio}=softmax\left(\frac{\mathrm{Q}^{\mathrm{EEG}}\left(\mathrm{K}^{Audio}\right)^{\mathrm{T}}}{\sqrt{d_k}}\right)\mathrm{V}^{Audio} \tag{3}$$

where $\mathrm{W}^{\mathrm{Q}},\mathrm{W}^{\mathrm{K}},\mathrm{W}^{\mathrm{V}}\in\mathbb{R}^{d\times d_k}$ are learnable weight matrices, and d_k denotes the dimension of the attention key vectors. The attention scores $\mathrm{S}=\frac{\mathrm{Q}^{\mathrm{EEG}}\left(\mathrm{K}^{Audio}\right)^{\mathrm{T}}}{\sqrt{d_k}}$ represent the weights indicating which parts of the audio sequence each time step in the EEG sequence should attend to.

Conversely, the cross-modal attention with audio as the query and EEG as the key and value is calculated similarly:

$$Q^{Audio}=F^{Audio}W^{Q'},\ K^{EEG}=F^{EEG}W^{K'},\ V^{EEG}=F^{EEG}W^{V'}$$
$$F^{Audio\to EEG}=softmax\left(\frac{Q^{Audio}\left(K^{EEG}\right)^{T}}{\sqrt{d_k}}\right)V^{EEG} \tag{4}$$

Thus, the bidirectional attention results $F^{EEG\to Audio}$ and $F^{Audio\to EEG}$ mutually complement the critical feature information from different modalities. The features extracted from both directions are concatenated to obtain the final fused representation F^{fusion}:

$$F^{fusion}=Concat\left(F^{EEG\to Audio},F^{Audio\to EEG}\right)W^{F}+b^{F} \tag{5}$$

where $W^{F}\in\mathbb{R}^{2d_k\times d_f}$ and $b^{F}\in\mathbb{R}^{d_f}$ are the weight matrix and bias vector for dimensionality reduction and transformation after fusion, where d_f is the final embedding dimension. This linear transformation compresses the integrated information and adjusts the dimension suitable for the next stage of the model.

Finally, the fused features $F^{fusion}\in\mathbb{R}^{T\times d_f}$ are flattened into a one-dimensional vector and fed into a MLP for emotion classification:

$$\hat{y}=MLP\left(Flatten\left(F^{fusion}\right)\right) \tag{6}$$

where $\hat{y}\in\mathbb{R}^{C}$ denotes the predicted probability vector for C emotion classes. The MLP consists of multiple fully connected layers, activation functions (e.g., ReLU), dropout, and normalization layers to enhance model expressiveness and prevent overfitting.

This BCMA-based fusion method dynamically reflects temporal and semantic correlations between the EEG and audio signals, substantially improving emotion recognition accuracy and robustness compared to simple concatenation of independent modality features.

3. EXPERIMENT AND RESULTS

3.1 Dataset and Validation Strategy for Multimodal Emotion Recognition

To evaluate the performance of the proposed multimodal emotion recognition method, we conducted experiments using the publicly available DEAP dataset [14]. DEAP is a benchmark database comprising EEG, audio, and peripheral physiological signals collected from 32 participants (16 males and 16 females, aged between 19 and 37). The dataset adopts a dimensional emotion model based on arousal, valence, dominance, and liking. During the experiment, each participant viewed 40 music video clips while EEG signals were recorded from 32 electrodes at a sampling rate of 512 Hz. Each session lasted approximately 63 seconds, after which participants rated their emotional states using the Self-Assessment Manikin (SAM) method on a continuous scale from 1 to 9 across the four dimensions.

In this study, EEG data from 22 participants (11 males and 11 females) were selected for analysis. The corresponding audio signals were segmented using a 1-second sliding window and temporally synchronized with the EEG sequences to obtain aligned audio samples.

Emotion classification was conducted based on four representative emotional states: happiness, sadness, anger, and relaxation. The proposed model was evaluated using the Leave-One-Subject-Out Cross Validation (LOSO-CV) [15] method, in which the data of one subject is used as the test set while the remaining subjects' data serve as the training set. This approach ensures subject-independent evaluation and is suitable for assessing the generalizability and robustness of the model.

To further assess the effectiveness of the proposed multimodal method, we compared its performance with various baseline and state-of-the-art emotion recognition approaches. These included single-modal deep learning models such as CNN, LSTM, and Convolutional Recurrent Neural Network (CRNN), as well as multimodal fusion techniques based on cross-modal attention and multi-head attention mechanisms [16].

To evaluate the performance of the proposed multimodal emotion recognition method (PM), we implemented and compared it with various baseline and alternative models based on single and multimodal approaches. The details of each comparison method are as follows:

- E-CNN + A-BiLSTM + Softmax: This model extracts spatial features from EEG signals using a CNN, while temporal features are extracted from audio signals using a Bidirectional LSTM (BiLSTM) network. The concatenated multimodal features are then passed through a Softmax-based classifier for emotion prediction.
- E-TCN + A-TCN + CF + MLP: TCNs are independently applied to both EEG and audio signals to extract temporal features without any spatial encoding. The fused features are then fed into a Conformer (CF) [17] module for enhanced temporal modeling, followed by a MLP for final emotion classification.
- E-ViT + A-CRNN + MLP: Visual Transformer (ViT) [18] is applied to the EEG signals to capture global spatial-temporal patterns, while a CRNN, combining CNN and RNN, is used for audio feature extraction. The fused features are input into an MLP classifier to recognize emotions.
- E-EEGNet + A-VGG + CF + MLP: EEG features are extracted using EEGNet, and audio features are obtained using a VGG network. These features are fused and modeled through a Conformer, followed by an MLP for classification. This method does not utilize TCN and focuses on spatial and perceptual feature representations.
- E-EEGNetTCN + MLP: This unimodal method focuses solely on EEG signals. EEGNet is used to extract spatial features, and TCN is subsequently applied to obtain spatiotemporal representations. The resulting features are passed to an MLP for emotion recognition.
- A-VGGTCN + MLP: In this unimodal model, EEG features are excluded. Audio signals are encoded using a combination of VGG and TCN to extract spatiotemporal features, which are then classified using an MLP.
- PM (Proposed Model): This is the model proposed in the present study.

3.2 Experimental Rsults

Table 1presents a comparative analysis of emotion recognition performance across various neural network architectures, including the proposed method. The experimental results clearly demonstrate that the proposed model achieves the highest classification accuracy of 92.3%, outperforming all other tested approaches.

Table 1 – Multimodal emotion recognition results using DEAP dataset (%)

Methods	Recognition Accuracy (%)
E-CNN+A-BiLSTM+Softmax	77.4
E-TCN+A-TCN+CF+MLP	84.5
E-ViT+A-CRNN+MLP	81.7
E-EEGNet+A-VGG+CF+MLP	82.4
E-EEGNetTCN+MLP	71.6
A-VGGTCN+MLP	61.8
PM	92.3

Notably, this performance is significantly superior to the results obtained from unimodal models that utilize only spatiotemporal features extracted from either EEG or audio signals. Models that rely solely on a single modality tend to overlook cross-modal interactions that are critical for accurate emotional state inference. Furthermore, approaches that simply combine spatial features extracted from EEG and audio signals, or model temporal features using only a Conformer module, show lower recognition accuracy compared to the proposed method. However, these methods still outperform traditional baseline models, indicating that partial spatiotemporal modeling and modality fusion contribute to some degree of performance improvement. These findings emphasize the importance of jointly leveraging spatiotemporal features from both EEG and audio modalities, as well as effectively capturing cross-modal interactions, to enhance emotion recognition performance.

Finally, the lowest recognition accuracy was observed in the model that employed only spatiotemporal features derived from audio signals without any EEG input. This result further emphasizes the critical role of EEG data in emotion recognition tasks, particularly in capturing nuanced internal emotional states that may not be fully represented in audio cues alone.

4. CONCLUSIONS AND FUTURE WORK

The proposed method effectively extracts meaningful feature maps from the audio signals embedded within videos and the EEG signals simultaneously recorded while subjects watch these videos. The extracted EEG and audio features are integratively processed through a model combining TCN and a BCMA mechanism, thereby faithfully capturing the complex spatiotemporal and complementary interactions between the two modalities for successful emotion recognition. Experimental results on the DEAP dataset demonstrate that the proposed model achieves superior accuracy compared to various existing emotion recognition approaches, clearly validating the effectiveness of multimodal integration of EEG and audio signals. This finding suggests that multimodal fusion plays a critical role in enhancing emotion recognition performance compared to unimodal approaches. Future work will focus on more sophisticated fusion strategies involving EEG features along with both audio and visual features from videos. In particular, we plan to incorporate contrastive learning—a technique that has recently shown remarkable success in multimodal learning—to further improve accuracy and generalization in emotion recognition. Building upon this, we aim to deeply explore the interrelationships among three key intrinsic factors: external stimuli, mood changes, and physiological arousal, thereby contributing not only to emotion recognition but also to the broader understanding of psychophysiological processes.

ACKNOWLEDGEMENTS

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. NRF-2023R1A2C1006756).

REFERENCES

1. Andalibi N, Buss J. The human in emotion recognition on social media: attitudes, outcomes, risks. Proc 2020 CHI Conference on Human Factors in Computing Systems; April 2020; p. 1–16.
2. Dubey A, Shingala B, Panara JR, Desai K, Sahana M. Digital content recommendation system through facial emotion recognition. Int. J. Res. Appl. Sci. Eng. Technol 2023; 11, p. 1272–1276.
3. Chakravarthi B, Ng SC, Ezilarasan M, Leung MF. EEG-based emotion recognition using hybrid CNN and LSTM classification. Front. Comput. Neurosci. 2022; 16, 1019776
4. Antoniadis P, Pikoulis I, Filntisis PP, Maragos P. An audiovisual and contextual approach for categorical and continuous emotion recognition in-the-wild. Proc IEEE/CVF International Conference on Computer Vision, October 2021; p. 3645–3651.
5. Zhang YH, Huang R, Zeng J, Shan S, Chen X. Multi-modal continuous valence-arousal estimation in the wild. Proc 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition, 2020; p. 632–636.
6. Mocanu B, Tapu R, Zaharia T. Multimodal emotion recognition using cross modal audio-video fusion with attention and deep metric learning. Image Vis. Comput. 2023; 133, 104676.
7. Lin, W.; Li, C. Review of studies on emotion recognition and judgment based on physiological signals. Appl. Sci. 2023, 13, 2573.
8. Karnati M, Seal A, Bhattacharjee D, Yazidi A, Krejcar O. Understanding deep learning techniques for recognition of human emotions using facial expressions: A comprehensive survey. IEEE Trans. Instrum. Meas. 2023; 72, p. 1–31.
9. Lawhern VJ, Solon AJ, Waytowich NR, Gordon SM, Hung CP, Lance BJ. EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces. Journal of Neural Engineering. 2018; 15 (5), 056013.
10. Pandey A, Wang D. TCNN: Temporal convolutional neural network for real-time speech enhancement in the time domain. Proc 2019 IEEE International Conference on Acoustics, Speech and Signal Processing. 2019; p. 6875-6879.
11. Diwakar MP, Gupta B. VGGish Deep learning model: audio feature extraction and analysis. Int. Conference on Data Management, Analytics & Innovation. 2024; p. 59-70.
12. Xu X, Wang T, Yang Y, Zuo L, Shen F, Shen HT. Cross-modal attention with semantic consistence for image–text matching. IEEE Transactions on Neural Networks and Learning Systems, 2020; p. 5412-5425.
13. Seo S, Kim C, Kim JH. Convolutional neural networks using log mel-spectrogram separation for audio event classification with unknown devices. Journal of Web Engineering. 2022; 21 (2), 497-522.
14. Koelstra S, Muhl C, Soleymani M, Lee J-S, Yazdani A, Ebrahimi T, Pun T, Nijholt A, Patras I. DEAP: A database for emotion analysis; using physiological signals. IEEE Trans. on Affective Computing. 2011; 3, p. 18-31.

15. Keelawat P, Thammasan N, Numao M, Kijsirikul B. Spatiotemporal emotion recognition using deep CNN based on EEG during music listening. 2019; arXiv preprint arXiv:1910.09719.
16. Xi C, Lu G, Yan J. Multimodal sentiment analysis based on multi-head attention mechanism. Proc 4th International Conference on Machine Learning and Soft Computing. 2019; p. 34-39.
17. Gulati A, Qin J, Chiu CC, Parmar N, Zhang Y, Yu J, Pang R. Conformer: Convolution-augmented transformer for speech recognition. 2020; arXiv preprint arXiv:2005.08100 .
18. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, Uszkoreit J, Houlsby N. An image is worth 16x16 words: Transformers for image recognition at scale. 2020; arXiv preprint arXiv, 2010.11929.

PROCEEDINGS of the
16th International Conference on Theoretical and Computational Acoustics

August 18 to 22, 2025 BEXCO, Busan Korea

Fast Multi-Frequency DOA Estimation Using Frequency-Difference Sparse Bayesian Learning

Xiaoyan Liu[1], Haiqiang Niu[2], Haibin Wang[3]

[1] Institute of Acoustics Chinese Academy of Sciences, People's Republic of China

[2] Institute of Acoustics Chinese Academy of Sciences, People's Republic of China

[3] Institute of Acoustics Chinese Academy of Sciences, People's Republic of China

ABSTRACT

To address the challenges of high background interference and limited spatial resolution in frequency-difference beamforming (FDB) for wideband signals, this paper proposes a novel Frequency-Difference Sparse Bayesian Learning (FDSBL) algorithm. The proposed method integrates sparse Bayesian learning (SBL) with the frequency-difference technique, effectively overcoming the limitations of conventional FDB approaches while significantly improving direction-of-arrival (DOA) estimation accuracy. To further address the computational complexity inherent in multi-frequency processing scenarios, we develop an advanced Fast Multi-Frequency FDSBL (Fast-MF-FDSBL) method. This approach employs a joint processing strategy that concatenates multi-snapshot data from different high-frequency pairs corresponding to the same difference frequency. By simultaneously exploiting the sparsity patterns across both frequency and snapshot dimensions, the method achieves substantial computational efficiency gains without compromising estimation performance. Comprehensive performance evaluation through both simulation results and Swellex96 experimental data demonstrates the superiority of the proposed methods. The results show that Fast-MF-FDSBL not only preserves the desirable characteristics of conventional SBL, including high resolution and low sidelobe levels, but also realizes an approximately 6.67-fold computational speedup compared to traditional MF-FDSBL implementations, making it particularly suitable for practical applications requiring both accuracy and efficiency.

[1] liuxiaoyan@mail.ioa.ac.cn

[2] nhq@mail.ioa.ac.cn

[3] whb@mail.ioa.ac.cn

An improved particle-based track-before-detect algorithm for underwater weak target tracking

Yifei LI[1,2,3]; Bin QI[1,2,3]; Guolong LIANG[1,2,3]
[1] National Key Laboratory of Underwater Acoustic Technology, Harbin Engineering University, Harbin 150001, China
[2] Key Laboratory of Marine Information Acquisition and Security (Harbin Engineering University), Ministry of Industry and Information Technology; Harbin 150001, China
[3] College of Underwater Acoustic Engineering, Harbin Engineering University, Harbin 150001, China

ABSTRACT

Due to the development of stealth technology and the complexity of the marine environment, the signal-to-noise ratio of the signal of underwater acoustic targets is getting lower and lower, making the detection of such weak targets more difficult. Especially when there are both high SNR targets and low SNR targets in the observation area, and their trajectories intersect, the performance of traditional Detect-before-track (DBT) is reduced due to the limitation of threshold judgment. This paper proposes improvements to particle filter-based track-before-detect algorithm. Based on the independent partition particle filtering (IP-PF), considering the influence between adjacent targets, all targets are divided into several groups according to the predicted angle, and the likelihood of all targets within the group is introduced when calculating the first-order weight. When dividing groups, modify the clustering threshold based on the estimated value of the target at the previous moment and the predicted value at this moment, making it change from a fixed value to an adaptive parameter that can adapt to different situations. By taking these measures, improve the problems of high false alarms, high missed detections, and substantial bearing estimation errors, enhance tracking performance during target proximity, and finally compare the performance with particle filter-based track-before-detect (PF-TBD).

Keywords: Weak Target, Track Before Detect, Particle Filter

1. INTRODUCTION

In the field of underwater target detection and tracking, quickly and accurately identifying targets and achieving continuous and stable tracking have always been the focus of research. Due to the continuous advancement of stealth technologies and the inherent complexity of marine environments, the signal-to-noise ratio (SNR) of underwater acoustic target signals continues to deteriorate, making effective detection increasingly challenging (1).

Traditional detect-before-track algorithm relies on applying a fixed detection threshold to single-frame observation data to obtain measurement points before performing data association and track formation (2). The setting of this detection threshold significantly impacts subsequent data processing, when detecting weak targets, an excessively high threshold leads to missed detections while an overly low threshold generates false alarms. This fundamental limitation makes conventional detect-before-track approaches particularly unsuitable for weak target detection scenarios, as the critical target information may be lost during the initial threshold stage before any temporal integration can occur (3, 4). Compared to DBT, TBD avoids threshold detection, preserving the original measurements of the target and improving tracking accuracy. Therefore, in recent years, TBD has gradually become a hotspot in the field of low detectable target detection and tracking (5). The basic idea of TBD is to comprehensively process multi-frame data, achieving positional correlation of the target between consecutive frames to obtain tracking results, while noise exhibits positional

[1] qibin@hrbeu.edu.cn

randomness across consecutive frames, which can be filtered out through multi-frame processing (6). Currently, TBD algorithms can be roughly divided into two methods: iterative processing and batch processing (7). Compared to batch processing algorithms, the PF-TBD algorithm in iterative processing has lower requirements for storage and computation (8). Therefore, this paper conducts research within the framework of the PF-TBD algorithm.

When dealing with more complex tracking scenarios, the PF-TBD algorithm encounters certain issues. For example, during multi-target tracking, problems such as target proximity, trajectory crossings, and model mismatch in maneuvering target tracking may arise. To mitigate the impact of these challenges, this paper proposes improvements to the particle filter algorithm in the context of passive sonar bearing-only multi-target tracking. The enhanced approach accounts for interactions between neighboring targets by adopting a parallel partitioning strategy. Specifically, all targets are divided into multiple groups based on predicted bearing angles, and targets within the same group are treated as a collective when estimating posterior probability distributions. This modification improves tracking performance, as demonstrated through comparative analysis with the conventional PF-TBD method.

2. STATE-SPACE MODEL

Track-before-detect (TBD) methods achieve weak target trajectory tracking through multi-frame accumulation (9). Compared with conventional detect-before-track algorithms, they yield superior tracking performance under low signal-to-noise ratio conditions.

2.1 Target Motion Model

The state equation describes the evolution of the target state over time, i.e., the target's motion model. Currently, commonly used target motion models include the Constant Velocity (CV) model, Constant Acceleration (CA) model, Singer model, among others. This paper selects the Constant Velocity (CV) model as the target's motion model (10). The state equation describes the process of how the target state changes over time, which is the motion model of the target. Where the variables $\boldsymbol{x}$ and $\dot{\boldsymbol{x}} = \mathrm{d}\boldsymbol{x}/\mathrm{d}t$ represent the target's angle information and angular velocity information. The target motion equation can be expressed as

$$\begin{bmatrix} \boldsymbol{x}_k \\ \dot{\boldsymbol{x}}_k \end{bmatrix} = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_{k-1} \\ \dot{\boldsymbol{x}}_{k-1} \end{bmatrix} + \omega_{k-1} \tag{1}$$

Where $\boldsymbol{x}_k$ denotes the state vector at frame k, T represents the sampling time interval, and ω_{k-1} is the process noise. The process noise covariance of the target motion satisfies

$$Q = E\left[\omega_{k-1}\omega_{k-1}^T\right] = q \times \begin{bmatrix} T^3/3 & T^2/2 \\ T^2/2 & T \end{bmatrix} \tag{2}$$

Where q is the intensity parameter of the process noise.

2.2 Measurement Model

The observation equation involves modeling the sensor observations, which is used to describe how observable data is obtained from the motion state of the target (11, 12). In this paper, the measurements are discrete resolution cells divided by azimuth angle, and the spatial spectral energy amplitude within each resolution cell is the measurement information being tracked.

As shown in Figure 1, when the bearing angle is divided into $[\theta_1, \theta_2, ..., \theta_B]$, the measurement at frame k can be expressed as $[z_1^k, z_2^k, ..., z_B^k]$, and the measurement information of the b-th resolution cell at frame k can be represented as z_b^k.

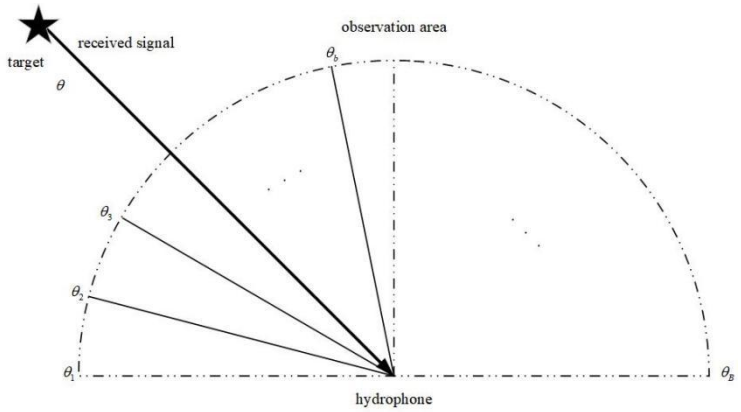


Figure 1 – Schematic diagram of passive sonar resolution cell division

2.3 Likelihood Ratio Function Construction

A Gaussian function is selected to construct the point spread function for modeling the target response, defined as

$$h^i = a_{k,m} \exp(-(\theta_i - \theta_{k,m})^2 \Phi(\theta_{k,m})) + w \tag{3}$$

Where h^i represents the target response in resolution cell i, $a_{k,m}$ denotes the signal intensity received through beamforming, $\theta_{k,m}$ indicates the predicted bearing of the m-th target at frame k, Φ characterizes the attenuation degree, which depends on both the beamforming algorithm and sonar system characteristics.

According to the point spread function model, the likelihood function of the i-th target in the k-th frame is

$$p(z_k \mid x_{k,m}) = \frac{1}{u_i} \exp(-z_i / u_i) \tag{4}$$

Where i represents the corresponding resolution cell for θ_i.

The likelihood ratio for the existence of the target and the non-existence of the target is

$$p(\boldsymbol{z}_k \mid \boldsymbol{x}_{k,m}) = \frac{p(\boldsymbol{z}_k^i \mid \boldsymbol{x}_{k,m}, E_k = 1)}{p(\boldsymbol{z}_k^i, E_k = 0)} \tag{5}$$

Where $E_k = 1$ indicates the target exists, and $E_k = 0$ indicates the target does not exist.

3. Research Based on Particle Filter Algorithm

3.1 Particle Filter-Based Track-Before-Detect Algorithm

The particle filter-based track-before-detect algorithm is a sequential processing method under the Bayesian filtering framework. It samples a set of random samples (particles) from a known distribution and assigns corresponding weights to the particles based on the likelihood function. These particles are then used to approximate the posterior probability density of the target, thereby enabling the estimation of the target state.

In Bayesian filtering, the calculation formula for the posterior probability density often cannot yield an analytical solution in practical applications, so importance sampling is introduced to approximate the solution. Its specific implementation uses a set of weighted particles to represent the posterior probability density distribution of the target for estimating the target's posterior probability density.

Considering the sequential case, introducing an important density function and decomposing it into

$$q(\boldsymbol{X}_k \mid \boldsymbol{Z}_k) \propto q(\boldsymbol{x}_k \mid \boldsymbol{X}_{k-1}, \boldsymbol{Z}_k) q(\boldsymbol{X}_{k-1} \mid \boldsymbol{Z}_{k-1}) \tag{6}$$

Using the Chapman-Kolmogorov equation, measurement model, and Bayesian formula, the posterior probability density $p(\boldsymbol{X}_k \mid \boldsymbol{Z}_k)$ is expanded as

$$p(\boldsymbol{X}_k \mid \boldsymbol{Z}_k) \propto p(z_k \mid \boldsymbol{x}_k) p(\boldsymbol{x}_k \mid \boldsymbol{x}_{k-1}) p(\boldsymbol{X}_{k-1} \mid \boldsymbol{Z}_{k-1}) \tag{7}$$

In importance sampling, the weight can be represented as

$$w_k^i \propto \frac{p(\boldsymbol{X}_k^i \mid \boldsymbol{Z}_k)}{q(\boldsymbol{X}_k^i \mid \boldsymbol{Z}_k)} \tag{8}$$

By substituting Eq.(8) and Eq.(9) into Eq.(10), the importance weight update formula can be obtained as

$$w_k^i \propto w_{k-1}^i \frac{p(z_k \mid x_k^i) p(x_k^i \mid x_{k-1}^i)}{q(x_k^i \mid X_{k-1}^i, Z_k)} \tag{9}$$

Since the evolution of the target state follows a first-order Markov process, it is further simplified to $w(x_k \mid X_{k-1}, Z_k) = w(x_k \mid X_{k-1}, z_k)$, Eq.(11) is transformed into

$$w_k^i \propto w_{k-1}^i \frac{p(z_k \mid x_k^i) p(x_k^i \mid x_{k-1}^i)}{q(x_k^i \mid x_{k-1}^i, z_k)} \tag{10}$$

The importance function is selected as the state transition probability distribution, i.e. $p(x_k^i \mid x_{k-1}^i)$, the weight calculation formula can be updated as

$$w_k^i \propto w_{k-1}^i \times p(z_k \mid x_k^i) \tag{11}$$

The posterior probability density can be approximated as

$$p(x_k \mid Z_k) \approx \sum_{i=1}^{N} w_k^i \delta(x_k - x_k^i) \tag{12}$$

In multi-target tracking, each particle samples the state of every target and calculates corresponding weights. Subsequently, the overall posterior probability distribution of the tracking area at the current time step is characterized based on these states and weights. Therefore, a single particle in multi-target particle filtering algorithm contains state samples and associated weights for all targets, referred to as a multi-target particle, as illustrated in Figure 2.

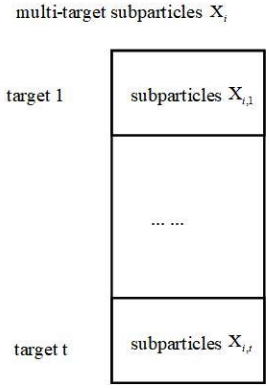


Figure 2 – Schematic diagram of multi-target particles

In this context, the weight calculated for a specific target state within a multi-target particle is termed the first-order weight. The combination of a first-order weight and its corresponding target constitutes a partition (also referred to as a sub-particle) within the multi-target particle. The weight computed in Eq.(13) corresponds precisely to this first-order weight. At the holistic multi-target particle level, each multi-target particle also possesses a distinct weight, known as the second-order weight, which is calculated as follows

$$W_k^i \propto W_k^i \times p(z_k \mid X_k^i) \tag{13}$$

The posterior probability density can be approximated as

$$p(X_k \mid Z_k) \approx \sum_{i=1}^{N} W_k^i \delta(X_k - X_k^i) \tag{14}$$

The importance function selected by the algorithm is $p(X_k \mid X_{k-1})$.

3.2 The Improved Particle-Based Track-Before-Detect Algorithm

In the context of multi-target tracking, the algorithm is enhanced by adopting a parallel partitioning approach to address the interference issues caused by proximate targets.

At frame $k-1$, assuming the number of targets is n_{k-1}, the angle information of all targets is first predicted. The predicted state of the i-th target at frame k is denoted as

$$\hat{x}_{k,m} = \sum_{i=1}^{N} w_{k-1}^i \cdot x_{k|k-1,m}^i \tag{15}$$

Where $x_{k|k-1,m}^i$ represents the predicted state of the m-th target in the i-th particle, w_{k-1}^i denotes the weight of the particle, and N indicates the total number of particles.

The predicted angles of all targets are used for grouping, where each target is assigned to exactly

one cluster.

$$\left|\hat{\theta}_{k|k-1,m_1} - \hat{\theta}_{k|k-1,m_2}\right| < \Lambda_\theta \tag{16}$$

Where $\hat{\theta}_{k|k-1,m_1}$ and $\hat{\theta}_{k|k-1,m_2}$ represent the predicted bearing angles of the m_1-th and m_2-th targets at frame k, Λ_θ denotes the clustering threshold.

To enhance tracking performance for proximate targets, post-clustering processing is performed within each group by incorporating the likelihoods of all group members during first-order weight computation. At frame k, the first-order weight of the i-th partitioned sub-particle for m-th target can be expressed as

$$w_k^{i,m} \propto p(\boldsymbol{z}_k \mid \boldsymbol{x}_{k,m}, \hat{C}_{k|k-1}^{g-\{m\}}) \tag{17}$$

Where g denotes the maximum number of targets within one group, $\hat{C}_{k|k-1}^{g-\{m\}}$ represents the set of predicted bearing angles for all targets in the cluster excluding target m.

As derived from Eq.(12), the importance density function is now specified as

$$q_m(\boldsymbol{x}_{k,m} \mid \boldsymbol{x}_{k-1,m}, \boldsymbol{z}_k) \propto p(\boldsymbol{z}_k \mid \boldsymbol{x}_{k,m}, \hat{C}_{k|k-1}^{g-\{m\}}) \times p(\boldsymbol{x}_{k,m} \mid \boldsymbol{x}_{k-1,m}) \tag{18}$$

At frame k, the second-order weight of the i-th multi-target particle is correspondingly updated as

$$W_k^i = W_k^i \times \frac{p(\boldsymbol{z}_k \mid \boldsymbol{X}_k^i)}{\prod_{m=1}^{t_{k-1}} p(\boldsymbol{z}_k \mid \boldsymbol{x}_{k,m}^i, \hat{C}_{k|k-1}^{g-\{m\}})} \tag{19}$$

Where m denotes the true target identifier and t_{k-1} represents the number of targets at frame $k-1$. After obtaining the second-order weights for all particles, normalization is performed followed by resampling to generate a new particle set. The state estimate for the target is ultimately computed as

$$\hat{\boldsymbol{x}}_{k,m} = \sum_{i=1}^{N} \boldsymbol{x}_k^i \tilde{W}_k^i \tag{20}$$

3.3 Main Steps of The Algorithm

The workflow of the improved algorithm is illustrated in the table 1 below.

Table 1 – IPF-TBD algorithm flow

Input: The spatial spectral energy amplitude $\boldsymbol{Z}_k$ within each resolution cell.

Step 1: Parameter initialization. Set initial parameters such as the number of particles, the values of various noises, the clustering threshold, etc.

Step 2: Particle state prediction. Let the particle set at frame $k-1$ be $[\boldsymbol{x}_{k-1}^i, E_{k-1}^i, W_{k-1}^i]$, $i = 1,2,...,N$. x_{k-1}^i represents the state vector of i-th particle. E_{k-1}^i represents the existence variable of i-th particle. W_{k-1}^i represents the weight of i-th particle. For particles with existence variable $E_k^i = 1$ at frame k, perform prediction and cluster targets based on the predicted state information.

Step 3: Sampling. Sample m partitioned particle substates and calculate first-order weights.

a. Sample the state of the i-th particle in the m-th partition at frame k through probability transition density function $p(\boldsymbol{x}_k^{i,m} \mid \boldsymbol{x}_{k-1}^{i,m})$.
b. Calculate the first-order weight $w_k^{i,m}$ of the i-th particle in the m-th partition using the likelihood function $p(\boldsymbol{z}_k \mid \boldsymbol{x}_{k,m}, \hat{C}_{k|k-1}^{g-\{m\}})$ of m-th partition.
c. Repeat the previous two steps to complete the sampling of all particles within the m-th partition and calculate their first-order weights.
d. Normalize the first-order weights of the m-th partition.
e. Perform resampling on the m-th partition.

Step 4: Repeat Step 2 to complete resampling for all partitions.

Step 5: Calculate the second-order weights.

Step 6: Perform resampling on multi-target particles.

Step 7: Estimate the states of all targets.

Output: The estimated states of all targets relative to the observation point.

4. Simulation and Analysis

4.1 Simulation Parameters

Simulate real passive sonar receiving signals based on target-oriented direction. The parameters of the passive sonar system are as follows: the operating frequency band is $f_0 = 500\text{Hz} \sim 2500\text{Hz}$, the signal sampling frequency is 20000Hz, and the number of uniform linear array hydrophones is $M = 30$. The distance between adjacent hydrophones is 0.5 times the wavelength. The spatial spectrum azimuth resolution is 0.1 degrees, with $B = 1800$ resolution units. The target motion time in the simulation scenario is 50 frames.

4.2 Simulation Performance Metrics: RMSE and OSPA

The evaluation criteria for multi-target tracking are key metrics for measuring the performance of tracking algorithms, used to assess their capability in tracking multiple targets. Given the inherent complexity of multi-target tracking algorithms and the diversity of evaluation criteria, this study employs two comprehensive metrics Root Mean Square Error (RMSE) and Optimal Sub-pattern Assignment (OSPA) to evaluate tracking performance.

The Root Mean Square Error is a widely used metric for quantifying the discrepancy between tracked trajectories and ground truth trajectories. RMSE is defined as the square root of the average squared difference between estimated values and true values. Its mathematical formulation is

$$RMSE=\sqrt{\frac{1}{K}\sum_{k=1}^{K}(\boldsymbol{x}_k-\hat{\boldsymbol{x}}_k)^2} \tag{21}$$

Where K is the total number of frames, representing the length of the processed data. $\boldsymbol{x}_k$ denotes the true bearing of the target, while $\hat{\boldsymbol{x}}_k$ represents the estimated bearing.

The Optimal Sub-Pattern Assignment (OSPA) metric is an evaluation criterion specifically designed for assessing multi-target tracking and estimation performance, particularly suited for scenarios with unknown and potentially varying numbers of targets. The OSPA measures the distance between two sets (such as predicted target sets and true target sets in multi-target tracking), simultaneously considering both the localization error of elements within the sets and the cardinality error (i.e., the difference in target counts) (13). The OSPA distance is designed to satisfy three fundamental properties: metric properties, symmetry, and the triangle inequality.

The OSPA distance consists of two components: (1) Localization error: This component reflects the average error between corresponding elements in the sets under optimal assignment. (2) Cardinality error: This component quantifies the impact of differences in set sizes, representing the discrepancy between the predicted number of targets and the actual number of targets.

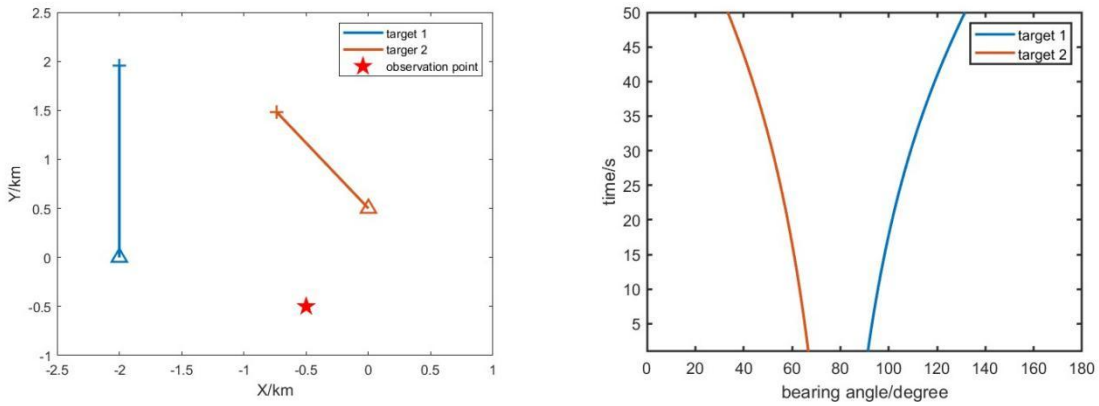
Let X and Y be two sets, where the elements represent the true target and the estimated target state, respectively. $d(x,y)$ is the distance between element x in X and element y in Y. Given a truncation parameter $c>0$ and an exponent $p\geq 1$, the OSPA distance is defined as

$$\text{OSPA}(X,Y)=\left(\frac{1}{n}\left(\sum_{i=1}^{n}\min(\min_{j}d(x_i,y_i),c)^p+\max(0,|X|\text{-}|Y|\cdot c^p\right)\right)^{\frac{1}{p}} \tag{22}$$

Where $n=\max(|X|,|Y|)$ ensures that the base error is reasonably considered. The truncation parameter c limits the maximum error between individual elements, preventing extreme values from having an excessive impact on the overall error; the exponent p controls the weight between different errors.

4.3 Simulation Experiment 1: Validating the Performance of IPF-TBD in Multi-Weak-Target Detection

The simulated scenario contains two targets, both moving at a constant linear velocity. Target 1 and Target 2 do not affect each other, and the simulation time is 50 seconds. As shown in Figure 3, it is the situational map of the two targets. In Figure (a), the observer's position is marked with a red "☆", and the starting positions of the three targets are marked with "+", while the ending positions are marked with "△". Figure (b) shows the azimuth angles of the three targets relative to the observer, covering a total of 50 seconds, with 1 second corresponding to 1 frame of data.



(a) Relative position between observation point and target

(b) Target bearing angle relative to the observation point

Figure 3 – Simulation scenario 1: target situation diagram

The signal-to-noise ratio of Target 1 is -27dB, and that of Target 2 is -26dB. The conventional beamforming process was applied to the simulated signal, and the resulting time-bearing history plot is shown in Figure 4.

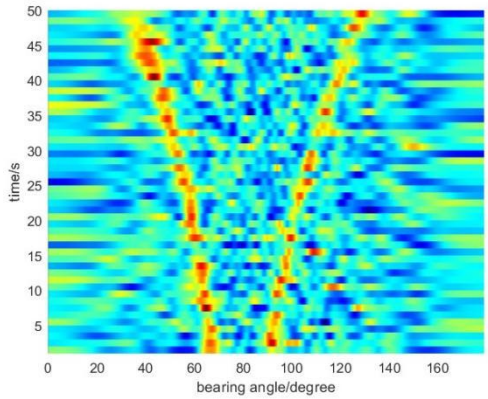


Figure 4 – Time-bearing history plot for dual targets

The traditional DBT approach first applies threshold selection and picking to the time-bearing history plot obtained through beamforming. The final threshold selection result is shown in Figure 5.

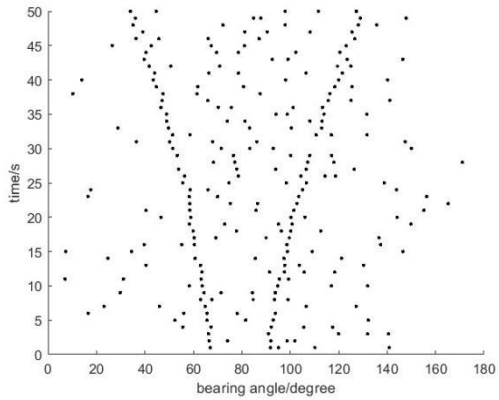


Figure 5 – The result graph after threshold selection on the time-bearing history plot

After processing the threshold-selected results with the Rao-Blackwellized track-after-detect algorithm, the final tracking results are shown in Figure 6. Due to the low signal-to-noise ratio, the targets are occasionally obscured by environmental noise. During threshold selection, it is impossible to completely filter out outlier points with energy levels too close to the targets. As a result, intermittent target trajectories, clutter-formed tracks, and overlapping trajectories occur, leading to poor tracking performance.

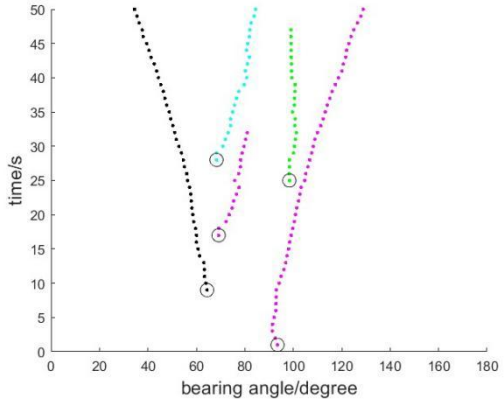


Figure 6 – Simulation scenario 1: detect-before-track algorithm tracking result

The IPF-TBD algorithm proposed in this study was applied directly to the raw data, i.e., tracking was performed on the time-bearing history plot shown in Figure 4. The resulting tracking output is presented in Figure 7.

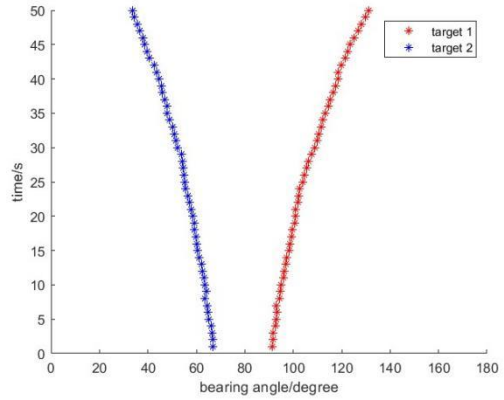
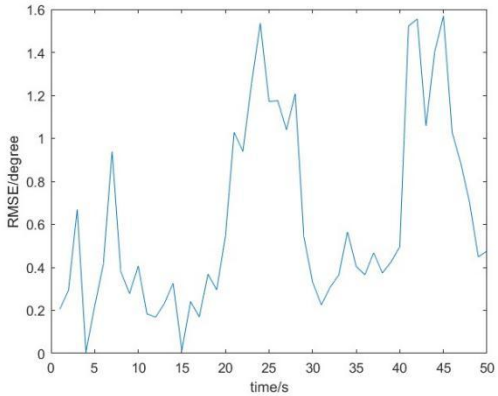
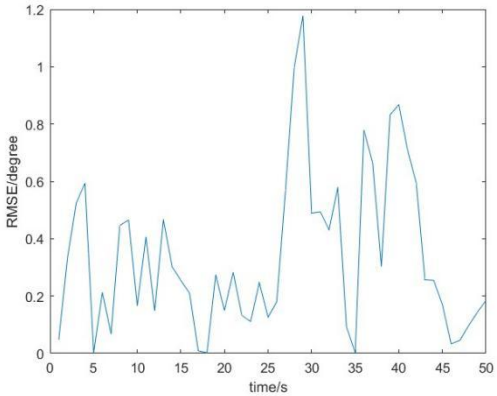


Figure 7 – Simulation scenario 1: IPF-TBD tracking result

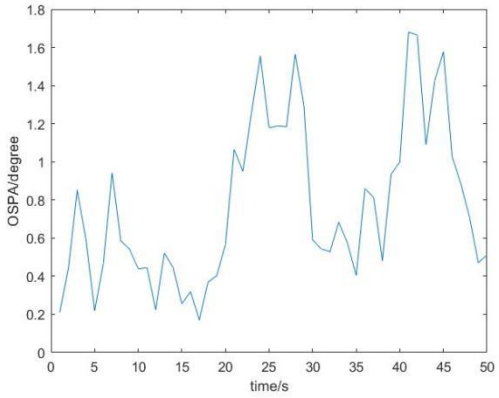
Figure 8 shows the tracking errors between the true trajectories and estimated trajectories of the two targets, along with the OSPA (Optimal Sub-pattern Assignment) error for multiple targets. The final root mean square (RMSE) errors for the two trajectories are 0.6 degrees and 0.5 degrees.



(a) Tracking Error for Target 1



(b) Tracking Error for Target 2



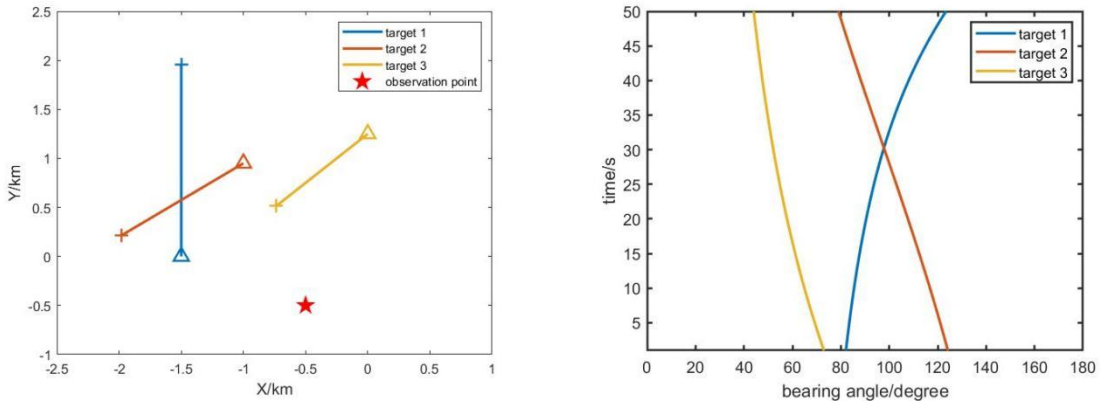
(c) OSPA error

Figure 8 – Tracking errors and OSPA errors obtained using IPF-TBD

4.4 Simulation Experiment 2: Performance Evaluation of IPF-TBD in Mixed Strong-Weak Target Crossing Scenarios

Test the performance of the improved algorithm and the traditional algorithm separately, forming RMSE curves and OSPA curves for comparing the performance of the two algorithms under different simulation scenarios.

The simulation scenario includes three targets, all of which move in uniform linear motion. Targets 1 and 2 have a neighboring cross situation, while Target 3 is unaffected by the other targets. The simulation time is 50 seconds. Figure 9 shows the situation map of the three targets. In Figure (a), the observer's position is marked with a red "☆", and the starting positions of the three targets are marked with "+", while the ending positions are marked with "△". Figure (b) shows the azimuth angles of the three targets relative to the observer, totaling 50 seconds, with 1 second corresponding to 1 frame of data.



(a) Relative position between observation point and target

(b) Target bearing angle relative to the observation point

Figure 9 – Simulation scenario 2: target situation diagram

The signal-to-noise ratio (SNR) of Target 1 is -22.5dB, Target 2 is -25dB, and Target 3 is -25dB. Conventional beamforming was applied to the simulated signal, and the resulting time-bearing history plot is shown in Figure 10.

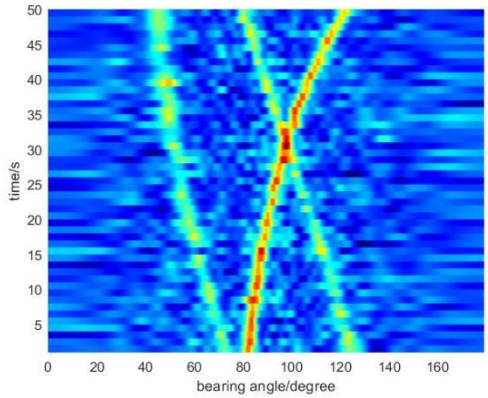
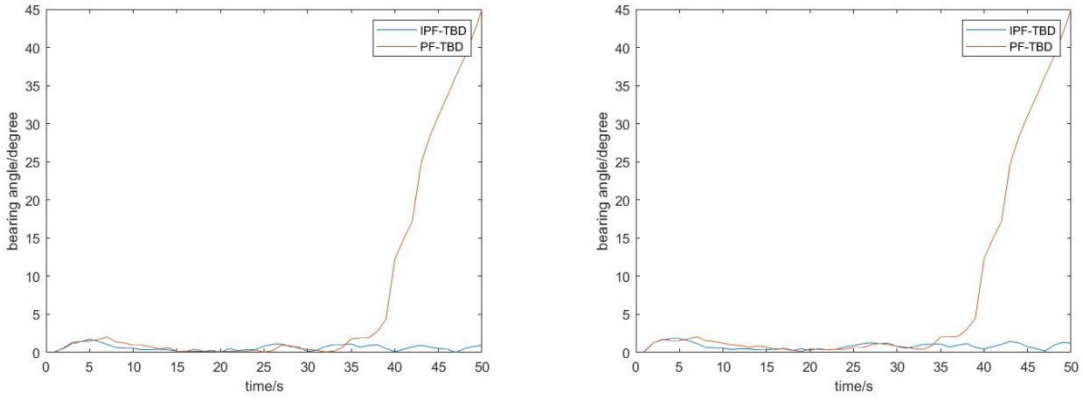


Figure 10 – Time-bearing history plot for dual targets

The simulation results of PF-TBD and IPF-TBD are shown in Figure 11.



(a) Target 2 Tracking error comparison plot (b) OSPA error comparison plot

Figure 11 – Tracking errors and OSPA metrics: comparative analysis of two algorithms

The final root mean square (RMSE) errors for the three trajectories are 0.2 degrees, 0.3 degrees and 0.5 degrees. For Target 1 (with high SNR) and Target 3 (whose trajectory does not intersect with the others), the tracking performance of both algorithms shows little difference. However, for Target 2 (with low-SNR and trajectory crossing a high-SNR target), the simulation results clearly indicate that IPF-TBD significantly outperforms PF-TBD in tracking performance.

5. CONCLUSIONS

This paper addresses the challenges of low signal-to-noise ratio underwater target detection and tracking in complex marine environments, particularly the performance degradation in scenarios involving nearby targets and trajectory crossings, by proposing an improved particle-based track-before-detect algorithm. By introducing an adaptive clustering mechanism and a group likelihood weighting strategy, the algorithm enhances performance in three aspects: (1) employing a dynamic target clustering method based on predictive angles to effectively capture spatial correlations among multiple targets; (2) replacing fixed clustering thresholds with adaptive parameters based on historical estimation and current prediction; (3) improving tracking performance for weak targets when strong and weak targets appear in proximity and are assigned to the same cluster, through intra-cluster joint likelihood computation. Simulation experiments demonstrate that, compared to traditional DBT and PF-TBD, the proposed IPF-TBD exhibits superior performance in detecting and tracking weak targets when strong and weak targets intersect trajectories.

ACKNOWLEDGEMENTS

This work is supported by the National Natural Science Foundation of China under Grant Nos.62271162.

REFERENCES

1. Jiaming Yang, Yue Pan, Qiang Wang, Huaigang Cao, Sunpei Gao. Research on deep learning method of underwater weak target tracking [J]. Acta Armamentarii, Volume 45, Issue 2, Pages 385-394, 2024.
2. Stone, Lawrence D. Conditions for MHT to be an exact bayesian solution to the multiple target tracking

problem [C]. 2018 21st International Conference on Information Fusion, FUSION 2018, Pages 982-988, 2018.
3. Leiru Bu, Bin Rao, Dan Song. A group target track-before-detect approach using two-stage strategy with maximum-likelihood probabilistic data association [J]. IET Radar Sonar and Navigation, Volume 18, Issue 8, Pages 1351-1363, 2024.
4. Yang Yue, Cui-E Zheng, DaJun Sun, Jian Xu. A track before detect algorithm for underwater acoustic continuous-wave signal based on particle filter [J]. Control and Decision, Volume 40, Issue 1, Pages 111-118, 2025.
5. Grossi Emanuele, Lops Marco, Venturino Luca. A novel dynamic programming algorithm for track-before-detect in radar systems [J]. IEEE Transactions on Signal Processing, Volume 61, Issue 10, Pages 2608-2619, 2013.
6. Yunfeng Zhang. Research on track before detect of underwater weak targets [D]. Harbin Engineering University, 2024.
7. Xuemin Wang, Mingming Sun, Xianzhong Liu. Track-before-detect algorithm for underwater targets: an overview [J]. Journal of Ordnance Equipment Engineerinng, Volume 45, Issue 12, Pages 99-108, 2024.
8. Northardt Thomas, Nardone Steven C. Track-before-detect bearings-only localization performance in complex passive sonar scenarios: a case study [J]. IEEE Journal of Oceanic Engineering, Volume 44, Issue 2, Pages 482-491, 2019.
9. Jun Wang, Junsheng Jiao. Track before detect for low frequency active towed array sonar [C]. ICSIDP 2019 - IEEE International Conference on Signal, Information and Data Processing, 2019.
10. Jianghua Fan. Passive multi-object tracking and location technology [D]. Harbin Engineering University, 2019.
11. Sarkka Simo. On unscented kalman filtering for state estimation of continuous-time nonlinear systems [J]. IEEE Transactions on Automatic Control, Volume 52, Issue 9, Pages 1631-1641, 2007.
12. Luxiao Xu. Research of underwater targets track-before-detect algorithm based on passive array [D]. University of Electronic Science and Technology, 2018.
13. Schuhmacher Dominic, Vo Ba-Tuong, Vo Ba-Ngu. A consistent metric for performance evaluation of multi-object filters [J]. IEEE Transactions on Signal Processing, Volume 56, Issue 8, Pages 3447-3457, 2008.
14. Zhiyao Du, Yu Hao, Longhao Qiu, Chenmu Li, Guolong Liang. Sparsity-based direction-of-arrival estimation in the presence of near-field and far-field interferences for small-scale platform sonar arrays [J]. The Journal of the Acoustical Society of America, Volume 156, Issue 5, Pages 2989-3005, 2024.
15. Guolong Liang, Boyu Zhang, Bin Qi, Yu Hao, Zhiyao Du, Xiang Li. Underwater multitarget fusion tracking method for passive sonar [J]. Acta Acustica, Volume 49, Issue 3, Pages 501-512, 2024.

PROCEEDINGS of the
16th International Conference on Theoretical and Computational Acoustics

August 18 to 22, 2025 BEXCO, Busan Korea

Virtual Sound Source Construction Based on Direct-to-Reverberant Ratio Control Using a Line Array Loudspeaker Consisted of Parametric and Electrodynamic Loudspeakers

Takumi IRIGUCHI [1]; Yuting GENG [1]; Masato NAKAYAMA [1]; Takanobu NISHIURA [1]
[1] Ritsumeikan University, Japan

ABSTRACT

Virtual sound source (VSS) construction method has been proposed based on direct-to-reverberant ratio (DRR) of room impulse responses (RIRs) using parametric array loudspeakers (PAL) and electrodynamic loudspeakers (EDL). This method focuses on the differences in DRRs of RIRs between PALs and EDLs, aiming to control the perception of VSS distance. The sharp directivity of PAL, which combines a large-amplitude direct sound and low-amplitude reverberant sound, results in a higher DRR than EDL, enabling VSSs construction at closer positions. However, this method requires a large space for placing multiple loudspeakers around the listener, making it inconvenient in practical usage. To address this, we propose a VSS construction method using a line array loudspeaker unit with PALs and EDLs, which realizes VSS distance presentation within a smaller space. The EDLs are configured to keep the DRR low, so that the VSS is perceived as farther away. Moreover, the sounds are weighed based on the VSS's construction position before emitting from the PALs and EDLs. This allows for the reproduction of the desired DRR at the listener's position and controlling the VSS's perceived distance. The effectiveness of the proposed method was confirmed through evaluation experiments.

Keywords: virtual sound source construction, line array loudspeaker, direct-to-reverberant ratio, parametric-array loudspeaker

1. INTRODUCTION

In recent years, communication using virtual reality devices and social networking services has become popular, making it easier to communicate with people in remote locations. However, amplitude panning, which is used for stereo sound reproduction in acoustic communication systems, has been insufficient in terms of VSS distance perception.

To understand the principles of spatial audio reproduction, it is necessary to first clarify the two key concepts of the sound image and the VSS. A sound image is the perception of a sound's location. A VSS reproduces a part of the cues of a real sound source to represent a specific sound image.

To achieve remote presence through sound spatiality, binaural systems using headphones have been proposed [1]. While binaural systems that consider the head-related transfer function (HRTF) [2] can represent sound images around the listener's head, it is difficult to represent sound images at a distance. As a method for representing distant sound images, sound image holography [3] using a loudspeaker pair, consisting of a parametric array loudspeaker (PAL) [4, 5] and an electrodynamic loudspeaker (EDL) [6] placed adjacent to each other, has been proposed. When a PAL emits amplitude-modulated signals at high pressure, it is possible to reproduce audible sound with high directivity because of self-demodulation of ultrasound in the air. As a result, compared to EDLs, the emitted sound has a higher direct-to-reverberant ratio (DRR) [7]. This system can construct a VSS at a position between the listener and the loudspeakers. This is achieved by arranging loudspeaker pairs to surround the listener and controlling the output levels of PAL and EDL. However, this method requires a large space in front, behind, and to the sides of the listener to surround the listener and control the output levels of the PAL and EDL.

We propose a method to construct a VSS based on the control of the DRR, and we devise a specific configuration for the integrated line array using an integrated line array consisting of two PAL-EDL pairs.

[1] {gr0758rr@ed, geng@fc, mnaka@fc, nishiura@is}.ritsumei.ac.jp

Therefore, the proposed configuration is designed to enable sound image perception even in a small space with compact equipment. To achieve this, the two PALs are arranged in front with their emission axes parallel and directed toward the listener.

The proposed method controls the DRR. The DRR is one of the perceptual cues of distance [8] and the interaural level difference (ILD) [9] is one of the perceptual cues of direction. We utilize HRTFs to control the ILD. The signals output from the PAL and EDL are subjected to distance weighting and convolution with HRTFs. In this method, for both PAL and EDL, the signals are weighted by coefficients according to the target distance of the VSS. Then, these signals are convolved with HRTFs to control the perceived direction. This process represents sound images at positions between the line array loudspeaker and the listener. We propose a method to construct a VSS that reproduces the perceptual cues of a real sound source, allowing a listener to perceive a sound image. The evaluation experiment results confirm the effectiveness of the proposed method.

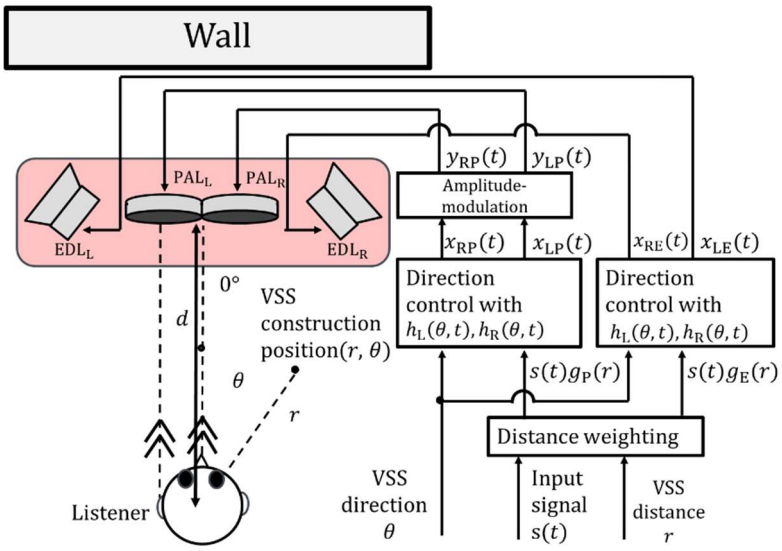


Figure 1 – Overview of the proposed method

2. PROPOSED METHOD

2.1 Overview of the Proposed Method

This research proposes a method for representing VSS through DRR control, using an array that includes two pairs of PALs and EDLs. The proposed method constructs the array by arranging two PALs in parallel, one for each ear, and two EDLs arranged in the left and right backward directions, all positioned in a line. This integrated arrangement allows for a smaller space required.

Figure 1 shows the overview of the proposed method. In this diagram, d is the distance from the listener to the loudspeaker, r is the distance from the center of the listener's head to the VSS, and θ is the direction of the VSS with the front direction as 0°. t is the time index. The input signal is denoted by $s(t)$. $g_E(r)$ and $g_P(r)$ are weighting coefficients in the distance control for the EDL and PAL which correspond to the VSS distance r. Similarly, $h_L(\theta,t)$ and $h_R(\theta,t)$ are the HRTFs for the left and right ears, when the VSS is constructed. $x_{LE}(t)$ is output signal from the left channel of EDL, to construct the VSS to the (r,θ) position. $x_{RE}(t)$ is output signal from the right channel of EDL, to construct the VSS to the (r,θ) position. The target signals for the PALs are $x_{LP}(t)$ and $x_{RP}(t)$, and the amplitude-modulated [10] outputs are $y_{LP}(t)$ and $y_{RP}(t)$. The formulation of the output signals from the two PALs and two EDLs is described in Section 2.2.

When an amplitude modulated signal from a PAL is emitted at high sound pressure, it self-demodulates in the air and becomes audible. The overall process involves controlling the signal so that both distance and direction of the VSS match those of the real sound source. When the VSS is closer than the loudspeaker array, the signals output to the PAL are subjected to distance weighting coefficients, used with HRTFs, and undergo amplitude-modulation. The signals output to the EDL are subjected to distance weighting coefficients and convolved with HRTFs. In this way, VSS is constructed by outputting signals with controlled distance and direction from the PAL and EDL.

2.2 Output Signal Design for VSS Construction

This section describes the output signal design for constructing a VSS in the range between the listener and the loudspeaker array, where the target distance satisfies the condition $0 \leq r \leq d$. When the VSS is at the position (r, θ), the output signal for EDL is weighted by $g_{\mathrm{E}}(r)$ and the signal for the PAL is weighted by $g_{\mathrm{P}}(r)$. Subsequently, the signals for the left and right channels are convolved with the respective HRTFs, $h_{\mathrm{L}}(\theta, t)$ and $h_{\mathrm{R}}(\theta, t)$.

The output signals from the EDL are as follows:

$$x_{\mathrm{LE}}(t) = g_{\mathrm{E}}(r) \cdot h_{\mathrm{L}}(\theta, t) * s(t), \tag{1}$$

$$x_{\mathrm{RE}}(t) = g_{\mathrm{E}}(r) \cdot h_{\mathrm{R}}(\theta, t) * s(t), \tag{2}$$

where * denotes the convolution operation. The equations for determining the weighting coefficients will be described later in Section 2.3.

When the VSS is at $(0 \leq r < d)$ the output signals from the PAL are as follows:

$$x_{\mathrm{LP}}(t) = g_{\mathrm{P}}(r) \cdot h_{\mathrm{L}}(\theta, t) * s(t), \tag{3}$$

$$x_{\mathrm{RP}}(t) = g_{\mathrm{P}}(r) \cdot h_{\mathrm{R}}(\theta, t) * s(t), \tag{4}$$

$$y_{\mathrm{LP}}(t) = \{1 + x_{\mathrm{LP}}(t)\} \cdot v_{\mathrm{C}}(t), \tag{5}$$

$$y_{\mathrm{RP}}(t) = \{1 + x_{\mathrm{RP}}(t)\} \cdot v_{\mathrm{C}}(t), \tag{6}$$

where, $v_{\mathrm{C}}(t)$ is the ultrasonic carrier wave, defined as follows:

$$v_{\mathrm{C}}(t) = \sin(2\pi F_C t). \tag{7}$$

Furthermore, the sound pressure levels of the EDL array and the PAL array are normalized to be equal at the listener position. In this way, by outputting signals with controlled distance and direction from the PAL and EDL, a VSS for each angle are presented.

2.3 Distance Weighting Coefficients for DRR Control

The distance control weighting coefficients $g_{\mathrm{E}}(r)$ and $g_{\mathrm{P}}(r)$ are as follows:

$$g_{\mathrm{E}}(r) = \sqrt{\frac{r}{d}}, \tag{8}$$

$$g_{\mathrm{P}}(r) = \sqrt{\frac{d-r}{d}}. \tag{9}$$

From Equation (8), when the VSS is at the listener's position $(r = 0)$, no signal is output from the EDL. From Equation (9), when the VSS is at the loudspeaker array's distance $(r = d)$, no signal is output from the PAL.

This distance weight is used to control DRR. This equation is an excerpt from the vector-based amplitude panning (VBAP) [11] equation with only the distance control section. A PAL is a loudspeaker that can reproduce ultrasonic waves. By emitting amplitude-modulated waves at high sound pressure, the signal becomes self-demodulated and audible due to the nonlinearity of the air. PALs and EDLs possess distinct directional characteristics. The ultrasonic wave emitted from a PAL has sharp directivity [12], resulting in a high direct sound with low reverberation [12]. In contrast, EDL has wide directivity, causing its sound to reflect and reverberate throughout the room's structure. Consequently, when used in the same environment, the DRR produced by a PAL is different from that of an EDL. This difference is utilized to control the overall DRR by adjusting the amplitude of the PAL and EDL signals. Through signal processing, the ratio of energy radiated from the PAL and the EDL during simultaneous playback is controlled. This method of directivity control makes it possible to control the DRR.

In this way, by using the two pairs of PALs and EDLs array and reproducing sound sources with controlled distance and direction from the PAL and EDL, VSS with different distances and angles are presented.

3. EVALUATION EXPERIMENTS

3.1 Experimental Setup and Conditions

To evaluate the effectiveness of the proposed method, we conducted objective evaluation experiments for DRR.

Table 1 – Experimental equipment

Parametric array loudspeaker	MEE, MSP-60E
Electrodynamic loudspeaker	ELAC, BS302
Power amplifier	YAMAHA, XM4180
Microphone	NEUMANN, KU100
D/A, A/D converter	RME, Fireface UFXII

Table 2 – Experimental conditions

Environment	Soundproof room (T_{60} = 0.24 s)
Ambient noise level	L_A = 29.8 dB
Sampling frequency	F_S = 192 kHz
Quantization	32 bits
Input signal	Time-stretched pulse [13]
Carrier wave frequency	F_C = 40 kHz
Modulation method	Double sideband amplitude modulation [10]
HRTF database	Nagoya University Head-Related Transfer Function Database [14]

Figure 2 –Photograph of the PAL and EDL array

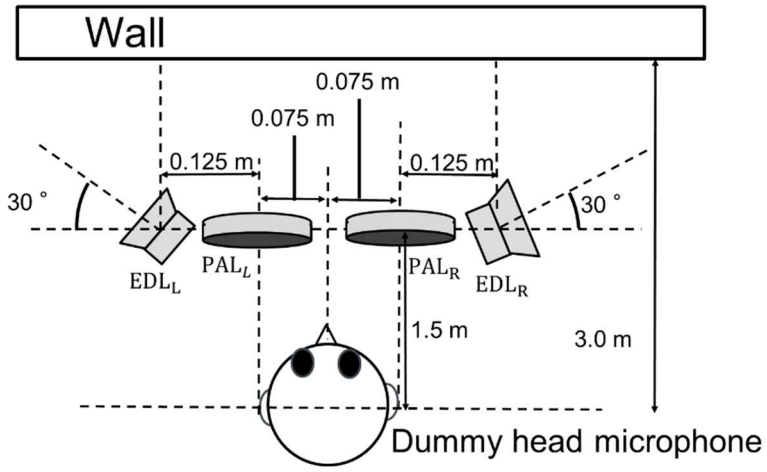


Figure 3 – Arrangement of the equipment

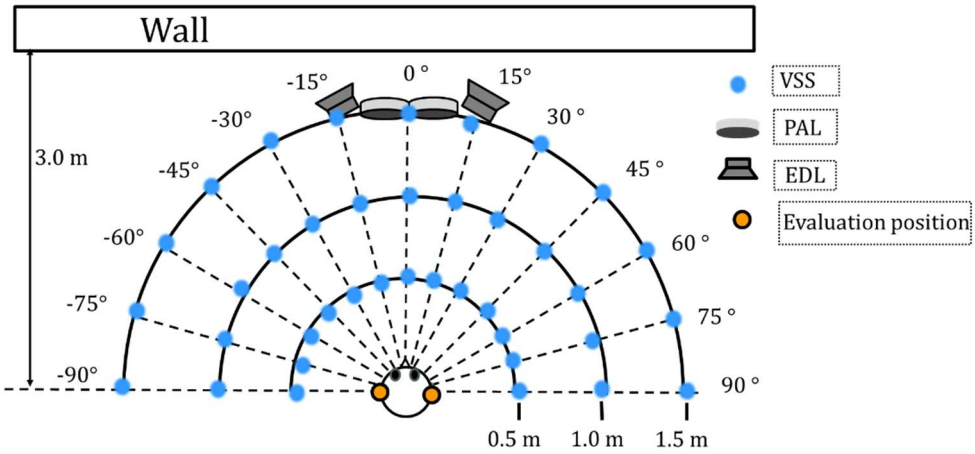


Figure 4 –VSS representation positions

Tables 1 and 2 show the experimental equipment and conditions. Figure 2 shows a photograph of the PAL and EDL array. Figure 3 shows arrangement of the equipment. To calculate reverberation using the line array, a microphone was placed instead of the listener at the listening position to calculate the impulse response. After using the calculated impulse response with distance weighting and direction control, a simulation of the array was performed by combining them. The impulse response was calculated via the time-stretched pulse (TSP) method [13], wherein the recorded signal, captured from an emitted TSP signal, was used with an inverse TSP. To compare with the impulse response of real sound sources, loudspeakers were actually placed at presentation distances and directions, and impulse responses were calculated and compared with the simulated synthesized ones. The comparison was performed using DRR. Figure 4 shows the VSS presentation positions used in this experiment to compare the real and simulated VSS. The comparison was performed under a total of 39 conditions, combining 13 angles (from -90 ° to 90 ° in 15° increments) and three distances (0.5 m, 1.0 m, and 1.5 m).

In these experiments, we use errors of DRR as evaluation metrics. DRR is one of the cues for perceiving distance. The formula for calculating the DRR from the power of each section of the direct sound and the reverberant sound is as follows:

$$\mathrm{DRR} = 10\log_{10}\left(\frac{\sum_{n=0}^{T_\mathrm{D}F_\mathrm{S}-1} p^2(n)}{\sum_{n=T_\mathrm{D}F_\mathrm{S}}^{T'F_\mathrm{S}-1} p^2(n)}\right), \tag{13}$$

where n is the sample index, $p(n)$ is impulse response, T_D is time of direct sound, F_s is sampling frequency, $T'$is time of reverberant sound. T_D is the length of the direct sound, which is an interval that does not include the initial reflection. T_D was set at 6 ms. T' is the length of the reverberant sound. T' was set to 240 ms. F_S is sampling frequency. In this paper, we calculated the error between the DRR of the real sound source and the DRR of the VSS. The error of DRR representing the discrepancy between the real sound source and VSS, can be calculated with Equation (13) as follows:

$$\mathrm{DRR}_\mathrm{error} = |\mathrm{DRR}_\mathrm{real} - \mathrm{DRR}_\mathrm{VSS}|. \tag{14}$$

$\mathrm{DRR}_\mathrm{real}$ denotes the DRR of real sound source, $\mathrm{DRR}_\mathrm{VSS}$ denotes the DRR of VSS, $\mathrm{DRR}_\mathrm{error}$ is the difference between $\mathrm{DRR}_\mathrm{real}$ and $\mathrm{DRR}_\mathrm{VSS}$. If the error in the DRR between the real sound source and the VSS is small, a human will perceive a sound image at the distance of the VSS.

3.2 Experimental Results and Discussions

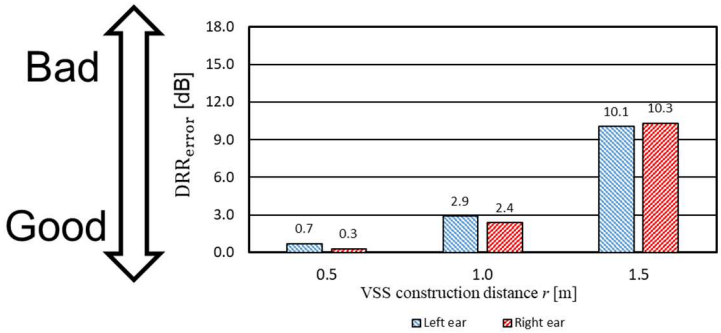
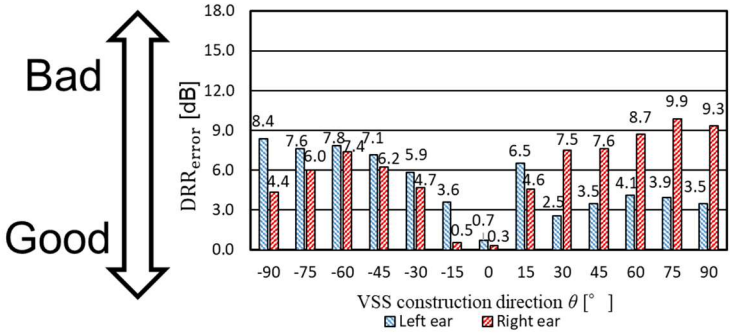
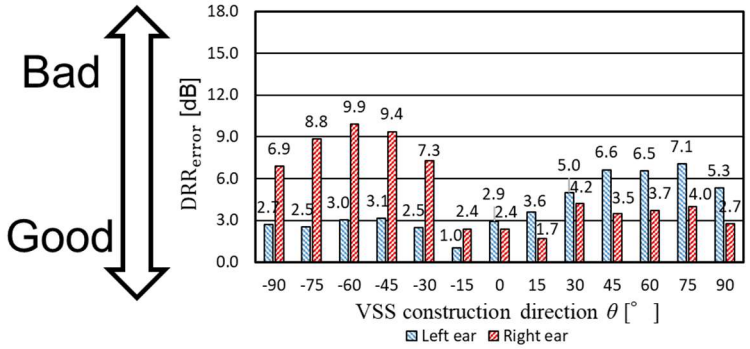


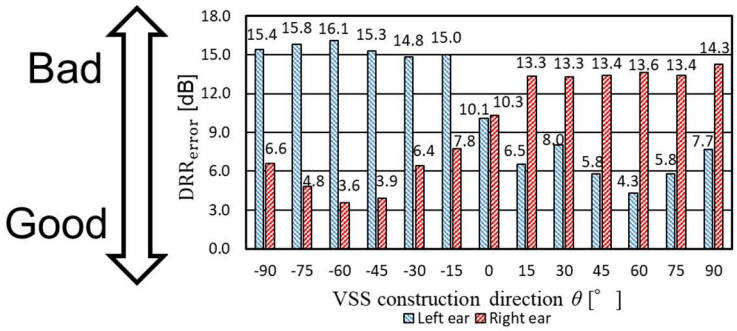
Figure 5 DRR errors of the proposed method in condition $\theta = 0°$



(a) Condition $r = 0.5$ m



(b) Condition $r = 1.0$ m



(c) Condition $r = 1.5$ m

Figure 6 DRR errors of the proposed in each condition of distance r

Figure 5 shows DRR errors of the proposed method in condition $\theta = 0°$. In each condition, the vertical axis shows DRR error, the horizontal axis shows VSS constructed position. From the results of Figure 5, which show the $\mathrm{DRR_{error}}$ when the presentation angle was set to 0 ° (front) and the presentation distance was changed. It is considered that the real sound source can be simulated because the error of DRR between the real sound source and the simulated sound source is less than 6 dB in the near-field range of 0.5 m to 1.0 m for sound source presentation distance. The large error of DRR at 1.5m is thought to be because only sound from the EDL is emitted at the 1.5 m position according to the distance weighting determination formula, and the reverberant sound component is strong when the EDL is directed backward, causing the reflected sound to be larger than the direct sound, leading to a larger error with the real sound source.

Figure 6(a), which shows the results for a presentation distance of 0.5 m, indicates that the error in the ipsilateral ear increases as the presentation angle deviates from 0 ° (front). In contrast, the error in the contralateral ear increases as the angle approaches 60 ° and −60 °. From the results of Figure 6(b), which show the $\mathrm{DRR_{error}}$ when the presentation distance was set to 1.0 m and the presentation angle was changed, it is considered that the real sound source can be simulated because the error is 6 dB or less in the range of −90 ° to 30 ° for the left ear and −15 ° to 90 ° for the right ear. As the sound source presentation angle approaches 60 ° or −60 °, the error increases in the contralateral ear. From the results of Figure 6(c), which show the $\mathrm{DRR_{error}}$ when the presentation distance was set to 1.5 m and the presentation angle was changed, it is considered that the real sound source can be simulated because the error is 6 dB or less when the sound source presentation angle approaches 60 ° or −60 °, the error increases in the contralateral ear.

The overall large error is thought to be because only sound from the EDL is emitted at the 1.5 m position according to the distance weighting determination formula, and the reverberant sound component is strong when the EDL is directed backward, causing the reflected sound to be larger than the direct sound, leading to a larger error with the real sound source.

From these results, when the conditions are within a range of 1.0 m in the 0 ° front direction, the VSS is effectively constructed. Therefore, the proposed method is effective in certain conditions.

4. CONCLUSION

In this research, we proposed VSS construction based on DRR control using a line array loudspeaker consisting of parametric and conventional loudspeakers. This method utilizes a specially configured line array that integrates two forward-facing PALs and two backward-facing EDLs. The proposed method applied distance weighting and direction filtering by HRTFs to the EDL signals. For the PAL signals, in addition to

the same processing, we performed amplitude modulation. The experimental evaluations demonstrated that the proposed method effectively constructed VSS in certain conditions. In the future, we will consider a compensation formula that can simulate the DRR of real sound sources and confirm the presentation performance.

ACKNOWLEDGEMENTS

This work was partly supported by Ritsumeikan University R-GIRO, ARC and RARA, and JSPS KAKENHI Grant Numbers JP23K28115, JP24K20803, JP25H01158 and JP25K15220.

REFERENCES

1. Henrik Møller, "Fundamentals of binaural technology," Applied Acoustics, vol. 36, no. 3-4, pp. 171-218, 1992.
2. Hironori Takemoto, Parham Mokhtari, Hiroaki Kato, Ryouichi Nishimura, Kazuhiro Iida, "Mechanism for generating peaks and notches of head-related transfer functions in the median plane," The Journal of the Acoustical Society of America, vol. 132, no. 6, pp. 3832-3841, 2012.
3. Masato Nakayama, Takuma Ekawa, Toru Takahashi, Takanobu Nishiura. "Virtual Sound Source Construction Based on Direct-to-Reverberant Ratio Control Using Multiple Pairs of Parametric-Array Loudspeakers and Conventional Loudspeakers," Applied Sciences, vol.15, no.7, pp.3744, 2025.
4. Woon-Seng Gan, Jun Yang, Tomoo Kamakura, "A review of parametric acoustic array in air," Applied Acoustics, vol. 73, no. 12, pp. 1211-1219, 2012.
5. Peter J. Westervelt, "Parametric acoustic array," The Journal of the Acoustical Society of America, vol. 35, no. 4, pp. 535-537, 1963.
6. Richard H. Small, "Closed-box loudspeaker systems Part I: Analysis," Journal of the Audio Engineering Society, vol. 20, no. 10, pp. 978-808, 1972.
7. Mehdi Zohourian, Rainer Martin, "Binaural direct-to-reverberant energy ratio and speaker distance estimation," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 28, pp. 92-104, 2019.
8. Pavel Zahorik, "Direct-to-reverberant energy ratio sensitivity," The Journal of the Acoustical Society of America, vol. 112, no. 5, pp. 2110-2117, 2002.
9. Stanley T. Birchfield, Rajitha Gangishetty, "Acoustic localization by interaural level difference," Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005, vol. 4, pp. 1109-1112, 2005.
10. Masahide Yoneyama, Jun-ichiroh Fujimoto, Yu Kawano, Shoichi Sasabe, "The audio spotlight: An application of nonlinear interaction of sound waves to a new type of loudspeaker design," The Journal of the Acoustical Society of America, vol. 73, no. 5, pp. 1532-1536, 1983.
11. Ville Pulkki, "Virtual sound source positioning using vector base amplitude panning," Journal of the Audio Engineering Society, vol. 45, no. 6, pp. 456-466, 1997.
12. Yuta Hatano, Chuang Shi, Yoshinobu Kajikawa. "Compensation for nonlinear distortion of the frequency modulation-based parametric array loudspeaker," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 25, no.8 pp. 1709-1717, 2017.
13. Nobuharu Aoshima, "Computer-generated pulse signal applied for sound measurement," The Journal of the Acoustical Society of America, vol. 69, no. 5, pp. 1484-1488, 1981.
14. Nagoya University Head-Related Transfer Function Database, URL: http://www.sp.m.is.nagoya-u.ac.jp/HRTF/index-j.html

Estimating Reverberation Time in Enclosed Spaces by Analyzing Spectrogram Variations in Stereo Recordings

Juseong KIM[1]; Joonwhi KIM[2]; Jungyu CHOI[3]; Seowoo KIM[4]; Jungin JU[5]; Sungbin IM[6]
Soongsil University, Republic of Korea

ABSTRACT

Reverberation time (RT60) is an acoustic parameter that varies depending on the characteristics of an indoor environment and can serve as a control indicator for enabling dereverberation algorithms to adapt to surrounding conditions. This paper proposes a signal processing-based framework for estimating RT60 by leveraging the difference between time–frequency spectrograms obtained from spatially separated stereo microphones.

Dry speech signals recorded in anechoic conditions are convolved with synthetic room impulse responses (Room Impulse Response) generated using Pyroomacoustics, in order to simulate diverse acoustic environments. The simulated rooms are rectangular with side lengths ranging from 3 m to 7 m, reflecting typical indoor settings. In all simulations, the sound source is positioned near the center of the room, while the two microphones are placed at distinct locations.

The model input consists of the magnitude difference between the STFT (short-time Fourier transform) spectrograms of the two microphone signals. The target RT60 values are computed using the Sabine formula, and a supervised regression model is trained accordingly. The generalization performance of the model is evaluated on test data including unseen audio sources and similar spatial configurations.

Experimental results demonstrate that the proposed model can reliably estimate RT60 under various acoustic conditions based on stereo spectrogram differences. These findings suggest that the proposed method has the potential to be employed as an environment-adaptive control signal within future dereverberation pipelines.

Keywords: Spectrogram, Reverberation, Machine Learning

1. INTRODUCTION

With the increasing prevalence of voice calls and video conferencing systems on personal computers (PCs), the need for accurate voice activity detection (VAD) and effective denoising algorithms has become more critical than ever. In particular, modern environments often incorporate noise cancellation technologies that, while beneficial, pose new challenges for VAD systems. One notable issue arises when speakers talk at low volumes—under such conditions, VAD systems frequently fail to detect speech segments accurately.

A common workaround is to lower the detection threshold to improve sensitivity. However, this approach can lead to unintended consequences, such as the capture of background noise and reverberation. These interferences significantly degrade the performance of VAD, resulting in unreliable detection. Consequently, users are often forced to manually adjust system parameters to adapt to their specific acoustic environments. This trial-and-error process is not only inconvenient but also impractical in dynamic or shared settings.

[1] bocjebi1557@gmail.com
[2] kbb@nate.com
[3] cjg@soongsil.ac.kr
[4] seowoo1188@gmail.com
[5] 1023jji@gmail.com
[6] sbi@ssu.ac.kr

In this context, there is an increasing demand for adaptive parameter control techniques that can dynamically adjust VAD and denoising settings in response to the speaker's position and the acoustic properties of the surrounding environment. Among various noise sources that degrade speech clarity, white noise and reverberation are the most prevalent. In particular, reverberation refers to the temporal smearing of sound caused by reflections from walls, ceilings, and objects after the original sound has ceased.

A widely used metric for characterizing reverberation is RT60 (Reverberation Time 60 dB), which denotes the time required for the sound energy to decay by 60 dB after the sound source stops. The RT60 value is influenced by the room's geometry, volume, and surface absorption coefficients, and it can be estimated using well-established formulas such as Sabine and Eyring equations (1) (2). Recent studies have leveraged RT60 as a control variable in dereverberation algorithms, enabling dynamic adjustment of filter coefficients or spectral masking parameters based on the estimated reverberation level (3) (4).

Building upon this background, the present study proposes a data-driven framework for estimating RT60 directly from speech signals under varying acoustic conditions. The core of the method is a supervised regression model that captures reverberation characteristics by analyzing time–frequency differences in two-channel stereo recordings. To validate the approach, a series of simulated datasets are constructed to represent a diverse range of indoor environments. The model demonstrates robust generalization performance, achieving accurate RT60 estimates even for unseen speakers and room configurations.

The remainder of this paper is structured as follows. Section 2 details the proposed estimation framework. Section 3 describes the experimental setup, including data generation and model architecture. Section 4 presents the results and analysis. Finally, Section 5 concludes the study and suggests directions for future work.

2. Proposed Method

The RT60 estimation framework proposed in this study aims to estimate reverberation time by analyzing spatial reverberation characteristics from stereo speech signals received via two microphones. The overall processing pipeline consists of three stages. First, spectrograms are generated using short-time Fourier transform (STFT) to obtain a time–frequency domain representation of the input signals. Next, generalized cross-correlation with phase transform (GCC-PHAT) features are extracted to compute the time delay between the two microphones. Finally, RT60 values are predicted using a convolutional recurrent neural network (CNN-RNN) based regression model. A flow diagram of the proposed method is shown in Figure 1.

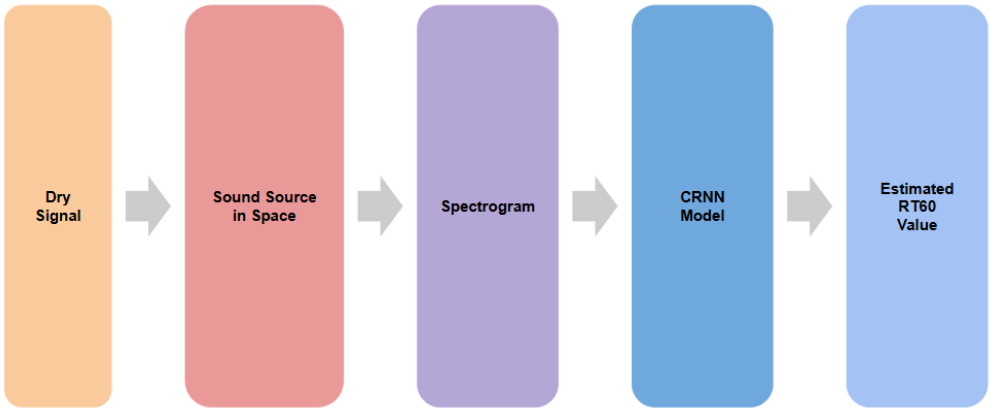


Figure 1 -System block diagram from input to output for the proposed approach

2.1 STFT-based Spectrogram Representation

The short-time Fourier transform (STFT) is applied to each microphone channel to obtain a time–frequency representation of the input signals (5). In this study, we use an FFT window size of n_fft = 1024 and a hop length = 512, which yields approximately 392 frames (corresponding to roughly 4.5 seconds) on the time axis, based on a sampling rate of 44.1 kHz.

As shown in Figure 2, each simulation sample generates two spectrograms—one per channel. we utilize the magnitude difference between the two spectrograms, which serves as the input feature. This approach effectively captures the spatial reverberation contrast between the microphones while reducing the input data size by half. As a result, it improves computational efficiency and reduces memory consumption during model training.

.

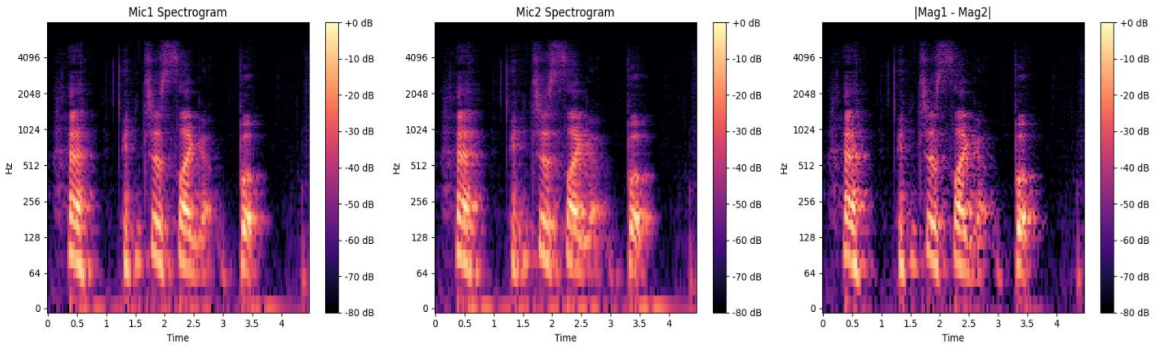


Figure 2 - Spectrograms for each channel and their magnitude difference

2.2 GCC-PHAT-based Temporal Feature

To capture the relative time delay between the sound source and each microphone, we extract time delay–related features for each simulation sample using generalized cross-correlation with phase transform (GCC-PHAT). GCC-PHAT is widely known as an effective feature for enhancing spatial resolution, particularly in tasks such as sound source localization. In this study, it is employed as an auxiliary input to the proposed model.

2.3 CRNN-based RT60 Estimation Model

The extracted spectrogram difference and time delay features are used as inputs to a CNN–LSTM–based regression model (7). First, a 2D convolutional neural network (CNN) summarizes local frequency–time patterns from the input. Then, a bidirectional long short-term memory (Bi-LSTM) layer learns temporal dependencies across the time axis to construct a sequential representation. Finally, a fully connected layer outputs a continuous RT60 value.

The model is trained using mean squared error (MSE) as the loss function. To prevent overfitting, early stopping and a learning rate scheduler are employed during training.

3. Experiment

3.1 Setup and Problem

To train and evaluate the proposed RT60 estimation model, we constructed a synthetic acoustic simulation dataset that reflects diverse indoor environments. The simulations were performed using the Pyroomacoustics library, under the assumption of shoebox-shaped rooms (rectangular cuboids) for all experiments.

To prevent overfitting to specific room sizes and ensure generalization performance, the room dimensions—length, width, and height—were randomly sampled between 3.0 m and 7.0 m, with an additional ±10% proportional distortion applied to each axis. This design allows the dataset to reflect a wide range of realistic indoor aspect ratios.

For each simulation, the sound source was randomly placed near the center of the room. A two-channel linear microphone array, aligned along the X-axis relative to the source, was used: the first microphone was located 25 cm from the source, and the second at 50 cm. Both microphones were positioned at least 10 cm away from the room boundaries to ensure physically plausible configurations. Figure 3 illustrates an example of a simulated room configuration implemented using the Pyroomacoustics library.

.

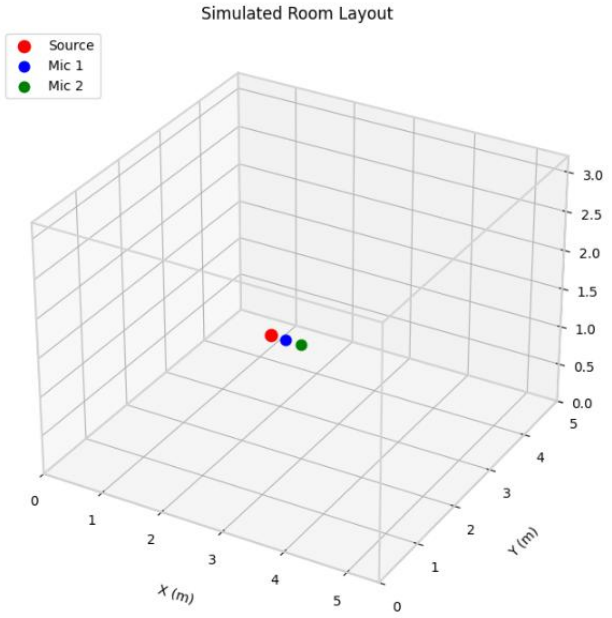


Figure 3 - Overview of a simulated room configuration implemented using the Pyroomacoustics module

To quantify the reverberation characteristics of each simulated room, the reverberation time RT60 was calculated using the Sabine formula as follows:

$$RT_{60} = \frac{0.161 \cdot V}{A} \tag{1}$$

, where V denotes the volume of the room, and A is defined as the total absorption area, calculated by multiplying the surface area of each wall by a uniform absorption coefficient α=0.2. The computed RT60 value was used as the regression target for supervised learning in each simulation sample.

The simulation output was stored as two-channel stereo audio files in WAV format. Associated metadata—including room dimensions, microphone and source positions, and the calculated RT60—were recorded for each sample. All metadata was integrated into a single CSV file, which was subsequently used for model training and performance analysis.

The generated audio samples were transformed into spectrograms via the short-time Fourier transform (STFT). For spectral analysis, the FFT size n_m to 1024 and hop length to 512. The FFT size determines the frequency resolution, while the hop length was set the temporal resolution. This configuration yield approximately 43 Hz frequency resolution with a 50% overlap. Such a setting allows for sufficient representation of low-frequency components, which often exhibit longer reverberation tails in indoor acoustic environments.

In addition, to incorporate time-delay-based acoustic features, the generalized cross-correlation with phase transform (GCC-PHAT) was computed for each sample and stored as an auxiliary input feature.

3.2 Dataset

To obtain dry speech signals suitable for acoustic simulation, we extracted speech segments from publicly available YouTube videos recorded in anechoic chamber environments. The selected recordings were manually verified to exhibit clear articulation and minimal reverberation, satisfying the conditions for dry source signals (8)(9). {The audio clips were used solely for non-commercial academic research, and the original copyrights remain with the respective video uploaders.}

Each speech file was trimmed to approximately 10 seconds using the REAPER digital audio workstation (DAW) and curated as a single-speaker audio sample for simulation. All files were resampled to 44.1 kHz to ensure consistency. The speaker pool consisted of three distinct adult male speakers categorized by age, including seven samples from a young male speaker, two samples from a middle-aged male speaker, and four samples from an older male speaker. These samples were used for model training and validation. For

testing, two previously unseen samples were employed—one from the same middle-aged speaker and another from the young speaker—in order to assess the model's generalization performance under novel speech and acoustic conditions.

.

Table 1 - CRNN Network Architecture

Layer	Type	Shape / Params	Output Dim
Input	-	(1, 513, 392)	(1, 513, 392)
Conv2D	(3×3), 16ch	stride=1, padding=1	(16, 513, 392)
MaxPool2D	(2×2)	-	(16, 256, 196)
Reshape	-	T=196, F=256×16	(T=196, 4096)
Bi-LSTM	hidden=128	bidirectional	(196, 256)
FC1	Linear(256→64)	ReLU	(64)
FC2	Linear(64→1)	-	(1)

3.3 Model Train & Test

A total of 6,500 training samples were generated by simulating each of the 13 dry speech recordings 500 times under random room conditions. All audio samples were resampled to 44.1 kHz, and transformed into spectrograms using the short-time Fourier transform (STFT). The FFT window size was set to n_m=1024, with a hop length of 512 samples. Given the sampling rate, the time resolution between STFT frames corresponds to approximately 512 / 44100 ≈ 11.6ms. We fixed the input length at 392 frames, resulting in each input covering approximately 4.55 seconds of audio (392×0.0116). Since 500 simulations were performed per dry sample, the total number of generated training samples was 6,500.

For each sample, the model input consists of the magnitude difference (mag_diff) between the spectrograms of the two microphone channels. In addition, GCC-PHAT–based time delay features were appended as auxiliary inputs.

Table 1 provides a summary of the model architecture. The model architecture follows a convolutional recurrent neural network (CRNN) structure that combines CNN and bidirectional LSTM layers. The CNN receives a 3D input tensor of shape (1, 513, 392), corresponding to one input channel, 513 frequency bins, and 392 time frames. The first convolutional layer uses a 3×3 kernel with 1 input channel and 16 output channels. The CNN output is then flattened along the time axis and passed to a bidirectional LSTM with 128 hidden units. The final RT60 value is predicted through a fully connected (FC) layer that maps the LSTM output to a single scalar.

.

$$\mathcal{L}_{\text{MSE}} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{2}$$

The model was trained using the mean squared error (MSE) as the loss function, and optimized with the Adam optimizer initialized with a learning rate of 0.001. A Reduce-LR-On-Plateau scheduler was applied to halve the learning rate when the validation loss failed to improve. In addition, early stopping was employed to terminate training when no further improvement in validation loss was observed for a set number of epochs.

The model parameters corresponding to the lowest validation loss were saved as a checkpoint, and this version was later used for testing and evaluation.

4. Experimental Results

4.1 Results

The performance of the model was evaluated using mean absolute error (MAE) and root mean square error (RMSE) as the primary metrics. On the training dataset, the model achieved a final MAE of 0.02384 seconds and an RMSE of 0.03094 seconds.

As shown in Figure 4, a scatter plot of the predicted versus ground truth RT60 values indicates that the model's predictions closely follow the identity line (y=x), demonstrating stable and reliable estimation performance.

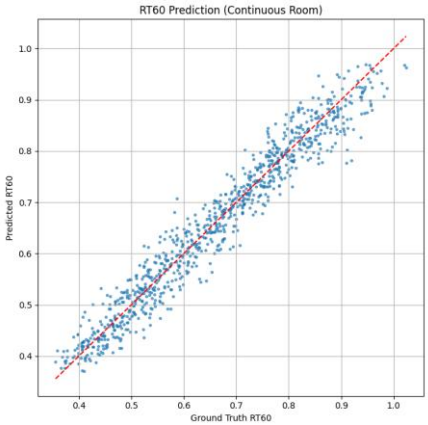


Figure 4 - RT60 prediction results on the training set (Ground Truth vs. Prediction)

The test dataset was composed of previously unseen speech samples and novel room configurations, distinct from those used during training. Figure 5 presents a scatter plot of the model's predictions on the test set. For each test utterance, 200 simulation samples were generated to evaluate model performance. The resulting scatter plot reveals a tendency to form two distinct trends, which are presumed to be related to the characteristics of the test utterances. Both trends exhibit a certain degree of systematic bias, suggesting the presence of speaker- or content-specific factors. Further investigation will be conducted in future work to analyze these biases and explore potential compensation strategies.

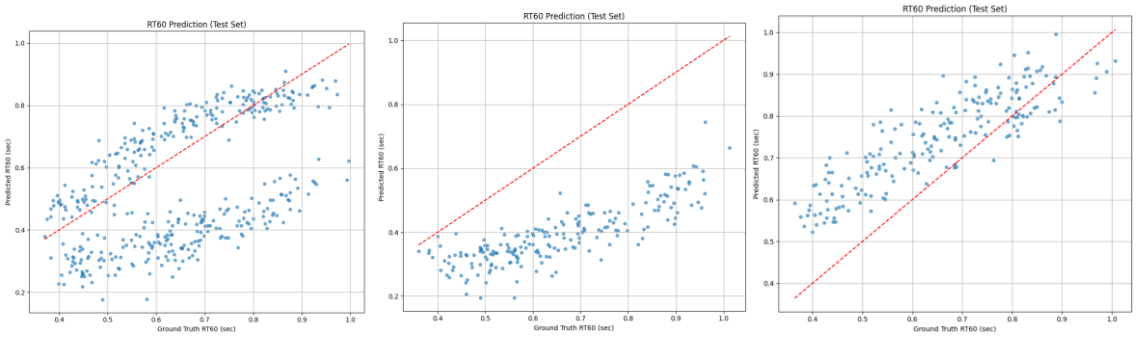


Figure 5 - RT60 prediction results on the test set
(from left to right: overall / middle-aged male / young male)

Additionally, the performance was separately analyzed for the two test utterances. For the young male speaker, the model achieved a test MAE of 0.10692 seconds and an RMSE of 0.1244 seconds, with predictions closely clustered around the identity line. In contrast, for the middle-aged male speaker, the model exhibited higher error, with a test MAE of 0.17874 seconds and an RMSE of 0.19628 seconds. This discrepancy may be attributed to the greater presence of young male speaker data in the training set, which likely allowed the model to better adapt to that voice profile.

These results suggest that increasing the quantity and diversity of training data could enhance the model's ability to generalize RT60 estimation to a wider range of speech inputs. Overall, the experimental results demonstrate that the proposed model is capable of estimating RT60 values with reasonable accuracy under a variety of room configurations and speech conditions.

4.2 Supplementary Experiments

To verify the necessity of each component used in the proposed framework, a series of controlled ablation experiments were conducted. These experiments involved removing one element at a time—such as using a single microphone instead of two, skipping the STFT-based spectrogram transformation, or replacing the magnitude difference computation with simple concatenation of the two spectrograms. Additionally, to confirm that the model's performance does not solely rely on GCC-PHAT, we included an experiment that excluded all audio transformations and used only the GCC-PHAT features as input. In Experiment 1, the model was trained using only one microphone channel as input. In Experiment 2, the STFT transformation was omitted, and the model was trained directly on the raw waveform. In Experiment 3, instead of computing

the difference between two spectrograms, the model received both spectrograms concatenated along the channel dimension. In Experiment 4, the input consisted solely of GCC-PHAT features, with no spectrogram-based representation. All four experiments used the same speech sources that had demonstrated strong performance in the original setup, ensuring consistency in evaluation.

In Experiment 1, although the training phase yielded similar performance to the original model, a significant drop in accuracy was observed during testing, as illustrated in Figure 6. This result highlights the importance of binaural information in achieving generalizable RT60 estimation.

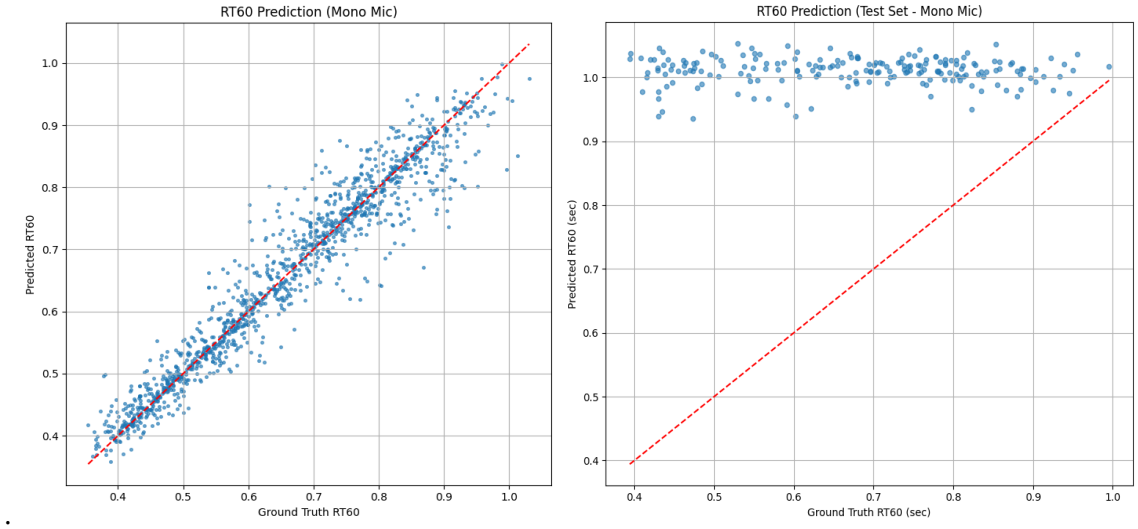


Figure 6 - Training and test results using data from a single microphone

In Experiment 2, where the STFT transformation was omitted and raw waveforms were used as input, the model's predictions were biased toward a narrow range, failing to reflect the actual spread of RT60 values. As shown in Figure 7, this indicates that time–frequency representation is essential to learn the variability present in reverberation characteristics.

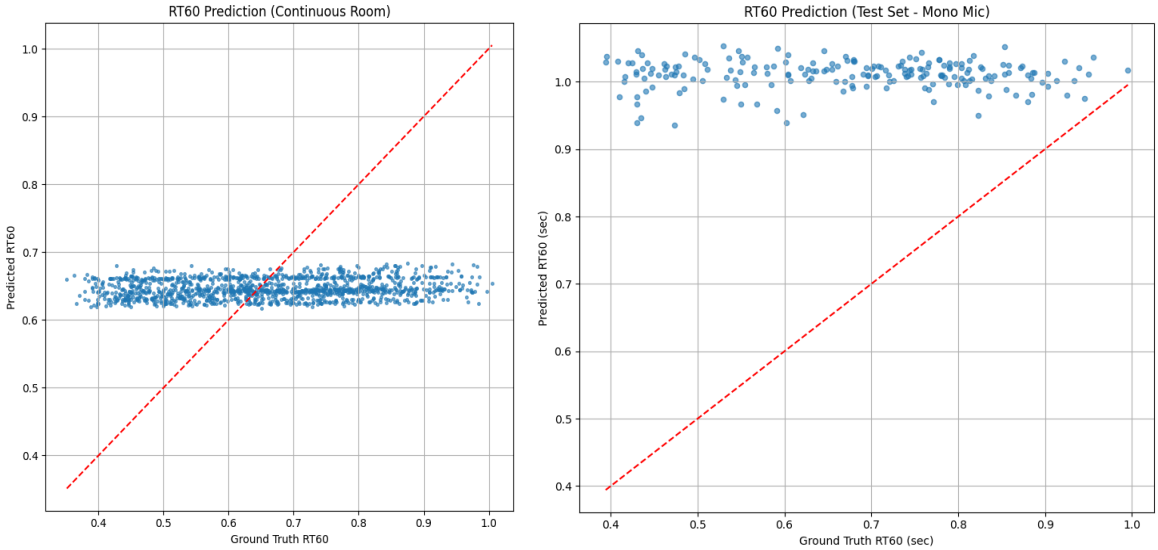


Figure 7 - Training and test results using data without STFT transformation

In Experiment 3, the model showed strong training performance, and although the test accuracy was lower than that of the original model, the predictions still exhibited a discernible trend. This suggests that while the magnitude difference is beneficial for generalization, simple spectrogram concatenation may still retain partial spatial information, as evidenced by the results in Figure 8.

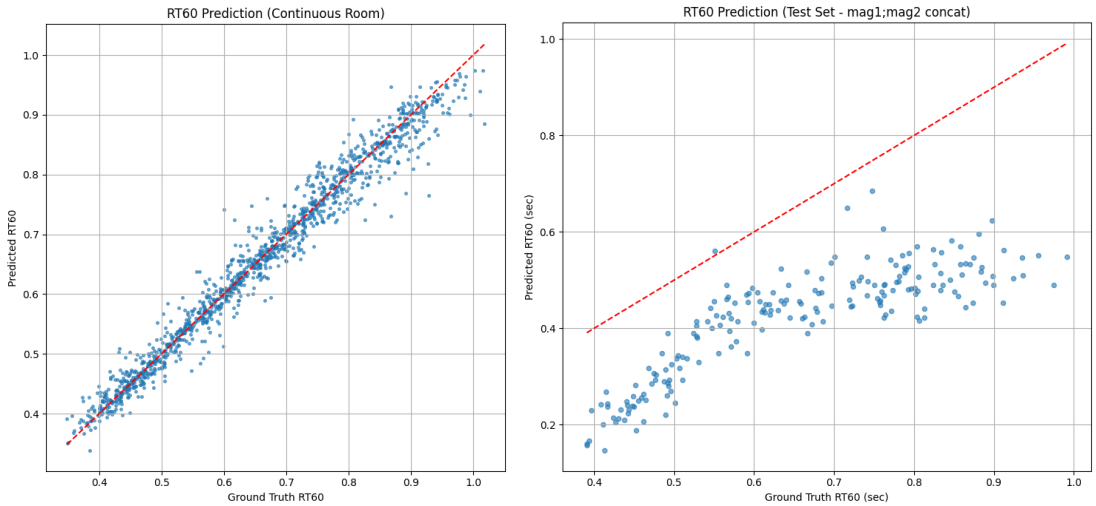


Figure 8 - Training and test results using concatenated spectrograms as input

As shown in Figure 9, the model produced extremely poor results in Experiment 4, outputting nearly a single constant value for all predictions in both training and testing. This confirms that GCC-PHAT alone is insufficient for reliable RT60 estimation in this setting.

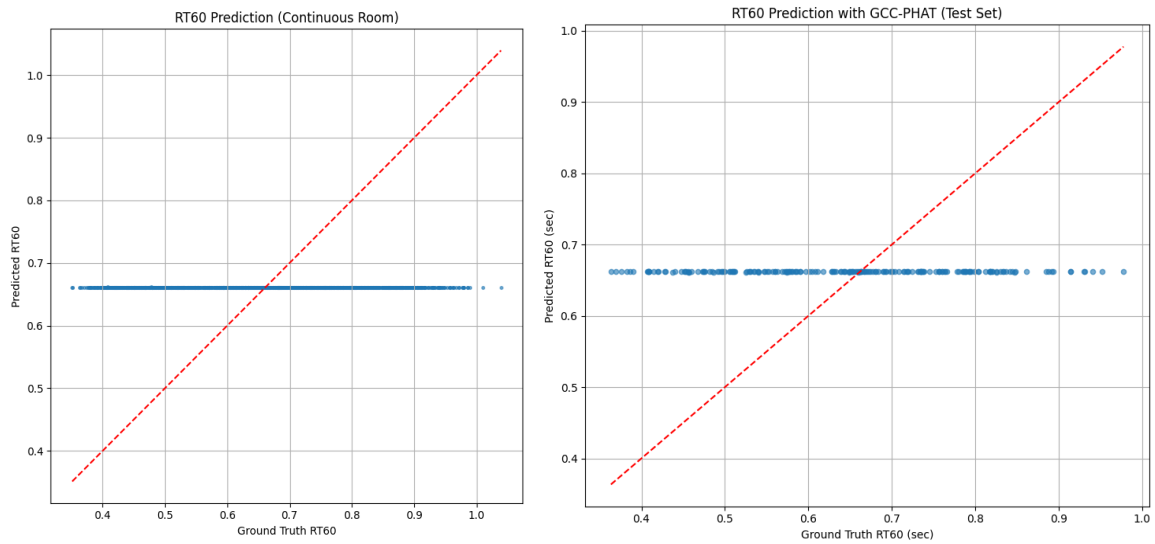


Figure 9 - Training and test results using GCC-PHAT only as input

5. CONCLUSIONS

In this study, we proposed a data-driven approach for estimating the reverberation time (RT60) of indoor environments by leveraging both the spectrogram magnitude difference between stereo microphone signals and GCC-PHAT–based time delay features. The model adopts a regression architecture combining convolutional neural networks (CNNs) with bidirectional long short-term memory (Bi-LSTM) layers, and was trained on speech data simulated in rooms of varying sizes and aspect ratios.

Experimental results demonstrated that the proposed method achieved low MAE and RMSE on both training and test datasets, maintaining reliable prediction performance across diverse room conditions. Notably, the model generalized well to unseen speech inputs, highlighting its potential for integration into environment-adaptive dereverberation systems as a real-time control signal. Additional ablation studies further reinforced the validity and reliability of the proposed architecture.

Moreover, the learned RT60 estimates could contribute to downstream tasks such as voice activity detection (VAD) and speaker localization in noisy environments. However, we observed a performance bias toward certain speaker profiles, which suggests the need for further model refinement. Future work will aim to enhance data diversity and incorporate real-world recordings to improve generalization across speakers

and conditions. Expanding the dataset and exploring lightweight architectures for real-time inference will also be key directions for continued research.

ACKNOWLEDGEMENTS

This work was supported by the Technology Innovation Program (20025005, Development of abnormal sound-voice detection system for victims at accident sites) funded By the Ministry of the Interior and Safety (MOIS), Korea and partly by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT), Korea (RS-2024-00355759, Factory facility anomaly detection and localization technology based on multi-noise removal).

REFERENCES

1. Kuttruff H. Room Acoustics. 5th ed. Boca Raton: CRC Press; 2009. Chapter 9.4, p. 255.
2. Sabine WC. *Collected papers on acoustics*. Cambridge (MA): Harvard University Press; 1922.
3. Ratnam R, Reich T, Nicol T, Buck J, Sullivan E, Turner J. Blind estimation of reverberation time. *J Acoust Soc Am*. 2003;114(5):2877–92.
4. Bistafa SR, Bradley JS. Predicting reverberation times in a simulated classroom. *J Acoust Soc Am*. 2000;108(4):1721–31.
5. Oppenheim AV, Schafer RW. *Discrete-time signal processing*. 3rd ed. Upper Saddle River (NJ): Pearson; 2009. Chapter 9.
6. Knapp CH, Carter GC. The generalized correlation method for estimation of time delay. IEEE Trans Acoust Speech Signal Process. 1976;24(4):320–7.
7. Ordonez FJ, Roggen D. Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition. Sensors (Basel). 2016;16(1):115.
8. Ola Englund - METAL IN AN ANECHOIC CHAMBER https://youtu.be/bJsuUOuEZ9Q
9. The Action Lab - The Loudest Sound In The Quietest Room https://youtu.be/EBClzOM_Lbw

Robust Time Delay Estimation via Probabilistic Weighting of Parameterized GCC-PHAT on Enhanced Signals

Eojin KIM[1] ; Sungbin IM[2]
Soongsil University, Republic of Korea

ABSTRACT

In this paper, we propose a lightweight and robust method for accurate Time Delay Estimation (TDE) in acoustically challenging environments characterized by strong noise and reverberation. Building upon the Generalized Cross-Correlation with Phase Transform (GCC-PHAT) framework, the proposed approach incorporates a Direction-of-Arrival (DOA)-independent Coherent-to-Diffuse Ratio (CDR) estimator to enhance input signals by suppressing diffuse noise components based on spatial coherence. A parameterized PHAT filtering scheme is applied to generate multi-scale cross-correlation representations, which are then fed into a Convolutional Neural Network (CNN) for regression-based probabilistic confidence weights that modulate the CNN output, improving the reliability of delay prediction under varying acoustic conditions. Experimental results on simulated reverberant environments demonstrate that the proposed model outperforms existing baselines in terms of Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Accuracy (ACC), while using significantly fewer parameters (682K). These findings highlight the effectiveness of combining the interpretability and efficiency of traditional signal processing methods with the representational power of deep learning. The proposed hybrid framework offers a robust and practical solution for speaker localization in acoustically complex environments.

Keywords: Sound Source localization, Time Delay Estimation, Generalized Cross Correlation, Coherent-to-Diffuse ratio, Convolutional Neural Network.

1. INTRODUCTION

Sound Source Localization (SSL) is a fundamental acoustic sensing technology that identifies the origin of sound by analyzing audio signals captured by microphone arrays. This technology plays a critical role across a wide range of application domains, including bioacoustics for tracking wildlife vocalizations (1), wind turbine noise evaluation (2), speech recognition and enhancement (3), medical diagnostics (4), and industrial automation (5). However, reverberation and complex background noise in real-world environments remain major challenges that hinder accurate localization. Approaches to SSL can be broadly classified into traditional signal processing techniques and modern deep learning-based methods. Classical approaches include Time Difference of Arrival (TDOA)-based estimation, Multiple Signal Classification (MUSIC) using spatial correlation matrices, and Steered Response Power (SRP) techniques that search over a spatial grid. (6, 7, 8). These methods offer stable performance under ideal conditions due to their clear physical modeling. Nevertheless, their accuracy significantly deteriorates in real environments with strong noise and reverberation, revealing a critical limitation.

To overcome these shortcomings, data-driven Deep Learning (DL) models have gained significant attention in SSL research (9, 10). Owing to their strong nonlinear feature extraction capability, deep learning models have shown superior performance in acoustically complex environments involving severe reverberation and the presence of multiple sound sources. However, they introduce new challenges such as the “black box” nature of their predictions and a strong dependency on large-scale labeled datasets.

In this paper, a parameterized weighting scheme is incorporated into the PHAT filter to generate a refined correlation vector, which is then input into a CNN architecture. The correlation vector serves as a probabilistic confidence weight map, which is multiplied element-wise with the CNN output to enhance robustness in the final TDOA estimation. This hybrid approach leverages the interpretability of classical signal processing and

[1] djwls1815@gmail.com
[2] sbi@ssu.ac.kr

the expressive power of deep learning, achieving resilient localization performance even in noisy and reverberant conditions.

The remainder of this paper is organized as follows. Section 2 describes the signal processing techniques, and proposed neural network architecture. Section 3 describes the experimental setup and results. Finally, Section 4 concludes the paper.

2. Methods

2.1 Speech Enhancement

To overcome the limitations of traditional signal processing methods, this study employs a coherence estimator for speech enhancement. In reverberant indoor environments, the spatial coherence is computed as follows

$$\hat{\Gamma}_X(m,k) = \frac{\hat{\Phi}_{X_1X_2}(m,k)}{\sqrt{\hat{\Phi}_{X_1X_1}(m,k)\hat{\Phi}_{X_2X_2}(m,k)}},$$

where $X_1(m,k)$ and $X_2(m,k)$ denote the STFT-domain representations of the signals received at the two microphones, where m and k represent the time frame and frequency bin, respectively. $\hat{\Phi}_{X_iX_j}(m,k)$ correspond to the auto- and cross-power spectral densities between the two channels.

If the ideal coherence values for the noise field and the direct-path signal are known, the CDR can be estimated, which reflects the diffuseness of the acoustic environment. However, such ideal coherence typically depends on the DOA, limiting its applicability in practical scenarios.

To address this issue, we utilize a DOA-independent coherence estimator to approximate the ideal coherence of the direct-path component (11).

$$\widetilde{CDR}(m,k) = \frac{\tilde{\Gamma}_N(k)\mathcal{R}\{\hat{\Gamma}_X(m,k)\} - |\hat{\Gamma}_X(m,k)|^2 - \bar{\Gamma}(m,k)}{|\hat{\Gamma}_X(m,k)|^2 - 1},$$

$$\bar{\Gamma}(\mathrm{m},\mathrm{k}) = \sqrt{\tilde{\Gamma}_N^2\mathcal{R}\{\hat{\Gamma}_X\}^2 - \tilde{\Gamma}_N^2|\hat{\Gamma}_X|^2 + \tilde{\Gamma}_N^2 - 2\tilde{\Gamma}_N\mathcal{R}\{\hat{\Gamma}_X\} + |\hat{\Gamma}_X|^2}.$$

The enhanced signal can be observed in Figure 1 via spectrogram comparison. As shown in Figure 1, temporal smearing of energy in the 2-6kHz range is significantly reduced, indicating that reverberant components caused by reflections have been effectively suppressed. Consequently, the temporal boundaries of each sound become more distinct, with clearer onsets and offsets. The difference map is dominated by blue regions, suggesting that the CDR-based filter successfully attenuated late reverberation while preserving the direct sound.

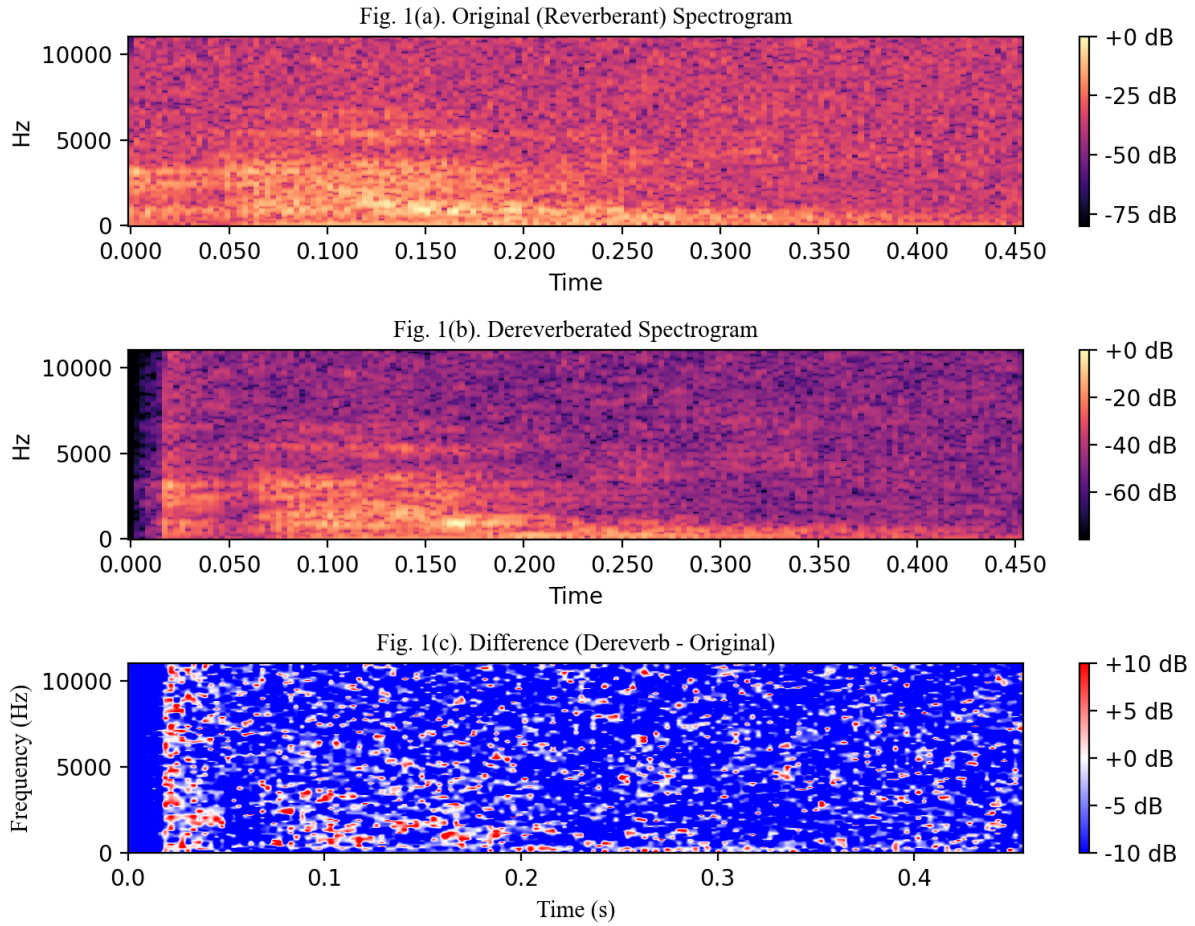


Figure 1 – Spectrogram comparison between original and dereverberated signals. (a) Original (reverberant) spectrogram. (b) Dereverberated spectrogram reconstructed via IFFT and STFT. (c) Difference map showing energy attenuation in dB.

2.2 TDOA estimation

The enhanced signals from the two microphones are processed in the frequency domain using the GCC-PHAT method to compute the cross-correlation function (12). To control the relative influence of the phase and magnitude components of the input signals, eleven scaling parameters uniformly spaced over the interval [0,1] are introduced into the PHAT filter (13).

$$\mathrm{R}(\beta,\tau,k) = \frac{1}{L}\sum_{f=0}^{L-1} \frac{X_1(f,k)X_2^*(f,k)}{|X_1(f,k)X_2^*(f,k)|^{\beta}} e^{j\frac{2\pi f\tau}{L}}.$$

As a result, cross-correlation matrices are generated under different normalization levels, corresponding to various scaling factors. This multi-scale representation enables the CNN model to learn robust time delay patterns, even in environments with significant variations in signal-to-noise ratio (SNR) across frequency bands.

2.3 Deep Learning Model

The deep learning model used in this study follows the CNN-based architecture proposed in (13). The overall structure consists of five convolutional layers, each followed by batch normalization and a ReLU activation function. The resulting feature maps are then flattened and passed through fully connected layers to perform regression-based estimation of the TDOA.

Due to residual noise present in the input audio signals, cross-correlation may lead to incorrect delay estimations. To address this issue, the proposed method uses the cross-correlation matrix not only as the input to the CNN, but also as a probabilistic weight map. Specifically, the cross-correlation values are used to perform element-wise multiplication with the CNN output, enabling the model to generate delay estimates that are adaptively weighted based on reliability. The number of model parameters is significantly reduced compared to (14), while still effectively incorporating cross-correlation information. The overall architecture of the proposed model is illustrated in Figure 2.

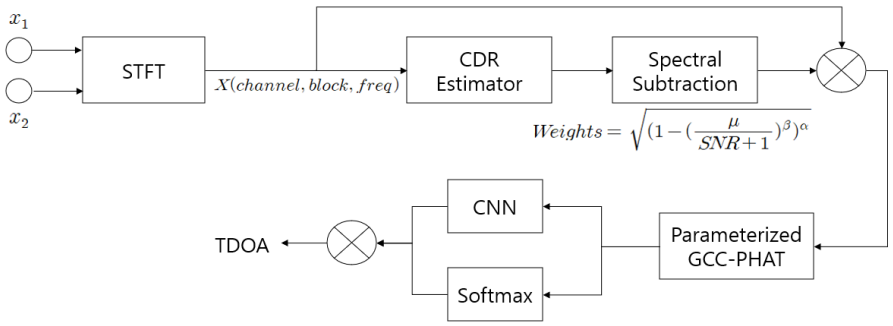


Figure 2 – Schematic diagram of the proposed model architecture.

3. Experiment and Results

3.1 Datasets

To precisely control and analyze the levels of noise and reverberation, this study simulated indoor acoustic environments for training using Pyroomacoustics (15). Speech signals were collected from the LibriSpeech dataset (16), with data from 36 speakers used for training, and data from 3 speakers each for validation and testing. Out of a total of 2,620 audio files, 2,152 were assigned to the training set, 214 to the validation set, and 254 to the test set. From each file, 2-second speech segments were extracted, and 8,000 segments were randomly sampled and used as input for model training.

3.2 Experimental Setup

The training simulation was conducted in a $7 \times 5 \times 3$m room, with microphones placed in the center of the room at $r_1 = [3.5,2.25,1.5]$m, $r_2 = [3.5,2.75,1.5]$m. This configuration allows for a maximum time delay of 23 samples. For each training instance, the source position was randomly sampled within the room following a uniform distribution, following the setup adopted from (17).

Testing was performed in a different room with dimensions $6 \times 4 \times 2.5$m, with microphones placed at $r_1 = [3.0,1.75,1.25]$m, $r_2 = [3.0,2.25,1.25]$m. During training, reverberation time (T60) and signal-to-noise ratio (SNR) were randomly sampled from uniform distributions in the ranges [0.2, 1.0]s and [0, 30]dB, respectively.

3.3 Results

The model was evaluated using three metrics, which are MAE, RMSE, and ACC. Accuracy is defined as the probability that the absolute error between the estimated and ground truth time delay is less than a predefined threshold, that is, $ACC = P(|\tau - \hat{\tau}| < \mu)$, where $\mu = 10$cm. Each recording was evaluated under various acoustic conditions, with SNRs of {0, 6, 12, 18, 24, 30}dB and T60 of {0.2, 0.4, 0.6, 0.8, 1.0}seconds. The evaluation results can be found in Figure 3, which summarizes the performance of the models across different acoustic conditions.

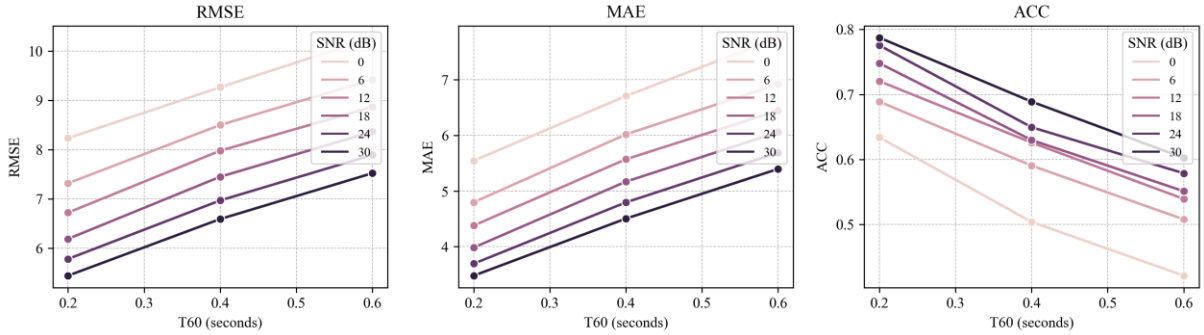


Figure 3 – Evaluation of TDOA estimation performance in terms of RMSE, MAE, and ACC in samples.

For comparison, the Parameterized GCC-PHAT (PGCC) model [13] and the CDR-based masking with CNN-enhanced GCC-PHAT (CDR-masked PGCC) model [14] were used as baselines. Under favorable conditions (SNR = 30dB, T60 = 0.2s), the proposed model achieved a Mean Squared Error (MSE) loss of 0.16, which is significantly lower than the baseline values of 0.88 and 0.44, respectively. Even under the most challenging scenario (SNR = 0dB, T60 = 0.6s), the proposed model maintained strong performance with a MSE loss of 0.31, outperforming the baselines (0.90 and 0.58). In addition, the model achieved a high overall

accuracy of 0.78, demonstrating robust prediction capabilities across varying acoustic conditions. The comparative results are illustrated in Figure 4.

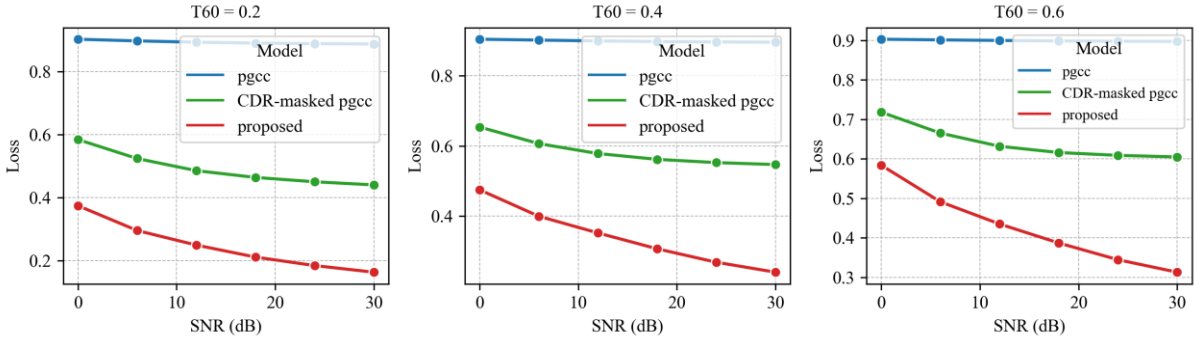


Figure 4 – Comparison of MSE loss per sample across different SNR levels

Furthermore, by incorporating a probabilistic weighting at the final CNN layer, the proposed model demonstrates a strong balance between model complexity and performance. It achieves this level of accuracy with only 682K trainable parameters, which is significantly fewer than those of typical deep neural network architectures.

This compact design makes the model particularly well-suited for real-time inference and deployment in resource-constrained environments, while still achieving comparable or even superior performance to heavier variants such as CDR-masked PGCC (11.5M parameters) under both clean and noisy acoustics conditions.

4. CONCLUSIONS

In this study, we proposed a lightweight deep learning-based framework for robust and accurate TDOA estimation in complex acoustic environments characterized by noise and reverberation. The proposed system first performs speech enhancement using a CDR estimator to suppress diffuse noise components based on spatial coherence. The enhanced signals are then processed using a GCC method with a parameterized PHAT filter to generate multi-scale correlation features.

These features are fed into a CNN model, where the cross-correlation vectors also serve as probabilistic weights applied to the output. This mechanism allows the model to reflect confidence levels in its predictions, effectively integrating the interpretability of signal processing with the expressive power of deep learning.

Experimental results demonstrate that the proposed model consistently outperforms existing baselines across all tested acoustic conditions, achieving accurate time delay estimation regardless of varying levels of noise and reverberation. The overall accuracy reached 0.78, confirming its robustness and reliability in diverse real-world scenarios.

In addition to its effectiveness, the model is lightweight, comprising only 682K training parameters. This compact design enables real-time inference and efficient deployment in resource-constrained environments without compromising accuracy. Taken together, the proposed approach offers a practical and interpretable solution for time delay estimation in real-world sound source localization tasks.

ACKNOWLEDGEMENTS

This work was supported by the Technology Innovation Program (20025005, Development of abnormal sound-voice detection system for victims at accident sites) funded By the Ministry of the Interior and Safety (MOIS), Korea and partly by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT), Korea (RS-2024-00355759, Factory facility anomaly detection and localization technology based on multi-noise removal).

REFERENCES

1. Smith, Bethany R., et al. "Acoustic localisation of wildlife with low-cost equipment: lower sensitivity, but no loss of precision." Wildlife Research 49.4 (2021): 372-381.
2. Ocker, Christof, et al. "Localization of wind turbine noise using a microphone array in wind tunnel measurements." Wind Energy 25.1 (2022): 149-167.
3. Subramanian, Aswin Shanmugam, et al. "Deep learning based multi-source localization with source splitting and its effectiveness in multi-talker speech recognition." Computer Speech & Language 75 (2022): 101360.
4. Padilla-Ortiz, Ana L., and David Ibarra. "Lung and heart sounds analysis: state-of-the-art and future

trends." Critical Reviews™ in Biomedical Engineering 46.1 (2018).
5. Jalayer, Reza, et al. "ConvLSTM-based Sound Source Localization in a manufacturing workplace." Computers & Industrial Engineering 192 (2024): 110213.
6. Wang, Han, et al. "3-D sound source localization with a ternary microphone array based on TDOA-ILD algorithm." IEEE Sensors Journal 22.20 (2022): 19826-19834.
7. Birnie, Lachlan I., Thushara D. Abhayapala, and Prasanga N. Samarasinghe. "Reflection assisted sound source localization through a harmonic domain music framework." IEEE/ACM Transactions on Audio, Speech, and Language Processing 28 (2019): 279-293.
8. Lai, Wei-Ting, et al. "Source Localization by Multidimensional Steered Response Power Mapping with Sparse Bayesian Learning." 2024 18th International Workshop on Acoustic Signal Enhancement (IWAENC). IEEE, 2024.
9. Zhao, Yunjie, et al. "Three-dimensional grid-free sound source localization method based on deep learning." Applied Acoustics 227 (2025): 110261.
10. Shams, Mahmoud Y., Tarek Abd El-Hafeez, and Esraa Hassan. "Acoustic data detection in large-scale emergency vehicle sirens and road noise dataset." Expert Systems with Applications 249 (2024): 123608.
11. Schwarz, Andreas, and Walter Kellermann. "Coherent-to-diffuse power ratio estimation for dereverberation." IEEE/ACM Transactions on Audio, Speech, and Language Processing 23.6 (2015): 1006-1018.
12. Azaria, Mordechai, and David Hertz. "Time delay estimation by generalized cross correlation methods." IEEE Transactions on Acoustics, Speech, and Signal Processing 32.2 (1984): 280-285.
13. Salvati, Daniele, Carlo Drioli, and Gian Luca Foresti. "Time Delay Estimation for Speaker Localization Using CNN-Based Parametrized GCC-PHAT Features." Interspeech. 2021.
14. Kim, Eojin, et al. "Robust Sound Source Localization Using CDR-Based Masking and CNN-Enhanced GCC-PRAT in Reverberant Environments." 2025 International Conference on Electronics, Information, and Communication (ICEIC). IEEE, 2025.
15. Scheibler, Robin, Eric Bezzam, and Ivan Dokmanić. "Pyroomacoustics: A python package for audio room simulation and array processing algorithms." 2018 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, 2018.
16. Panayotov, Vassil, et al. "Librispeech: an asr corpus based on public domain audio books." 2015 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, 2015.
17. Berg, Axel, et al. "Extending gcc-phat using shift equivariant neural networks." arXiv preprint arXiv:2208.04654 (2022).

PROCEEDINGS of the
16th International Conference on Theoretical and Computational Acoustics

August 18 to 22, 2025 BEXCO, Busan Korea

Optimized Microphone Array Configurations for TDOA-Based Direction Finding

Jae-hyoun Ha[1]

[1] Agency for Defense Development

ABSTRACT

Microphone arrays are widely used to estimate the direction of arrival (DOA) of sound sources. When the array aperture is limited and the incoming wavefronts are overlapped, time difference of arrival (TDOA)-based methods provide an effective approach for direction finding. In three-sensor configurations, the sensors are restricted to a single plane, limiting the performance of DOA estimation in three dimensions. In contrast, four-sensor arrays allow for fully three-dimensional configurations, such as the tetrahedral geometry, which enables improved DOA estimation performance. In this study, we analyze array configurations using the Cramér–Rao Lower Bound (CRLB) and present an optimal design for direction finding based on both three- and four-sensor geometries

[1] hahyoun@hanmail.net

PROCEEDINGS of the
16th International Conference on Theoretical and Computational Acoustics

August 18 to 22, 2025 BEXCO, Busan Korea

Low-Latency Android-Based Deployment of Neural Speech Enhancement for Hearing Aid Applications

Jin Tae Seok[1]; Hyeon Jae Shin[2]; Seon Man Kim*
School of Computing and Artificial Intelligence, Hanshin University, South Korea

ABSTRACT

This study presents the design, training, and real-time implementation of a deep learning-based speech enhancement model tailored for hearing aid applications within an Android-based low-latency audio streaming environment. The proposed neural network employs a 32-channel discrete Fourier transform (DFT) filterbank and complex-valued convolutional layers, achieving an algorithmic processing delay of 8 ms, thereby satisfying the strict latency requirements of hearing aids. To ensure total system latency remains below 10 ms, limitations of traditional Android SDK and NDK-based OpenSL ES frameworks were analyzed. In response, a real-time audio streaming pipeline was developed using Google's OBOE library. Experimental validation on a Galaxy S24 Ultra device demonstrated that the OBOE-based implementation significantly outperforms previous approaches in both latency and perceptual audio quality. The deep learning model, originally developed in PyTorch, is integrated with the Android Neural Processing Unit (NPU) to enable efficient real-time inference. This work confirms the feasibility of commercializing a smart hearing assistance system capable of on-device, low-latency speech enhancement.

Keywords: Real-time, Low-Latency, On-device, Enhancement

1. INTRODUCTION

Due to recent changes in living environments, the problem of hearing loss among modern people is increasing across all groups. Accordingly, research on smart hearing assistive devices that support listening through real-time sound enhancement is becoming active. Although deep learning-based speech enhancement technology shows good performance, hearing aid applications require that the overall round-trip latency be kept below 10ms. Moreover, implementing such systems on mobile platforms while satisfying this latency constraint remains a significant challenge.

However, most existing state-of-the-art speech enhancement models fail to meet this strict 10 ms latency requirement, which is essential for real-time hearing aid applications. For example, as seen in Table 1, widely used real-time-capable models such as RNNoise, PercepNet, and NSNet2 exhibit latencies of approximately 20 ms, making them unsuitable for seamless, natural conversations [1][2][3]. Even the more advanced DeepFilterNet family (including DeepFilterNet2 and DeepFilterNet3) achieves excellent enhancement quality with PESQ scores above 3.0 and STOI over 94%, but still suffers from around 40 ms latency, which exceeds practical limits for real-time use [4][5]. This clearly shows that despite significant

Table 1 - Comparison of Models by PESQ and STOI Scores on the VCTK/DEMAND Test Dataset

Model	Param. Size	PESQ	STOI	latency
RNNoise [1]	0.06 M	2.29	92.74%	20 ms
PercepNet [2]	8 M	2.73	93.43%	40 ms
NSNet2 [3]	6.17 M	2.47	92.2%	32 ms
DeepFilterNet [4]	1.8 M	2.81	94.2%	40 ms
DeepFilterNet2 [5]	1.8 M	3.08	94.3%	40 ms
DeepFilterNet3 [5]	-	3.17	94.4%	40 ms

[1] sjt2765@hs.ac.kr
[2] guswotls@hs.ac.kr
*smkim@hs.ac.kr

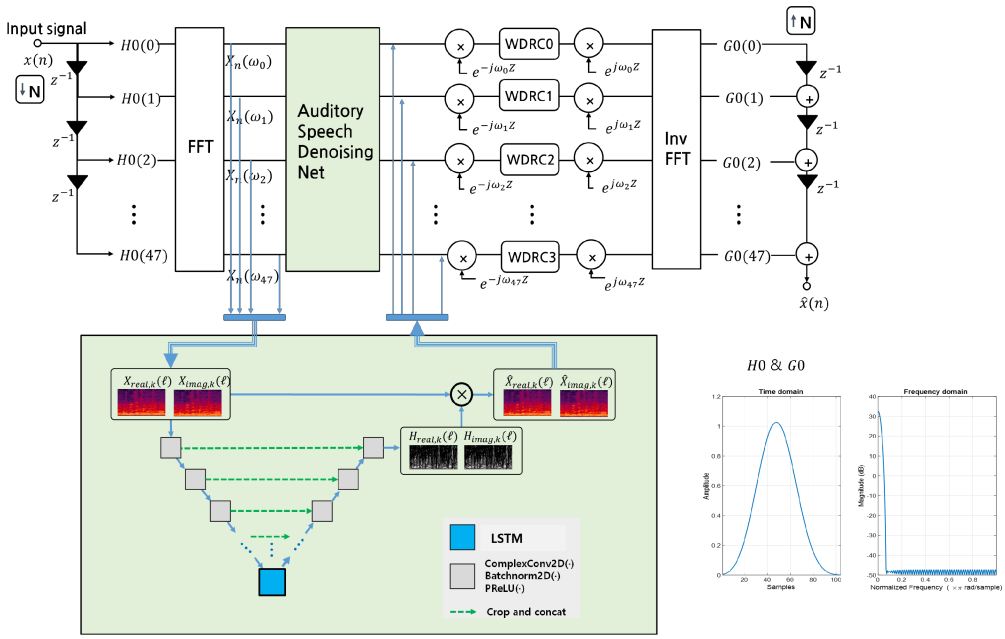


Figure 1 - Previously Proposed Neural Network Architectures for Hearing Enhancement

advances in enhancement quality, existing models fail to meet the critical 10 ms latency requirement necessary for practical and effective hearing aid deployment.

Previous research achieved an algorithmic processing latency of 8 ms, substantially lower than the 20-40 ms observed in existing models, with comparable or slightly reduced enhancement quality [6]. However, to maintain the overall system latency at around 10 ms, the hardware-level audio I/O path must be kept within 2-3 ms. To deploy the proposed low-latency model on an actual mobile platform while meeting these stringent latency requirements, we selected Android smartphones equipped with NPUs, specifically the Galaxy S24 Ultra, as the target platform.

Android devices are widely available, easy to develop for, and ship with high-performance CPUs and NPUs, making them the most straightforward environment for deploying deep learning-based hearing aid models. In this study, we propose a low-latency hearing aid application that controls hardware audio I/O path latency to 2-3 ms using Google's OBOE library (a C++-based native library) and employs Android's NNAPI to utilize the NPU for operation on modern smartphones such as the Galaxy S24 Ultra.

2. Related work of Deep Neural Networks for Hearing Enhancement

The speech enhancement model proposed in prior research was specifically designed for real-time applications in hearing assistance devices, with its core architecture comprising a complex-valued neural network-based denoising model (DCCRN) and a 32-channel uniform Discrete Fourier Transform (DFT) filterbank, as illustrated in Figure 1 [6]. This filterbank decomposes the input audio into 32 distinct frequency bands along the analysis path, allowing for fine-grained extraction of band-specific features. Each band is then processed by complex-valued neural network modules that perform denoising while preserving phase information to enable accurate reconstruction of the clean speech signal.

The model's bottleneck layers consist of two stacked complex LSTM layers with 128 units each, designed to maintain the temporal context of speech and improve the naturalness of the enhanced output. The training loss function calculates a weighted average of a primary SI-SNR–based loss on the full-band reconstructed signal and auxiliary SI-SNR–based losses between the restored and target signals for each band. This multi-loss design aims not only to enhance overall denoising performance but also to address issues of imbalance between frequency bands during restoration [6].

Furthermore, the entire system was optimized to achieve an algorithmic latency of under 8 ms, meeting the strict low-latency requirements for hearing aid use. Experimental evaluations on real test data demonstrated that the model outperformed traditional approaches in objective quality metrics such as PESQ, STOI, and Covl [6]. However, despite these advances in algorithmic design and performance, the model has not yet been implemented for real-time operation on mobile platforms such as Android. Crucially, issues related to platform-specific audio I/O path latency and hardware optimization remain unaddressed, highlighting the need for further research to enable practical deployment in hearing assistive devices.

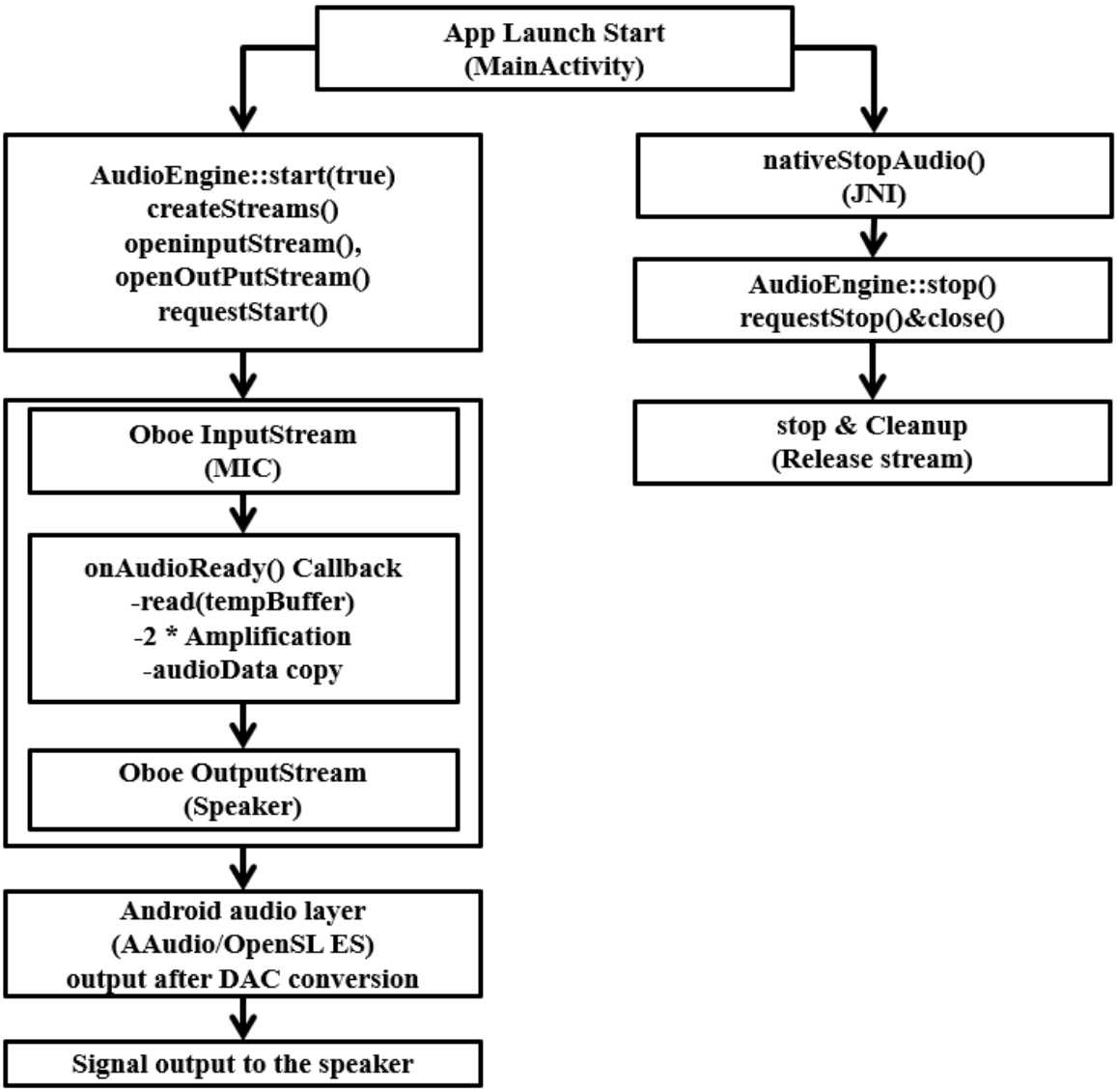


Figure 2 - Audio Input-Output Processing Pipeline Diagram using OBOE Library

3. Audio Input-Output Pipeline Latency Measurement on Android

This study aimed to evaluate and compare audio input-output pipeline latency on the Android platform to meet the stringent requirement of keeping overall system latency under 10 ms for hearing aid applications. Experiments were conducted on a Samsung Galaxy S24 Ultra running Android 14, equipped with a Snapdragon 8 Gen 3 SoC, Adreno 750 GPU, and Hexagon v75 NPU. To systematically analyze latency differences, we implemented three audio streaming paths, Android SDK, NDK (OpenSL ES), and Google's OBOE (C++ Native), each configured as a simple audio bypass from microphone input to speaker output [7][8]. Latency was measured using a loopback approach. A 500 Hz sine wave was generated on-device and played through the speaker while simultaneously being recorded via the device's microphone. The time difference between the generated and recorded signal peaks was analyzed to estimate round-trip audio latency.

All paths used consistent settings, including a standard sampling rate of 48 kHz and a frames-per-burst size of 96 frames. Additional tests enforced a downsampled condition of 16 kHz with the same buffer size to examine resampling overhead.

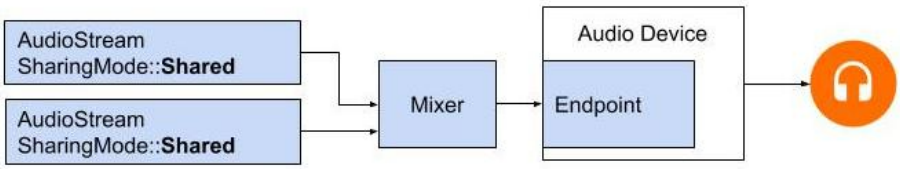


Figure 3 - OBOE SharingMode Flow : Shared

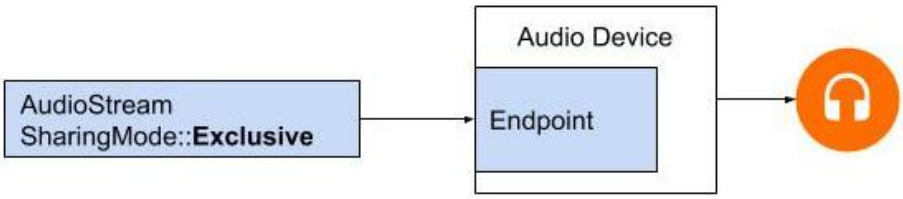


Figure 4 - OBOE SharingMode Flow : Exclusive

4. Android Audio Pipeline Latency Results

Loopback testing revealed that the Android SDK (Java) pathway exhibited approximately 20–30 ms round-trip latency, with clearly perceptible echo during listening tests (as shown in Figure 5). This latency level far exceeds the acceptable threshold for hearing aid applications, largely due to the overhead introduced by the high-level application framework.

The OpenSL ES pathway significantly reduced round-trip latency to approximately 4–7 ms (as shown in Figure 6). However, this came at the cost of degraded audio quality. Listening tests identified noticeably high-frequency losses and increased background noise, which are critical issues for hearing assistance. These artifacts are likely caused by inconsistent resampling behavior, buffer underrun risks, and variability in device-specific HAL implementations. When forced to operate at 16 kHz, additional resampling overhead led to slight latency increases within the same 4–7 ms range.

In contrast, the OBOE-based implementation achieved round-trip latencies of approximately 2–3 ms using PerformanceMode=LowLatency and SharingMode=Exclusive settings [9]. As shown in Figure 3 and Figure 4, Shared mode routes audio through the system mixer (adding mixing overhead), whereas Exclusive mode connects the AudioStream directly to the endpoint, bypassing the mixer and minimizing latency [10]. Tests conducted at 48 kHz with 96-frame bursts showed consistently low latency without perceptible artifacts such as echo or noise, yielding superior audio quality. Even under forced 16 kHz conditions, while slight overhead was observed, OBOE maintained significantly better latency and audio consistency compared to SDK and OpenSL ES.

As shown in Table 2, these results demonstrate that OBOE effectively overcomes the inherent limitations of standard Android audio APIs. Specifically, the OBOE-based pipeline can achieve hardware-level I/O latencies below 3 ms, making it a viable solution to meet the combined system latency budget of approximately 10 ms when integrated with a neural speech enhancement model designed for ~8 ms algorithmic latency.

Table 2 - Audio Loopback Test Results Across Android SDK, OpenSL ES, OBOE (48 kHz, 96 Frames)

Type	latency
SDK	20-30 ms
OpenSL ES	4-7 ms
OBOE	2-3 ms

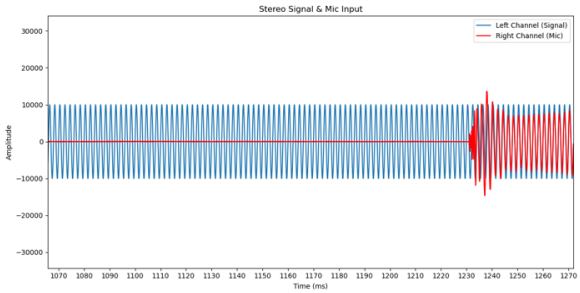


Figure 5 - Audio Loopback Test Results Android SDK (48 kHz, 96 Frames)

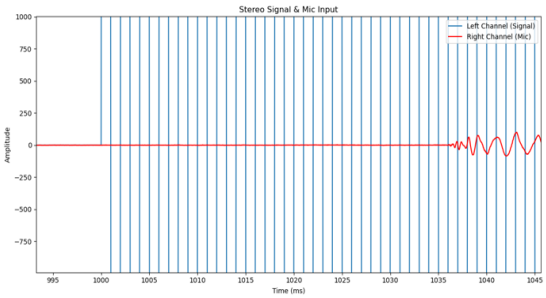


Figure 6 -Audio Loopback Test Results OpenSL ES (48 kHz, 96 Frames)

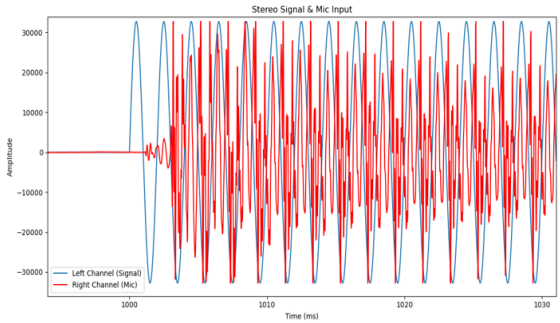


Figure 7 - Audio Loopback Test Results OBOE (48 kHz, 96 Frames)

5. CONCLUSIONS

This study investigated the practical feasibility of achieving ultra-low-latency audio input–output processing for hearing aid applications on modern platforms.

First, loopback testing on an Android device demonstrated that traditional SDK (Java) audio paths incurred approximately 20–30 ms round-trip latency, with clearly perceptible echo that exceeds the sub-10 ms requirement essential for real-time hearing aid use. The NDK (OpenSL ES) pathway reduced latency to around 4–7 ms but introduced audio artifacts, including high-frequency loss and increased background noise, attributed to inconsistent resampling and buffer management issues. In contrast, the OBOE-based implementation achieved approximately 2–3 ms round-trip latency at 48 kHz without perceptible artifacts, confirming its suitability for high-quality, low-latency audio streaming on Android. These findings demonstrate that state-of-the-art audio pipelines like OBOE can provide the necessary low-level I/O performance around 3 ms latency, enabling systems that meet the critical 10 ms end-to-end budget for hearing aid applications.

Future work will focus on leveraging on-device acceleration frameworks such as NNAPI and vendor-specific NPUs on Android. This will enable efficient, power-conscious, real-time inference directly on smartphones, making practical, high-quality, low-latency hearing assistance broadly accessible on mobile platforms.

ACKNOWLEDGEMENTS

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (NRF-2022R1A2C2010614).

REFERENCES

1. VALIN, Jean-Marc. A hybrid DSP/deep learning approach to real-time full-band speech enhancement. In:2018 IEEE 20th international workshop on multimedia signal processing (MMSP). IEEE, 2018. p. 1-5.
2. VALIN, Jean-Marc, et al. A perceptually-motivated approach for low-complexity, real-time enhancement of fullband speech.arXiv preprint arXiv:2008.04259, 2020.
3. SCHRÖTER, Hendrik, et al. Deepfilternet2: Towards real-time speech enhancement on embedded devices for full-band audio. In:2022 International Workshop on Acoustic Signal Enhancement (IWAENC).IEEE, 2022. p. 1-5.
3. SCHROTER, Hendrik, et al. DeepFilterNet: A low complexity speech enhancement framework for full-band audio based on deep filtering. In:ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2022. p. 7407-7411
4. SCHRÖTER, Hendrik, et al. Deepfilternet: Perceptually motivated real-time speech enhancement.arXiv preprint arXiv:2305.08227, 2023.
5. KIM, Seon Man. Auditory filterbank denoising neural network for speech enhancement in wearable auditory device.Electronics Letters, 2024, 60.10: e13228.
6. https://developer.android.com/ndk/guides/audio/opensl/opensl-for-android?hl=ko
7. https://github.com/google/oboe
8. chomium.googlesource.com/external/github.com/google/oboe/+/refs/heads/1.1-stable/docs/FullGuide.md

Low-Latency Deep Learning-Based Denoising of Self-Generated Footstep Noise in Quadruped Robots for Remote Acoustic Situational Awareness

Hyeon Jae Shin[1]; Jin Tae Seok[2]; So Yeong Park[3]; Seon Man Kim[*]
School of Computing and Artificial Intelligence, Hanshin University, South Korea

ABSTRACT

In disaster response scenarios, quadruped robots are increasingly deployed to support remote situational awareness by transmitting real-time audiovisual data from hazardous environments. However, their onboard microphones often capture high-amplitude, periodic footstep noise generated by the robot itself, which severely degrades the intelligibility of critical ambient sounds. To address this, this research focuses on developing a low-latency, deep learning-based denoising algorithm suitable for on-robot deployment. The proposed method modifies the Fair-Denoiser architecture, a powerful U-Net based model, to optimize for real-time performance and computational efficiency. We introduce targeted architectural changes and a novel loss function to enhance performance. The model was trained and evaluated on the Valentini benchmark dataset to establish its general-purpose denoising capabilities under controlled conditions. Evaluation results demonstrate that our proposed model surpasses strong baselines like the original Fair-Denoiser and CleanUNet, achieving a PESQ score of over 3.0 and a STOI score of over 0.94. This work validates the model's suitability for real-time applications, establishing a critical foundation for future deployment and fine-tuning on specialized robotic noise datasets.

Keywords: Denoising, Quadruped Robot, Robot Noise

1. INTRODUCTION

The deployment of quadruped robots in disaster response and hazardous environment monitoring has marked a significant advancement in remote situational awareness. These platforms can navigate complex and dangerous terrains inaccessible to human first responders, providing critical audiovisual data to remote operators. While visual information is essential, real-time audio transmission is equally paramount, as it can carry determinative cues such as the voices of trapped survivors, calls for assistance, or the sounds of shifting debris indicating structural instability. The clarity of this acoustic channel can directly influence the speed and efficacy of rescue operations.

However, a fundamental challenge arises from the robot's own operation. The mechanical action of the robot, particularly the high-amplitude, periodic impact of its footsteps and the whine of its motors, generates significant "ego-noise." This self-generated noise is captured by the onboard microphones and can easily mask the subtle, yet vital, ambient sounds that are the objective of the acoustic monitoring mission. The structured and impulsive nature of this footstep noise presents a unique and difficult challenge for traditional denoising algorithms.

The ultimate goal of this research is to develop a comprehensive acoustic situational awareness system that can intelligently separate critical environmental sounds from the robot's own ego-noise, as illustrated in Figure 1. In a typical disaster scenario, the raw audio captured by the robot is a complex mixture of target sounds (e.g., emergency speech) and noise (e.g., robot footsteps and motor whine). To provide a clear audio feed to the remote operator, a multi-stage process is envisioned: first, the dominant, structured ego-noise from footsteps and motors must be removed. Following this, a sound source separation module can distinguish between enhanced speech and other crucial environmental sounds. This paper focuses on the first and most

[1] sjt2765@hs.ac.kr
[2] guswotls123@gmail.com
[3] kong02931@gmail.com
* kobem30002@gmail.com

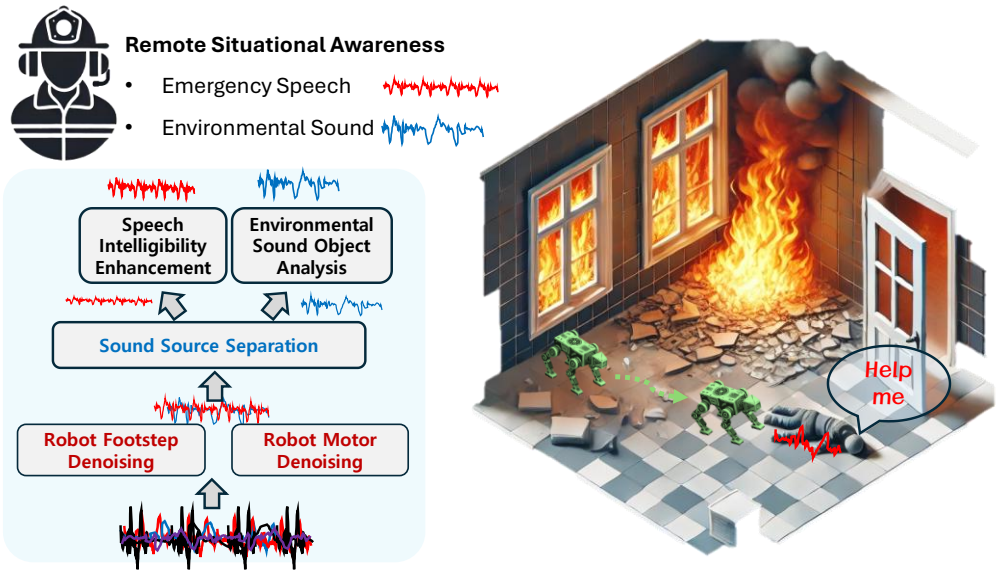


Figure 1 – The overall concept for achieving remote acoustic situational awareness. The system first denoises the robot's self-generated noise (footsteps, motor) from the raw audio signal before separating critical sounds like emergency speech.

critical stage of this pipeline: the effective and real-time removal of robot footstep noise. Addressing this primary source of contamination is a necessary prerequisite for any subsequent analysis and is therefore the central justification for the present work.

For any on-robot processing system intended to support a human operator, latency is a non-negotiable constraint. To be effective in a dynamic disaster scenario, the audio denoising process must occur in real-time. This project operates under a strict processing delay requirement, rendering many computationally intensive, state-of-the-art denoising models unsuitable for deployment. This necessitates a focused effort on developing lightweight and computationally efficient model architectures. While the ultimate goal is to suppress robotic noise, the initial phase of this research focuses on creating a model with superior denoising performance and real-time capabilities, validated on a standardized public dataset. A specialized dataset of robotic footstep noise is currently under construction and was not used in this phase.

In recent years, deep learning-based speech enhancement has been dominated by end-to-end models that operate directly on the raw audio waveform [1]. Architectures based on the U-Net structure, such as Fair-Denoiser [2] and CleanUNet [3], have demonstrated remarkable performance. This paper presents a novel denoising model specifically tailored for real-time performance. Our primary contribution is a modified U-Net architecture, derived from the Fair-Denoiser model, which has been systematically optimized for both performance and computational efficiency. We detail architectural changes and introduce a novel loss function, demonstrating superior performance against established baselines on the Valentini benchmark dataset [4].

2. The Fair-Denoiser Base Architecture

The foundation of our proposed model is the Fair-Denoiser, a causal, real-time speech enhancement model that operates directly on the raw audio waveform [2]. This model is an adaptation of the DEMUCS architecture, a U-Net-like convolutional structure originally developed for music source separation, and its end-to-end, waveform-to-waveform approach makes it well-suited for high-fidelity audio processing tasks [5].

The model is structured as a fully convolutional encoder-decoder network featuring a recurrent bottleneck and U-Net-style skip connections [2]. The process begins with the encoder network, which is composed of a series of L layers designed to progressively downsample the input waveform and extract hierarchical feature representations. Each encoder layer consists of a strided one-dimensional convolution (Conv1D) with kernel size K and stride S, followed by a Rectified Linear Unit (ReLU) activation. This is then passed through a 1x1 convolution and a Gated Linear Unit (GLU), which acts as a learned gating mechanism to selectively pass or block information. At the deepest level of the architecture, the compressed feature representation is processed by a 2-layer Long Short-Term Memory (LSTM) network, which serves as the model's bottleneck [2]. The

Table 1 – Architectural comparison of base and modified models

Parameter	Base Model (Fair-Denoiser)	Proposed Model (Mod 1)	Lightweight Alternative (Mod 2)
Layer Depth	8	8	5
CNN Bias	Local	Global	Global
CNN Kernel Size & Stride	8 & 4	8 & 4	4 & 2
Encoder Activation	GLU	GLU	GLU
Decoder Activation	GLU	ReLU	ReLU
Max Channels	768	768	576
Resampling	Enabled	Enabled	Disabled
Loss Function	Multi-resolution STFT	Multi-dynamic range & Multi-res STFT	Multi-dynamic range & Multi-res STFT

LSTM is critical for capturing the long-range temporal dependencies and contextual information within the audio signal. The decoder network then mirrors the encoder's structure, tasked with upsampling the feature representation to reconstruct the clean audio waveform using transposed convolutions to increase the temporal resolution. To ensure a high-fidelity output, skip connections—a hallmark of the U-Net architecture—link the output of each encoder layer directly to the input of its corresponding decoder layer [5]. This is crucial for preserving fine details in the audio waveform and preventing the information loss that can occur in deep networks.

The original Fair-Denoiser model is trained using a composite loss function that operates in both the time and frequency domains. The total loss is a weighted sum of an L1 loss on the raw waveform and a multi-resolution Short-Time Fourier Transform (STFT) loss [2]. This hybrid approach ensures that the model not only matches the waveform sample-by-sample but also preserves the spectral structure of the speech, which is highly correlated with human perception of audio quality.

3. Proposed Modified Denoiser for Real-Time Performance

The primary design objective was to enhance denoising performance and efficiency for real-time applications. This goal guided a series of deliberate architectural changes to the Fair-Denoiser baseline, as summarized in Table 1. Our final proposed model, "Modification 1," focuses on improving performance through targeted changes that do not alter the core network complexity. For applications where latency and model size are the absolute highest priorities, we also present "Modification 2" as a more aggressive optimization.

3.1 Proposed Model (Modification 1)

Our final proposed model focuses on improving performance and efficiency through targeted changes that do not alter the core complexity of the network. One such change is the replacement of the Gated Linear Unit (GLU) with the standard Rectified Linear Unit (ReLU) in the decoder layers. While GLUs offer a more powerful, learned gating mechanism, this capability comes at a computational cost; by switching to the faster ReLU activation, we gain a tangible increase in execution speed.

A second key innovation is the introduction of the "Multi-dynamic range & Multi-resolution STFT" (MDRR-STFT) loss function. This loss addresses a key limitation of standard STFT-based losses, which tend to be dominated by high-energy spectral components, causing the model to neglect quieter but perceptually important details like unvoiced consonants. The MDRR-STFT loss mitigates this by partitioning the spectrogram of each frame into multiple dynamic range bands (e.g., high, mid, low energy) based on their relative decibel levels. The loss is then calculated independently for each band i as a sum of spectral convergence (L_{sc}) and log-magnitude (L_{mag}) losses :

$$L_{\text{STFT}}^{(i)} = L_{\text{sc}}^{(i)} + L_{\text{mag}}^{(i)} \tag{1}$$

By computing the loss across these separate energy bands, the model is forced to learn from both high-energy and low-energy speech components, preventing weaker signals from being overlooked during optimization and leading to improved overall intelligibility and preservation of fine spectral details.

Table 2 – Denoising performance comparison on the Valentini test set

Model	Parameters, M	PESQ (WB)	STOI,%
Noisy	-	1.94	92.7
FAIR-denoiser	73.70	2.93	94.7
CleanUNet	180.0	2.99	94.9
Proposed Model (Mod 1)	73.70	3.04	94.9
Proposed Model (Mod 2)	**31.0**	2.89	94.6

3.2 Lightweight Alternative (Modification 2)

For applications where latency and model size are the absolute highest priorities, we propose a second, more lightweight model that incorporates all the changes from Modification 1 and adds further complexity reductions. To minimize the computational load, we reduced the number of encoder-decoder layers from 8 to 5, the maximum channel count from 768 to 576, and the convolutional kernel size from 8 to 4, with the stride correspondingly halved from 4 to 2. These changes collectively yield a substantial reduction in the model's total parameter count and computational demands. Furthermore, the computationally intensive pre/post-processing resampling step is disabled, eliminating this overhead entirely and further reducing end-to-end latency.

4. EXPERIMENTAL SETUP

To validate our proposed models, we conducted a comprehensive experiment on the Valentini dataset from the Voice Bank corpus, a benchmark designed to evaluate performance in domain-mismatched scenarios [4]. All audio, originally sampled at 48 kHz, was resampled to 16 kHz. The training set utilized 28 speakers and consisted of 11,572 clean/noisy pairs mixed with 10 different noise types at SNR levels of 0, 5, 10, and 15 dB [4]. The test set, ensuring variability, used 2 different speakers and comprised 824 pairs mixed with 5 distinct noise types at SNRs of 2.5, 7.5, 12.5, and 17.5 dB. For training, 4.5-second clips were randomly extracted and augmented using Remix, Band-mask, and Revecho techniques. The models were trained for a maximum of 400 epochs with a batch size of 64 and an early-stopping parameter of 300, using the Adam optimizer (β_1 =0.9, β_2 ==0.999) on six NVIDIA A40 GPUs. A linear warmup with a cosine annealing learning rate schedule was employed, with a maximum learning rate of 2×10−4 and a 5% warmup ratio, optimizing a full-band short-term Fourier transform loss. Performance was quantitatively assessed using two industry-standard metrics: Perceptual Evaluation of Speech Quality (PESQ) to measure perceptual quality and Short-Time Objective Intelligibility (STOI) to measure speech intelligibility [6].

5. RESULTS AND DISCUSSION

The effectiveness of our proposed modifications was evaluated against the base Fair-Denoiser and the state-of-the-art CleanUNet model, with the results summarized in Table 2 demonstrating the success of our approach. Our final proposed model, "Modification 1," achieved the highest performance across all tested configurations, delivering a PESQ score of 3.03537 and a STOI of 0.94891, surpassing both the original Fair-Denoiser and the strong CleanUNet baseline. This result empirically validates the effectiveness of our architectural changes, particularly the new multi-dynamic range loss function and the switch to ReLU in the decoder, which together provide superior denoising quality without increasing model complexity.

In contrast, the "Lightweight Alternative (Modification 2)" underwent aggressive complexity reductions, leading to an expected slight decrease in the PESQ score to 2.88561. Crucially, however, its STOI score remained exceptionally high at 0.94606, indicating almost no degradation in speech intelligibility. This outcome represents a successful and intentional engineering trade-off, as the substantial reduction in architectural complexity is precisely what enables the model to achieve minimal latency, making it a compelling option for deployment on hardware with extreme resource constraints. Therefore, the choice between Modification 1 and 2 depends on the specific application's priority: maximum perceptual quality versus minimum latency.

6. CONCLUSIONS

This paper presented the design and evaluation of a deep learning-based denoising algorithm optimized for real-time performance, with the ultimate goal of deployment on resource-constrained robotic platforms. Through targeted modifications to the Fair-Denoiser architecture, specifically by introducing a novel multi-dynamic range loss function and simplifying decoder activations, we developed a model that outperforms strong baselines on the standardized Valentini benchmark, achieving a superior balance of performance and efficiency with high PESQ and STOI scores. Additionally, we presented a lightweight alternative model that prioritizes minimal latency and model size through significant complexity reductions. While this alternative exhibits a minor drop in perceptual quality, it maintains excellent speech intelligibility, making it a viable candidate for the most demanding real-time applications. This research successfully validates our architectural modifications and establishes a robust, high-performing, and efficient denoising model. The technology developed provides a critical foundation for enhancing the acoustic situational awareness of robots, and future work will focus on deploying this model onto a physical quadruped robot and fine-tuning it using a custom-built dataset of robotic ego-noise to transition from a general-purpose denoiser to a specialized system for real-world operational environments.

ACKNOWLEDGEMENTS

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (NRF-2022R1A2C2010614).

REFERENCES

1. Benesty J, Makino S, Chen J. Speech Enhancement. New York, USA: Springer; 2005.
2. Défossez A, Synnaeve G, Adi Y. Real Time Speech Enhancement in the Waveform Domain. Proc Interspeech 2020; 25-29 October 2020; Shanghai, China. 2020. p. 3291-5.
3. Kong Z, Ping W, Dantrey A, Catanzaro B. Speech Denoising in the Waveform Domain with Self-Attention. Proc ICASSP 2022; 22-27 May 2022; Singapore. 2022. p. 7772-6.
4. Stoller D, Ewert S, Dixon S. Wave-U-Net: A Multi-Scale Neural Network for End-to-End Audio Source Separation. Proc 19th International Society for Music Information Retrieval Conference; 23-27 September 2018; Paris, France. 2018. p. 334-40.
5. Valentini-Botinhao C, Wang X, Takaki S, Yamagishi J. Speech Enhancement for a Noise-Robust Text-to-Speech Synthesis System Using Deep Recurrent Neural Networks. Proc Interspeech 2016; 8-12 September 2016; San Francisco, USA. 2016. p. 352-6.
6. Taal CH, Hendriks RC, Heusdens R, Jensen J. An algorithm for intelligibility prediction of time-frequency weighted noisy speech. IEEE Transactions on Audio, Speech, and Language Processing. 2011;19(7):2125-36.

PROCEEDINGS of the
16th International Conference on Theoretical and Computational Acoustics

August 18 to 22, 2025 BEXCO, Busan Korea

Study on railway rolling sound calculation model

Hyo-In Koh[1], Anders Nordborg[2]
[1] Korea Railroad Research Institute
[2] Sound View Instruments

ABSTRACT

This study presents a model for the calculation of railway rail/wheel rolling noise, formulated in both the frequency and time domains. The rail-wheel interaction is computed in the time domain, enabling the inclusion of nonlinear effects and parametric excitation. Additionally, the model further accounts for noise radiation from the rail. In this paper, the results of calculations on the factors are presented that generate rolling sound when a train wheel rolls, an example of calculating the sound power radiated from the train pass-by is also introduced and discussed.

[1] hikoh@krri.re.kr
[2] anders.nordborg@soundview.se

An automatic target extraction algorithm for active acoustic detection based on image processing

Fan Yin[1]; Yiming Gu [1]; Haibin Wang [1]; Chao Li [1]; Leixin Nie [1]; Hao Yin [1]
[1] State Key Laboratory of Acoustics and Marine, Institute of Acoustics, Chinese Academy of Sciences, Beijing, China

ABSTRACT
In the field of active underwater acoustic detection, target detection results are typically represented as two-dimensional time-delay versus angle images. These images often contain a mixture of reverberation, clutter, and actual targets of interest, making accurate target detection a challenging task. Traditional detection algorithms frequently struggle with this issue, as they tend to misclassify clutter or reverberation as valid targets, or conversely, fail to detect interested targets, leading to degraded detection performance. To address these challenges, This paper improves the yolov11 framework to dual-head architecture to enable more effective extraction of targets of interest, which can autonomously detects and marks target within the detection images. By effectively distinguishing actual targets from clutter and reverberation, the algorithm enhances detection accuracy and reliability, and the Precesion(B) value for this experiment is 0.429. Furthermore, it provides precise target position and range (time-delay) information, which is essential for subsequent tracking tasks. The improved method has been rigorously tested using various of simulated data, demonstrating robust and consistent performance across varying underwater environments. These results highlight the algorithm's potential for improving the effectiveness of active underwater acoustic detection systems in practical applications, ultimately contributing to more reliable underwater monitoring.

Keywords: Active Acoustic Detection, Underwater Target Detection, YOLO

1. INTRODUCTION

In the field of active acoustic detection, target echoes are typically represented as bright spot echoes on a time-delay versus azimuth two-dimensional image. To accurately detect targets and return the position of the targets of interest, the algorithm needs to simultaneously perform a fusion of both target detection and target recognition tasks.

Traditional signal processing methods typically employ a two-stage processing approach. In the first stage, target detection is performed using Constant False Alarm Rate Detector (CFAR) [1] class algorithms to determine the presence of targets in the image. In the second stage, clustering algorithms, such as Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [2], are used for target classification and identification. Similar two-stage detection algorithms also exist in machine learning, such as the Deformable Parts Model (DPM) [3] and R-CNN [4] algorithms. To further reduce detection time and improve accuracy, one-stage detection algorithms like YOLO [5] have emerged and been widely applied in both academic and industrial fields. YOLO-type algorithms efficiently and rapidly detect targets in images, mark their locations, and calculate the confidence of the target being real, thus providing a basis for subsequent decision-making processes.

In 2016, Ali Farhadi et al. [5] introduced a one-stage natural image detection algorithm—YOLO, and after years of iteration and optimization, it has evolved into YOLOv11[6]. This algorithm has achieved significant improvements in both target detection and labeling accuracy. The contribution of this paper is the application of the YOLOv11 network to detect interest target echoes in active underwater acoustic detection, and its performance is compared with traditional two-stage signal processing algorithms, demonstrating the practical feasibility of the method. Furthermore, this paper improves the multi-scale detection heads of YOLO and establishes a small, non-public dataset with simulated data for active underwater acoustic detection.

[1]yinfan@mail.ioa.ac.cn

2. RELATED WORK

In one-stage target detection algorithms, the YOLO series is renowned for its high accuracy and low latency, making it widely used in various natural image detection tasks. YOLOv1 introduced the first end-to-end one-stage target detection framework and brought new ideas to real-time target detection. Subsequently, the YOLOv2 and YOLOv3 versions [7, 8] were released, further optimizing algorithm performance. YOLOv4 [9] made significant improvements to the algorithm's framework. Although it retained the one-stage detection concept, the architecture was divided into three main components: Backbone, Head, and Neck. The Backbone is responsible for feature extraction, typically composed of multiple layers of Convolutional Neural Networks (CNNs) for cross-layer learning and image feature extraction. The Neck refines and combines these features to generate rich feature representations, which are then used by the Head to make predictions, outputting the target's bounding box and its confidence score.

Since then, the YOLO series has continued to evolve. YOLOv6 [10] focused on industrial applications, enabling high-precision real-time outputs even on low-end hardware. YOLOv11 aims to improve computational efficiency without sacrificing detection accuracy. It introduces new components such as the C3k2 block and C2PSA block, significantly enhancing the algorithm's performance while reducing the model's parameter count.

3. PROPOSED METHOD

To enhance the accuracy and adaptability of active target echo detection algorithms, this paper introduces the YOLOv11 algorithm for data processing and compares its performance with those of traditional two-stage signal processing algorithms. A private, custom-built dataset was used in this study, which includes various type of simulated data.

3.1 Traditional algorithm

As shown in Figure 1, traditional echo detection algorithms typically use the CFAR algorithm for target detection and apply a clustering algorithm, such as DBSCAN, for classification, as used in this paper. Both the CFAR and clustering stages require parameter tuning. For different simulated underwater environments, such as deep-sea, shallow-sea, varying seabed geology, and different frequencies of transmitted signals, specific detection parameters need to be applied. Therefore, the effectiveness of the algorithm in practical applications relies on certain prior knowledge.

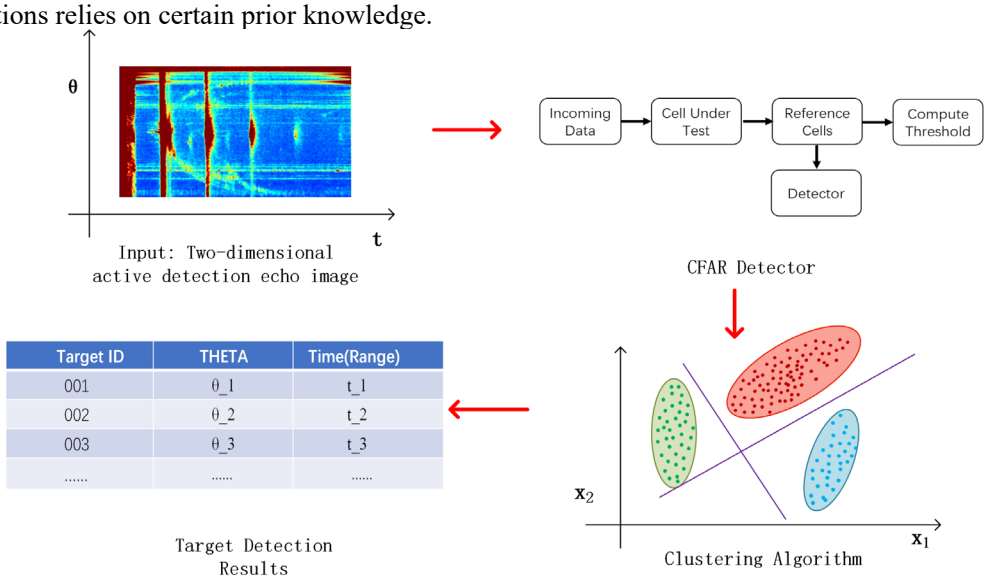


Figure 1 – Traditional Algorithm Framework

In traditional methods, the focus is typically on detecting target echoes. After CFAR and clustering operations, the detection results may be influenced by large areas of reverberation, which are commonly present in active sonar data, leading to false alarm detections. To reduce the probability of false alarms, a clustering area constraint is often applied. However, the parameters for this area constraint need to be adjusted for different scenarios, which poses a challenge for traditional two-stage signal processing algorithms. These algorithms struggle to adapt automatically to various echo scenarios, making them less effective in real-world applications.

3.2 Improved YOLOv11 Dual-Head Architecture

As shown in Figure 2, this is the complete framework of the improved YOLOv11. Based on the understanding of the active target echo detection task mentioned before, this paper posits that the detection of reverberation areas under the target is not critical, while more attention should be given to detecting large clutter areas and small target echoes of interest. Therefore, the original three-detection-head framework of YOLOv11 was modified to a two-detection-head design, specifically removing the detection head designed for large objects (which is typically associated with the lowest resolution feature map, downsampled by a factor of 32). In terms of algorithm execution, what is most important is not what to do, but what to avoid doing. The modified dual-detection-head YOLO framework can more effectively focus on the detection and differentiation of target echoes and clutter, while minimizing excessive attention to reverberation areas.

Considering the network's real-time requirements, this paper chose the nano version of YOLOv11 as the network framework due to its lightweight nature. During training, no pre-trained parameters were used; instead, the model was trained on the custom dataset from scratch.

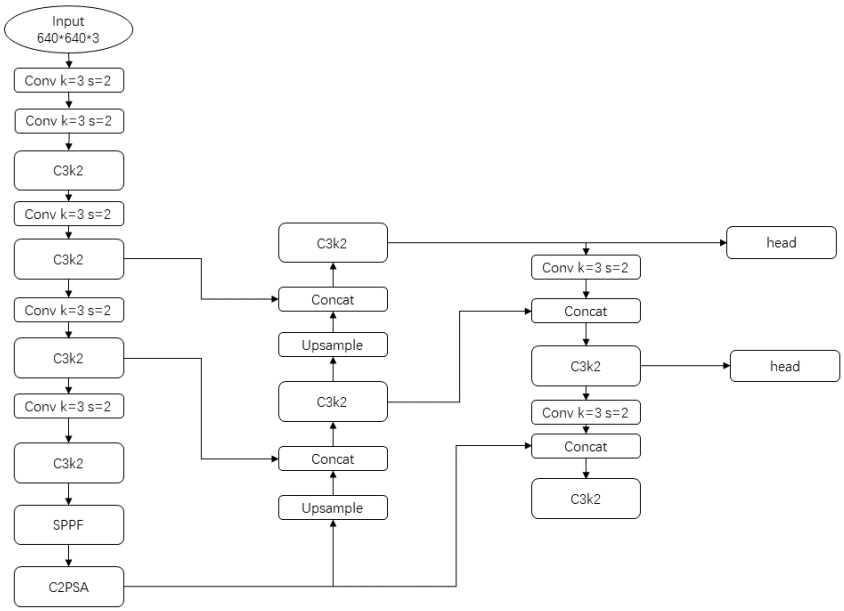


Figure 2 – Improved YOLOv11 Dual-Head Architecture Framework

As shown in Figure 2, C3k2 is a feature extraction and fusion module that divides the input feature map into two parts: one part is passed directly, while the other part undergoes a series of convolutions and Bottleneck modules. Finally, the two feature maps are concatenated. Here, k2 indicates that the convolutional layer within this module uses two kernels with a kernel size of k. The C2PSA (Cross-Stage Local Spatial Attention) module enhances the spatial attention within the feature map, improving the model's focus on important parts of the image. Through spatial pooling features, the model can more effectively concentrate on specific regions of interest. The SPPF (Spatial Pyramid Pooling Fast) module enhances the model's multi-scale perception ability by performing pooling operations at multiple scales.

In Figure 2, after the feature map is processed by the C3k2 module, it undergoes an upsampling operation to increase its size, then concatenates with feature maps from the backbone network or other shallower layers. This process aims to fuse features from different layers, especially passing semantic information from deeper layers to shallower ones to enhance the representation capability of shallow features. Through multiple upsampling and concatenation operations, the model constructs feature maps with varying resolutions. Feature maps with higher resolution (i.e., larger size) have smaller receptive fields, making them more suitable for detecting small objects in the image. In contrast, feature maps with lower resolution (i.e., smaller size) have larger receptive fields, making them better suited for detecting larger objects in the image.

On the right side of the figure, the outputs from different feature layers are shown. Some feature maps are directly connected to the head through C3k2, while others are downsampled via a convolution with a stride of 2 (Conv k=3, s=2), followed by concatenation and C3k2 processing before being passed to the head. For instance, the topmost head typically receives the feature map with the highest resolution, which has been fused after multiple upsampling operations. These high-resolution feature maps contain more detailed information and have smaller receptive fields, making them more suitable for detecting small targets in the image.

4. EXPERIMENTS

4.1 Dataset and evaluation metrics

In this experiment, we used a custom dataset that contains 150 images of simulated active underwater detection results. The dataset includes 120 training samples and 30 test samples. The goal is to validate the proposed approach for this task and provide a reference for future research. In the test set, there are a total of 290 reverberation instances, 22 clutter instances, and 246 target instances. We will evaluate the algorithm's performance after training. To prevent overfitting and increase the diversity of the data, we mixed various detection scenarios in the experiment. The received echo data includes signals with durations of 64 seconds and 32 seconds, as well as detection signals at different frequencies, such as high, medium, and low frequencies. This diversity aims to process the relatively small dataset in as many different ways as possible.

To evaluate the performance of the target detection model, we used three key metrics that assess the model's ability to recognize objects in images and accurately localize them: Precision measures the proportion of correctly predicted positive samples out of all samples predicted as positive by the model. In target detection, precision indicates how many of the detected objects are correct. A higher precision means that the model has a stronger ability to avoid false positives, i.e., most of its detection results are correct with fewer false alarms. Recall measures the proportion of true positive samples that are correctly detected by the model. In target detection, recall indicates how many of the true objects in the image are successfully "recalled" or "discovered" by the model. A higher recall means the model has a stronger ability to avoid missing detections, i.e., it can detect more true targets in the image with fewer missed detections. mAP (Mean Average Precision) is the most commonly used and important comprehensive evaluation metric in target detection. It measures the precision at different recall levels and averages the results across all classes. A higher mAP50 (B) value means that the model can provide relatively accurate predictions while maintaining a high recall rate for most classes, with the predicted bounding boxes meeting the basic requirements of Intersection over Union (IoU > 0.5).

The training process in this study directly uses active detection echo images for input. The labels for the input images were manually annotated, relying on expert knowledge to distinguish between echoes, clutter interference and reverberation in the images.

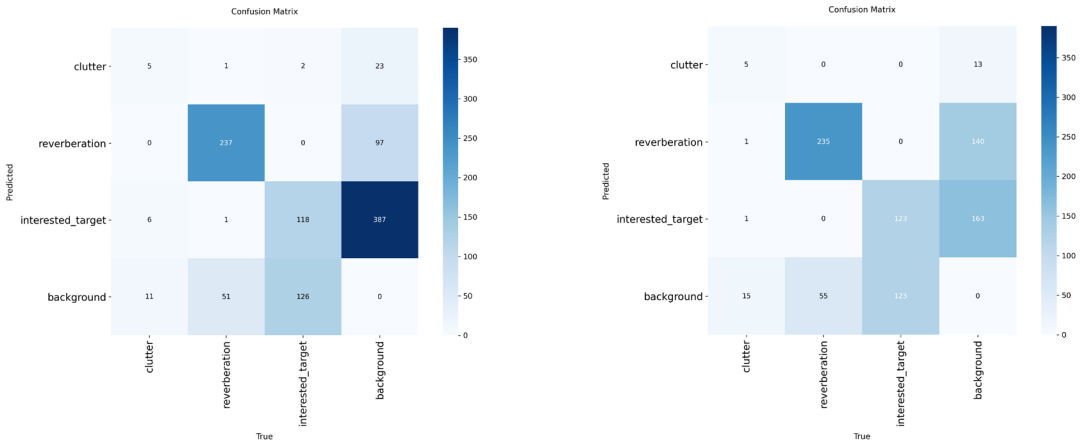


Figure 3 – Comparison for detection results (left: original, right: improved)

Table 1 – The performance of related methods on the custom test dataset.

Method	Precesion(B)	Recall(B)	mAP50(B)
Traditional	0.326	0.402	0.311
Original-yolov11	0.230	0.479	0.328
Proposed method	**0.429**	**0.500**	**0.388**

Figure 3 shows the confusion matrix of the detection results. As seen in the figure, in the original network architecture, a large amount of background and clutter was incorrectly identified as target of interest, leading to a high false alarm rate. In contrast, with the proposed method, the number of background instances misclassified as targets was reduced to 40% of the original, and the number of clutter instances misclassified as targets of interest was reduced to 16% of the original. Additionally, the proposed method also showed an improvement in recall.

As shown in Table 1, the improved method achieved the highest scores across all three evaluation

metrics. It is worth noting that for Precision (B), the original network architecture even scored lower than traditional signal processing algorithms.

4.2 Results and discussion

In the domain of active echo detection, we first implement the real-time capable YOLOv11 network. Using a proprietary database, we have demonstrated its superior detection performance compared to conventional two-stage signal processing techniques. A key advantage of our approach is its inference consistency; the trained network parameters can be deployed without requiring data-specific recalibration.

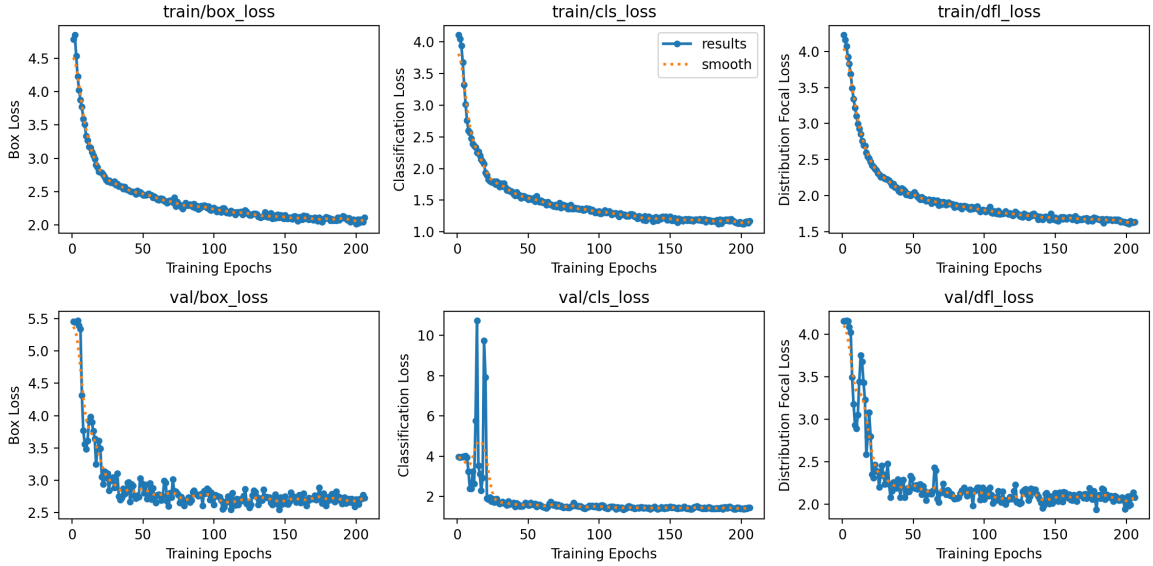


Figure 4 – training process for proposed method

Next, informed by an analysis of echo and reverberation patterns, we adapted the multi-scale, triple-detector head architecture of YOLOv11. We enhanced the network's focus on relevant signals by retaining only the two smaller detector heads. This modification yields a higher success rate in identifying target echoes and substantially lowers the incidence of background reverberation being mistaken for a target. Consequently, the overall false alarm rate is reduced to a level viable for engineering applications. The complete training and testing progression, spanning 210 epochs, is depicted in Figure 4. The data shows a sharp initial decrease in all loss functions, followed by a smooth convergence, while performance metrics consistently trended upward before stabilizing. The optimizer used during training is AdamW, with an initial learning rate of 0.01 and a weight decay of 0.0005.

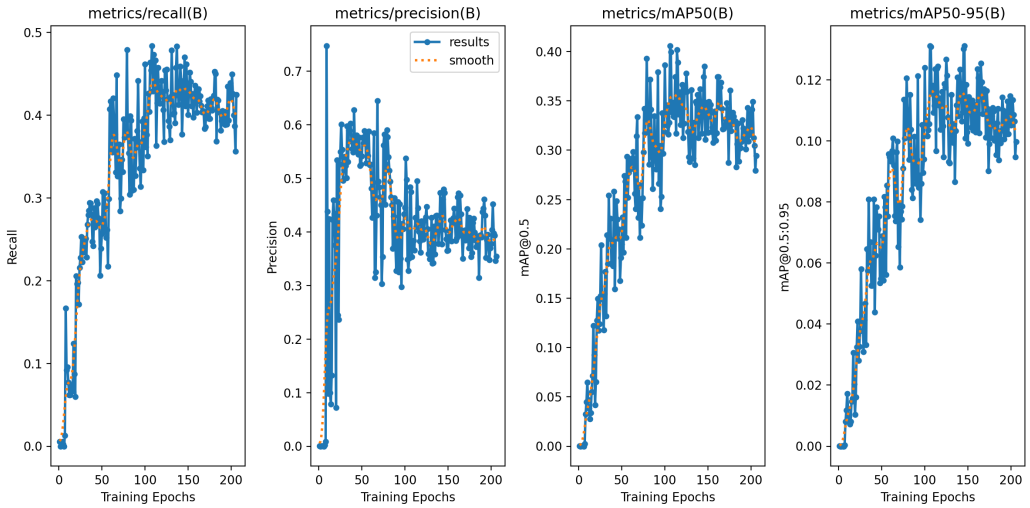


Figure 5 – results for proposed method

As shown in Figure 5, this set of figures illustrates the trends of four key performance metrics of the proposed method over 210 training epochs. From left to right, these metrics are Recall, Precision, Mean Average Precision at an IoU threshold of 0.5 (mAP@0.5), and Mean Average Precision averaged over IoU thresholds from 0.5 to 0.95 with a step size of 0.05 (mAP@0.5:0.95). The Recall and mAP@0.5 curves exhibit rapid initial growth, stabilizing after approximately 100 epochs, indicating significant improvement

in target identification and overall performance under a lenient IoU threshold. Precision shows greater fluctuations but generally stabilizes. The mAP@0.5:0.95, being a more stringent metric, demonstrates a slower but positive upward trend, suggesting that while the model improves in precise localization, there is still further room for enhancement.

Lastly, to further enhance the model's robustness, our future efforts will focus on expanding the dataset with a wider variety of echo scenarios. We also identify multi-frame joint processing as a critical path to greater detection accuracy. This method allows the algorithm to leverage more prior information, improving its decision-making. For instance, while the confidence of an echo exhibiting sudden morphological or positional changes in a single frame may be downgraded, the confidence for a target with uniform motion across consecutive frames would be boosted. This multi-frame approach will also provide significant performance gains in low signal-to-noise ratio (SNR) conditions.

5. CONCLUSIONS

We have constructed a custom dataset of simulated active underwater target echoes, classifying the targets into three distinct categories: reverberation, clutter, and targets of interest. By introducing a single-stage, real-time detection algorithm into the active echo processing domain, our experimental results clearly demonstrate its superior performance over traditional methods, especially in terms of target detection accuracy capabilities. Furthermore, to address the inherent limitations of conventional algorithms when dealing with large, continuous reverberation targets, we improved the algorithm furtherly by removing the detection head designed for large objects. This enhancement enables it to effectively disregard such extended targets, thereby directing its focus more intensely towards the detection of highlight echoes, which has led to a significant improvement in detection efficacy. Our overarching research goal is to deliver a practical detection algorithm for active sensing applications that offers both excellent real-time performance and superior operational effectiveness. It is clear that the approach presented in this paper possesses strong practical viability, and future work will involve further optimization and the promotion of its real-world deployment.

REFERENCES

1. Rohling H. Radar CFAR thresholding in clutter and multiple target situations[J]. IEEE transactions on aerospace and electronic systems, 2007 (4): 608-621.
2. Birant D, Kut A. ST-DBSCAN: An algorithm for clustering spatial–temporal data[J]. Data & knowledge engineering, 2007, 60(1): 208-221.
3. Felzenszwalb P, McAllester D, Ramanan D. A discriminatively trained, multiscale, deformable part model[C]//2008 IEEE conference on computer vision and pattern recognition. Ieee, 2008: 1-8.
4. Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 580-587.
5. Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 779-788.
6. Glenn J, and Jing Q. Ultralytics YOLO11, 2024
7. Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 7263-7271.
8. Redmon J, Farhadi A. Yolov3: An incremental improvement[J]. arXiv preprint arXiv:1804.02767, 2018.
9. Bochkovskiy A, Wang C Y, Liao H Y M. Yolov4: Optimal speed and accuracy of object detection[J]. arXiv preprint arXiv:2004.10934, 2020.
10. Li C, Li L, Jiang H, et al. YOLOv6: A single-stage object detection framework for industrial applications[J]. arXiv preprint arXiv:2209.02976, 2022.

PROCEEDINGS of the
16th International Conference on Theoretical and Computational Acoustics

August 18 to 22, 2025 BEXCO, Busan Korea

Passive source localization by two hydrophones in direct arrival zone and shadow zone

Kang Zheng[1], Jixing Qin[2], Shuanglin Wu[3]

[1] State Key Laboratory of Acoustics, Institute of Acoustics, Chinese Academy of Sciences

[2] State Key Laboratory of Acoustics, Institute of Acoustics, Chinese Academy of Sciences

[3] State Key Laboratory of Acoustics, Institute of Acoustics, Chinese Academy of Sciences

ABSTRACT

A broadband source localization method based on two kinds of multipath time delays using dual vertical hydrophones suitable for both direct arrival zone (DAZ) and shadow zone (SZ) in deep water is presented. We analyze the multipath time delays in the DAZ and SZ for sound source near the surface, which are related to source depth and range. Thus, source position can be estimated using these time delays. However, there is currently a problem of source position ambiguity in the previous localization methods, especially misjudgment of sound zones. Our approach addresses the above issues from two perspectives. On the one hand, two types of time delays are used for localization to avoid the ambiguous position results caused by using a single time delay. On the other hand, we use dual hydrophones to overcome the problem of misjudgment in sound zones. Four positioning results can be obtained by locating the signals received by two vertical hydrophones using time delays calculated in the DAZ and SZ, and the average of the two results with the highest similarity is the accurate source position. The simulation results and the experimental data indicate that the method has good localization performance and robustness.

[1] zhengkang@mail.ioa.ac.cn

[2] qjx@mail.ioa.ac.cn

[3] wushuanglin@mail.ioa.ac.cn

PROCEEDINGS of the
16th International Conference on Theoretical and Computational Acoustics

August 18 to 22, 2025 BEXCO, Busan Korea

Deep-sea active detection method based on vertical phased array

Naibin CHEN[1,2]; Dayong PENG[1]; Li MA[1]; Tao HU[1]
[1] Key Laboratory of Underwater Acoustic Environment, Institute of Acoustics, Chinese Academy of Sciences, Beijing 100190, China
[2] College of Physical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China

ABSTRACT

When the target is in the deep-sea acoustic shadow zone, the target echo signal energy received by traditional monostatic active sonar is low, making it difficult to detect targets. Phased transmission technology is used to effectively avoid the deep-sea acoustic shadow zone effect at the transmitting end. This paper presents a method for detecting targets that avoids the deep-sea acoustic shadow zone effect from both the transmitting end and the receiving end. At the transmitting end, active beamforming technology is applied to focus acoustic energy at specified angle parameters, and the deep-sea acoustic shadow zone is "illuminated" through the bottom reflection mode to stimulate the target to produce scattered echoes. At the receiving end, the vertical receiving array is deployed at a great depth, and a reliable acoustic path is utilized to receive the scattered echo of the target. Finally, the detection of targets in the deep-sea acoustic shadow zone is achieved by processing the echo signal. The verification results based on sea trial data show that the low acoustic energy problem in the deep-sea acoustic shadow zone is effectively addressed from both the transmitting end and the receiving end, significantly improving the deep-sea active target detection capability.

Keywords: Phased emission, Acoustic shadow zone, Target detection, Active sonar

Analysis of Vertical Correlation Characteristics of Deep-Sea Acoustic Fields in Cold Eddy Environments Based on Remote Sensing Data

Pengxi Zhou[123]; Tao Hu[12]; Zhen Wang[12]; Fujin Yang[12]

[1] Institute of Acoustics, Chinese Academy of Sciences, China

[2] Key Laboratory of Underwater Acoustic Environment, Chinese Academy of Sciences, China

[3] University of Chinese Academy of Sciences, China

ABSTRACT

Based on remote sensing data, a cold eddy model is established to analyze the vertical coherence characteristics of deep-sea acoustic fields under the influence of the cold eddy. The ray theory method is used to explain the impact of the arrival time differences of eigenrays and multipath time spreading on the vertical correlation characteristics. The results indicate that periodic variations in the phase differences caused by changes in the dominant eigenray arrival time differences lead to oscillatory behavior in the vertical correlation coefficient with increasing receiver depth. Within the shadow zone, the periodic oscillation pattern of multipath spreading in the dominant acoustic pulses aligns well with the variations in the vertical correlation coefficient. In the convergence zone, the periodic oscillation of received acoustic energy with increasing receiver depth also shows consistency with the changes in the vertical correlation coefficient.

Keywords: Mesoscale eddy, Vertical coherence, Multipath arrivals

PROCEEDINGS of the
16th International Conference on Theoretical and Computational Acoustics

August 18 to 22, 2025 BEXCO, Busan Korea

Analysis of spatial correlation of non complete deep-sea acoustic axis sound field

Zhifei FANG[12]; Xinyi GUO[1];Li MA[1]

1 Key Laboratory of Underwater Acoustic Environment, Institute of Acoustics, Chinese Academy of Sciences, Beijing 100190, China

2 College of Physical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China

ABSTRACT

The propagation of sound signals along the axis of incomplete deep-sea acoustic channels is affected by various propagation conditions such as refraction and reflection, resulting in a certain degree of complexity in the sound field. Studying the sound field characteristics and spatial correlation of sound sources located at depths along the axis of incomplete deep-sea acoustic channels is of great significance for underwater communication, target detection, and other directions. This article focuses on analyzing the characteristics of the signal arrival structure near the middle channel axis with distance in deep-sea incomplete channel propagation experiments, and through theoretical analysis of its generation mechanism. At the same time, the influence of different path sound rays on the depth spatial correlation of the channel axis is studied, and the main channel factors affecting the correlation under different conditions are analyzed and summarized. In this paper, the influence of sound lines with different propagation conditions on the correlation of axis sound field in the deep sea channel is discussed theoretically and experimentally, which provides theoretical support for the analysis of physical characteristics of underwater sound field, the optimization of underwater acoustic communication array, and the improvement of target detection performance.

Keywords: Spatial Correlation; Non-Ideal Deep-Sea Sound Channel; Sound Channel Axis; Arrival Structure; Acoustic Field Structure.

Feasibility of seawater temperature observation using coastal acoustic tomography in Yeosu Bay of South Korea

Wan-Gu KIM; Young Geul YOON; Ho Youn JI; Min-Seok CHOI; Seong Hyeon KIM;
Bok Kyoung CHOI; Byoung-Nam KIM[1]
Korea Institute of Ocean Science & Technology, Busan 49111, Republic of Korea

Siwoo LEE
Underwater Survey Technology 21, Inc., Incheon 21999, Republic of Korea

ABSTRACT
Coastal acoustic tomography techniques can be used to estimate the seawater temperature distribution as well as the sea current field in the area of interest through the transmission and reception of acoustic signals. It has great oceanographic significance as it can replace some of the traditional cumbersome on-site measurement. We have been operating five coastal acoustic tomography stations in Yeosu Bay since 2023 to investigate the validity of ocean temperature observations using the ocean acoustic tomography technique. This study explores the feasibility of seawater temperature observations using ocean acoustic tomography by analyzing tomographic data acquired during the first quarter of 2024. We estimated the sound speed from the pair transmission data for each of the seven lines between the five stations, from which we calculated the seawater temperature. Mackenzie's formula was used to calculate the seawater temperature from the sound speed, with the salinity data measured at the same period as the formula's input data. Comparing this estimated seawater temperature with CTD data measured during the same period shows a good agreement in the trend of temperature change over time. However, regarding absolute magnitude, it showed some difference of up to 4°C. This appears to be an error caused by the fact that the length of the sound ray was calculated by assuming it is straight when performing the tomographic inversion for sound velocity. To compensate for this error, we calculated a sound velocity correction for each line based on data from the first few days of measurements. With this correction, we were able to estimate the seawater temperature in the Yeosu Bay with a fairly high degree of accuracy over the total measurement period of the quarter, the one and a half month.
Keywords: coastal acoustic tomography (CAT), Yeosu Bay, seawater temperature, CTD, length of the sound ray

ACKNOWLEDGEMENTS
This research was supported by Korea Institute of Marine Science & Technology Promotion (KIMST) funded by the Ministry of Oceans and Fisheries (RS-2021-KS211521).

REFERENCES
1. A. Kaneko, X.-H. Zhu, and J. Lin, Coastal acoustic tomography (1st edition, Elsevier, Chennai, 2020).
2. W. Munk, P. Worcester, C. Wunsch, Ocean Acoustic Tomography (Cambridge University Press, New York, 1995).
3. C. Zhang, A. Kaneko, X.-H. Zhu, and N. Gohda, "Tomographic mapping of a coastal upwelling and the associated diurnal internal tides in Hiroshima Bay, Japan," J. Geophys. Res. Oceans **120**, 4288-4305 (2015).

[1] bnkim@kiost.ac.kr

PROCEEDINGS of the
16th International Conference on Theoretical and Computational Acoustics

August 18 to 22, 2025 BEXCO, Busan Korea

Implementation and Sea Trial Validation of an OFDM-based Broadband Underwater Acoustic Communication System for Seismic Data Transmission

Seung-Gyu Kim[1]; Sae-Yong Park[2]; Tae-Ho Im[3]
[1,2,3] Department of Information and Communication Engineering, Hoseo University, Republic of Korea.

Abstract

This study presents an OFDM-based underwater acoustic communication system designed as a practical wireless solution for real-time seismic monitoring. The system is designed to continuously transmit a 4096-bps data stream from a seafloor seismometer to a surface buoy using the 12-28 kHz frequency band. We developed and implemented a complete end-to-end prototype, featuring an optimized 1-second frame structure to handle the continuous data flow. The system's performance was validated through sea trials, where it achieved highly reliable data transmission with a Packet Error Rate (PER) of 0.00387 under realistic channel conditions (100-m range, 35-m depth). Our results demonstrate a robust and viable alternative to traditional cabled or autonomous systems.
Keywords: Sound, Transmission

1. Introduction

Seafloor seismic monitoring plays a pivotal role in early earthquake warnings and tsunami forecasting. Although most global earthquakes occur along oceanic plate boundaries, particularly around the Pacific Ring of Fire where approximately 90% of all earthquakes are concentrated [1], the majority of seismometers are installed on land. This imposes significant limitations on detecting submarine seismic events due to their distance from offshore epicenters and restricted azimuthal coverage [2]. Currently, methods for collecting seafloor seismic data can be broadly categorized into two types. The first is the autonomous Ocean Bottom Seismometer (OBS), which operates independently for 3 to 12 months before being retrieved to download the recorded data [3]. This approach, however, cannot provide real-time monitoring and is thus unsuitable for early warning systems. The second method utilizes cabled systems, such as Japan's S-net and DONET networks, which enable real-time data transmission through submarine cable networks. A major drawback of this approach is that installation costs increase exponentially with distance and water depth [4,5]. To address these challenges and limitations of traditional seafloor seismic monitoring methods, various advanced technologies have been explored. Recently, techniques such as Distributed Acoustic Sensing (DAS) utilizing fiber optic cables are emerging as a promising new direction in marine geosciences and seafloor monitoring [6,7]. In parallel with such developments in the field of seafloor observation technology, this study proposes and implements a robust wireless solution for real-time seismic data transmission based on underwater acoustic communication. We have developed a robust system employing Orthogonal Frequency-Division Multiplexing (OFDM), a technique resilient to the multipath fading and Doppler effects prevalent in underwater environments [8]. Through sea trials conducted in realistic marine environment, we have verified the system's capability for highly reliable seismic data transmission. This paper presents the system design methodology considering underwater channel characteristics, implementation techniques for real-time processing, and the performance evaluation results from our sea trials.

[1] ksg980920@naver.com
[2] dyd2634@naver.com
[3] taehoim@hoseo.edu

2. OFDM System Design and Implementation

2.1 Overall System Architecture

The system implemented in this study consists of a Seafloor Observatory and an Ocean Buoy, as illustrated in Figure 1. Both systems are equipped with identical Underwater Acoustic (UWA) modems, enabling bi-directional communication. This architecture allows the Ocean Buoy to transmit downlink signals—such as control commands, status checks, and time synchronization signals—to the Seafloor Observatory. The primary objective of this research is the real-time transmission of seismic data, for which the system primarily operates in an uplink-centric manner, transmitting data from the seafloor to the surface. The uplink data processing follows a systematic flow. Initially, the Seismic Sensor within the Seafloor Observatory detects seismic activity and generates digital data. Subsequently, this data is transferred to the UWA modem via a wired interface. The modem's embedded Single Board Computer (SBC) serves as the core processing unit, handling the incoming digital seismic data and modulating it into an OFDM signal, which is well-suited for the underwater channel environment. The digital signal processed by the SBC is then converted into an analog waveform by a Digital-to-Analog Converter (DAC). This signal is subsequently filtered and amplified by the Analog Amp/Filter stage to match the transmission frequency band. Finally, the resulting analog acoustic signal is transmitted into the underwater channel through the UWA Sensor. At the Ocean Buoy, the UWA Sensor receives the acoustic signal propagating through the water. The received analog signal passes through an Analog Amp/Filter stage before being converted back into a digital signal by an Analog-to-Digital Converter (ADC). Lastly, the SBC processes this digital signal, performing complex signal processing tasks including synchronization, channel estimation, OFDM demodulation, and error correction to successfully recover the original seismic data. A stable power supply for all components in each unit is provided by a battery.

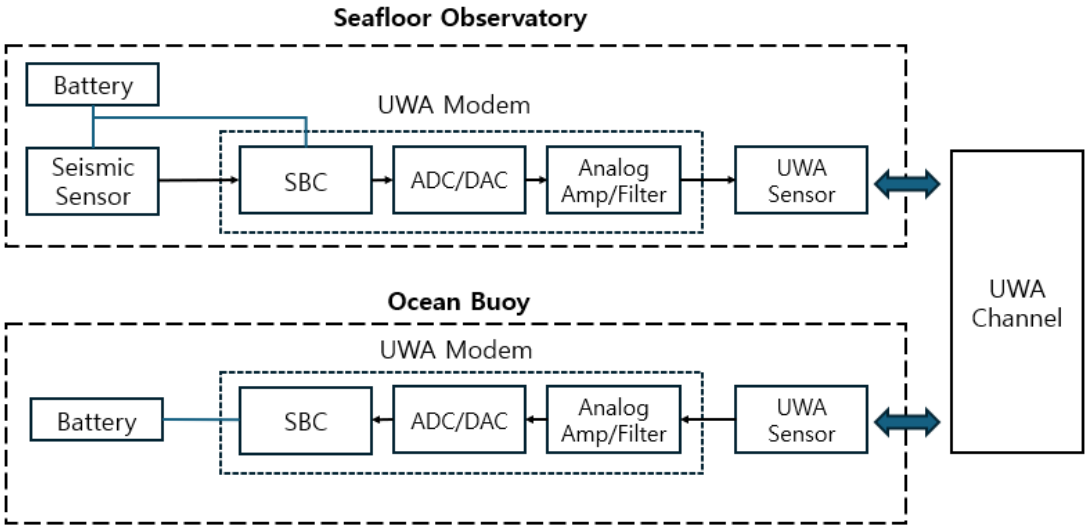
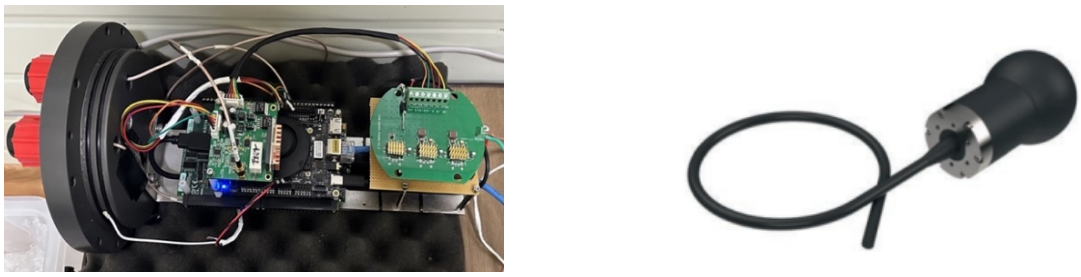


Figure 1. System architecture of the OFDM-based underwater acoustic communication system.

2.2 Hardware and Seismic Data Interface Implementation

The hardware of our system consists mainly of the transceiver modem and the underwater acoustic sensor. The modem is composed of a control unit, an ADC/DAC board, an analog amplifier and filter board, and a power board. Figure 2 shows photographs of the transceiver modem and the underwater acoustic sensor. Figure 2(a) displays the modem, which is responsible for transmitting and receiving signals to and from the sensor. Figure 2(b) shows the acoustic sensor, which serves as the acoustic transducer for signal transmission and reception in the water; for this, a Neptune Sonar MODEL D/26/BB was used [9].



(a) Interior view of the underwater acoustic communication modem

(b) Underwater acoustic sensor for signal transmission and reception

Figure 2 – Photographs of the transceiver modem and the underwater acoustic sensor.

To implement the interface with the seismic sensor, data is acquired via a wired RS232 connection. The seismometer generates data for three axes (X, Y, Z) by sampling at 100 Hz. Each axis produces data at a rate of 512 bytes (4096 bits) per second. In this study, to enhance monitoring efficiency and accommodate communication bandwidth constraints, we selectively acquire and transmit only the data from the vertical axis (Z-axis), which is the most critical of the three. Consequently, the actual amount of data transmitted by the communication system is 512 bytes (4096 bits) per second, sent in 1-second intervals. Figure 3 illustrates the interface and procedures between the seismic sensor and the transceiver modem. The modem decodes the received seismic data to obtain the pure seismic information, which amounts to 512 bytes. Note that the actual received data packet may be larger than 512 bytes due to redundancy added by the interface protocol.

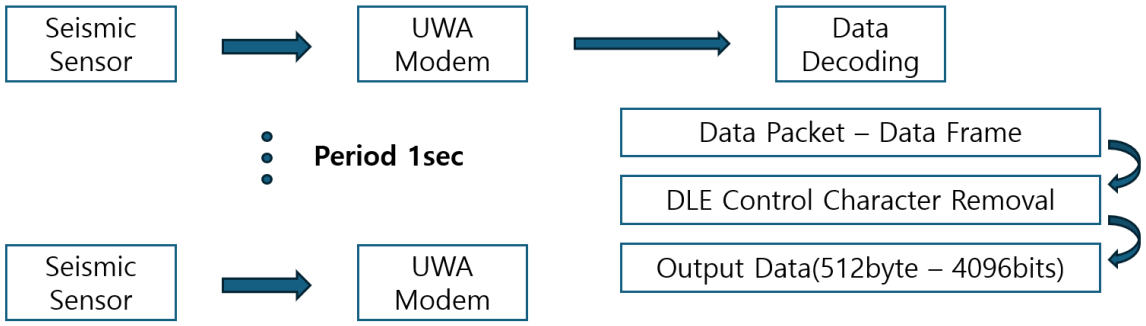


Figure 3 – Data interface between seismic sensor and UWA modem.

2.3 OFDM System Design

The frame structure of the OFDM-based underwater communication system was designed as shown in Figure 4. The total frame length was set to 1 second, consisting of a preamble and the OFDM signal. The preamble is a Hyperbolic Frequency Modulated (HFM) signal with a duration of 0.125 seconds, which plays a crucial role in time synchronization and Doppler shift estimation algorithms [10]. Following the preamble, the OFDM signal has a total duration of 0.875 seconds and is composed of seven OFDM symbols. Each OFDM symbol has a total duration of 0.125 seconds (including CP), consisting of a useful symbol duration of 102.4 milliseconds and a cyclic prefix of 22.6 milliseconds.

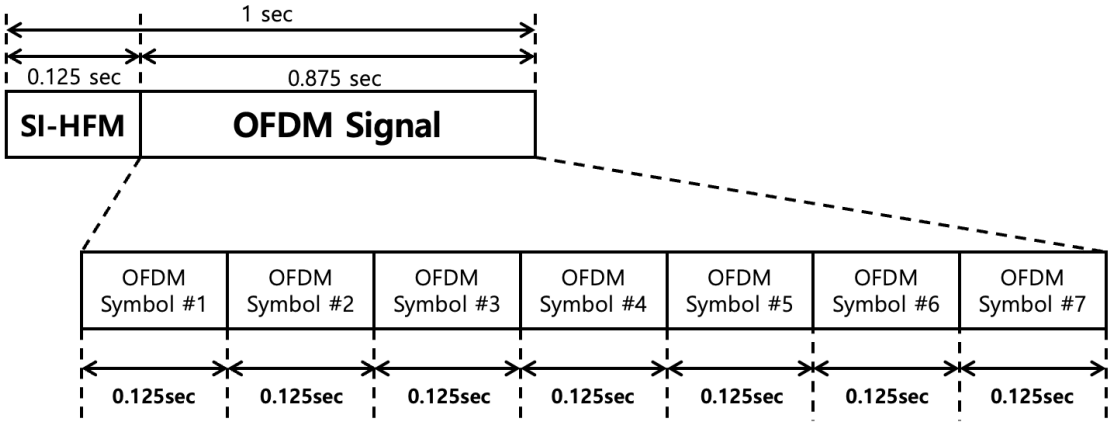


Figure 4 – OFDM frame structure for underwater acoustic communication.

The Cyclic Prefix (CP) length for the OFDM symbols was designed based on the measured RMS delay spread of the underwater channel, following the principle that CP length should exceed the channel delay spread to eliminate inter-symbol interference [11]. Table 1 details the parameters of the OFDM system. The OFDM FFT size was set to 2048, with 1768 effective subcarriers after accounting for guard bands on both sides. This results in a bandwidth of 16 kHz (from 12 kHz to 28 kHz). Each OFDM symbol consists of 884 pilot carriers and 884 data carriers. In total, the 6188 data carriers can carry 12,376 bits using QPSK modulation. For Forward Error Correction (FEC), Turbo coding was employed to make the system robust in the harsh underwater channel environment [12]. With these OFDM parameter settings, the system implemented in this study achieves an information rate of approximately 4.1 kbps, which is sufficient to support the seismometer's data transmission target of 4096 bps.

Table 1 – Detailed parameters of the OFDM system

Parameter Name	Parameter Value
Center Frequency	20kHz
Bandwidth	16kHz [12kHz – 28kHz]
OFDM FFT Size	2048
Subcarrier Spacing	9.7656Hz
CP Length	22.6ms
Useful Symbol Duration	102.4ms
Modulation	QPSK
Coding Rate	1/3 (Turbo)
Pilot Spacing Freq	2
Pilot Spacing Time	1
Sampling Freq	200kHz

3. Experimental Design and Results

3.1 Sea Trial Scenario

Prior to designing the sea trial scenario, we selected the coastal waters of Yeonmyeong, Sanyang-eup, Tongyeong-si, Gyeongsangnam-do, Korea (near Buso breakwater) as the experimental site. The scenario was designed with a horizontal distance of 100 meters between the transmitter and receiver. The transmitter sensor was deployed at a depth of 35 meters, while the receiver sensor was set at a depth of 5 meters. Figure 5(a) shows a photograph of the UWA modem and sensor integrated with the seismometer. Figure 5(b) depicts the UWA sensor being deployed from the breakwater, simulating a surface buoy installation.

(a) Integrated UWA modem and sensor coupled with the seismometer

(b) UWA sensor deployed for reception, simulating a surface buoy installation

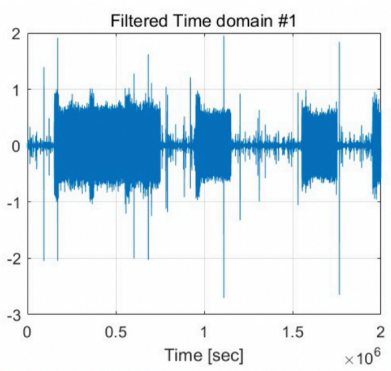
Figure 5 – Transceiver modem setup and sensor deployment for the sea trial.

Figure 6 shows the seafloor observatory system, which integrates the seismometer and the UWA modem, being deployed to a depth of 35 meters at a location 100 meters away from the receiver (at the breakwater). The observatory system was fixed to the seafloor. To characterize the underwater channel environment, we used a Sound Velocity Profiler (SVP) to measure sound speed and temperature profiles with depth. The sound speed at the seafloor (35-m depth) was 1521–1522 m/s, while at the sea surface it was 1520–1520.5 m/s. The water temperature ranged from 20–21°C at the seafloor to 21–23°C at the surface. Following these measurements, the underwater communication experiment was conducted.

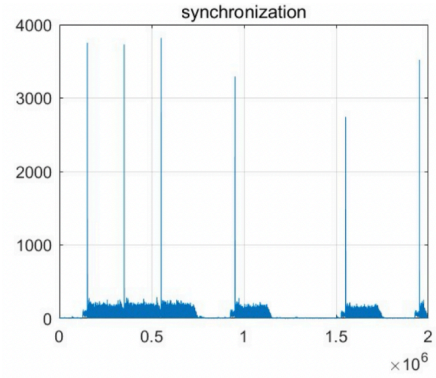
Figure 6 – Deployment of the integrated seafloor observatory system.

3.2 Experimental Results

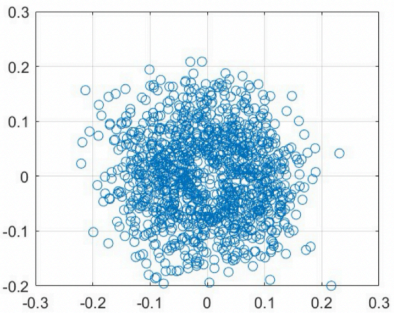
The experimental results were displayed using a Graphical User Interface (GUI) that plots graphs and outputs performance metrics. Figure 7 presents the data acquisition results from the experiment. Figure 7(a) shows the time-domain signal of 10 consecutive frames over a 10-second interval. Figure 7(b) displays the result of the correlation with the HFM signal, indicating successful synchronization. Figure 7(c) shows the constellation diagram of the demodulated OFDM symbols before channel equalization. Finally, Figure 7(d) presents the constellation diagram after channel equalization based on pilot data, showing clearly distinguished symbols.



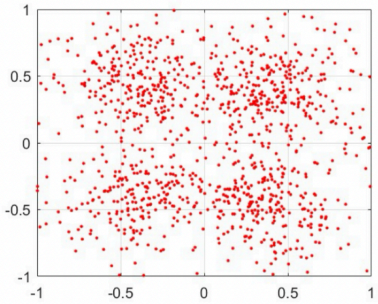
(a) Received signal in the time domain



(b) Synchronization result from HFM correlation



(c) Constellation diagram of demodulated OFDM symbols before channel equalization.



(d) Constellation diagram after pilot-based channel equalization

Figure 7 – Performance results of the underwater acoustic communication system.

In this experiment, the standard Bit Error Rate (BER) metric could not be used because the real-time data generated by the seismometer is random, making it impossible to compare transmitted and received data against a reference. Therefore, the Packet Error Rate (PER) was calculated based on an 8-bit CRC attached to each data packet. The experimental results showed that out of 1293 transmitted packets, a PER of 0.00387 was achieved. Figure 8 visualizes the received seismometer data, showing the Z-axis data sampled at 100 Hz.

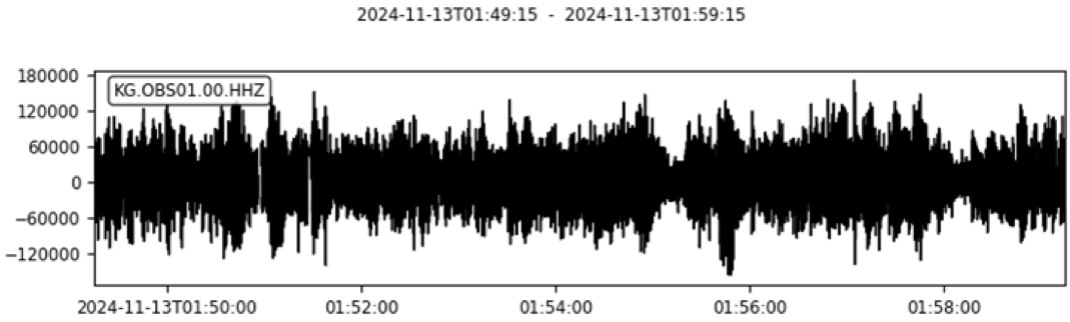


Figure 8 – Received seismic data from the vertical Z-axis.

4. Conclusion

In this study, we designed, implemented, and verified an OFDM-based underwater acoustic communication system for the real-time transmission of seafloor seismic data through sea trials conducted off the southern coast of Korea. The key achievements of this research are as follows. First, by adopting a strategy to selectively transmit data from the vertical axis (Z-axis) among the multi-axis data generated by the seismometer, we successfully achieved real-time data transmission at 4096 bps over the 12–28 kHz frequency band using an optimized OFDM system. Second, we demonstrated robust data transmission performance in a realistic marine environment, achieving a Packet Error Rate (PER) of 0.00387 under conditions of 35-meter water depth and a 100-meter transmission distance. Third, by establishing a complete

end-to-end system—from hardware implementation to signal processing algorithms—we have confirmed the practical viability of the system for real-world applications.

ACKNOWLEDGEMENTS

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean Government (MSIT) (NRF-2022M3J9A107887613). This work was supported by the Institute of Information & Communications Technology Planning & Evaluation (IITP)-Innovative Human Resource Development for Local Intellectualization program grant funded by the Korean government (MSIT) (IITP-2025-RS-2024-00436765).

REFERENCES

1. USGS (2020). "Where do earthquakes occur?" U.S. Geological Survey.
2. Strutz, T., & Curtis, A. (2024). The Roles of Low-Noise Stations, Arrays and Ocean-Bottom Seismometers in Monitoring UK Offshore Seismicity. arXiv preprint.
3. Mànuel, A., Roset, X., Del Rio, J., et al. (2012). Ocean Bottom Seismometer: Design and Test of a Measurement System for Marine Seismology. Sensors, 12(3), 3693-3719.
4. Kanazawa, T., Uehira, K., Mochizuki, M., et al. (2016). S-net project: Construction of large scale seafloor observatory network for tsunamis and earthquakes along the Japan Trench. In Proceedings of the AGU Fall Meeting 2016.
5. Favali, P., Beranzoli, L., De Santis, A., Kawaguchi, K., Kaneko, S., Nishida, T., & Komine, T. (2015). Construction of the DONET real-time seafloor observatory for earthquakes and tsunami monitoring. *Seafloor Observatories: A new vision of the Earth from the Abyss*, 211-228.
6. Xiao, H., et al. (2024). "Detection of Earthquake Infragravity and Tsunami Waves With Underwater Distributed Acoustic Sensing." *Geophysical Research Letters*, 51(6), e2023GL106767.
7. Chen, X., et al. (2024). "Distributed acoustic sensing technology in marine geosciences." *Intelligent Marine Technology and Systems*, 6, 39.
8. Zhou, Y., Tong, F., & Zhang, G. (2019). Distributed compressed sensing estimation of underwater acoustic OFDM channel. *Applied Acoustics*, 150, 230-237.
9. Neptune Sonar Limited. (2020). *Transducer Product Catalogue*. Neptune Sonar Ltd., East Yorkshire, UK.
10. Wang, K., Chen, S., Liu, C., et al. (2015). Doppler estimation and timing synchronization of underwater acoustic communication based on hyperbolic frequency modulation signal. In *2015 IEEE 12th International Conference on Networking, Sensing and Control* (pp. 75-80).
11. Al-jzari, A.A., Kostanic, I., & Mabrok, K.H.M. (2015). Effect of Variable Cyclic Prefix Length on OFDM System Performance over Different Wireless Channel Models. Universal Journal of Communications and Network, 3(1), 7-14.
12. Yang, L., Wang, H., Chen, P., & Liu, X. (2022). Research and Implementation of Turbo Coding Technology in High-Speed Underwater Acoustic OFDM Communication. *Journal of Robotics*, 2022, 2576303.

Smart Manufacturing-Oriented Dataset for Early Detection of Bearing Defects

Yunsu Kim[1]; Seungwoo Lee[2]; Yujin Bang[3] Jongwon Seok[4]

[1,2,3,4] Changwon National University, Republic of Korea

ABSTRACT

In the manufacturing industry, production interruptions due to equipment failures represent a significant challenge, necessitating the implementation of predictive maintenance strategies that anticipate malfunctions for proactive repair. Recent advancements in machine learning and deep learning technologies have accelerated research in this area, emphasizing the critical role of high-quality datasets. However, conventional datasets are often limited to representing only normal and defective states, without accounting for the intermediate states that may occur as equipment transitions from a normal to a defective condition. To address this limitation, this paper introduces a novel predictive maintenance dataset that includes data representing a "warning" state – a transitional phase between normal operation and failure. By training a deep learning model on this expanded dataset, we develop a simulated predictive maintenance system capable of more nuanced fault detection. The proposed system achieved 99% accuracy in distinguishing between normal and defective states on test datasets, and 89% accuracy when incorporating cautionary state data. This work contributes to the development of a truly predictive maintenance framework, facilitating preemptive actions rather than reactive measures, by leveraging a transitional dataset and an advanced fault prediction model.

Keywords: Deep Learning, Signal Processing, Manufacturing Sound, Fault Detection

1. INTRODUCTION

In the manufacturing industry, production interruptions caused by equipment de-fects result in substantial time and cost inefficiencies, posing a critical challenge. To miti-gate these issues, predictive maintenance, which anticipates and proactively addresses equipment malfunctions, is of paramount importance. Recent advancements in machine learning and deep learning have significantly influenced various modern technological fields, prompting active research into their integration within predictive maintenance frameworks. These models operate primarily by extracting patterns and rules from data, making their effectiveness highly dependent on the availability of large, high-quality da-tasets that are specifically tailored for their intended applications.

In response, various universities and research institutions have been developing environments that simulate both defective and normal conditions of components with low durability, such as bearings in manufacturing equipment. These simulated environments enable the measurement of vibration data to create datasets that support the training and testing of AI models for practical application in manufacturing settings. Notable datasets for detecting bearing defects in electric motor equipment include those from Case Western Reserve University (CWRU) [1], Paderborn University [2], and the FEMTO-ST Institute (Franche-Comté Electronics, Mechanics, Thermal Science, and Optics - Science and Technology) [3]. Existing bearing fault predictive maintenance datasets [4], including those from CWRU, Paderborn University, and FEMTO-ST, typically focus on data that distinguishes between normal and defective states. While some datasets, such as CWRU's, provide detailed labeling for various types of defects, they do not offer data that simulates or labels the intermediate states between normal and defective conditions over extended periods. Predictive maintenance aims to identify potential equipment or component failures in advance. However, datasets

[1] kor8780@gs.cwnu.ac.kr
[2] sw203080@naver.com
[3] bangbang1731@naver.com
[4] jwseok@changwon.ac.kr

limited to distinguishing only between normal and defective states are inadequate for achieving true predictive maintenance. This paper proposes a novel predictive maintenance system that employs a transitional dataset designed to simulate intermediate states and trains machine learning models to detect potential risks before defects fully manifest. The proposed system consists of two main components: a signal analysis module based on frequency domain techniques and a classification model training module utilizing deep learning methods.

2. Materials and Methods

2.1 Manufacturing Datasets

2.1.1 CWRU

The CWRU dataset differentiates between normal and faulty states with a particular focus on various types of bearing faults. These faults include defects in the inner race, ball, and outer race of the bearings, with outer race defects further categorized based on their specific location. The data collection process utilizes two sensor positions: the drive-end and the fan-end of the motor. To capture detailed sensor readings under different fault conditions, bearings with fault sizes of 7, 14, 21, 28, and 40 mils (1 mil = 0.001 inches) are alternately mounted on the equipment, which is then operated to measure and record the corresponding data.

2.1.2. Paderborn University

The Paderborn dataset was collected by Paderborn University using an electric motor, where six normal bearings and 26 faulty bearings were alternately installed to gather vibration data. Among the 26 faulty bearings, 12 were artificially damaged to simulate specific fault conditions for the experiment, while the remaining 14 bearings were naturally damaged, allowing for the collection of data under realistic wear and tear conditions.

2.1.3. FEMTO-ST

The FEMTO-ST dataset comprises recordings from sensors that continuously monitored the vibrations of bearings until they reached the point of failure. This approach enabled a comprehensive capture of the bearing degradation process, providing valuable data on the progression from normal operation to complete failure.

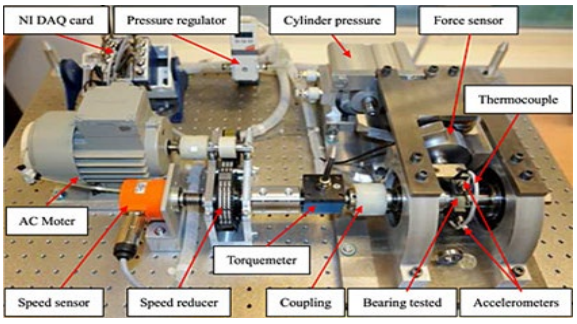


Figure 1 – Bearing test rig for cloud bearing system

Figure 1 illustrates the equipment used to collect the FEMTO-ST dataset, which includes vertical and horizontal vibration data, timestamps, and temperature readings. The dataset simulates three distinct real-world manufacturing conditions based on the rotation speed of the bearings and the loads applied indirectly to the outer race. The first condition involves a rotation speed of 1800 RPM with a load of 4000 N, the second condition operates at 1650 RPM with a load of 4200 N, and the third condition is set at 1500 RPM with a load of 6000 N. The point of bearing failure is determined arbitrarily, considering the cumulative damage to the equipment over time. Despite the growing need for extensive datasets driven by advancements in machine learning and deep learning technologies, the FEMTO-ST dataset remains relatively small and was developed using international infrastructure, posing challenges for domestic implementation of predictive maintenance systems based on this dataset.

2.2 Bearing Fault Detection Dataset

2.2.1 Binary Bearing Fault Detection Dataset

To address the challenges mentioned earlier, the rotational speed of the equipment was adjusted using motors and reducers, and a powder brake was employed to apply direct loads to the motor, allowing for the

simulation of various defect signals and data collection. The rotation speeds were set at 200, 600, and 1000 RPM, and for each speed, loads of 0, 1, and 2 N·m were applied to the motor. By operating the entire setup and measuring sensor values, both normal and various defective conditions were simulated. The data were categorized according to the rotation speed and the presence or absence of a load on the motor, effectively distinguishing between normal operations and fault conditions.

Table 1 – Data categorizing of conventional bearing fault detection dataset

Parameters		Data Category
Rotational Speed[RPM]	Load[N·m]	200-Normal
200	Normal(0 load)	600-Normal
		1000-Normal
600	Fault(1-2 load)	200-Fault
		600-Fault
1000		1000-Fault

Table 1 illustrates the categorization process of the conventional bearing fault detection dataset, which has been designed to increase the data volume compared to existing datasets and utilizes domestically produced equipment to create a more locally adaptable configuration. However, this dataset still primarily distinguishes between normal and faulty states, aligning it more closely with post-event analysis rather than true predictive maintenance. To address this limitation, this paper introduces an augmented dataset that includes transitional category data, where the load gradually increases over time. By training a deep learning model with this enhanced dataset, a predictive maintenance system has been developed. This approach aims to move beyond simply distinguishing between normal and faulty states by enabling the detection of potential failures before they fully develop.

2.2.2 Bearing Fault Detection Dataset Incorporating Warning States

In this paper, a new dataset was introduced to the existing bearing fault predictive maintenance dataset, incorporating a novel category of signals collected by incrementally increasing the load on the motor over a period of 30 seconds. As depicted in Figure 2, this new data was obtained by gradually adjusting the tension controller's dial from 0 N·m to 1 N·m or from 0 N·m to 2 N·m. This additional data type is categorized as a "warning" state in the reconstituted dataset. The collected and reconstituted data are further differentiated based on the machine's rotational speed and the load applied to the motor, which is controlled by a reducer to three specific speeds: 200, 600, and 1000 RPM. The degree of load applied to the motor, regulated by a powder brake, is classified into three categories: a "normal" state with no load or a light load of 1 N·m, where the motor operates stably; a "warning state," where the load increases gradually over 30 seconds from 0 N·m to 1 N·m or from 0 N·m to 2 N·m, indicating a need for attention to potential faults; and a "fault" state, where a load of 2 N·m signifies a severe defect.

Figure 2 – Increasing load using tension controller

Figure 3 presents a graph illustrating the intensity of the load applied to the motor over time, segmented by operational state. The horizontal axis represents time, while the vertical axis shows the magnitude of the load. The lines are color-coded to differentiate between various states. Cautionary period, marked on the time axis, represents a 30-second interval during which the load on the motor is progressively increased using a powder brake in the simulated environment. This visual representation effectively demonstrates how load variations correspond to different operational states of the equipment.

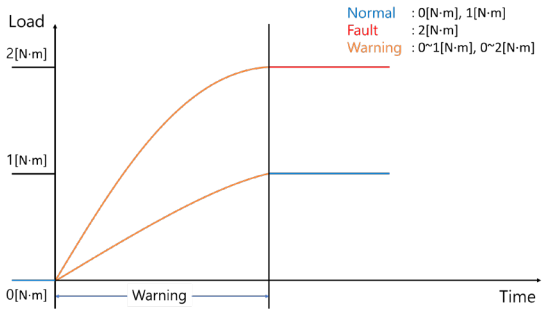


Figure 3 – Load variation by type of data

Table 2 outlines the composition of the dataset, with each 30-second interval considered as a single case. Each entry in the table represents a specific scenario under simulated conditions. For instance, "RPM200_Normal" denotes signals simulated under normal operating conditions at a rotational speed of 200 RPM. This method of data representation provides a systematic organization of various scenarios within the dataset, facilitating easier analysis and application in predictive maintenance research. Each case captures different operating conditions, including various RPM settings and states (normal, warning, fault), to comprehensively represent the full range of potential operational scenarios.

Table 2 – Dataset configuration

Class	number of files
RPM200_Normal	120
RPM200_Warning	58
RPM200_Fault	29
RPM600_Normal	145
RPM600_Warning	62
RPM600_Fault	30
RPM1000_Normal	148
RPM1000_Warning	60
RPM1000_Fault	30

2.3 Time-Frequency Signal Analysis Methods

2.3.1 Continuous Wavelet Transform

The Continuous Wavelet Transform (CWT) [5] is a signal processing technique used for frequency domain analysis that retains time-domain information, in contrast to the Fourier Transform, which loses time information during transformation. Compared to the Short Time Fourier Transform (STFT), CWT offers higher temporal resolution at higher frequencies and lower temporal resolution at lower frequencies, providing a balanced approach to time and frequency resolution. The CWT operates by passing the signal through a "mother wavelet" and calculating the correlation coefficients. The mathematical representation is as follows:

$$\mathrm{CWT(t,s)} = \frac{1}{\sqrt{s}} \int x(\tau)\psi^*(\frac{t-a}{a})dt \quad (1)$$

where $\mathrm{x}(\tau)$ is the signal to be analyzed, $\psi^*(\cdot)$ represents the complex conjugate of the mother wavelet, t denotes the translation parameter (time shift), and s is the scale parameter (frequency scaling). This equation enables the analysis of the signal's frequency components while preserving temporal details across various scales [6,7].

2.4 Signal Analysis

2.4.1 Analysis of Normal and Fault Signal

To analyze warning state signals, where the load increases over time, it is effective to use signal analysis techniques that provide insights into both time and frequency domain information. Notable time-frequency domain analysis methods include spectrogram analysis using the Short-Time Fourier Transform (STFT) and scalogram analysis using the Continuous Wavelet Transform (CWT). These techniques enable the

visualization of differences between normal and faulty states in the signals.

A spectrogram visually represents the spectrum of frequencies in a signal over time, with the horizontal axis representing time, the vertical axis representing frequency, and color indicating the amplitude of the frequency components. Figure 4 shows a randomly selected spectrogram of signals from both normal and faulty states. While the differences between normal and faulty conditions may not be immediately apparent, an overall increase in energy is observed throughout the signal under an increasing load, particularly above 3000 Hz in the higher frequency bands. This trend suggests that the energy in higher frequencies becomes more pronounced under fault conditions, likely due to the increased mechanical stress and vibrations associated with bearing defects.

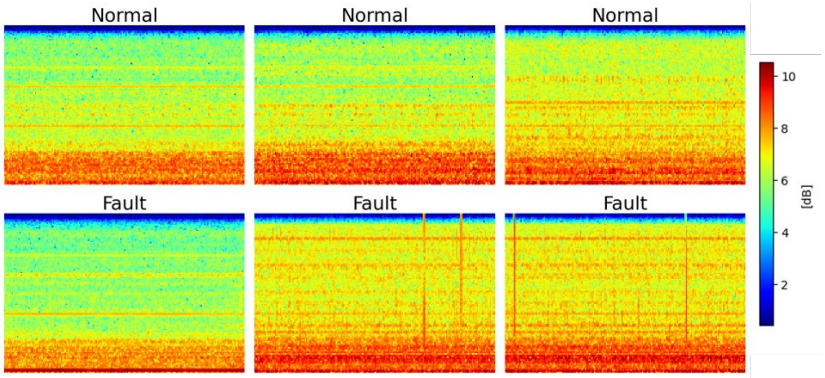


Figure 4 – Spectrogram of Normal and Fault data

Figure 5 presents a scalogram of randomly selected signals from both normal and faulty states. In this visualization, the horizontal axis represents time, while the vertical axis corresponds to the wavelet bank, set at 256 levels. In Figure 6, it can be observed that in the normal state, the areas of bright, concentrated energy are relatively small, whereas in the faulty state, these areas are significantly larger. This indicates an overall increase in energy due to the load, similar to the observations made from the spectrogram analysis.

The scalogram, utilizing a wavelet transform, provides a different perspective on signal characteristics by varying the scale of analysis. This method is particularly effective in highlighting transient features and non-stationary aspects of the signal, which may not be as apparent in a spectrogram. When comparing the two, the distinctions between normal and faulty states are more pronounced in the scalogram. This is because wavelets excel at capturing localized changes in both frequency and time, making them especially useful for detecting anomalies in signals with defects, particularly under varying load conditions. The enhanced visibility of signal characteristics in faulty conditions, as shown in the scalogram, can provide critical insights for developing more effective predictive maintenance strategies.

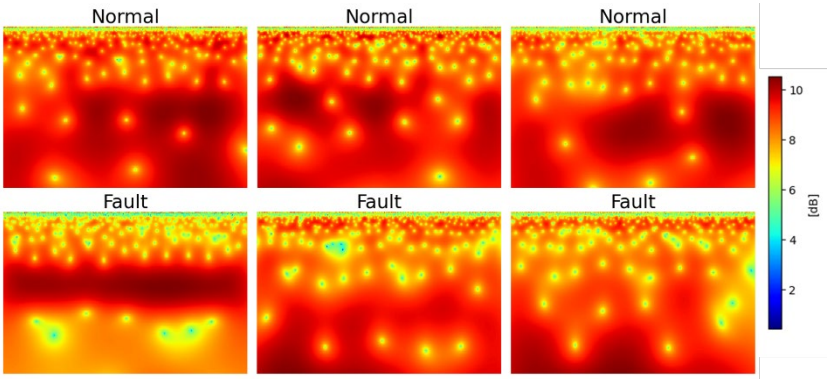


Figure 5 – Scalogram of Normal and Fault category

2.4.2 Analysis of Warning Signal

Next, we analyze the waveform and spectrogram of both normal and warning state signals over a 30-second period.

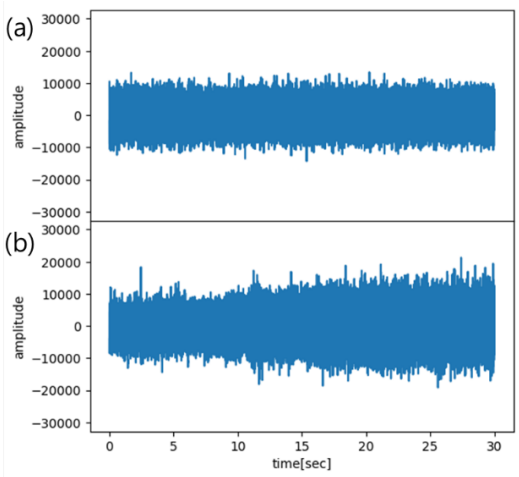


Figure 6 – (a) Waveform of normal category (b) Waveform of warning category

Figure 6 shows the waveforms for both the normal and warning states at a rotational speed of 600 RPM. In the normal state, the waveform maintains a consistent amplitude throughout. In contrast, in the warning state, the envelope of signal generally increases due to the applied load and continues to rise progressively over time as the load increases.

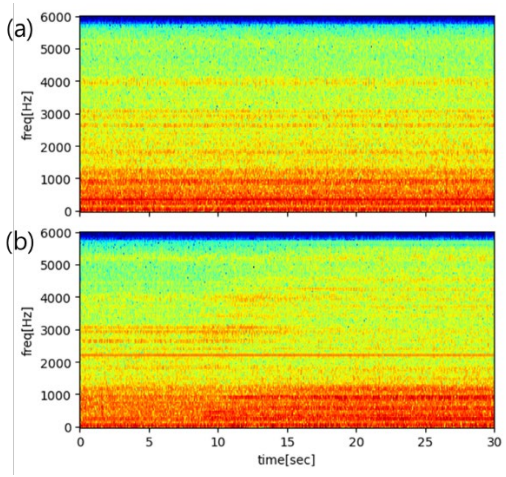


Figure 7. (a) Spectrogram of normal category (b) Spectrogram of warning category

Figure 7 presents the spectrograms for both the normal and warning states at a rotational speed of 600 RPM. In the normal state signal, the energy variation over time is minimal and remains relatively consistent. In contrast, for the warning state, there is little change in the first 10 seconds; however, an increase in energy is observed in the frequency region above 3000 Hz thereafter. Unlike the spectrogram, the scalogram has limitations due to input length constraints, making it challenging to capture the entire cautionary period.

3. Experiments

In this study, we designed experiments to evaluate the recognition performance of a deep learning model trained on signals collected from normal, warning, and faulty states. Unlike in previous studies [4], the inclusion of a warning state, where the load progressively increases over time, enables the experiment to capture temporal changes in frequency components more effectively through the use of spectrograms and scalograms generated from signal analysis. Figure 9 illustrates the schematic of the experimental methodology employed in this study. The experimental framework was divided into four distinct classification tasks:

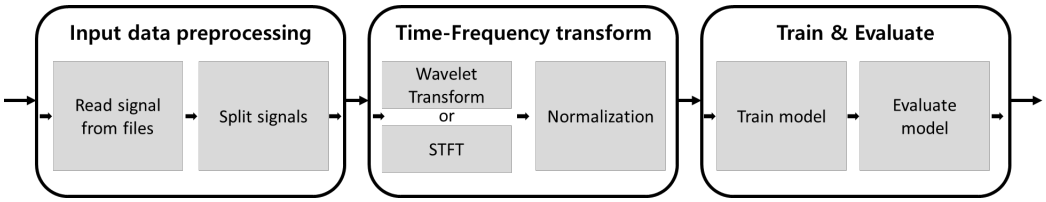


Figure 8. Experimental flow chart

1. A binary classification experiment leveraging spectrograms to discriminate between data from normal and faulty states.
2. A ternary classification experiment utilizing spectrograms to differentiate among normal, warning, and faulty state data.
3. A binary classification experiment employing scalograms to discern between normal and faulty states.
4. A ternary classification experiment using scalograms to categorize normal, warning, and faulty states.

3.1 Preparation of Input Data

The data were processed according to the time-frequency domain analysis methods described earlier, using the following parameters. The processed signals were labeled according to the specified scenarios as normal, faulty, or warning states. The labeled data were then preprocessed using the previously outlined signal analysis methods. Tables 3 and 4 provide the distribution of data counts for each label.

The parameters for the spectrogram are as follows:

- Input length : 24000 [samples]
- FFT window : hann
- FFT length : 256 [samples]
- FFT overlap : 50 [%]

Table 3 – Number of spectrogram data by category

category	number of data
Normal	60144
Warning	25872
Fault	12432

The parameters for the scalogram are as follows:

- Input length : 2048 [samples]
- Wavelet function : Morlet
- Number of Wavelet filter bank : 256

Table 4 – Number of scalogram data by category

category	number of data
Normal	146601
Warning	63063
Fault	30303

To prepare the dataset for training, the data were divided into training, validation, and test sets in a ratio of 3:1:1, respectively. Additionally, the distribution of data across each label was configured to be equal, ensuring a balanced dataset for model training and evaluation.

3.2 Classification Model Structure

The deep learning model used for classification follows a standard Convolutional Neural Network (CNN) architecture. Figure 10 presents a simplified schematic of the model structure employed in this study. Each CNN block consists of 2D convolutional layers, followed by 2D max-pooling layers. The number of filters in the 2D convolutional layers varies depending on the classification model: the spectrogram-based model uses 16 filters, while the scalogram-based model uses 8 filters. The filter size is consistently set to (5,5), and padding is applied to maintain the same input and output dimensions for the CNN layers. Additionally, in the spectrogram classification model, a Dropout layer with a rate of 0.25 is included after the first Dense layer following the Flatten layer to mitigate overfitting.

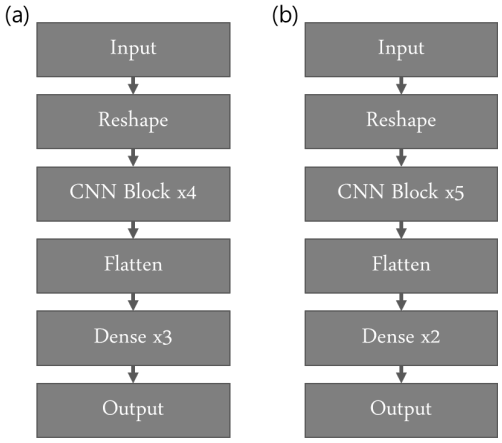


Figure 9. (a) Classification model architecture for spectrogram data (b) Classification model architecture for scalogram data

3.3 Training and Performance Validation

In this study, all deep learning models were trained over 200 epochs using the Adam optimizer. During the training process, the model that achieved the highest accuracy on the validation dataset was saved, and its performance was subsequently evaluated on the training, validation, and test sets. For the three-category classification, a confusion matrix was also generated for the test set to provide a more detailed analysis; this matrix was normalized to display proportions rather than absolute data counts. Table 5 presents the accuracies for both the binary (two-category) and ternary (three-category) classifications using spectrogram data, while Table 6 shows the accuracies for these classifications using scalogram data.

The predictive maintenance system using spectrogram input demonstrated stable performance, achieving a test set accuracy of approximately 99% for the two-category classification. However, the three-category classification showed a somewhat less stable performance, with a test set accuracy of about 89%. This indicates that while the model was effective at distinguishing between normal and faulty loads, it had difficulty classifying the warning state data, highlighting the complexity of this task. In contrast, with scalogram input, the two-category classification maintained a stable accuracy of 99% on the test set, while the three-category classification achieved a slightly lower accuracy of around 80% compared to the spectrogram input. Nevertheless, it still exhibited a respectable recognition performance.

Table 5 – Classification performance for spectrogram data (unit : %)

Dataset	Binary category	Ternary category
Train	100	99
Validation	100	90
Test	99	89

Table 6 – Classification performance for scalogram data (unit : %)

Dataset	Binary category	Ternary category
Train	99	80
Validation	99	78
Test	99	78

Tables 7 and 8 present the confusion matrices for the three-category classifications using spectrogram and scalogram data, respectively, expressed as proportions. In both matrices, instances where data in a normal state were misclassified as faulty, or vice versa, were close to zero, indicating that the classification performance for normal and faulty data remained stable despite the inclusion of warning state data.

A notable observation is that with the spectrogram, there was a higher tendency to misclassify data in a normal state as a warning state, or vice versa. For the scalogram, there were more instances of misclassifying normal state data as a warning state and actual faulty state data as a warning state, compared to the spectrogram. This suggests that the scalogram data, due to its considerably shorter analysis segments, may

not provide sufficient length to adequately distinguish the warning state as effectively as the spectrogram.

Table 7 – Confusion matrix of test set for ternary category spectrogram data classification (unit : %)

Pred \ True	Normal	Warning	Fault
Normal	91	11	00
Warning	09	86	06
Fault	00	03	93

Table 8 – Confusion matrix of test set for ternary category scalogram data classification (unit : %)

Pred \ True	Normal	Warning	Fault
Normal	80	28	01
Warning	19	68	20
Fault	00	04	79

3.4 Effectiveness verification

In addition to collecting data on normal and defective states, as done in existing methodologies, this paper also introduces the collection of data from an intermediate "warning" state. A simple neural network was designed and trained to perform binary classification, achieving recognition accuracies of 99% for binary classification and 89% for three-class classification. However, it is important to note that the data used in this study were simulated based on a controlled environment and predefined scenarios, rather than collected from an actual process environment. Therefore, additional validation is required to determine whether the model trained on the proposed dataset is versatile enough for practical application in real-world settings. To evaluate the actual applicability of the proposed dataset, this study will compare the recognition performance of the same neural network using the CWRU dataset, which has been widely applied [8] and analyzed [9] in various fields of fault detection.

To evaluate the effectiveness, the recognition experiment configuration for the CWRU dataset was designed to closely match the label structure used in the proposed dataset. For the experiments conducted in this paper, three defect types—Ball, Inner Race (IR), and Outer Race at the 6:00 position in the load zone (OR@6)—were selected, with defect diameters ranging from 00 (normal state) to 07, 14, and 21 mils, respectively. Based on these criteria, the dataset was grouped into four labels across three experimental groups.

Table 9 – Transformation parameters for each dataset (unit : samples)

Parameter	proposed	CWRU
# of input samples	24000	1280
FFT length	256	512
FFT overlap	128(50%)	32
Output shape	256x186	65x10

The models used in the recognition experiments all employed the same lightCNN architecture as depicted in Figure 9(a). The input data conversion method utilized the Fast Fourier Transform (FFT), which demonstrated superior performance in previous experiments, with the spectrogram used as input. For the proposed dataset, the transformation was applied using the same parameters outlined in Section 3.1. For the CWRU dataset, the transformation was performed following the analysis in Yoo, Y. (2020) [10] to ensure adequate temporal resolution for the recognition process. Table 9 presents the transformation parameters for each dataset and the shape of the resulting feature maps after transformation. The output shape becomes the input shape of the model. The distribution of labeled data for the CWRU dataset is shown in Table 10.

All training was conducted for 50 epochs, and the same process was repeated a total of 10 times, with the average accuracy across these runs used as the final result.

Table 10 – Distribution of the CWRU dataset by label

Label	Ball	IR	OR@6
Normal	761	766	766
Fault(07)	764	766	764
Fault(14)	761	759	761
Fault(21)	761	761	763

Table 11 – Average prediction accuracy of proposed and CWRU dataset (unit : %)

Subset	Proposed (binary)	Proposed (ternary)	CWRU (Ball)	CWRU (IR)	CWRU (OR@6)
Train	100	99	100	100	92.50
Validation	100	90	100	100	92.51
Test	99	89	99.85	100	92.50

Table 11 presents the results of training the same lightCNN model using both the proposed dataset and the CWRU dataset. Despite the fundamental difference that the proposed dataset consists of signals collected via a microphone, while the CWRU dataset comprises signals obtained from an accelerometer sensor directly attached to the component, the recognition accuracy ranges from 90% to 97% on average. This indicates that the influence of the two types of data is not significantly different when the same neural network evaluates the results based on data patterns.

However, the collection cost for CWRU data is relatively high due to the use of sensors directly attached to the part, which provides overall stable recognition performance of more than 92%. In contrast, the data collected in this study were obtained through a microphone without direct attachment to the machine. This method is more susceptible to background noise, which resulted in a drop in test accuracy to 89% for the ternary classification that includes warning signals. This suggests that more careful tuning is required when designing predictive maintenance systems in future research to account for noise and other external factors.

4. Conclusions

This paper proposes a predictive maintenance system that utilizes a transitional dataset, designed for preemptive rather than reactive processing. This dataset was developed by incorporating and labeling data that progressively transitions from a normal to a faulty state with an increasing load over time. To enable effective distinction of the warning state, the dataset was preprocessed using time-frequency domain signal analysis techniques, specifically spectrograms and scalograms. The preprocessed data was then employed to train deep learning models to implement the predictive maintenance system.

The results demonstrated that the model using spectrogram inputs achieved an accuracy of 99% for binary classification between normal and faulty states and 89% for ternary classification, which included the warning state. For the ternary classification, the model with spectrogram inputs outperformed the model using scalogram inputs. This performance discrepancy is attributed to the shorter analysis segments of scalograms, which can be mitigated by extending the time duration analyzed by scalograms. With longer input times, it is expected that the scalogram-based deep learning model would also show improved performance. This research advances the field of predictive maintenance by employing a dataset configured for preemptive actions rather than post-facto analysis, contributing to the development of a truly anticipatory maintenance system.

ACKNOWLEDGEMENTS

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea(RS-2022-NR075671)

REFERENCES

1. Case Western Reserve University Bearing Data Center. *Apparatus and Procedures*. [Internet]. Available from: https://engineering.case.edu/bearingdatacenter/apparatus-and-procedures
2. Konstrucktions – Und Antriebstechnik (KAt) Datasets. *Data Sets and Download*. [Internet]. Available from: https://mb.uni-paderborn.de/en/kat/main-research/datacenter/bearing-datacenter/data-sets-and-download
3. Femto-st-Sciences & Technologies. *The Institute – Introduction*. [Internet]. Available from: https://www.femto-st.fr/en/The-Institute/introduction
4. Kim YS, Bae SH, Seok JW. Building bearing fault detection dataset for smart manufacturing. *J Korean Inst Electr Eng (IKEEE)*. 2022;26(3):488-93.
5. Peng Z, Chu F, He Y. Vibration signal analysis and feature extraction based on reassigned wavelet scalogram. *J Sound Vib*. 2002;253(5):1087-100.
6. Baliunas S, Frick P, Sokoloff D, Soon W. Time scales and trends in the central England temperature data (1659–1990): A wavelet analysis. *Geophys Res Lett*. 1997;24(11):1351-4.
7. De Moortel I, Mulday SA, Hood AW. Wavelet analysis: the effect of varying basic wavelet parameters. *Sol Phys*. 2004;222(2):203-28.
8. Neupane D, Seok J. Bearing fault detection and diagnosis using Case Western Reserve University dataset with deep learning approaches: a review. *IEEE Access*. 2020;8:93155-78.
9. Raj KK, Kumar S, Kumar RR, Andriollo M. Enhanced fault detection in bearings using machine learning and raw accelerometer data: A case study using the Case Western Reserve University dataset. *Information*. 2024;15:259. Available from: https://doi.org/10.3390/info15050259
10. Yoo Y, Jo H, Ban SW. Lite and efficient deep learning model for bearing fault diagnosis using the CWRU dataset. *Sensors*. 2023;23:3157. Available from: https://doi.org/10.3390/s23063157

A Shallow-water Source Localization Neural Network Based on Source Depth Classification

Jing Guo[1,2]; Juan Zeng[1]
[1] Institute of Acoustics, Chinese Academy of Sciences, China
[2] University of Chinese Academy of Sciences, China

ABSTRACT

Deep learning has been widely applied in source localization, where neural networks extract meaningful features from underwater acoustic measurements to localize sources. In theory, incorporating physical constraints can effectively guide network optimization and improve convergence, especially in the complex and dynamic shallow-water environment where measured data are often limited. In this study, the relationship between the structure of the localization ambiguity surface and source depth in shallow-water passive localization is analyzed. Based on this relationship, a multi-task learning neural network is proposed, where source depth classification serves as an indirect constraint to enhance range regression estimation. Using sparse grids of model parameters to generate acoustic data, networks based on depth regression and depth classification are trained and evaluated on the same training and test sets. Results show that depth classification significantly improves range estimation, demonstrating that incorporating physical constraints enhances localization performance, particularly when training data are limited.

Keywords: Deep learning, Source localization, shallow water, Range-dependent sound speed profile

Unambiguous Localization Method for Long-Baseline Acoustic Sources Based on Consistent Likelihood Decision

Dong XU[1,2,3]. Yan WANG[1,2,3]. Bin QI[1,2,3]†. Binchao YANG[1,2,3]. Xiang LI[1,2,3]

1. National Key Laboratory of Underwater Acoustic Technology, Harbin Engineering University, Harbin 150001, China;

2.Key Laboratory of Marine Information Acquisition and Security (Harbin Engineering University), Ministry of Industry and Information Technology; Harbin 150001, China;

3.College of Underwater Acoustic Engineering, Harbin Engineering University, Harbin 150001, China.

ABSTRACT

Long-baseline (LBL) acoustic source localization systems face range ambiguity when the time delay of received pulse signals exceeds the synchronization period. To address this issue, this paper proposes an improved anti-ambiguity method based on consistent likelihood decision. After setting the maximum ambiguity period, the approach constructs a likelihood function that integrates time-delay measurements and unambiguous period constraints from the perspective of time-delay parameter estimation. By evaluating the spatial consistency of localization results across all possible unambiguous combinations among multiple array buoys, potential unambiguous combinations are screened. Subsequently, the optimal unambiguous period combination is determined through consistent likelihood decision. Compared to existing methods, this approach transforms the anti-ambiguity problem into a nonlinear optimization problem constrained by unambiguous period counts, effectively resolving distance ambiguity while ensuring computational efficiency and achieving high-precision localization of moving acoustic sources.

Keywords: Long-baseline acoustic localization. range ambiguity. likelihood-based consistency decision

1. INTRODUCTION

Underwater target positioning technology plays an important role in marine development, marine scientific research, and military applications, and is an important topic in marine engineering. For the vehicles that need long-term underwater navigation, they need a long-term and high-precision positioning system to serve them. Inertial navigation technology inevitably has the disadvantage of error accumulation, which makes the navigation error of vehicles using this technology increase with time. At the same time, due to the rapid attenuation of electromagnetic wave in the water, satellite positioning systems using electromagnetic wave, such as GPS and GNSS, cannot work in the deep sea, so underwater acoustic positioning technology plays a key role.

At present, most underwater acoustic positioning systems are acoustic source positioning systems that rely on geometric principles. According to the baseline length, they can be divided into long baseline system (LBL), short baseline system (SBL) and ultra-short baseline system (USBL). The baseline length of long baseline system ranges from several hundred meters to several thousand meters, the baseline length of short baseline system ranges from several meters to tens of meters, and the baseline length of ultra-short baseline system is less than or equal to half wavelength.

The long baseline positioning system uses acoustic trigger to receive the acoustic pulse time delay, each array buoy is arranged in a certain geometric way, and the position of the target is calculated by measuring the distance between the target and each array buoy (1). However, the measurement delay measured by the positioning system is the remainder of the acoustic signal propagation delay and the pulse synchronization period. For the long baseline positioning system with large array buoy spacing and small pulse synchronization period, the absolute delay of acoustic signal propagation is greater

[1] hsudong@hrbeu.edu.cn

[2] qibin@hrbeu.edu.cn

than the pulse synchronization period, and the measurement delay will differ from the acoustic signal propagation delay by one or more pulse synchronization periods, resulting in multivalued positioning results and distance ambiguity (2).

For the distance ambiguity problem of long baseline positioning and navigation, many researchers have proposed two kinds of solutions for the source of the distance ambiguity problem (3,4). The first kind of signal design scheme increases the unambiguous distance of the acoustic signal by designing the acoustic signal parameters that distinguish different pulse synchronization periods. However, this scheme needs to transform the sound source and signal processing algorithm, and the technical implementation is more complex.

The second is software logic decision scheme, which uses a series of decision logic to modify the measurement delay and find the true solution. The reference position method suppresses the distance ambiguity before solving, which can be called a priori anti ambiguity method, and the required priori information is the reference position of the target. The reference position is also called the initial position. It needs to be preset a priori. In addition, the reference position can be automatically updated at each time in the future. However, this method relies heavily on the correctness of the initial position calibration. The algorithm uses the initial position to determine the ambiguous period of each array buoy. If the initial position is wrong, the correct result cannot be solved. The second method of show of hands completely uses the file information built in the process of system implication or tracking to distinguish the true solution from many ambiguous solutions. The essence of the algorithm is like "election", which lists all possible solutions. Each solution evaluates its weight according to the criteria composed of various information. Finally, the one with the largest number of votes and the highest quality of the solution with the largest weight is selected as the true solution. As the mitigation of delay ambiguity is conducted after the positioning estimation, this approach can be referred to as a post-processing anti-delay ambiguity method. After generating all possible ambiguous solutions, redundant array buoys and historical trajectories are used to judge the true solution from all possible ambiguous solutions. However, in practical applications, the amount of calculation increases exponentially with the maximum number of ambiguous cycles (5).

In this paper, an improved method of consistent likelihood test based on time delay measurement is proposed. The likelihood function is constructed by using the number of ambiguous cycles and the measurement delay. The possible number of ambiguous cycles is selected by calculating the aggregation degree of multiple array buoys in the same ambiguous cycle number solution, and then the global optimal number of ambiguous cycles through the likelihood function test is recorded as the unambiguous propagation delay of underwater acoustic signal. Finally, the steepest descent method is used to solve the minimum mean square error of the actual distance and estimated distance between the array buoy and the target, to obtain the best estimation of the target position and realize the positioning of the underwater target.

The organizational structure of this paper is as follows: in the second section, the long baseline positioning model is given, and the problem of distance ambiguity in long baseline positioning is introduced. Then the algorithm derivation of the long baseline unambiguous location method based on consistent likelihood decision is given, and the implementation of the algorithm is given, that is, the consistent constraint is used to divide the unambiguous delay state, and the delay likelihood function is used to estimate the unambiguous delay of each array buoy. Finally, the sound source target is accurately located by the obtained unambiguous time delay. In the third section, the effectiveness of the algorithm is verified by simulation.

2. Long baseline positioning model and distance ambiguity

2.1 Long baseline positioning model

Long baseline underwater acoustic localization uses the absolute propagation delay of the signal to estimate the relative distance between the target and each array buoy, and combines the position information of the array buoy to solve the target position. The measurement system used in this system is a buoy array. The target and the measurement system are equipped with high-precision clock synchronization equipment, which works in the same synchronization cycle. The target is a mobile sound source equipped with cooperative beacons, which emits pulse signals outward at a certain synchronization period T. when each buoy of the buoy system receives and detects the target signal, it records the relative delay value of the signal arrival time, and packages and sends the signal information to the display and control computer for calculation (6). The working principle of long baseline underwater acoustic positioning is shown in Figure 1:

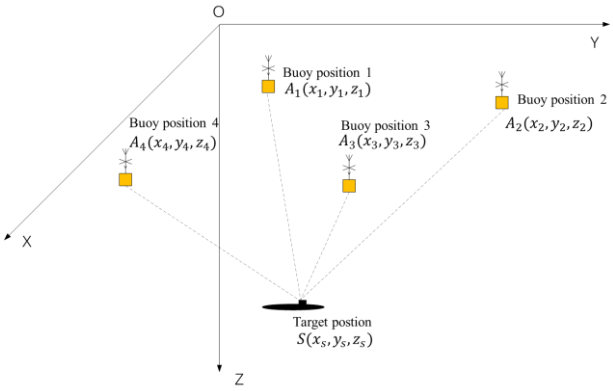


Figure 1 Schematic diagram of long baseline underwater acoustic positioning

Long baseline underwater acoustic localization determines the spatial distance between the two by estimating the signal propagation delay from the array buoy to the target, and calculates the target position by using the distance and the array buoy position. For the long baseline navigation system composed of N_p buoys, the coordinate position A_i of each buoy is expressed as:

$$A_i = (x_i, y_i, z_i), i = 1, 2, ..., N_p \tag{1}$$

The long baseline underwater acoustic positioning system needs to obtain the time delay information of multiple array buoys to calculate the position of underwater target. The position of underwater target in each frame is expressed as $S = (x_s, y_s, z_s)$. the solution equations composed of estimated time delay and array buoy coordinate information are

$$(x_s - x_i)^2 + (y_s - y_i)^2 + (z_s - z_i)^2 = (c \cdot t_m^i)^2 \tag{2}$$

Where (x_s, y_s, z_s) is the target coordinate to be solved, (x_i, y_i, z_i) is the coordinate of the i array buoy, c is the underwater sound velocity, and t_m^i is the propagation delay from the signal transmitted by the i array buoy to the target.

Because the baseline of the horizontal dimension is a large scale, compared with the vertical direction, there is almost no scale. If the target is in three dimensions by using the geometric relationship directly according to the buoy measurement time delay and other information, it can only ensure the two-dimensional positioning accuracy in the horizontal direction, which is not enough to ensure the positioning accuracy of the system in the vertical direction. Therefore, the depth measurement in the vertical direction is independent of the horizontal direction in the design of the system.

A common method of depth acquisition of sound source target is to add a synchronous signal section on the target. The beacon uses a double pulse mechanism to synchronously transmit pulse signals. The beacon section is equipped with a pressure sensor, which can adjust the distance of the emitted double pulses by measuring its own depth, to carry the depth information of the target. In the measurement system, if the delay information of double pulses can be accurately extracted, the depth measurement can be realized.

In the double pulse sounding model, the beacon periodically transmits double pulses with a pulse spacing of. The double pulse spacing τ is the minimum when the target is at zero depth ($H_{\min}$), the double pulse spacing $\tau_{\max}$ is the maximum when the target is at the maximum depth ($H_{\max}$), and the intermediate depth (H) corresponds to a linear relationship. The double pulse spacing is

Among them, $\tau_{\min}$, $\tau_{\max}$, $H_{\min}$ and $H_{\max}$ are known quantities. By measuring the double pulse spacing at the receiving end of the array buoy, the depth of the sound source target can be measured according to the above formula. Therefore, the unknown quantity of equation group in Eq.(2) is only (x_s, y_s), and the position of the target in this frame can be obtained by simultaneous equations.

2.2 Distance ambiguity

For the long baseline positioning system, the acoustic beacon carried by the target emits synchronous acoustic signals according to a certain period. The synchronization period of the acoustic signal is. If the absolute propagation delay from the target to the array buoy is less than the synchronization period, it can be solved normally. If the absolute propagation delay of the signal from the target to the submerged target is greater than the synchronization period, it is necessary to restore

the relative propagation delay to the absolute delay to calculate the true position of the target, as shown in Figure 2.

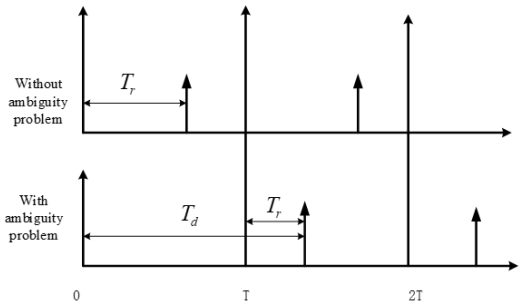


Figure 2 Schematic diagram of distance ambiguity

In synchronous mode, if the maximum operating distance of a single buoy is $R_{\max}$, the maximum ambiguity period of the system is:

$$N_{\max} = \lfloor R_{\max} / cT \rfloor \tag{3}$$

Where $\lfloor \cdot \rfloor$ is the downward rounding function. While $N_{\max} \geq 1$, the unknown quantity of the sound source localization problem will include the target position coordinate (x_s, y_s) and the number of ambiguous periods corresponding to each array buoy. For the sound source localization system that only uses time delay information to locate, the number of unknown quantities is always two more than the number of equations when solving its position quantity from Eq.(2). Theoretically, the unique solution cannot be directly obtained. Therefore, it is necessary to solve the number of ambiguous periods corresponding to each array buoy with the help of a specific anti distance ambiguity algorithm. Suppose that the effective time delay sequence received by the N_p buoys in frame k is $T_r(k)$

$$T_r(k) = [t_r^1(k), t_r^2(k), ..., t_r^i(k), ..., t_r^{N_p}(k)] \tag{4}$$

Where $t_r^i(k)$ AA is the delay information received by the i second buoy in the k frame. Then the relationship between the propagation delay T_m of the signal from the underwater target to the array buoy and the time delay T_r recorded by the system is:

$$T_m(k) = T_r(k) + N_j(k) \cdot T \tag{5}$$

In Eq.(5), $N_j(k)$ is the j -th unambiguous number combination of k frame, which is the compensation of the long baseline positioning system for the propagation delay when the distance ambiguity problem occurs. If n_i is the unambiguous number of the i array buoy, the unambiguous number combination can be expressed as:

$$N_j = [n_1, n_2, ..., n_i, ..., n_N], n_i = 0, 1, ..., N_{\max} \tag{6}$$

For N array buoys receiving the target acoustic signal, there are $J = (N_{\max} + 1)^{N_p}$ unambiguous combinations in each frame. For the j -th combination, $j = 1, 2, ..., J$ 。 The purpose of the anti-range ambiguity algorithm is to find the correct combination from the Unambiguous combinations in J, to obtain the correct propagation delay T_m, and then carry out the localization of the sound source target.

2.3 Unambiguous location method based on consistent likelihood decision

The method introduced in this paper represents a non-ambiguous localization approach suitable for scenarios with unknown initial positions. In contrast, the process of the non-ambiguous localization method via a show of hands involves the following steps: utilizing the delay information from redundant array buoys, when solely employed for judgment without participation in the solution process, we calculate the propagation delay of the pulse signal from the calculated target coordinates to the array buoys not involved in the solution. This delay is then converted into relative time delay and compared with the received delay recorded by the respective array buoy. If the difference falls within the error threshold, the solution is deemed accurate. However, as the number of solutions increases, the computational load escalates exponentially, necessitating the repetition of this calculation for each positioning processing frame to sustain the positioning trajectory.

When Equation (2) contains only two unknowns (x_s, y_s), a unique positioning solution can be obtained using time delay measurements from at least three buoys in a single frame. The remaining

buoy measurements serve as redundant delays. Thus, when N buoys provide N time delays, there exist $N_S = \frac{N(N-1)(N-2)}{6}$ possible combinations of measurements. The set of N_s computed position solutions is denoted as X_j:

$$X_j = [\{x_{s1}, y_{s1}\}, \{x_{s2}, y_{s2}\}, ..., \{x_{sN_s}, y_{sN_s}\}] \tag{7}$$

Where, $\{x_{s1}, y_{s1}\}$ represents the first position solution in this combination, $\{x_{s2}, y_{s2}\}$ denotes the second position solution, and $\{x_{sN_s}, y_{sN_s}\}$ stands for the N_s-th position solution.

The time delay resolution for this group is determined by N_s position solutions X_j, specifically by employing consistency constraints to partition unambiguous delay states and using a time-delay likelihood function to estimate the unambiguous delays for each buoy. At this stage, the set of position solutions X_j forms a two-dimensional convex polygon, where the area $S(N_j)$ of this convex polygon is expressed as:

$$S(N_j) = \frac{1}{2}\left|\sum_{l=1}^{N_s}(x_{sl} \cdot y_{sl+1} - x_{sl+1} \cdot y_{sl})\right| \tag{8}$$

For a correct Unambiguous combination, its set of position solutions X_j should exhibit strong spatial consistency. Therefore, initial pruning of ambiguity count combinations N_j can be performed based on the area $S(N_j)$ of their corresponding 2D convex hull. Given a threshold Λ_s, if J' ambiguity-free combinations satisfy condition $S(N_{j'}) < \Lambda_s, j' = 1, 2, ..., J'$, a delay likelihood function is constructed for the pruned combinations $N_{j'}$ for secondary decision-making.

For the remaining Unambiguous combinations $N_{j'} = [n_1', n_2', ..., n_i', ..., n_N'], n_i' = 0, 1, ..., N_{\max}$, the trend function $f_i(t_r^i(k))$ representing the change in ambiguity period for the i-th buoy in the k-th frame is defined as:

$$f_i(t_r^i(k)) = \begin{cases} n_i'(k) & k = 1 \\ n_i'(k-1) - \dfrac{t_r^i(k) - t_r^i(k-1)}{T} & k > 1 \end{cases} \tag{9}$$

Where, $t_r^i(k)$ denotes the propagation delay received by the $i-th$ buoy in the k-th frame. $t_r^i(k-1)$ represents the propagation delay received by the i-th buoy in the k-th frame. $n_i'(k)$ is the ambiguity number for the i-th buoy in the k-th frame. and $n_i'(k-1)$ is the ambiguity number for the i-th buoy in the $k-1$-th frame.

Based on the ambiguity period trend function $f_i(t_r^i(k))$ for the i-th buoy in the k-th frame, the time-delay likelihood function $f_i(t_r^i(k))$ for the i-th buoy in the k-th frame is derived as:

$$L(n_i'(k) \mid t_r^i(k)) = \exp\left(-\frac{(n_i'(k) - f_i(t_r^i(k)))^2}{2}\right) \tag{10}$$

Eq.(11) reflects the consistency between the unambiguous number $n_i'(k)$ of the k −th frame buoy i and the predicted value of the trend function for the unambiguous number $n_i'(k)$. Further, the consistent joint likelihood function $L(N_{j'} \mid T_r)$ including N_p buoys is established as:

The unambiguous number combination $N_{j'}$ which maximizes the joint likelihood function $L(N_{j'} \mid T_r)$ is selected as the unambiguous number combination of this frame. Based on the unambiguous number combination $N_{j'}$ and the propagation delay sequence T_r of the k −th frame, the unambiguous delay estimation of the k −th frame is obtained. Combined with the array buoy coordinate A_i, the target equation is constructed as follows:

$$E = \frac{1}{N}\sum_{i=1}^{N}(R_i - D_i)^2 = \frac{1}{N}\sum_{i=1}^{N}(\sqrt{(x_s - x_i)^2 + (y_s - y_i)^2 + (z_s - z_i)^2} - c \cdot t_m^i)^2 \tag{11}$$

The steepest descent method seeks to iteratively find the minimum of the objective function E. Based on the principles of the steepest descent method, the iterative solution formula is constructed

and expressed as:

$$p^{k+1} = p^k + \mu^k g^k \tag{12}$$

Where p^k is the target position (x_s^k, y_s^k, z_s^k) of iteration k, μ^k is the step size of iteration k, g^k is the negative gradient of the objective function at v^k, and the expression of g^k is:

$$g^k = \begin{bmatrix} \frac{\partial E(x_s^k, y_s^k, z_s^k)}{\partial x_s^k} \\ \frac{\partial E(x_s^k, y_s^k, z_s^k)}{\partial y_s^k} \\ \frac{\partial E(x_s^k, y_s^k, z_s^k)}{\partial z_s^k} \end{bmatrix} = 2 \begin{bmatrix} \sum_{i=1}^{N} \frac{(x_s^k - x_i)(D_i - R_i^k)}{R_i^k} \\ \sum_{i=1}^{N} \frac{(y_s^k - y_i)(D_i - R_i^k)}{R_i^k} \\ \sum_{i=1}^{N} \frac{(z_s^k - z_i)(D_i - R_i^k)}{R_i^k} \end{bmatrix} \tag{13}$$

The steepest descent method iteratively searches for the minimum of the objective function according to the equation, proceeding along the local negative gradient direction of the objective function (7,8). Based on practical conditions, a mean squared error threshold and a maximum iteration count are set. The iteration stops when the mean squared error falls below the threshold, and the solution at that point is taken as the target position. If the maximum iteration count is reached without the mean squared error falling below the threshold, the final iteration result is still used as the target position. This yields the position of the underwater target in the $k-th$ frame, thereby determining the target's motion trajectory and achieving high-precision underwater target localization.

3. Simulation and Results

To verify the effectiveness of the proposed algorithm, an underwater acoustic source target motion scenario was constructed. In this scenario, the time delays between the target and the long baseline (LBL) buoys were processed for range ambiguity resolution and correction, ultimately achieving underwater target localization.

In the simulation experiment, the spacing between LBL buoys was set to 5000 m, and the underwater target moved at a constant velocity along a straight path. The acoustic beacon mounted on the target transmitted interrogation signals at a period of 2 s. Under these conditions, the system's time delay measurements could introduce up to three cycles of range ambiguity, posing a more challenging test for the positioning algorithm. This setup was designed to validate the algorithm's effectiveness in scenarios involving multiple ambiguity periods.

3.1 Simulation Conditions

The simulated scenario consists of one moving acoustic source target and four long baseline (LBL) buoys, with their coordinates listed in Table 1. The target moves within the array's coverage area, following a trajectory that includes both straight-line and curved motions. In the defined coordinate system, the target's initial position is set at (200, 2500) m, with a depth of 60 m.

The detailed simulation conditions are provided in Table 2. The time delay measurement errors include: Gaussian-distributed noise as the primary error source. A 5% probability of signal miss-detection, introducing an error of up to 20 ms. A 5% probability of multipath interference, causing an additional 20 ms delay error. Furthermore, the buoy position measurement errors and sound speed estimation errors both follow Gaussian distributions, with standard deviations of 10 m and 10 m/s, respectively.

Table 1 Array buoy coordinates

	X(m)	Y(m)	Z(m)
Buoy 1	0	0	40
Buoy 2	5000	0	40
Buoy 3	5000	5000	40
Buoy 4	0	5000	40

Table 2 Simulation conditions

Time delay measurement error	5ms
Time delay missed detection probability	5%
Time delay false alarm probability	5%
Maximum time delay false alarm error	20ms
Buoy position measurement error	10m
Sound velocity measurement error	10m/s

The simulation scenario is conducted in a 5000 m × 5000 m area with four array buoys deployed at optimized positions to ensure observability and localization accuracy. The target initiates its motion from the initial position. Its trajectory consists of the following phases: Straight-line motion: The target moves at a constant velocity along a linear path. Two maneuvering turns: The target performs two consecutive turns with predefined radii and angular velocities to simulate realistic dynamic behavior. Final straight-line motion: After the maneuvers, the target resumes linear motion to complete the trajectory. Figure 3 illustrates the target's trajectory,. The trajectory is designed to test the robustness of the long-baseline localization system under both steady-state and dynamic conditions.

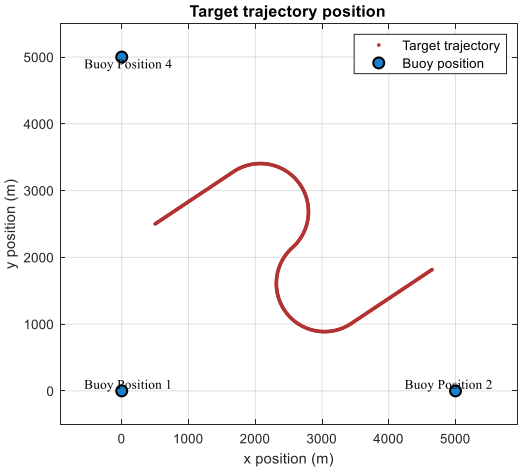


Figure 3 Target trajectory

3.2 Simulation analysis

Figure 4 presents the measured time delays across the four channels of the long baseline (LBL) hydrophone array. These time-delay measurements undergo tracking-based interpolation to enable stable acoustic source localization.

Figure 5 demonstrates the proposed anti-ambiguity algorithm for resolving distance ambiguities. In this figure, each geometric marker represents a unique ambiguity number combination, while the red dot denotes the selected combination for the current frame. This selection is achieved through a likelihood-based consistency test that evaluates the temporal evolution of the ambiguity combinations. Notably, the correct ambiguity number combination exhibits a distinct and coherent distribution in the two-dimensional convex hull domain, whereas incorrect combinations manifest as irregular outliers.

The proposed method leverages a consistency likelihood function to dynamically track the convex hull variation pattern associated with the correct ambiguity combination. By exploiting this regularity, the algorithm achieves ambiguity-resistant localization of a moving acoustic source.

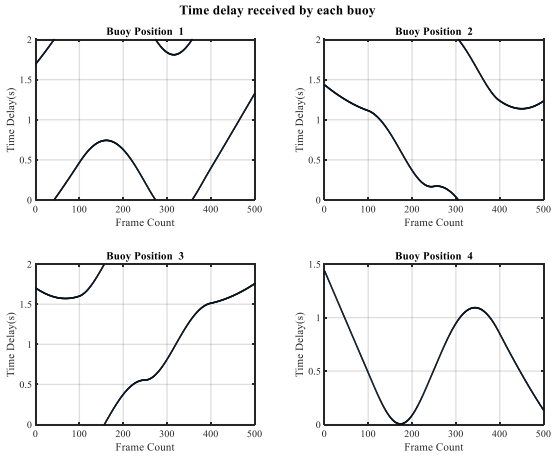


Figure 4 Receiving delay of each array buoy

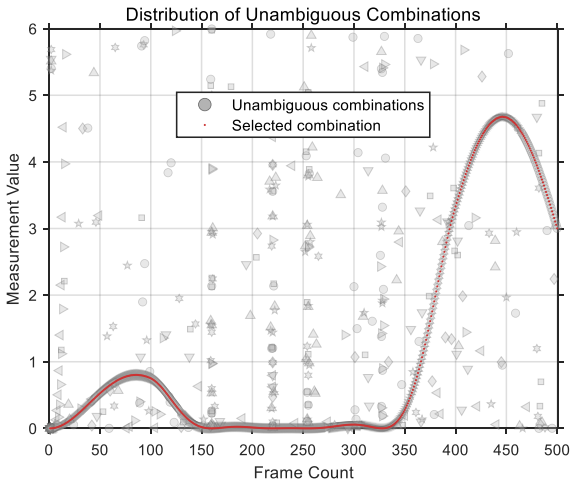


Figure 5 Distribution of unambiguity number combinations

After obtaining the correct ambiguity number combination, the precise position of the acoustic source can be determined by solving the time delays measured across all buoys using the steepest descent method. This iterative optimization algorithm minimizes the residual error function derived from the time-difference-of-arrival (TDOA) equations, efficiently converging to the global solution when initialized within the vicinity of the true source location.

To evaluate the performance of the steepest descent algorithm in practical scenarios, Figure 6 illustrates both the search path trajectory and the spatial distribution of localization errors across a 5000 m × 5000 m operational area. The search path demonstrates the algorithm's convergence behavior, typically characterized by orthogonal steps toward the gradient-descent direction, while the error distribution reveals systematic biases induced by buoy geometry and stochastic noise in delay measurements. Notably, errors are observed to amplify near the periphery of the buoy array due to geometric dilution of precision (GDOP), whereas sub-meter accuracy is achievable near the array center where sensor coverage is optimal.

The steepest descent solver's robustness against local minima in this context is attributed to the convexity of the TDOA cost function after ambiguity resolution, though its convergence rate remains sensitive to the initial guess. This analysis underscores the method's suitability for large-scale underwater acoustic localization, if ambiguity numbers are pre-resolved and sensor configurations are well-conditioned.

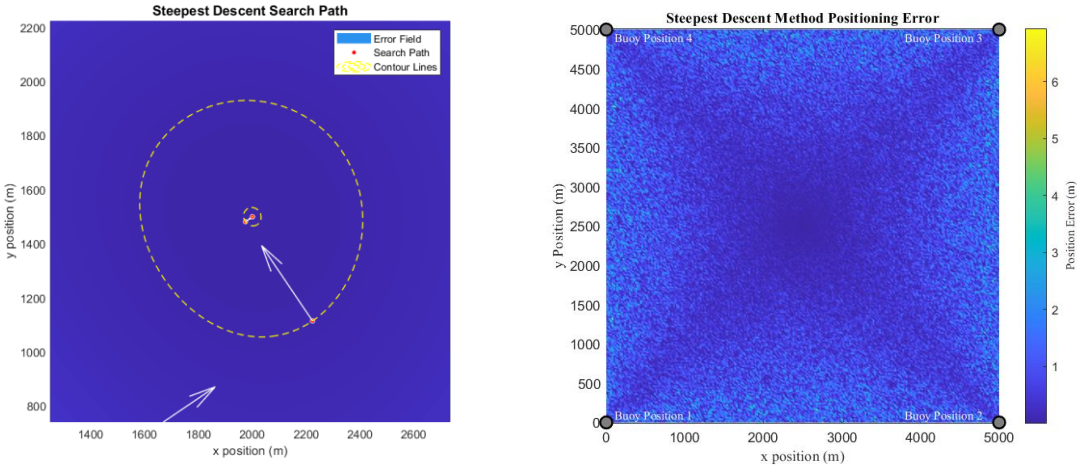
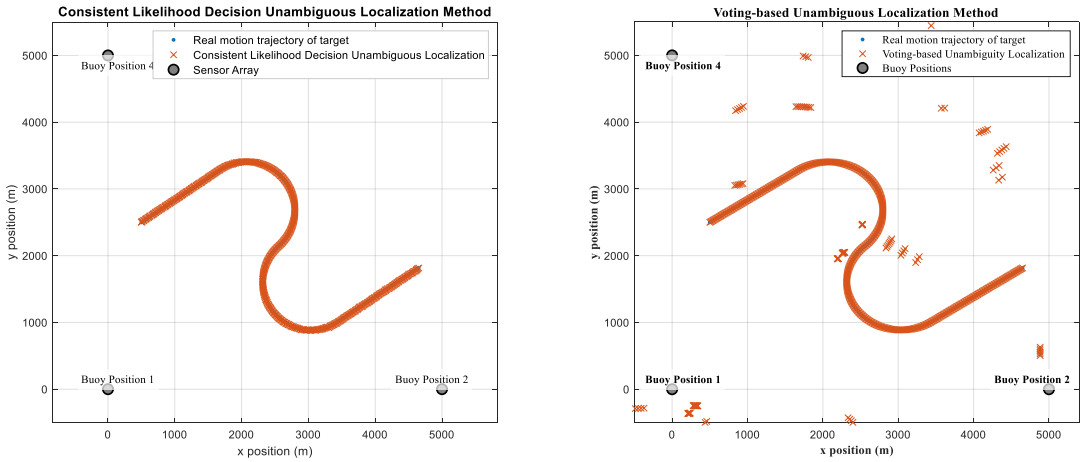


Figure 6 Steepest descent search path



(a) (b)

Figure 7 Comparison diagram of positioning trajectory
(a)Consistent likelihood decision method. (b)Voting-based method.

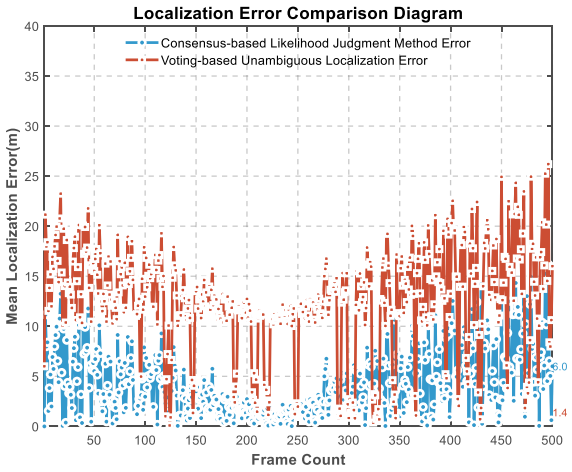


Figure 8 Comparison diagram of positioning error

Figure 7 presents a comparative analysis of the positioning results between the proposed algorithm and the voting-based ambiguity resolution method. As shown in Figure 7(b), the voting method, which relies solely on redundant buoys' time delays for ambiguity resolution, exhibits two significant limitations: high computational complexity throughout the entire navigation cycle, and susceptibility to erroneous judgments. Multiple instances of incorrect ambiguity number determinations can be observed at various time points. Furthermore, the multi-value effect within the buoys' detection range leads to numerous outliers in the positioning trajectory, ultimately degrading the algorithm's overall positioning performance.

A comparison of positioning errors between the proposed method and the voting-based ambiguity resolution method is presented in Figure 8. The results demonstrate that under identical conditions, the proposed method achieves superior positioning accuracy compared to the reference algorithm. This improvement stems from two key factors: reduced misjudgment in ambiguity period determination, and minimized multi-value effects in subsequent positioning calculations.

4. CONCLUSIONS

This paper proposes an ambiguity-free localization method for long baseline acoustic sources based on consistency likelihood decision. To address the range ambiguity problem in long baseline acoustic source positioning, the proposed algorithm employs consistency constraints to partition unambiguous delay states and utilizes a time-delay likelihood function to estimate the unambiguous delays of each buoy. By maintaining multiple hypotheses at the initial stage and performing consistency likelihood tests at subsequent stages, the method resolves the true combination of ambiguity numbers, transforming the anti-ambiguity problem into a nonlinear optimization problem constrained by unambiguous cycle counts. This approach effectively resolves range ambiguity while

maintaining computational efficiency. Simulation experiments demonstrate the effectiveness of the proposed method in achieving ambiguity-free localization for underwater acoustic targets under conditions where the initial position is uncalibrated or subject to significant calibration errors.

ACKNOWLEDGEMENTS

This work is supported by the National Natural Science Foundation of China under Grant Nos. 62271162.

REFERENCES

1. X. Li, Y. Wang, B. Qi, et al, “Long baseline acoustic localization based on track-before-detect in complex underwater environments,” IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 1-14.
2. G. L. LIANG, S. LI, N. ZOU, et al, “Locational lot number association algorithm for multi-station joint observation of non-cooperative targets,” ACTA ACUSTICA, 2024, 49(5): 1005-1018.
3. KORSHUNOV A Y, FRIDMAN L B, SINITSIN E A. Effectiveness of range ambiguity mitigation in primary radars when using signals with high pulse-repetition frequency[C]/ / Proceedings of the 23rd international conference on microwave and telecommunication technology. Washington DC,2013: 1154-1155.
4. RIEGER P. Range ambiguity resolution technique applying pulse-position modulation in time-of-flight scanning lidar applications[J]. Optical engineering,2014,53(6): 061614.
5. HUI Juan, LIANG Guolong. Technology of resolving range ambiguities of active spherical positioning model combined with passive hyperboloid ranging model[J]. Journal of Harbin engineering university,2007,28(6): 665-668.
6. DIAMANT R, LAMPE L. Underwater localization with time synchronization and propagation speed uncertainties[J]. IEEE transactions on mobile computing,2013, 12(7): 1257-1269.
7. CARROLL P, MAHMOOD K, ZHOU Shengli, et al. On-demand asynchronous localization for underwater sensor networks [J]. IEEE transactions on signal processing,2014,62 (13): 3337-3348.
8. X. Li, Y. Wang, B. Qi, et al, “Underwater multi-target passive acoustic localization based on multi-sensor collaboration,” ACTA ACUSTICA, 2024, 49(1): 16-27.

High-precision underwater source localization based on motion compensation and ray tracing

Binchao YANG[1,2,3]; Yan WANG[1,2,3]; Bin QI[1,2,3]; Dong XU[1,2,3]; Xiang LI[1,2,3]
[1] College of Underwater Acoustic Engineering, Harbin Engineering University, China
[2] National Key Laboratory of Underwater Acoustic Technology, Harbin Engineering University, China
[3] Key Laboratory of Marine Information Acquisition and Security(Harbin Engineering University), Ministry of Industry and Information Technology, China

ABSTRACT
In complex underwater acoustic environments, source localization accuracy is often limited by time-delay errors caused by target motion and acoustic ray-bending effects due to variations in the sound speed. To address the mismatch between signal transmission and reception positions in long baseline (LBL) systems operating in interrogation–response mode, this paper proposes a high-precision localization algorithm that integrates radial velocity compensation and ray tracing. The algorithm first estimates the one-way travel time based on the target's radial velocity, aligning the localization point with the start of the synchronization cycle. A Kalman filter is employed to estimate the target's state vector, and the radial velocity is reconstructed using geometric relationships. An adaptive error feedback mechanism is introduced to further enhance the accuracy of radial velocity estimation. To mitigate the effects of acoustic ray bending, a ray tracing table-lookup method based on a stratified sound speed profile is employed, significantly reducing distance estimation errors caused by acoustic refraction. Simulation results demonstrate that the proposed method substantially improves localization accuracy compared to conventional navigation-based algorithms, while maintaining the real-time performance required for practical underwater acoustic applications.

Keywords: Underwater source localization, Radial velocity compensation, Kalman filter

1. INTRODUCTION
Long Baseline (LBL) systems employ networks of seafloor-deployed beacons to provide reliable three-dimensional positioning for underwater targets. With their increasing use in deep-sea operations, growing demands on positioning accuracy and response speed arise due to faster and deeper ocean developments. Therefore, improving the localization accuracy and environmental adaptability of unmanned underwater vehicles (UUVs) has become a key research focus.

For effective autonomous navigation, UUVs must acquire precise real-time spatial position and motion data using high-accuracy underwater localization methods. However, the continuous motion of underwater targets results in discrepancies between signal transmission and reception locations in interrogation–response mode. Ignoring the target's displacement during signal propagation may introduce substantial localization errors, compromising navigation reliability.

To mitigate this issue, various methods have been proposed, including target displacement prediction and beacon clock synchronization. Nevertheless, displacement prediction is highly dependent on sensor accuracy, while clock synchronization demands additional hardware, increasing system cost and complexity (1,2).

To address motion-induced localization errors, this paper introduces a compensation method based on radial velocity information. Leveraging the timing structure of the interrogation–response mode, the method estimates the target's radial velocity relative to each beacon from system-measured time delays and constructs a corresponding correction term. Without requiring clock synchronization or additional sensors, it enables dynamic correction of the one-way travel time, thereby enhancing localization accuracy. This strategy further incorporates a Kalman filter for state vector estimation,

[1] yangbc926@163.com
[2] qibin@hrbeu.edu.cn

with radial velocity precision enhanced via an adaptive error feedback mechanism. To mitigate acoustic ray bending caused by sound speed gradients, ray tracing is integrated into the system to correct the signal propagation path.

This approach improves localization accuracy while satisfying real-time navigation requirements, demonstrating superior adaptability and feasibility compared to existing methods.

2. System Architecture and Problem Formulation

2.1 Localization Principle of the LBL System

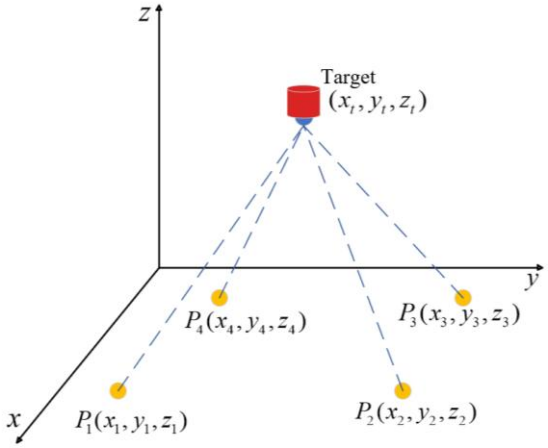


Figure 1 – Schematic Diagram of the LBL System Principle

The core principle of the LBL system lies in multi-beacon acoustic ranging and geometric positioning. As illustrated in Figure 1, the system establishes a stable reference coordinate frame by deploying multiple acoustic beacons with known positions on the seafloor. The target node engages in two-way acoustic communication with these beacons to measure signal travel times and estimate its distances to each beacon, thereby enabling position calculation (3).

The target's depth can be obtained in real time using a depth sensor. With valid distance measurements from at least three beacons, its position can then be determined using the spherical intersection method.

Establish a set of localization equations based on beacon positions and distance measurements:

$$(x_t - x_n)^2 + (y_t - y_n)^2 + (z_t - z_n)^2 = l_n^2 \quad (n = 1,2,3) \tag{1}$$

In Eq. (1), (x_t, y_t, z_t) denotes the true location of the underwater target, (x_n, y_n, z_n) is the known coordinate of the nth beacon, and l_n is the distance between the target and the nth beacon. Typical solution methods include least squares (LS) and weighted least squares (WLS).

2.2 Timing Structure of the Interrogation–Response Mode

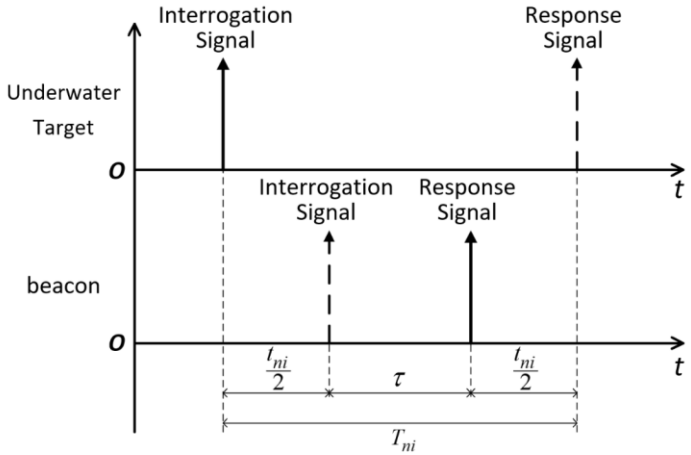


Figure 2 – Timing Structure of the Interrogation–Response Mode

Interrogation–response mode enables ranging and localization without strict clock synchronization among system devices (4). The basic process proceeds as follows:

1. Interrogation Signal Transmission: At the start of the ith synchronization cycle, the target node

transmits an interrogation signal to the seafloor beacons.

2. Signal Reception and Delayed Response: Upon receiving the interrogation signal, each beacon replies with a response signal after a preset delay τ.

3. Response Signal Reception: The target node receives the response signal from the nth beacon and records the total elapsed time T_{ni} from transmission to reception.

4. One-way Distance Calculation: The two-way travel time t_{ni} is obtained by subtracting the beacon's delay τ from the total elapsed time T_{ni}. Assuming the target displacement is negligible, the one-way distance to the beacon is calculated as $l_n = ct_{ni} / 2$, where c denotes the sound speed.

The timing structure of this operation mode is illustrated in Figure 2, depicting the signal interaction process between the target node and the beacons. Since accurate localization relies on the precise calculation of one-way travel distances, any factor that affects the acoustic propagation path or travel time may introduce localization errors. The next section analyzes the sources of such system errors in detail, followed by a focused discussion on corresponding compensation methods in subsequent chapters.

2.3 Analysis of Localization Error Sources

The LBL system estimates the distance between the target and seafloor beacons by measuring acoustic signal travel time, thereby enabling three-dimensional localization. In practical scenarios, the ranging process is influenced by various sources of error. The primary error sources can be categorized into the following two types.

1. Geometric Errors Induced by Target Motion

As shown in Figure 3, the target transmits the interrogation signal at the start of the synchronization cycle. Since the target continues to move during this period, there is a mismatch between its positions at the times of signal transmission and response reception. Directly applying the one-way distance calculation from Section 2.2 leads to significant geometric errors in the ranging result.

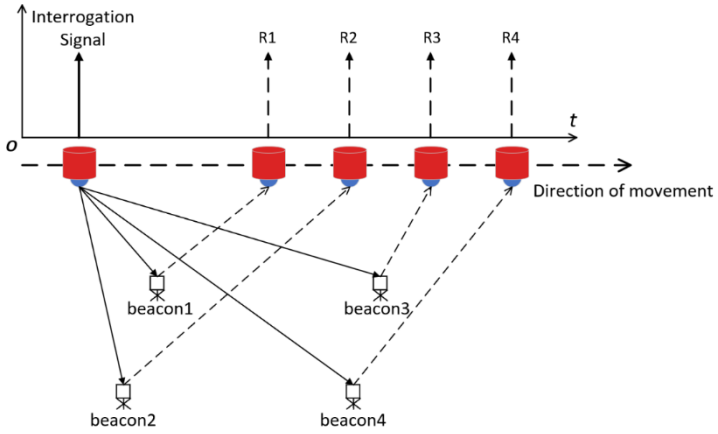


Figure 3 – Target Movement Variation During Signal Reception

2. Propagation Path Deviation due to Sound Speed Variation

The sound speed in water is influenced by factors such as temperature, salinity, and pressure, typically forming a depth-dependent gradient that causes acoustic ray bending (5). If a straight-line propagation model is used in localization calculations, deviations caused by ray bending can degrade ranging accuracy.

To enhance the localization accuracy of LBL systems in dynamic target tracking and non-uniform sound speed environments, the next chapter will focus on radial velocity compensation methods and ray tracing-based path correction methods.

3. Motion Compensation and Ray Tracing Techniques

To address geometric errors induced by target motion, the target's radial velocity relative to each beacon is introduced to compensate for the one-way propagation time of acoustic signals. Meanwhile, ray tracing techniques are applied to correct the bending of acoustic rays due to sound speed variations.

3.1 Radial Velocity Estimation Method

Figure 4 illustrates the target's signal transmission and reception during adjacent synchronization cycles. r_{i1}, r_{j1}, r_{k1} represent the ranges between the target and the beacons at the time of interrogation

transmission, while r_{i2}, r_{j2} denote the ranges at response reception. T denotes the synchronization cycle. Based on the geometric relationships between the target and the beacons, the radial velocity of the target relative to each beacon is given by:

$$v_r = (r_{k1} - r_{i1}) / 2T \quad (k = i + 2) \tag{2}$$

However, in practical applications, the system cannot predict the distance between the target and the beacon after two synchronization cycles. Therefore, the term r_{k1} in Eq. (2) cannot be obtained in real time and must be approximated using currently available information.

In deep-sea environments, LBL beacons are typically deployed at depths of 1000 – 4000 m, resulting in long acoustic propagation paths. A typical UUV cruises at 3 – 4 m/s under high-speed operation. Assuming a synchronization cycle of 2 s, the target's displacement within a cycle is approximately 6 – 8 m. Compared to propagation distances of several kilometers, this displacement is negligible. Therefore, the radial velocity expression can be linearly approximated.

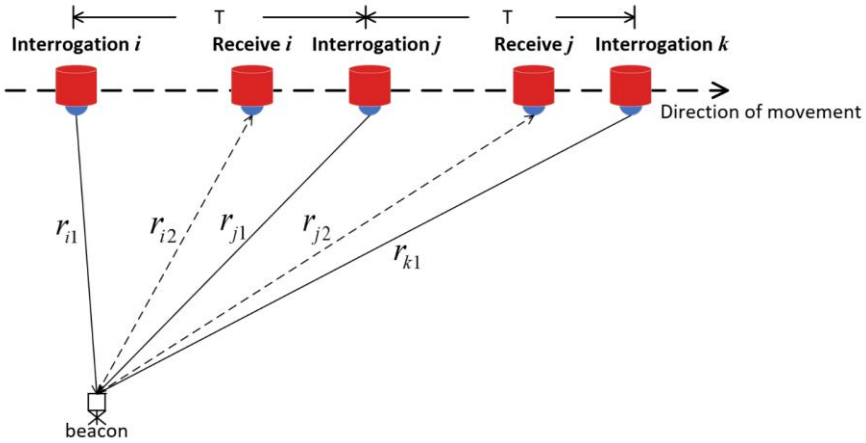


Figure 4 – Signal Interaction During Two Cycles

Assuming relatively steady target motion during localization, the radial displacement over two synchronization cycles can be approximated from the geometry in Figure 4 as:

$$r_{k1} - r_{i1} = (r_{k1} - r_{j1}) + (r_{j1} - r_{i1}) \approx (r_{j2} - r_{i2}) + (r_{j1} - r_{i1}) = R_j - R_i \tag{3}$$

Substituting Eq. (3) into Eq. (2) leads to an approximate expression for the radial velocity. In Eq. (3), R_i and R_j represent the two-way travel distances of the signals. In Eq. (4), t_i and t_j represent the two-way travel times of the acoustic signals.

$$\hat{v}_r = (t_j - t_i)c / 2T \quad (j = i + 1) \tag{4}$$

In summary, given the target's limited displacement within a synchronization cycle, its radial velocity relative to each beacon can be approximated as a function of the two-way travel time. This approximation allows the construction of a velocity estimation model based on observable data. The next section presents the calculation procedure for one-way travel time correction.

3.2 One-Way Travel Time Correction Algorithm

Based on the radial velocity approximation model, this section presents a one-way travel time correction algorithm to compensate for geometric errors caused by target motion. The algorithm estimates the target's radial velocity relative to each beacon using two-way travel times from two consecutive synchronization cycles and corrects the one-way travel time by incorporating the target's motion trend.

The expression for the two-way travel distance R_i is derived from the target's trajectory during the ith synchronization cycle, as shown in Figure 4:

$$R_i = r_{i1} + r_{i2} = r_{i1} + [r_{i1} + (t_i + \tau)] = 2r_{i1} + (t_i + \tau)v_r \tag{5}$$

Variables in Eq. (5) are as previously defined.

As shown in Figure 3, the target is typically at different positions when receiving response signals from different beacons. To unify the time reference, the localization point is set at the start of the synchronization cycle, where the one-way distance is defined as r_{i1}. Substituting the estimated radial velocity $\hat{v}_r$ into the expression of R_i and simplifying provides the corrected one-way travel distance.

$$r_{i1} = c[t_i / 2 + (t_i + \tau)(t_i - t_j) / 4T] \tag{6}$$

Dividing both sides of Eq. (6) by the sound speed c gives the corrected one-way travel time t_{ri}.

$$t_{ri} = t_i / 2 + (t_i + \tau)(t_i - t_j) / 4T \tag{7}$$

At this point, the derivation of the one-way travel time correction algorithm is complete. By accounting for the target's motion during the signal round-trip, the algorithm improves the spatiotemporal consistency of localization data.

3.3 Ray Tracing in Stratified Media

To model the complex sound speed profile in water, this paper adopts a stratified medium model that divides the profile into layers with nearly constant sound speed gradients (6). In each layer, acoustic rays are approximated as arcs, enabling reconstruction of propagation paths for improved calculation and analysis.

In stratified media, acoustic ray propagation follows Snell's law:

$$\frac{c_0}{\cos\theta_0} = \frac{c_1}{\cos\theta_1} = \cdots = \frac{c_h}{\cos\theta_h} = \cdots = \text{const} \tag{8}$$

In Eq. (8), c_0 and θ_0 denote the sound speed and grazing angle at the ray's origin, while c_h is the sound speed in the hth layer, and θ_h is the grazing angle at the lower boundary of the hth layer.

3.4 Calculation of Ray Propagation Parameters

To obtain the initial grazing angle, this paper employs underwater acoustic simulation software to model the sound field and extract the earliest-arriving direct ray at the target point. The angle between the incident direction of this ray and the horizontal plane is taken as the initial grazing angle for ray tracing. Based on the known sound speed profile, grazing angles at any depth can be calculated using Snell's law (7).

In a constant sound speed gradient layer, the acoustic ray follows a circular arc, and the corresponding horizontal distance is:

$$x_h = \frac{c_0}{\cos\theta_0 \cdot g_h} \left| \sin\theta_{h-1} - \sin\theta_h \right| \tag{9}$$

The travel time of the acoustic ray in this layer is:

$$t_h = \frac{1}{2g_h} \left| \ln\frac{1+\sin\theta_{h-1}}{1-\sin\theta_{h-1}} - \ln\frac{1+\sin\theta_h}{1-\sin\theta_h} \right| \tag{10}$$

In Eq. (9) and Eq. (10), g_h denotes the sound speed gradient in the hth layer, while θ_{h-1} and θ_h represent the grazing angles at the upper and lower boundaries, respectively.

Dividing the sound speed profile into H layers, the total horizontal propagation distance and travel time of the ray can be expressed as:

$$x = \sum_{h=1}^{H} x_h \ , \quad t = \sum_{h=1}^{H} t_h \tag{11}$$

Eq. (11) shows the relationship between ray travel time and horizontal distance, enabling the creation of a ray correction table. Using the corrected one-way travel time from Section 3.2, the corresponding horizontal distance can be retrieved from the table. Since the table is discrete, interpolation methods can be applied in practice to improve lookup accuracy.

After obtaining the horizontal distance, the slant range between the target and each beacon can be determined by combining the known beacon depth and the target depth measured by the depth sensor. Additionally, the equivalent sound speed can also be derived. These slant ranges are then substituted into the localization equations to solve for the target's three-dimensional position.

The ray correction table can be generated offline and stored in the system before localization, avoiding repeated calculations and ensuring real-time computational efficiency.

4. State Estimation and Error Feedback

Although the target's 3D position has been obtained, frame-by-frame localization is susceptible to noise, compromising trajectory continuity and stability. To enhance the continuity of the localization trajectory, this paper adopts a Kalman filter for recursive estimation of the target's motion state.

4.1 Kalman Filter-Based State Estimation

In underwater localization scenarios, the target's vertical motion can be approximated as static, with movement primarily in the horizontal plane. This section focuses on estimating the two-

dimensional motion state, applying joint filtering to its horizontal position and velocity. The state vector is defined as $\mathbf{X}_m = [x_m, \dot{x}_m, y_m, \dot{y}_m]^T$. x_m and y_m represent the target's planar coordinates at frame m, and $\dot{x}_m$, $\dot{y}_m$ denote the corresponding velocity components (8).

Underwater targets are typically non-maneuvering and can be approximated as moving at constant velocity over short time scales. Accordingly, a constant-velocity model is used to construct the state-space representation, which is given by:

$$\begin{cases} \mathbf{X}_m = \mathbf{F}\mathbf{X}_{m-1} + w_{m-1} \\ \mathbf{Z}_m = \mathbf{H}\mathbf{X}_m + v_m \end{cases} \tag{12}$$

In Eq. (12), $\mathbf{X}_m$ is the state vector at frame m, $\mathbf{Z}_m$ is the observation vector, w_{m-1} denotes process noise, v_m denotes observation noise. $\mathbf{F}$ and $\mathbf{H}$ are the state transition and observation matrices, respectively.

The target's initial state $\mathbf{X}_0$ and error covariance matrix P_0 are typically known beforehand and used to initialize the Kalman filter. The filtering process involves two steps: state prediction and measurement update, carried out recursively as follows:

Prediction stage:

$$\hat{\mathbf{X}}_m^- = \mathbf{F}\hat{\mathbf{X}}_{m-1} \tag{13}$$

$$P_m^- = \mathbf{F}P_{m-1}\mathbf{F}^T + Q \tag{14}$$

Update stage:

$$K_m = P_m^-\mathbf{H}^T(\mathbf{H}P_m^-\mathbf{H}^T + R)^{-1} \tag{15}$$

$$\hat{\mathbf{X}}_m = \hat{\mathbf{X}}_m^- + K_m(\mathbf{Z}_m - \mathbf{H}\hat{\mathbf{X}}_m^-) \tag{16}$$

$$P_m = (\mathbf{I} - K_m\mathbf{H})P_m^- \tag{17}$$

In these equations, $\hat{\mathbf{X}}_m^-$ denotes the predicted state vector at frame m, and $\hat{\mathbf{X}}_m$ is the updated state vector. P_m^- and P_m represent the predicted and updated error covariance matrices, respectively. K_m is the Kalman gain. The process noise covariance matrix Q and the observation noise covariance matrix R characterize the statistical properties of model uncertainty and measurement noise.

The system now provides dynamic estimation of the underwater target's position and velocity.

4.2 Error Feedback Correction Algorithm

To improve radial velocity estimation, this section introduces an error feedback mechanism that dynamically adjusts the estimate using the state vector output by the Kalman filter. The following describes the system's processing procedure within a single synchronization cycle, with time subscripts omitted for clarity.

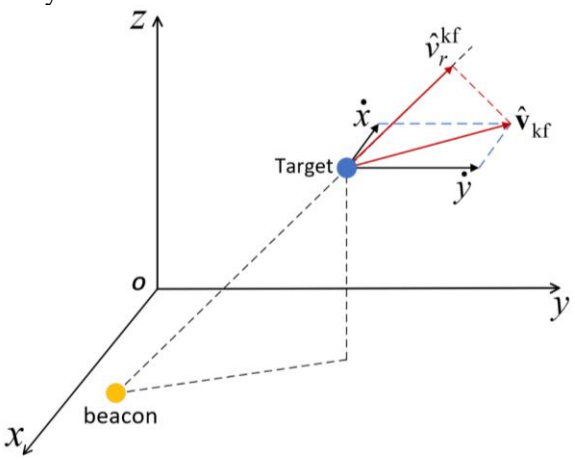


Figure 5 – Schematic of Radial Velocity Reconstruction

As shown in Figure 5, the target's planar velocity vector $\hat{\mathbf{v}}_{kf}$ is constructed using the filter output components $\dot{x}$ and $\dot{y}$. Since the target's vertical velocity is negligible, $\hat{\mathbf{v}}_{kf} = [\dot{x}, \dot{y}, 0]$. Assuming the target's spatial position at the current frame is $\mathbf{P}_t = [x, y, z]^T$, and the beacon's spatial position is

$\mathbf{P}_b = [x_b, y_b, z_b]^T$, the unit vector along the line connecting them is given by:

$$\mathbf{u}_r = \frac{\mathbf{P}_b - \mathbf{P}_t}{\|\mathbf{P}_b - \mathbf{P}_t\|} = \frac{1}{\sqrt{(x_b - x)^2 + (y_b - y)^2 + (z_b - z)^2}} \begin{bmatrix} x_b - x \\ y_b - y \\ z_b - z \end{bmatrix} \tag{18}$$

By projecting the velocity vector $\hat{\mathbf{v}}_{\text{kf}}$ onto the unit vector $\mathbf{u}_r$, the radial velocity component at the current frame can be reconstructed as:

$$v_r^{\text{kf}} = \mathbf{u}_r^T \cdot \hat{\mathbf{v}}_{\text{kf}} = \frac{(x_b - x)\dot{x} + (y_b - y)\dot{y}}{\sqrt{(x_b - x)^2 + (y_b - y)^2 + (z_b - z)^2}} \tag{19}$$

Based on the previously obtained radial velocity estimate $\hat{v}_r$ and the reconstructed radial velocity $\hat{v}_r^{\text{kf}}$, the error feedback correction model is constructed as follows:

$$v_r^* = \hat{v}_r + \eta \cdot \Delta v_r = \hat{v}_r + \eta(\hat{v}_r^{\text{kf}} - \hat{v}_r) \tag{20}$$

In Eq. (20), v_r^* is the corrected radial velocity, and η is the feedback gain factor, determined by the radial velocity error Δv_r and the estimation uncertainty. It is specifically defined as:

$$\eta = \frac{|\Delta v_r|}{|\Delta v_r| + \lambda} \tag{21}$$

The parameter λ quantifies the uncertainty associated with velocity estimation in the Kalman filter and is calculated as:

$$\lambda = \sqrt{\text{Var}(\dot{x}) + \text{Var}(\dot{y})} \tag{22}$$

In the expression for the parameter λ , $\text{Var}(\dot{x})$ and $\text{Var}(\dot{y})$ are the variances of the velocity components $\dot{x}$ and $\dot{y}$ at current frame, extracted from the diagonal elements of the error covariance matrix P in the Kalman filter. This adaptive feedback mechanism dynamically adjusts the correction magnitude according to the current velocity estimation error, enhancing radial velocity accuracy while ensuring stability throughout the correction process.

5. Experiments and Simulation Analysis

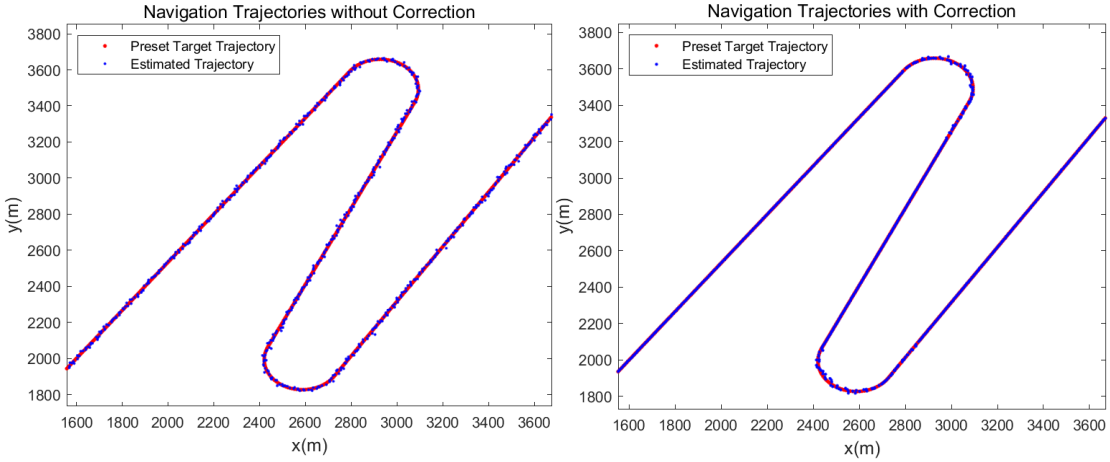
To validate the effectiveness of the proposed algorithm in typical underwater localization scenarios, this paper constructs a simulation system and evaluates its performance under representative parameter settings. The main configuration parameters of the simulation system are summarized as follows. The system operates with a synchronization cycle of 2 s and an observation duration of 1800 s. The target is located at a depth of 300 m, and four seabed beacons are arranged in a square with 5 km spacing at a depth of 3 km. The target's initial state is set as $\mathbf{X}_0 = [400, 3, 400, 4]$, and its true trajectory is shown in Figure 6. The simulation includes a 1 ms time-delay error, 1 m/s sound speed error, and 0.5 m beacon position error, all modeled as zero-mean Gaussian noise. A typical deep-sea sound speed profile is used.

Figure 6 – Preset Target Trajectory in Simulation

The simulation results presented in Figures 7 and 8 demonstrate that the proposed algorithm

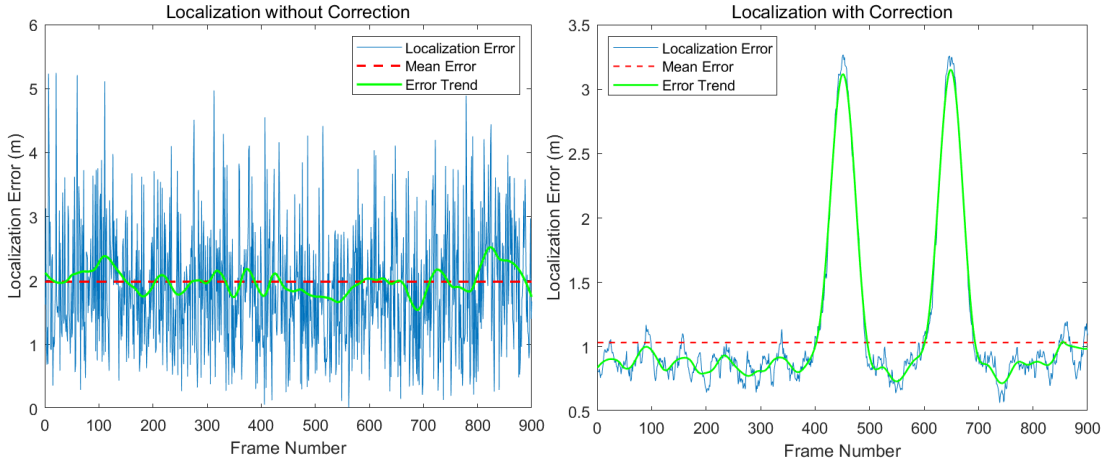
performs well in overall localization accuracy and effectively mitigates the effects of various error sources. The traditional algorithm exhibits significant fluctuations in localization error, which seriously affect the continuity of the navigation trajectory. Compared to the conventional method, the improved algorithm maintains a consistently lower error level for most of the time. Specifically, the average localization error of the conventional algorithm is about 2 m, while the improved algorithm achieves an average error of approximately 1 m, resulting in a smoother and more reliable navigation trajectory.

Nevertheless, due to the constant-velocity model used in the Kalman filter, model mismatch occurs when the target performs curved motion or experiences rapid state changes within a short period. In such cases, the filter's prediction accuracy degrades, causing temporary increases and fluctuations in localization error. Then, once the target returns to a relatively stable motion state, the filter corrects the errors and the system restores high-precision tracking capability. Overall, proposed localization method based on radial velocity estimation demonstrates enhanced stability and accuracy.



(a) Navigation Trajectory Before Correction (b) Navigation Trajectory After Correction

Figure 7 – Comparison of Navigation Trajectories



(a) Localization Error Before Correction (b) Localization Error After Correction

Figure 8 – Comparison of Localization Errors

6. CONCLUSIONS

This paper proposes a high-precision underwater target localization method for LBL systems, addressing localization errors caused by target motion and sound speed variations through the integration of motion compensation, one-way travel time correction, a Kalman filter with error feedback, and ray tracing. Simulation results indicate that the proposed method significantly outperforms traditional approaches in localization accuracy while enhancing trajectory continuity and system robustness. Overall, the proposed method delivers high localization accuracy in complex acoustic environments and suits typical applications such as UUV navigation and underwater target tracking. Future work will address model mismatch by exploring interacting multiple model algorithms to improve adaptability to varying motion patterns.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the support of the National Natural Science Foundation of China (Grant No. 62271162), which contributed importantly to the advancement of this research.

REFERENCES

1. Feng G, Shan Z, Xiang W. A localization algorithm based on modified observations for underwater mobile nodes. Journal of Chinese Inertial Technology. 2020,28(5): 603–7.
2. Wang X, Li J, Huang M, Feng H. Time delay correction algorithm for underwater positioning based on clock error modeling. Technical Acoustics. 2021,40(2): 194–8.
3. Chen H, He Z, Wang J. A high-precision real-time distance difference localization algorithm based on long baseline measurement. Journal of Marine Science and Engineering. 2024,12(10):1724.
4. Yu X, Qin H D, Zhu Z B. Underwater localization of AUVs in motion using two-way travel time measurements with unknown sound velocity. IEEE Transactions on Vehicular Technology. 2023,72(9): 11358–73.
5. Zhao S, Wang Z J, Liu H M. Investigation on underwater positioning stochastic model based on sound ray incidence angle. Acta Geodaetica et Cartographica Sinica. 2018,47(09): 1280–9.
6. Berger C R, Zhou S, Willett P. Stratification effect compensation for improved underwater acoustic ranging. IEEE Transactions on Signal Processing, 2008, 56(8): 3779–83.
7. Liu B S, Huang Y W. Principles of Underwater Acoustics. Beijing: Science Press, 2019.
8. Huang Y, Zhang Y, Xu B. A new adaptive extended Kalman filter for cooperative localization. IEEE Transactions on Aerospace and Electronic Systems, 2017, 54(1): 353–68.

PROCEEDINGS of the
16th International Conference on Theoretical and Computational Acoustics

August 18 to 22, 2025 BEXCO, Busan Korea

Analysis of uncertainty aspects for multi-megawatt turbines in acoustic measurements

Eunkuk Son [1]; Seungjin Kang[2]; Jinjae Lee [3]; Minsang Kang [4]; Songjune Lee [5]; Gwang-Se Lee[6]

[1,2,3,4,5,6] Wind Energy Research Department, Jeju Global Research Center (JGRC),

Korea Institute of Energy Research

ABSTRACT

Underwater acoustic detection performance is affected by the variability of ocean noise level, which fluctuates with factors such as ship density and weather. However, measuring ocean noise level in real time is time-consuming and often impractical in time-sensitive operations such as anti-submarine warfare. This study investigates ocean noise level variability associated with ship density and its impact on detection range, aiming to support reliable detection range estimation when direct ocean noise level measurements are unavailable

The predicted received level at a single receiver was calculated by summing acoustic intensities radiated from ships within the defined area. For this, the source level of each ship was modeled using the Ross power-law based on its specifications. Transmission loss from each ship to the receiver was computed using a parabolic equation-based acoustic propagation model. To evaluate the reliability of the predicted received level, its similarity to the measured received level was assessed by comparing the two, which showed a high degree of similarity. Therefore, it was used to assess detection range variation as a function of ship density. Detailed discrepancies and limitations will be discussed in the presentation.

PROCEEDINGS of the
16th International Conference on Theoretical and Computational Acoustics

August 18 to 22, 2025 BEXCO, Busan Korea

Review of 3D acoustic intensimetry for the sound source localization based on a compact microphone array

In-Jee JUNG[1]; Wan-Ho CHO[2]
[1]Korea Research Institute of Standards and Science, South Korea
[2]Korea Research Institute of Standards and Science, South Korea

Sound source localization based on three-dimensional(3D) acoustic intensimetry can be applied to various fields because it has the advantage of configuring a compact microphone array. The conventional technique is based on the active intensity calculated between the signals measured from the two adjacent microphones. However, it has the limitation that the estimation is possible in the low Helmholtz number range due to the finite difference error that occurs in the process of estimating the particle velocity. In addition, a large error occurs due to the spectral fluctuation caused by the coherent noise in the narrow frequency band. In this study, the results of studies to improve the accuracy of source localization through 3D sound intensimetry by compensating for the inherent errors are reviewed. Also, it is described that the case studies of implementing a compact microphone array and applying it to a real-world sound source localization problem. Finally, the considerations that should be taken into account when implementing 3D sound intensimetry in practice are discussed.

Keywords: acoustic intensimetry, sound source localization, microphone array

[1] injee@kriss.re.kr
[2] chowanho@kriss.re.kr

Pork adulteration detection in beef using an electronic nose system

Tajmal HUSSAIN [1]; Jongwon SEOK [2, *]
Department of Information and Communication Engineering, Changwon National University, Korea

ABSTRACT

Beef meat is a highly perishable food product that is often adulterated with pork, raising significant concerns related to food fraud, public health, and halal authentication, particularly for individuals who avoid pork for religious or other reasons. Therefore, it is important to ensure the purity of the meat before consumption. This study investigates a publicly available dataset collected using an electronic nose system and analyzed with machine learning algorithms to predict adulteration in beef meat. The dataset includes seven mixture combinations: 100% pure beef and beef mixed with pork at 10%, 25%, 50%, 75%, 90%, and 100% pure pork. The proposed hybrid model, combining one-dimensional convolutional neural networks (1D-CNN) and long short-term memory (LSTM) networks, achieved an accuracy of 98.80% in detecting adulteration. Overall, the findings suggest that the electronic nose system, combined with appropriate data handling and machine learning techniques, can effectively detect adulteration, providing a valuable tool for food quality control and the meat industry.

Keywords: E-nose, beef, adulteration detection, 1D-CNN+LSTM

1. INTRODUCTION

Food adulteration, or food fraud, refers to the intentional alteration of food products for economic gain, posing significant health risks such as diarrhea, allergies, and heart disease (1). Globally, foodborne illnesses affect approximately 600 million people annually and cause an estimated 56 million deaths (2). Meat is widely consumed for its flavor and nutritional benefits (3). However, adulteration practices, such as mixing beef with cheaper meats, result in substantial economic losses. For instance, in Europe, the adulteration of beef with horse meat led to financial losses amounting to $45 million (4). The widely publicized 2013 horse meat scandal, involving the fraudulent substitution of horse meat in beef products (5), raised serious concerns regarding food safety, ethics, and public trust, highlighting the urgent need for stronger measures to prevent fraud and adulteration. Currently, detecting illegal meat adulteration has become a critical focus in ensuring food quality. Although analytical techniques such as liquid chromatography and gas chromatography are available for detection, these methods are often complex, time-consuming, costly, and lacking in portability.

The electronic nose, which mimics the human olfactory system, has proven to be an effective approach for detecting adulteration in meat products (6). Its growing popularity in the food quality industry is largely due to its affordability, ease of use, rapid detection capabilities, and adaptability (7). An electronic nose typically consists of three main components: a gas sensor array, a data preprocessing unit, and a pattern recognition system (8). Several studies have investigated the use of electronic nose technology for detecting meat adulteration (6, 7, 9). However, these studies primarily rely on traditional machine learning models for pattern recognition. Such models often struggle to extract complex features from raw e-nose data. Moreover, the application of deep learning approaches remains limited in the existing literature.

This study aims to detect pork adulteration in beef using a hybrid model that combines 1D-CNN and LSTM networks. The performance of the proposed hybrid model is compared with that of traditional machine learning models and individual deep learning models, demonstrating superior accuracy, robustness, and overall effectiveness.

[1] tajmalhussain20237609@gmail.com
[2] jwseok@changwon.ac.kr (Corresponding author)

2. METHODOLOGY

The block diagram of the proposed system is presented in Figure 1. The system operates through four main stages: data collection, preprocessing, pattern recognition, and output generation.

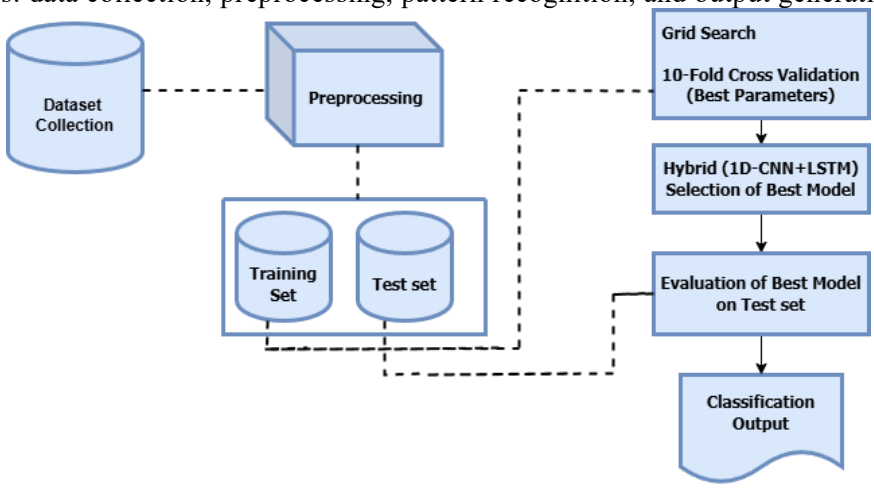


Figure 1 – Block Diagram of Proposed Methodology

2.1 Details of Dataset and Preprocessing

In this study, an electronic nose dataset developed by Sarno et al. (10) was employed to investigate the adulteration of beef with pork. The adulteration process involved mixing pork into beef at various concentration levels: 0%, 10%, 25%, 50%, 75%, 90%, and 100%. Volatile organic compounds emitted from the meat samples were captured using a metal oxide-based gas sensor array consisting of MQ138, MQ137, MQ136, MQ135, MQ9, MQ6, MQ4, and MQ2 sensors. In addition to detecting VOCs, the DHT22 sensor was used to record environmental parameters, specifically temperature and humidity.

To ensure data quality and enhance model performance, preprocessing was applied to the raw sensor dataset, which had a shape of (420, 60, 10), corresponding to 420 samples, each comprising 60-time steps and 10 sensor channels. Outliers were addressed using the Interquartile Range (IQR) (11), in which values outside the defined range were replaced with the median of the corresponding sensor's distribution. Subsequently, noise reduction was performed using the Discrete Wavelet Transform (DWT) (12). Each sample was processed using the 'sym4' wavelet function, which is well-suited for time-series signal processing. Soft thresholding was applied to the wavelet coefficients to suppress high-frequency noise while preserving essential signal characteristics. These preprocessing steps resulted in a cleaner and more robust dataset, providing a reliable foundation for subsequent feature extraction and classification tasks.

2.2 Proposed Hybrid 1DCNN-LSTM Model

The proposed approach detect adulteration by combining 1D-CNN (13) and LSTM (14). The model accepts input with a shape of (60, 10), corresponding to 60-time steps and 10 sensor channels. Initially, a 1D convolutional layer with 64 filters and a kernel size of 3 was applied to extract local features, followed by batch normalization and max-pooling to reduce dimensionality and enhance training stability. The output was then passed to two stacked LSTM layers with 128 and 64 units, respectively, to capture temporal dependencies within the sequence. To mitigate overfitting, dropout layers with a rate of 0.5 were incorporated. The dense layers consisted of a fully connected layer with 64 neurons activated by ReLU, followed by another dropout, and a final output layer with 7 neurons using softmax activation for multi-class classification. The architecture of the hybrid model is shown in Figure 2.

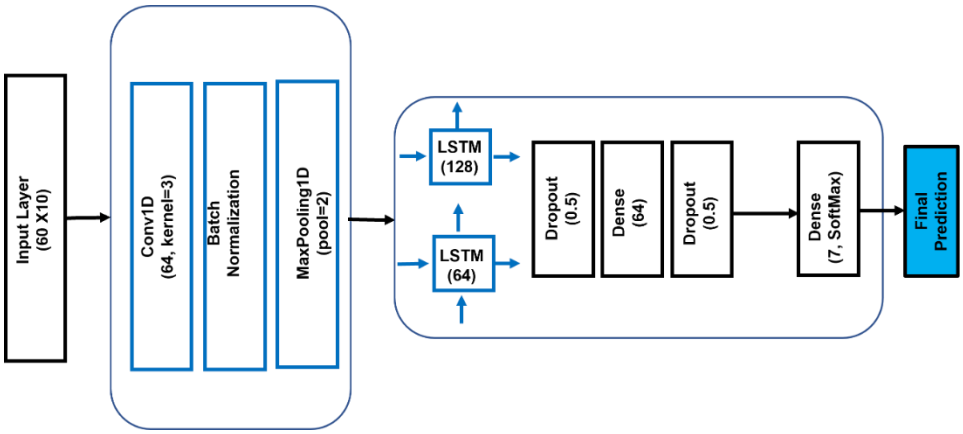


Figure 2 – A proposed hybrid model based on 1D-CNN and LSTM

3. RESULTS, DISCUSSION, AND EXPERIMENTAL SETUP

For model development and evaluation, the dataset was initially divided into training and testing sets in an 80:20 ratio, ensuring that the test set remained completely unseen during training to provide an unbiased assessment of the model's performance. The training set, comprising 80% of the data, was further utilized for cross-validation.

The proposed hybrid model consisted of convolutional layers for feature extraction, followed by LSTM layers to capture temporal dependencies, with dropout layers incorporated to prevent overfitting. A softmax-activated dense layer was used for final classification into seven classes. The model was wrapped using the keras classifier and optimized through 10-fold cross-validation, with a grid search conducted over batch sizes of 16 and 32. It was compiled using the Adam optimizer with a learning rate of 0.0001 and trained using categorical cross-entropy as the loss function. The best-performing model, determined based on mean cross-validation accuracy, was selected for final evaluation.

This selected model was then evaluated on the 20% test set to obtain an unbiased estimate of its performance on unseen data. Table 1 summarizes the performance of various models in terms of accuracy, precision, recall, and F1-score. The hybrid 1D-CNN + LSTM model achieved the highest performance across all metrics, with an accuracy of 96.43%. Logistic Regression (LR) also performed well, achieving 95.24% accuracy. The K-Nearest Neighbors (KNN) algorithm showed moderate performance with an accuracy of 92.86, while Naive Bayes yielded the lowest performance, with an accuracy of 89.29%. Overall, the deep learning-based hybrid approach outperformed the traditional machine learning classifiers.

Table 1 – Performance Comparison of Classification Models Based on Evaluation Metrics

Models	Accuracy	Precision	Recall	F1_Score
KNN	92.86%	94.70%	92.86%	93.00%
LR	95.24%	96.04%	95.24%	94.98%
Naive Bayes	89.29%	91.84%	89.29%	87.61%
LSTM	94.05%	95.33%	94.05%	93.63%
Proposed	96.43%	96.86%	96.43%	96.29%

Table 2 presents a comparative analysis of the proposed 1D-CNN+LSTM hybrid model and existing studies by Sarno (9) and Malikhah (7), all of which were evaluated using 10-fold cross-validation on similar datasets. The proposed hybrid deep learning model achieved the highest mean accuracy of 98.80%, outperforming both the optimized SVM, which achieved 98.10%, and the ensemble method, which achieved 98.33%. These results highlight the effectiveness of combining 1D-CNN and LSTM networks for capturing both spatial and temporal features, leading to improved classification performance in meat adulteration detection.

Table 2 – Comparative Performance of Proposed Model with Existing Studies

Studies	Methodology	Validation Split	Mean Accuracy
Sarno (9)	Optimized SVM	10-Fold	98.10%
Malikhah (7)	Ensemble Learning	10-Fold	98.33%
Proposed	1DCNN+LSTM	10-Fold	98.80%

4. CONCLUSIONS

This study proposed a hybrid model combining 1D-CNN and LSTM networks for detecting pork adulteration in beef using electronic nose sensor data. The proposed 1D-CNN+LSTM model outperformed traditional machine learning classifiers such as LR, KNN, and Naive Bayes, achieving a test accuracy of 96.43% and an F1-score of 96.29%. Furthermore, when compared with existing studies conducted on similar datasets, the proposed model achieved a superior mean accuracy of 98.80%, surpassing the performance of both the optimized SVM and ensemble-based method. These findings highlight the potential of the proposed model in effectively capturing both spatial and temporal patterns in sensor data. Future work may focus on further optimization strategies, real-time implementation, and the application of the model to more diverse or complex datasets. Additionally, while this study utilized only publicly available data from metal oxide gas sensors, future research will explore the use of alternative sensor technologies such as surface acoustic wave sensors and optical sensors, which may offer improved sensitivity and specificity for adulteration detection.

ACKNOWLEDGEMENTS

The authors thank the National Research Foundation of Korea (NRF) and the Ministry of Education for their financial support through the Basic Science Research Program (NRF-2022R1I1A3063493).

REFERENCES

1. Mahanti N K, Shivashankar S, Chhetri K B, Kumar A, Rao B B, Aravind J, et al. Enhancing food authentication through E-nose and E-tongue technologies: Current trends and future directions. Trends Food Sci Technol. 2024; 150: 104574.
2. Anwar H, Anwar T, Murtaza S. Review on food quality assessment using machine learning and electronic nose system. Biosens Bioelectron: X. 2023; 14: 100365.
3. Rahmati S, Julkapli N M, Yehye W A, Basirun W J. Identification of meat origin in food products–A review. Food control. 2016; 68: 379-90.
4. Chen X, Lu L, Xiong X, Xiong X, Liu Y. Development of a real-time PCR assay for the identification and quantification of bovine ingredient in processed meat products. Sci Rep. 2020; 10: 2052.
5. Premanandh J. Horse meat scandal–A wake-up call for regulatory authorities. Food control. 2013; 34: 568-9.
6. Pulluri K K, Kumar V N. Qualitative and quantitative detection of food adulteration using a smart e-nose. Sensors. 2022; 22: 7789.
7. Malikhah M, Sarno R, Sabilla S I. Ensemble Learning for Optimizing Classification of Pork Adulteration in Beef Based on Electronic Nose Dataset. Int J Intell Eng Syst. 2021; 14.
8. Peng P, Zhao X, Pan X, Ye W. Gas classification using deep convolutional neural networks. Sensors. 2018; 18: 157.
9. Sarno R, Triyana K, Sabilla S I, Wijaya D R, Sunaryono D, Fatichah C. Detecting pork adulteration in beef for halal authentication using an optimized electronic nose system. IEEE Access. 2020; 8: 221700-11.
10. Sarno R, Sabilla S, Wijaya D, Sunaryono D, Fatichah C. The Dataset for Pork Adulteration from Electronic Nose System. Data Brief. 2020; 32: 106139.
11. Dash C S K, Behera A K, Dehuri S, Ghosh A. An outliers detection and elimination framework in classification task of data mining. Decis Anal J. 2023; 6: 100164.
12. Rioul O, Vetterli M. Wavelets and signal processing. IEEE Signal Process Mag. 1991; 8: 14-38.
13. Kiranyaz S, Avci O, Abdeljaber O, Ince T, Gabbouj M, Inman D J. 1D convolutional neural networks and applications: A survey. Mech Syst Signal Process. 2021; 151: 107398.
14. Lindemann B, Müller T, Vietz H, Jazdi N, Weyrich M. A survey on long short-term memory networks for time series prediction. Procedia Cirp. 2021; 99: 650-5.

PROCEEDINGS of the
16th International Conference on Theoretical and Computational Acoustics

August 18 to 22, 2025 BEXCO, Busan Korea

Lightweight Neural Codec for DIFAR Sonobuoy Signal Transmission Using Residual Quantization and Pruning

Yeonjin Park[1]; Jungpyo Hong[2];
[1] Department of Information and Communication Engineering, Changwon National University, Republic of Korea
[2] Department of Information and Communication Engineering, Changwon National University, Republic of Korea

ABSTRACT

We present a lightweight autoencoder-based neural codec tailored to Directional Frequency Analysis and Recording (DIFAR) sonobuoys. The encoder, intended for battery-powered buoys, compresses short acoustic segments into compact index vectors via residual vector quantization (RVQ), while a compensation module mitigates reconstruction errors introduced during the quantization process. To accommodate the limited energy budget on board, the encoder is thinned through unstructured pruning, reducing computational cost without significantly impacting reconstruction quality. The decoder, hosted on a maritime patrol aircraft, restores the signal from the transmitted bitstream. The system is evaluated using simulated continuous-wave (CW) and linear-frequency-modulated (LFM) signals in a bistatic sonar setting. The evaluation demonstrates that the proposed codec significantly reduces the information required for transmission and preserves high spectral fidelity. The proposed approach reduces transmission bandwidth while maintaining reliable reconstruction, offering a practical balance between compression efficiency, signal fidelity, and on-board resource usage for passive Anti-Submarine Warfare (ASW) applications.

Keywords: Sonobuoy, Autoencoder, Signal Compression

1. INTRODUCTION

Directional Frequency Analysis and Recording (DIFAR) sonobuoys are critical assets in passive Anti-Submarine Warfare (ASW). As disposable, air-launched sensors, their role is to detect underwater acoustic signatures and relay them to a Maritime Patrol Aircraft (MPA) for real-time analysis. The conventional data transmission method, based on Frequency Division Multiplexing (FDM), is suboptimal for modern applications. It requires transmitting the entire raw acoustic data, leading to inefficient use of the limited bandwidth and creating a communication link with low security (1).

The emergence of neural audio codecs, driven by advances in deep learning, offers a paradigm shift for such constrained communication scenarios (2, 3). An autoencoder architecture, in particular, can learn an efficient, compressed representation (latent vector) of an audio signal. This approach not only enables significant data reduction but also enhances security, as the proprietary decoder, trained in conjunction with the encoder, is necessary for accurate signal reconstruction. While a basic autoencoder has been shown to be feasible for this task (3), its reconstruction fidelity often falls short of operational requirements.

State-of-the-art neural codecs have demonstrated remarkable performance by employing sophisticated techniques like Residual Vector Quantization (RVQ) and dedicated compensation modules (4, 5). RVQ provides a pathway to high-ratio compression with fine-grained detail, and compensation modules can effectively correct the errors inherent in the quantization process. However, these models are typically designed for unconstrained hardware and are too computationally intensive for battery-powered sonobuoys.

This paper introduces a lightweight autoencoder-based neural codec specifically designed for the operational realities of DIFAR sonobuoys. We engineer a solution that finds a practical balance between compression efficiency, signal fidelity, and on-board resource usage. Our key contributions are:

1. An asymmetrical system design featuring a highly efficient encoder tailored for sonobuoy hardware and a high-fidelity decoder for the MPA.

[1] 20247090@gs.cwnu.ac.kr
[2] hansin@changwon.ac.kr

2. The integration of RVQ and a novel compensation module to maximize reconstruction quality from a highly compressed bitstream.
3. The application of unstructured pruning to create an ultra-lightweight encoder, minimizing its energy footprint without significant performance degradation.

2. PROPOSED METHOD

The architecture of our proposed system is fundamentally asymmetrical, a design choice dictated by the constraints of the sonobuoy platform. A computationally lean encoder operates on the buoy, while the powerful, unconstrained decoder is hosted on the MPA. This ensures that the heavy lifting of signal reconstruction does not impact the sonobuoy's limited energy budget.

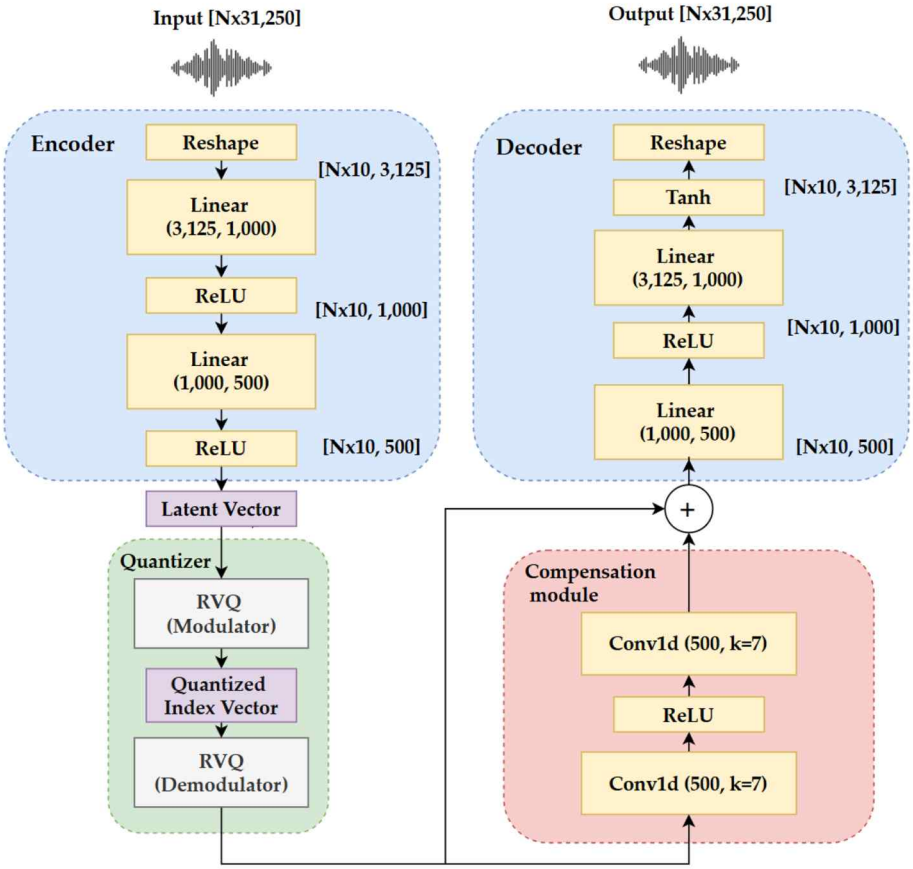


Figure 1 – The proposed codec architecture, illustrating the flow from the input signal through the lightweight encoder, RVQ, compensation module, and the high-fidelity decoder.

2.1 Asymmetrical Codec Architecture

The model, as shown in Figure 1, comprises four key stages: the encoder, a multi-stage quantizer, a compensation module, and the decoder.

- **Encoder:** Sonobuoy signals of interest, like CW and LFM pulses from active sonar, are characterized by sparse time-frequency patterns. In contrast to general-purpose audio, these signals do not require complex convolutional layers for effective feature extraction. Our encoder leverages this property, using a streamlined architecture of fully connected linear layers. It processes short acoustic segments (0.1 seconds) and maps them into a 500-dimensional latent vector, providing a rich feature representation.
- **Residual Vector Quantization (RVQ):** To compress the latent vector into a compact bitstream, we employ RVQ (4). This technique quantizes the vector in an iterative, multi-stage process. Over 10 successive stages, a codebook (containing 1024 entries) quantizes the residual error from the preceding stage. This hierarchical refinement process transforms the 500-dimensional vector into a small set of codebook indices. The final output is a compact index vector that can be serialized into a bitstream for transmission.
- **Compensation Module (CM):** Quantization is inherently lossy. To mitigate information loss, our

architecture includes a dedicated compensation module. This lightweight module, consisting of two shallow 1D-convolutional layers, takes the de-quantized vector as input and predicts the quantization error. This predicted error is then added back to the de-quantized vector, creating a refined representation that more closely matches the original latent vector from the encoder.

- **Decoder:** Hosted on the MPA, the decoder mirrors the encoder's structure. It takes the corrected latent vector and projects it back to the original signal dimension through a series of up-sampling layers, restoring the acoustic waveform from the transmitted bitstream.

2.2 Encoder Thinning via Unstructured Pruning

The primary hardware limitation for a sonobuoy is its finite battery life. To ensure a viable operational duration, the on-board encoder must be energy-efficient. To achieve this, we "thin" the encoder through unstructured pruning (6), a technique that removes unnecessary connections from a trained neural network. The process is:

1. **Train:** The full model is first trained to convergence.
2. **Prune:** A fixed percentage of the weights with the smallest absolute magnitudes within the encoder are permanently removed. These connections are deemed least influential to the network's output.
3. **Retrain:** The thinned network is then retrained for a few epochs. This allows the remaining weights to adapt and compensate for the removed connections, restoring performance.

This cycle of pruning and retraining is repeated until the encoder reaches a target level of sparsity. This disciplined thinning process drastically reduces the encoder's computational cost (FLOPs) and memory footprint, making it suitable for deployment on low-power hardware without a major sacrifice in reconstruction quality.

3. PERFORMANCE EVALUATION

The proposed system was evaluated using a dataset of simulated sonar signals, including both Continuous Wave (CW) and Linear Frequency Modulation (LFM) types. These signals were generated to mimic a bistatic sonar setting, representative of a passive ASW scenario (3). The model's reconstruction performance was benchmarked against a conventional autoencoder baseline using standard audio quality metrics: Spectrogram Mean Squared Error (MSE), Log Spectral Distance (LSD), and Signal-to-Noise Ratio (SNR).

3.1 Reconstruction Fidelity and Efficiency

The quantitative results, summarized in Table 1, highlight the effectiveness of our proposed approach. The "Proposed" codec, with its RVQ and compensation module, outperforms the "Conventional" baseline. For CW signals, it achieved a 60.58% reduction in Spectrogram MSE, while for LFM signals, the reduction was 55.25%. This result highlights the effectiveness of the architectural improvements.

Crucially, the "Proposed-Pruned" model, after undergoing significant encoder thinning, maintained most of these performance gains. Its reconstruction quality remained superior to the conventional method, confirming that our lightweighting strategy provides an trade-off between resource usage and signal fidelity. For example, its Spectrogram MSE for LFM signals (0.00719) was still approximately 50% better than the baseline. This result validates that a highly efficient on-board encoder can be achieved without compromising the core mission of delivering high-quality acoustic data.

Table 1 – Quantitative performance comparison

Method	Signal Type	Spectrogram MSE	LSD (dB)	SNR (dB)
Conventional [3]	CW	0.00277	1.19	25.93
	LFM	0.01428	1.19	17.66
Proposed	CW	0.00109	1.13	29.19
	LFM	0.00639	1.10	21.56
Proposed–Pruned	CW	0.00128	1.15	28.59
	LFM	0.00719	1.12	21.08

The qualitative improvements are visually confirmed in the spectrograms shown in Figure 2. For both CW and LFM signals, the reconstructions from the conventional method (c, d) are corrupted by significant background artifacts, visible as horizontal noise bands across the spectrum. These artifacts degrade the overall signal clarity and can interfere with detailed analysis.

In stark contrast, our proposed method (e, f) almost eliminates this patterned noise, producing a clean

reconstruction that is visually close to the original "Clean" signals (a, b). The primary spectral components are preserved with high fidelity. The spectrograms from the proposed-pruned method (g, h), while showing some degree of variation compared to their unpruned counterparts, still show improved reconstruction quality compared to the conventional method.

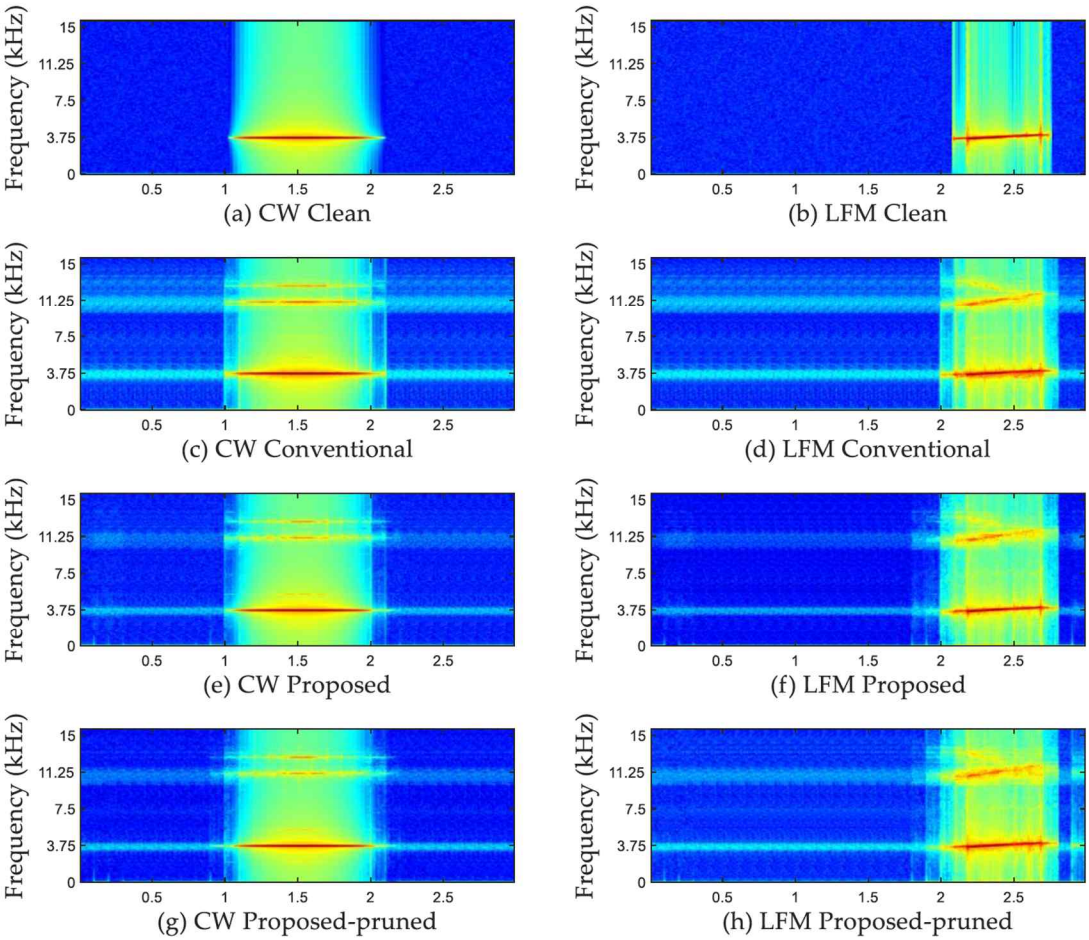


Figure 2 – Visual comparison of spectrograms for both Continuous Wave (CW) and Linear Frequency Modulation (LFM) signals. The plots compare the original 'Clean' signals (a, b) against reconstructions from the 'Conventional' method (c, d), the 'Proposed' method (e, f), and the lightweight 'Proposed-pruned' method (g, h).

4. CONCLUSIONS

This paper has introduced a lightweight, autoencoder-based neural codec tailored to the specific needs of DIFAR sonobuoys. The proposed approach finds a practical balance between compression efficiency, signal fidelity, and on-board resource usage. By integrating residual vector quantization and a compensation module, the codec achieves a high compression ratio while preserving the spectral fidelity of important acoustic signals. Furthermore, through unstructured pruning, we have demonstrated that the encoder can be significantly thinned to accommodate a limited energy budget without substantially impacting reconstruction quality.

Our end-to-end evaluation, benchmarked against previous methods, suggests that this system can offer an effective and reliable solution for modernizing passive ASW data links. It works to reduce the required transmission bandwidth and aims to ensure that the reconstructed signal quality is dependable under typical channel conditions, representing a notable advancement for real-world maritime surveillance operations.

ACKNOWLEDGEMENTS

"This research was funded by Changwon National University 'Outstanding Research Support Project' in 2025."

REFERENCES

1. Kuzu A, Aksit M, Cosar A, Guldogan MB, Gunal E. Calibration and test of DIFAR sonobuoys. In: Proceedings of the 2011 IEEE International Symposium on Industrial Electronics (ISIE); 27-30 June 2011; Gdansk, Poland. 2011. p. 1276-1281.
2. Baldi P. Autoencoders, unsupervised learning, and deep architectures. In: Proceedings of the ICML Workshop on Unsupervised and Transfer Learning; 2 July 2011; Bellevue, WA, USA. 2012. p. 37-49.
3. Park J, Seok J, Hong J. Autoencoder-based signal modulation and demodulation methods for sonobuoy signal transmission and reception. Sensors. 2022;22(17):6510.
4. Zeghidour N, Luebs A, Omran A, Skoglund J, Tagliasacchi M. SoundStream: An end-to-end neural audio codec. IEEE/ACM Trans Audio Speech Lang Process. 2022;30:495-507.
5. Xu L, Wang J, Zhang J, Xie X. LightCodec: A high fidelity neural audio codec with low computation complexity. In: Proceedings of the ICASSP 2024; 14-19 April 2024; Seoul, Republic of Korea. 2024. p. 586-590.
6. Han S, Pool J, Tran J, Dally WJ. Learning both weights and connections for efficient neural networks. arXiv. 2015;arXiv:1506.02626.

High-Speed Moving Object Tracking Using Deep Reinforcement Leaning

Jinyong Ha[1]; Jungpy Hong[2]; Jongwon Seok[3]
[1] Department of Information and Communication Engeenring, Changwon National University, Changwon, Korea
[2] Department of Information and Communication Engeenring, Changwon National University, Changwon, Korea
[3] Department of Information and Communication Engeenring, Changwon National University, Changwon, Korea

ABSTRACT

High-speed moving object tracking is critical in domains such as ocean exploration, autonomous navigation, and aerospace operations. Among the various tracking methods, the Kalman Filter(KF) and its variants have been widely used for estimating the states of dynamic targets in real time. However, these KF-based methods are fundamentally limited in nonlinear environments due to their reliance on linear motion assumptions and their sensitivity to noise, especially to the configuration of system and measurement noise covariance matrices. To overcome these limitations, this study proposes a deep reinforcement learning(DRL)-based tracking system utilizing a Deep Q-Network (DQN) architecture integrated with Pareto optimization. The proposed method formulates the tracking task as a Markov Decision Process (MDP), incorporating detection probability and velocity estimation to enable stable learning in uncertain environments. Pareto optimization is employed to balance detection accuracy and false positives, enhancing the reliability of network updates. Experiments conducted in both linear and nonlinear 3D motion scenarios demonstrate that the DQN algorithm significantly outperforms conventional KF-based methods, achieving over 50% improvement in Root Mean Square Error (RMSE). These results confirm the proposed method's adaptability and accuracy in complex environments, making it a viable alternative for real-time high-speed object tracking applications.

Keywords: Target tracking, Deep reinforcement learning, Deep Q-Network

1. INTRODUCTION

The importance of high-speed moving object tracking technology has been growing across various industries, playing a crucial role in applications that require precise real-time tracking and rapid response. Representative applications include ocean and space exploration, autonomous driving, and drone navigation systems. These technologies are essential for continuously tracking the locations of moving objects and calculating optimal routes in real time across various domains. For example, in ocean exploration and underwater robot operations, tracking the trajectories of fast moving marine organisms or submarines is crucial for both research and practical applications(1-3). High-speed targets in underwater environments are often difficult to detect and track with existing tracking technology. In general, a speed of 20 m/s (40 knots) or higher is defined as high speed, which encompasses the speeds of nuclear-powered submarines and high-speed torpedoes.

Traditionally, the Kalman Filter(KF) and its variants, such as the Extended KF(EKF) and the Unscented KF(UKF), have been widely used for high-speed object tracking. KF is an algorithm that continuously estimates a target's location based on observed data and is known to be highly effective, especially in linear systems. However, in real-world scenarios, the trajectories of moving objects often exhibit complex nonlinear motion and unpredictable behavior(4). In such an environment, existing KF-based methods may experience severe performance degradation. In particular, under highly nonlinear conditions, estimation accuracy deteriorates significantly as tracking errors in KF accumulate.

To overcome these limitations, this study proposes a high-speed moving object location estimation system based on Deep reinforcement learning(DRL). DRL is a technique that employs deep neural networks to

[1] realmir13@naver.com
[2] hansin@changwon.ac.kr
[3] jwseok@changwon.ac.kr

optimize decision making in complex environments and has recently attracted significant attention for its efficient learning and adaptability across various fields. In this work, deep Q-Network(DQN) algorithms are applied to achieve more accurate tracking performance compared to existing methods. Unlike conventional KF-based methods, DQN learns through interactions with the environment and offers the advantage of effectively handling nonlinear object motion beyond simple linear models. In addition, Pareto Optimization techniques are applied to optimize the target's detection and tracking performance to improve the neural network's update process, thereby securing more stable and reliable tracking performance.

The structure of this paper is as follows: Section 2 Related Works reviews previous research, Section 3 DRL-based Target Tracking Method describes the structure and learning mechanism of the proposed DRL-based tracking method, and Section 4 Experiments presents the experimental setup, and performance analysis. Finally, Section 5 Conclusion summarizes the research findings and discusses future research directions

2. Related Works

2.1 Kalman Filter

The Kalman Filter (KF) is a recursive algorithm designed to predict the state of a linear dynamic system and correct it using measurement data. It has proven highly effective in real-time state estimation through its prediction and correction mechanisms, and has been widely applied in various fields such as finance (5), aerospace (6), and autonomous vehicle systems (7).

For example, Kalman filters have been applied in acoustic navigation systems of Autonomous Underwater Vehicles (AUVs) to estimate underwater positions in real time. Experimental results have confirmed that KF can provide stable position estimates using received acoustic signals (8). Furthermore, algorithms utilizing the Extended Kalman Filter (EKF) have been proposed for tracking underwater moving targets, effectively handling nonlinear underwater dynamics (9). In cases where sonar systems face uncertainty regarding the presence of target signals, detection-probability-based update strategies have also been introduced. For instance, some methods apply observation updates to the KF only when the detection probability exceeds a predefined threshold; otherwise, only the predicted state is retained (10,11). This approach helps suppress unnecessary state updates under uncertainty and enhances tracking stability.

Despite its strengths, the Kalman Filter faces several limitations. First, its performance is highly sensitive to the configuration of the system noise and measurement noise covariance matrices, making it difficult to determine appropriate values. Incorrect tuning can disrupt the balance between prediction and observation reliability. Second, KF is fundamentally based on linear assumptions, making it less suitable for nonlinear systems involving rotational motion or fluid dynamics. Although EKF and the Unscented Kalman Filter (UKF) have been developed to address this issue, they still suffer from performance limitations and high computational costs (12,13). Therefore, extending the applicability of KF to nonlinear systems requires parameter optimization and adaptive approaches, calling for the development of alternative algorithms.

2.2 Deep Reinforcement Learning

Reinforcement Learning (RL) is a method by which an agent learns an optimal policy that maximizes cumulative rewards through interactions with its environment. It is fundamentally based on the Markov Decision Process (MDP), where the agent selects an action according to the current state, receives a reward from the environment, and transitions to a new state. Through repeated interactions, the agent gradually learns a policy that maximizes the expected long-term reward. RL has been applied to various fields, including robot control and autonomous driving systems.

Deep Reinforcement Learning (DRL) combines reinforcement learning with deep neural networks to enable effective learning even in high-dimensional and complex environments. Traditional RL methods, such as Q-learning, suffer from performance degradation as the size of the state space increases. In contrast, DRL utilizes neural networks to efficiently explore and learn from large-scale state spaces. Representative DRL techniques include Deep Q-Networks (DQN), policy gradient methods, and actor-critic approaches. DRL provides a more powerful and flexible learning framework than conventional RL, making it well-suited for complex real-world applications.

DRL has been applied in various domains such as healthcare , autonomous underwater vehicle (AUV) control , and radar-based target tracking. For example, DRL has been used to enhance radar tracking performance by compensating for the sensitivity of EKF to initial values and uncertainties in noise models (14). It has also been employed to adapt radar parameters in real-time according to environmental conditions, thereby improving tracking accuracy in complex multi-target scenarios (15).

3. DRL-based Target Tracking Method

3.1 Design of MDP for Tracking

This study aims to accurately track the movement path of a target based on its location data. To achieve this, a tracking system was designed using the MDP, as shown in Figure 1, incorporating an approach that minimizes uncertainty by incorporating detection probability into the process. The designed MDP consists of an agent, states, actions, and rewards, and is structured to effectively capture both the movement characteristics of the target and the detection performance of the sensor. In particular, by integrating detection probability into the state update process and selectively utilizing only highly reliable observed values, this approach aims to achieve more stable and accurate tracking performance, even in environments with high uncertainty.

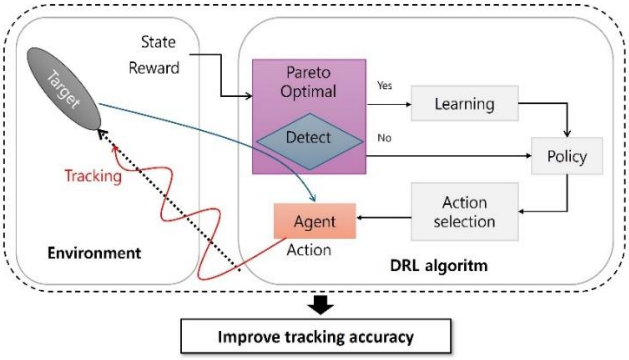


Figure 1 – Conceptual diagram of the proposed tracking method based on DRL

In this paper, the agent's target location tracking problem is formulated as follows:

$$L_{t+1} = L_t + \hat{V}_t \cdot 0.5, \quad (1)$$

where L is the location and V is the velocity. The agent utilizes the predicted velocity to update its status and rewards, and proceeds with learning to improve target location tracking accuracy. At time t, the action(a_t) is an element of the set A. Set A is the range of velocity that the tracking system can estimate and is defiend as follws:

$$A = \{-VC + 0.1i | i = 1, \dots, 36\}, \quad (2)$$

where VC is a velocity candidate, and may have a specific value according to the experimental environment. Within a range of 36 values at 0.1 m/s intervals based on a specific speed, the agent select the optimal speed to accurately track the location of the traget. Moreover, state is denoted as,

$$s_{t+1} = (\Delta V_t, \hat{V}_t), \quad (3)$$

where ΔV is a velocity error that is the difference between the measured velocity and the predicted velocity. Through this, the agent may be provided with basic information that may accurately recognize the current state and determine an appropriate behavior. Reword is a value obtained when an agent performs an action at a specific time, and is an important factor in determining the agent's learning direction. In particular, Reword Function is formulated as follows:

$$r = -\left|V_{true} - \hat{V}\right|. \quad (4)$$

In order to induce the learning process to minimize the velocity error by the agent, '-' was added to the reword to take a negative form.

The agent selects the optimal behavior in consideration of the uncertainty of the observed data and adjusts it more precisely through an iterative learning process. Through this, the agent improves tracking performance over time and provides a robust tracking system that can operate reliably even in changing environments.

3.2 Pareto Opptimality

Pareto optimization is an effective method for balancing multiple objectives, aiming to find an optimal solution where improving one performance does not come at the expense of another. This approach is particularly useful when optimizing conflicting criteria simultaneously(16).

Pareto optimization is applied in various fields, including neural network training(17), autonomous vehicle route optimization(18), and robot control(19). For example, in autonomous vehicle path optimization, travel time and fuel consumption often conflict. By applying Pareto optimization, it is possible to find an optimal path that minimizes fuel consumption while ensuring the shortest travel time. In robot control, both movement accuracy and robust control performance must be considered simultaneously. By applying Pareto optimization, various aspects of control performance can be improved. This allows for the design of a control

strategy that balances multiple factors rather than maximizing a single performance metric.

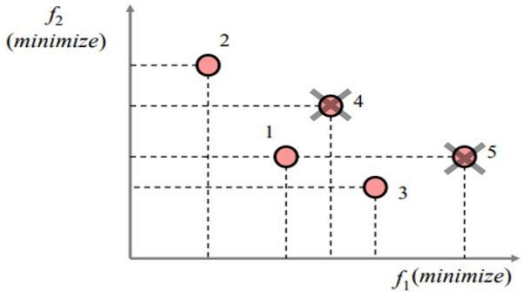


Figure 2 –Parreto Optimality Curve

Figure 2 visually illustrates the concept of Paretooptimal solutions. The objective is to minimize the two functions, f1 and f2, shown in the graph. Each point represents a specific solution and can be compared with others based on the coordinates connected by dotted lines. A Pareto optimal solution consists of solutions for which no better alternative exists, forming what is known as the Pareto front. In the Figure 2, points 1, 2, and 3 belong to the Pareto optimal set, meaning that improving one objective inevitably degrades the other. In contrast, points 4 and 5 are dominated solutions, meaning they are excluded from the optimal set because superior alternatives already exist.

Pareto optimization simultaneously considers multiple objective functions and focuses on maintaining a balance so that improving one performance metric during the optimization process does not degrade the other. In this study, Pareto optimization was applied to optimize the balance between detection probability and false positive probability in a high-speed mobile tracking system. Traditional tracking algorithms are often designed to maximize a single performance indicator (e.g., detection probability); however, this approach can lead to an increase in false positives, reducing stability in real-world environments. Therefore, this study adopts a multi objective optimization strategy to enhance detection probability while minimizing false positives.

To achieve this, a reinforcement learning based tracking agent measures the detection probability as it progresses through a new episode and compares it with the probability values forming the existing Pareto front. If the new probability value outperforms the existing one, the Pareto front is updated, and a new optimal solution is constructed. Subsequently, the learning process is continuously optimized by incorporating the newly selected detection probability as an update condition for DRL. Unlike traditional methods that optimize a single performance indicator, this approach contributes to building a more robust tracking system by employing an optimization strategy that balances multiple objective functions.

3.3 DQN tracking algorithm

DQN is a DRL algorithm that combines traditional Q-learning and deep learning technologies, enabling effective learning in a high dimensional state space. DQN uses deep neural networks to approximate state-action value functions and increase the efficiency of learning through innovative technologies such as experience reproduction and target networks. It overcomes the limitations of existing Q-learning and is used in various fields such as robotics(20), medical field(21), and autonomous driving(22).

Figure 3 shows the operation process of the DQN-based tracking algorithm. Agents select actions through the Appsilon greed policy, which combines random behavior and learned policies in a balanced way. Initially, the frequency of random behavior selection is high, and the speed of the target is estimated incorrectly, resulting in a large error in location prediction. However, as learning progresses, the frequency of random behavior selection gradually decreases, and the agent estimates the speed of the target more accurately while following the learned policy. This improves the accuracy of estimating the location of the target.

The agent receives reward and next state information from the environment and updates the weights of the Q-network and target network accordingly. Using the detection probability obtained through the Pareto optimization technique, the networks are trained only when the target is detected. Undetected experiences are excluded from learning due to their low reliability, whereas detected experiences, which occur in the presence of the actual target, are considered relatively reliable. Therefore, learning efficiency is maximized by training the network with highly reliable experiences.

Meanwhile, the movement of the target in the algorithm's learning process was defined as a single episode. The agent tracks the target every 0.5 seconds, and the algorithm is repeatedly executed as the target moves. In each scenario, the agent estimates the target's location and gradually refines its policy through learning to achieve optimal tracking performance. This approach enables more precise and reliable tracking compared to traditional Kalman filter based methods.

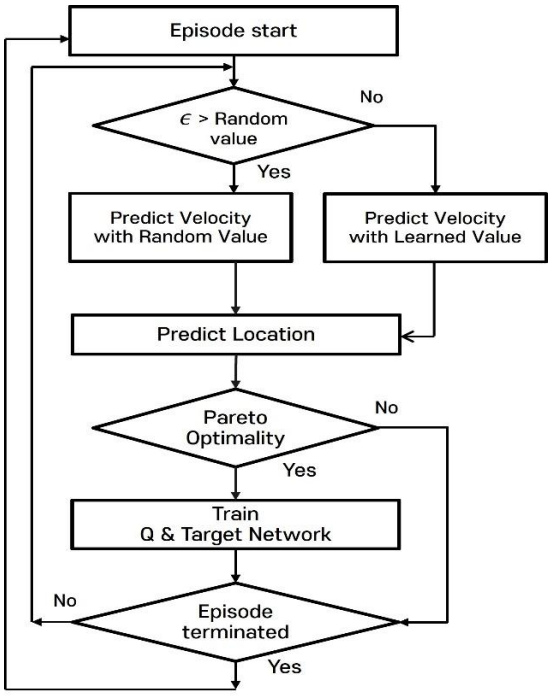


Figure 3 –DQN Tracking Algorithm

4. Experiments

4.1 Experimental Setup

In this study, both linear and nonlinear underwater environments were designed to evaluate the target tracking performance of the proposed deep reinforcement learning-based system. As shown in Figure 4, the simulation includes two types of three-dimensional target motion trajectories: a linear path in the left panel and a nonlinear path in the right panel.

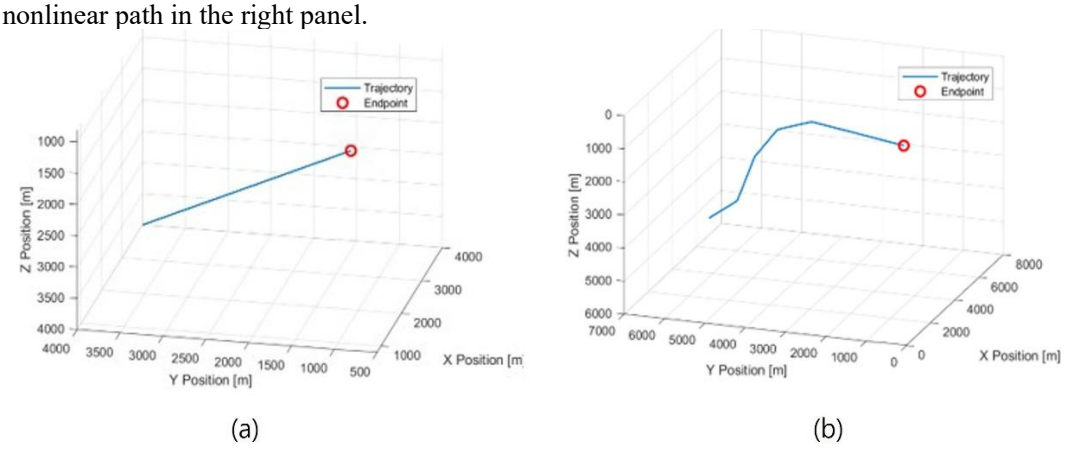


Figure 4 –target trajectories for experimental scenarios: (a)linear motion and (b)nonlinear motion

In the linear scenario, the target was configured to move at a constant speed of 20 meters per second along a straight path in a three-dimensional space. The initial position was set at a distance of 4000 meters from both the transmitter and receiver, and the total simulation duration was 150 seconds. This setup was used to validate the baseline tracking accuracy of the algorithm in a controlled and predictable environment.

In the nonlinear scenario, the target's trajectory included abrupt speed changes and arbitrary curved movements, reflecting realistic underwater conditions such as water currents, evasive maneuvers, and propulsion fluctuations. This setup enabled an in-depth analysis of the tracking system's adaptability and robustness under complex and dynamic motion patterns.

These experimental conditions provided a comprehensive basis for evaluating the performance of the deep reinforcement learning-based tracking algorithm in both structured and unstructured underwater environments.

Table 1 shows the parameters used in the learning process of DQN. Table 2 is the structure of the DQN

neural network and is designed based on MDP. The number of input nodes of the neural network is set to two, which is the number of variables indicating the environmental state. The number of nodes in the output layer is 100 matching the size of the action space. The output node represents the Q-Value for each possible action, and based on this, the agent selects the optimal action.

Table 1 – DQN Experimental Parameters

Parameter	Description	Value
γ	Discount factor	0.99
α	Learning rate	0.001
ϵ_I	Initial value of epsilon	1
η	Epsilon decay rate	0.999
D	Replay memory size	2000

Table 2 – DQN Neural Network Architecture

		Node	
Input	Dense Layer	6	Relu
Hidden		128	
		128	
		256	
Ouput		36	Linear
Optimizer		Adam	
Loss		MSE	

4.2 Experimental Results

The performance of various tracking algorithms was compared for linearly moving targets. Table 3 shows the tracking performance of each algorithm. The KF based algorithm performed 1,000 experiments through Monte Carlo simulation and DQN also performed 1,000 episodes to analyze the performance. As a result of the experiment, the root mean square error (RMSE), which is the error between the actual location and the estimated location, was calculated and evaluated. It was confirmed that DQN recorded the lowest RMSE value and could accurately track the location and speed of the target. As a result of analyzing the path of the target in the 3D space in Figure 5, DQN estimated the path of the target more accurately than the existing KF based algorithm. Through this, it was confirmed that the reinforcement learning based approach is an effective alternative to the high-speed mobile object tracking problem.

Table 3 – RMSE results for tracking in a Linear scenario

3-dimensional axis	KF	EKF	UKF	DQN
X	26	26.6	29.8	12.5
Y	26.2	27	29.8	12.3
Z	27.2	28.1	30	12.6

Table 4 – RMSE results for tracking in a Nonlinear scenario

3-dimensional axis	KF	EKF	UKF	DQN
X	89.8	89.2	99.4	18.3
Y	89.8	89.2	98	18.7
Z	142.4	143.3	129.8	19.3

The tracking performance of each algorithm was compared in an environment where the target moves along a nonlinear trajectory. As a result of the experiment in Table 4, the KF could not accurately track the trajectory of the target in nonlinear movements. In particular, in the section where the target speed fluctuates rapidly or the direction changes abnormally, the RMSE tended to increase rapidly.

The KF showed relatively stable performance in a linear environment, but errors accumulated because it could not adapt quickly to conditions such as nonlinear motion. The EKF and the UKF were designed to handle nonlinear models, but the experimental results showed that they could not effectively respond to sudden changes in motion at a specific point in time.

On the other hand, the DQN algorithm recorded a lower RMSE compared to other algorithms even in nonlinear target movements. DQN succeeded in accurately tracking the speed and position of the target by learning adaptively even in a nonlinear environment through a neural network designed based on reinforcement learning.

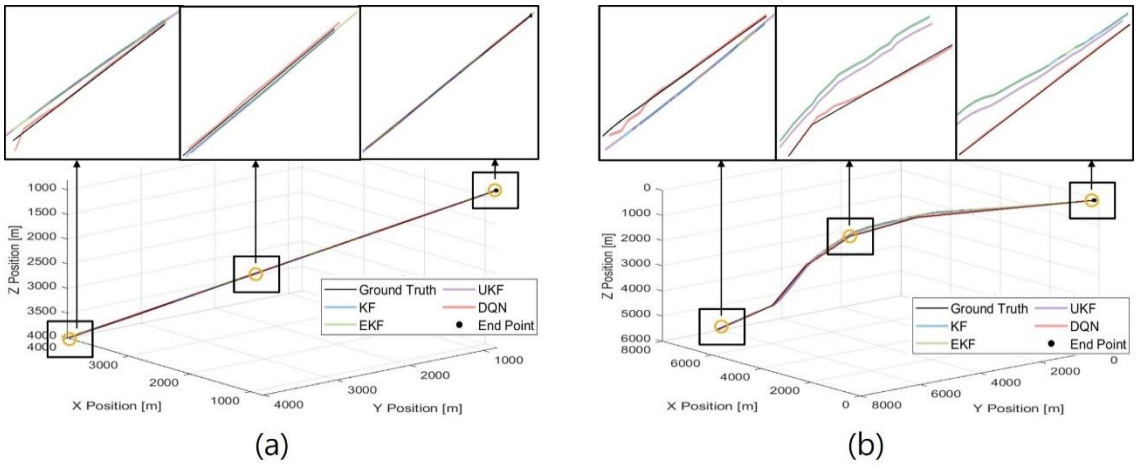


Figure 5 –Estimated Trajectories in experimental scenarios: (a)linear motion and (b)nonlinear motion

As a result of visually analyzing the target's movement path in a 3D space as shown in Figure 5, the tracking performance of DQN was stable even in the section where the target accelerates rapidly or draws a curved path. This shows that the DQN algorithm can be effectively learned and applied even in nonlinear movements. As a result, this experiment suggested that the DQN algorithm shows excellent performance not only in a simple linear environment but also in a complex nonlinear environment and can replace the existing KF based algorithms.

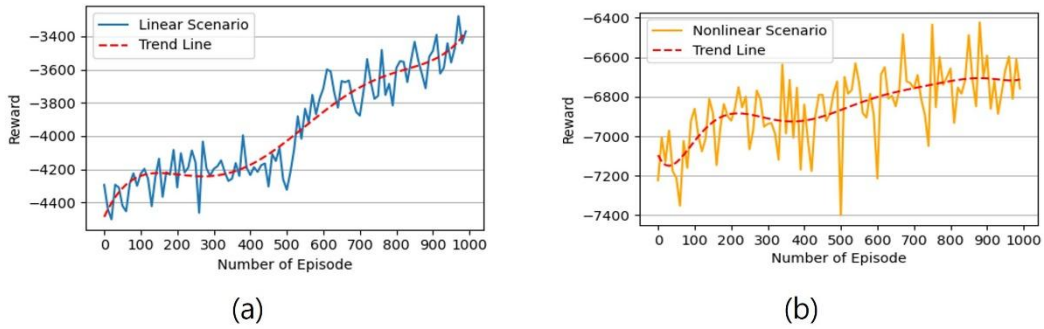


Figure 6 –Reward Results for DQN Training for experimental scenarios: (a)linear and (b)nonlinear

The reward function represents the difference between the actual speed and the predicted speed, and as the reward value increases, the agent predicts the speed of the target more accurately. Figure 6 shows the results of the reward function during the training process. In the linear scenario, the reward values show a general upward trend, indicating that the agent progressively learns better policies as training continues and improves its tracking performance. Similarly, in the nonlinear scenario, the reward values gradually increase compared to the initial stages, demonstrating that the agent is capable of optimizing its policy even in a complex and dynamic environment.

These reward trends in both scenarios confirm that the Deep Q-Network (DQN) algorithm has achieved stable learning. Through continuous interaction with the environment, the agent learns to accurately predict the target's velocity and position over time.

5. CONCLUSIONS

In this study, a high-speed mobile object position estimation system based on DRL was proposed and its performance was compared with the existing KF-based method. As a result of the experiment, the proposed DQN algorithm recorded a lower RMSE than the existing method in linear and nonlinear environments, showing an average performance improvement of more than 70%.

In particular, through tracking performance analysis in a nonlinear environment, it was confirmed that the proposed algorithm can overcome the limitations of the existing KF and adapt to environmental changes. In addition, as a result of the reward function analysis, it was shown that the reward value increases stably as learning progresses, and the speed and position of the target can be estimated more accurately.

The results of this study suggest that DRL-based tracking technology can be an effective alternative to existing methods. In future studies, research will be conducted in the direction of further improving real-time

tracking performance by applying a more complex environment and various algorithms.

ACKNOWLEDGEMENTS

This work was supported by Korea Research Institute for defense Technology planning and advancement(KRIT) grant funded by the Korea government(DAPA(Defense Acquisition ProgramAdministration)) (No. KRIT-CT-22-052, Physics-guided Intelligent Sonar Signal Detection ResearchLaboratory, 2025)

REFERENCES

1. He B, Ran H, Yang K, Huang L, Ren C. Path planning and tracking for autonomous underwater vehicles. In: 2009 International Conference on Information and Automation; 2009. p. 728–33.
2. Wang C, Yu W, Zhu S, Song L, Guan X. Safety-critical trajectory generation and tracking control of autonomous underwater vehicles. IEEE J Ocean Eng. 2022;48(1):93–111.
3. Ding F, Wang R, Zhang T, Zheng G, Wu Z, Wang S. Real-time trajectory planning and tracking control of bionic underwater robot in dynamic environment. Cyborg Bionic Syst. 2024;5:0112.
4. Adzemović M, Tadić P, Petrović A, Nikolić M. Beyond Kalman filters: deep learning-based filters for improved object tracking. Mach Vis Appl. 2025;36(1):1–22.
5. Zhuang Q, Chen L. Dynamic prediction of financial distress based on Kalman filtering. Discrete Dyn Nat Soc. 2014;2014(1):370280.
6. Rafi M, Steck JE, Watkins J. Kalman-filter-based adaptive control: flight testing on general aviation aircraft. J Guid Control Dyn. 2017;40(9):2307–15.
7. Farag W. Kalman-filter-based sensor fusion applied to road objects detection and tracking for autonomous vehicles. Proc Inst Mech Eng Part D J Automob Eng. 2021;235(7):1125–38.
8. Ji CL, Xu M, Wang H, Zhang Z. Application of Kalman filter in AUV acoustic navigation. Appl Mech Mater. 2014;525:695–701.
9. Huang J. An underwater target tracking algorithm based on extended Kalman filter. Mob Inf Syst. 2023;2023:9916531.
10. Kim G, Lim S, Park J, Kwon Y. Design of generalized sinusoidal frequency modulated pulse train waveform to improve tracking performance of high duty cycle sonar systems. IEEE Access. 2022;10:106543–55.
11. Park J, Kim G, Kwon Y, Lim S. Pulsed active sonar using generalized sinusoidal frequency modulation for high-speed underwater target detection and tracking. IEEE Access. 2023;11:143081–91.
12. LaViola JJ. A comparison of unscented and extended Kalman filtering for estimating quaternion motion. In: Proceedings of the 2003 American Control Conference; 2003. Vol. 3. p. 2435–40.
13. Julier SJ, Uhlmann JK. Unscented filtering and nonlinear estimation. Proc IEEE. 2004;92:401–22.
14. Tang Y, Lin Y, Yang X, Wang Z. Reinforcement learning compensated extended Kalman filter for attitude estimation. In: 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems; 2021. p. 6854–9.
15. Huang Y, Li J, Zhang X, Wang Y, He D. Deep reinforcement learning based radar parameter adaptation for multiple target tracking. IEEE Trans Aerosp Electron Syst. 2024;60.
16. Censor Y. Pareto optimality in multiobjective problems. Appl Math Optim. 1977;4:41–59.
17. Shu T, Huang K, Zhao M, Han J. Learning Pareto set for multi-objective continuous robot control. arXiv preprint arXiv:2406.18924. 2024.
18. Pan X, Li J, Huang W, Yang Y. A convex optimal control framework for autonomous vehicle intersection crossing. IEEE Trans Intell Transp Syst. 2022;24:163–77.
19. Taherkhorsandi M, Alizadeh M, Baloochestani S, Shoorehdeli MA. Pareto design of an adaptive robust hybrid of PID and sliding control for a biped robot via genetic algorithm optimization. Nonlinear Dyn. 2015;79:251–63.
20. Escobar-Naranjo J, Jimenez-Vargas S, Gómez-García C, Rojas-Pinzon A. Autonomous navigation of robots: optimization with DQN. Appl Sci. 2023;13:7202.
21. Liu N, Yin S, Wang J, Li Y, Hu W. Learning the dynamic treatment regimes from medical registry data through deep Q-network. Sci Rep. 2019;9:1495.
22. Pérez-Gil O, Calderon-Aparicio J, Garcia-Sanchez AJ, Garcia-Sanchez F. DQN-based deep reinforcement learning for autonomous driving. In: Advances in Physical Agents II: Proceedings of the 21st International Workshop of Physical Agents (WAF 2020), Alcalá de Henares, Madrid, Spain. Springer International Publishing; 2021. p. 60–76.

PROCEEDINGS of the
16th International Conference on Theoretical and Computational Acoustics

August 18 to 22, 2025 BEXCO, Busan Korea

Web-based real-time speech conversion system for Dysarthria

Kwanghyun PARK[1]; Yujin JI[2]; Kangmin LIM[3]; Jungpyo Houng[4]
[1] Department of Information and Communication Engineering, Changwon National University, Korea
[2] Department of Information and Communication Engineering, Changwon National University, Korea
[3] Department of Information and Communication Engineering, Changwon National University, Korea

ABSTRACT

Dysarthria is a motor speech disorder caused by impaired control of the articulatory muscles, resulting in unclear and effortful speech that hinders effective communication. Many existing assistive technologies for dysarthria rely on external devices, manual selection, or word-level generation, which can be slow and unnatural in real-world conversations. This study introduces a real-time speech conversion system that transforms sentence-level Korean dysarthric speech into fluent, intelligible speech. The system comprises two sequential modules. First, a conformer-based automatic speech recognition (ASR) model is trained on dysarthric utterances to improve transcription quality using advanced sequence modeling. Second, a text-to-speech (TTS) model synthesizes natural Korean speech using FastSpeech2 and a neural vocoder. To ensure accessibility and ease of use, the entire system is modularized and implemented through the Python Flask web framework. Users can either record their own dysarthric speech or speak directly via the web interface, and the converted speech is returned through real-time audio playback. This web-based approach enables practical use in daily environments without requiring specialized hardware. The proposed system provides an effective and accessible tool for supporting individuals with dysarthria through real-time speech conversion.

Keywords: Dysarthria, Speech conversion, Automatic Speech Recognition (ASR)

1. INTRODUCTION

Dysarthria, a motor speech disorder stemming from impaired neurological control over the articulatory muscles, significantly compromises speech intelligibility and fluency (1). This condition poses substantial communication barriers for affected individuals in their social and professional lives. While various assistive technologies, such as Augmentative and Alternative Communication (AAC) devices, have been developed, they often have limitations. Many systems require manual input, operate at a word-by-word level, or lack the real-time processing capability needed for natural, conversational interactions (2).

Furthermore, much of the research in dysarthric speech recognition has been concentrated on English, leaving a significant gap for other languages like Korean, which has distinct phonetic and structural characteristics (3). Existing Korean dysarthric speech recognition models have shown limitations in computational efficiency and performance, particularly with long or complex sentences.

To address these challenges, this paper presents a web-based, real-time speech conversion system designed specifically for sentence-level Korean dysarthric speech. Our goal is to provide a practical and accessible communication aid that directly converts a user's dysarthric utterance into clear, non-dysarthric speech. The system integrates a high-performance dysarthric Automatic Speech Recognition (ASR) model with a high-quality Text-To-Speech (TTS) synthesizer, delivered through a simple web interface to maximize usability without specialized hardware.

2. PROPOSED SYSTEM ARCHITECTURE

The proposed system operates through a cascaded architecture consisting of two core modules: a Korean dysarthric ASR model and a Korean TTS model. These modules are integrated into a web server using the

[1] 20247089@gs.cwnu.ac.kr
[2] 20193036@gs.cwnu.ac.kr
[3] Km1282@gs.cwnu.ac.kr
[4] hansin@changwon.ac.kr

Python Flask framework, enabling real-time processing. The user provides dysarthric speech input via a microphone on the web interface; the ASR module transcribes it into text, and the TTS module immediately synthesizes it into clear, audible speech. Figure 1 illustrates this overall workflow.

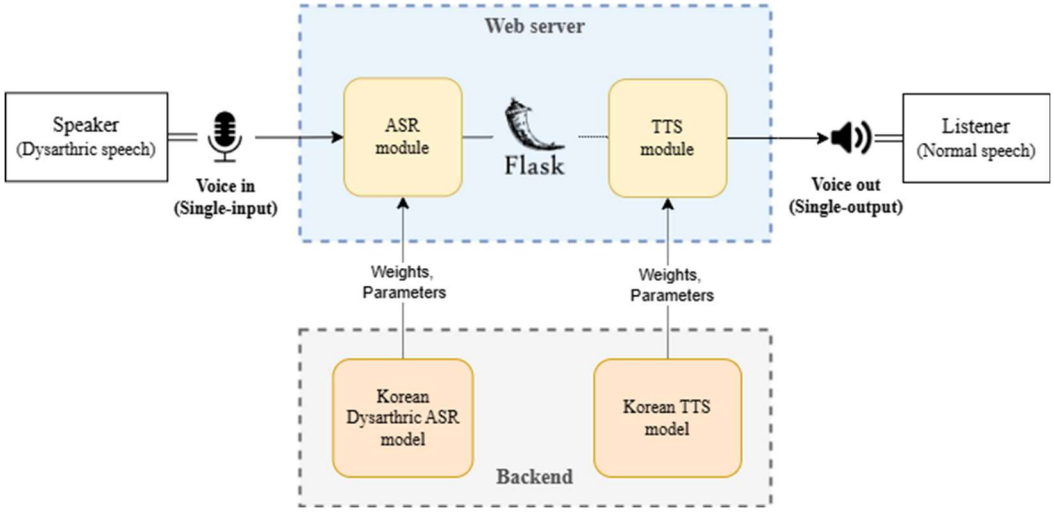


Figure 1 – A diagram showing the overall architecture of the proposed system

2.1 Dysarthric Speech Recognition (ASR) Model

To accurately transcribe the highly variable and irregular patterns of Korean dysarthric speech, we developed an ASR model based on a conformer architecture (4). Unlike traditional models that separate local feature extraction (e.g., CNN) and global context modeling (e.g., Transformer), the conformer encoder effectively combines convolution and self-attention mechanisms. This structure allows the model to simultaneously capture fine-grained local acoustic features and long-range contextual dependencies within a single, efficient framework.

For sequence alignment and loss calculation, we employed the Graph Transducer Network-based Connectionist Temporal Classification (GTN-CTC) algorithm (5). The graph-based approach of GTN-CTC is particularly well-suited for Korean, as it flexibly models the complex syllable structures and phonological variations inherent in the language, leading to improved recognition accuracy compared to conventional CTC methods.

2.2 Normal Speech Synthesis (TTS) Model

For the speech synthesis component, we selected the JETS (Jointly Training FastSpeech2 and HiFi-GAN for End-to-End Text-to-Speech) model (6). JETS integrates a fast and robust acoustic model (FastSpeech2) with a high-fidelity vocoder (HiFi-GAN) into a single, end-to-end training process. This joint optimization allows the model to generate highly natural and clear Korean speech. Its architecture, which includes variance adaptors for pitch and energy, is adept at capturing the prosodic nuances of Korean, resulting in synthesized speech that is superior in quality to that of models like Tacotron2 or Transformer TTS.

3. EXPERIMENTS AND RESULTS

3.1 Experimental Setup

The ASR model was trained on approximately 100 hours of Korean dysarthric speech data from the AI-Hub dataset (7), a large-scale corpus collected from over 1200 patients. The TTS model was trained on the KSS dataset (8), a 12-hour corpus of a single professional female Korean speaker. All experiments were conducted on a server equipped with an NVIDIA RTX A6000 GPU, using the ESPnet toolkit (9). ASR performance was evaluated using Character Error Rate (CER) and Word Error Rate (WER), while TTS quality was assessed using the Mean Opinion Score (MOS).

3.2 ASR Performance

The proposed conformer-based ASR model demonstrated significant performance improvements over the baseline hybrid CTC/attention transformer model. As shown in Table 1, our model achieved a CER of 9.1% and a WER of 11.6%. This represents a relative error reduction rate (ERR) of 41.7% in CER and 38.3% in WER, confirming the superiority of the conformer architecture for this task.

Table 1 - Performance evaluation results of the proposed ASR model.

Model	CER (%)	WER (%)	ERR(CER) (%)	ERR(WER) (%)
Baseline Transformer model (7)	15.6	18.8	-	-
Proposed Conformer model	9.1	11.6	41.7	38.3

3.3 TTS Performance

The JETS model's performance was compared against two other common TTS models, Tacotron2 and Transformer TTS. In a subjective MOS evaluation with 10 participants rating 20 audio samples per model, JETS achieved the highest score, as detailed in Table 2. Listeners rated the JETS-generated speech as significantly more natural in both intonation and voice quality compared to the more mechanical-sounding outputs of the other models.

Table 3 - Evaluate MOS comparisons for different TTS models.

TTS Model	MOS (1–5)
JETS (6)	4.64 ± 0.06
Tacotron2 (10)	3.03 ± 0.07
Transformer TTS (11)	2.51 ± 0.09

3.4 Integrated System Performance

The integrated system demonstrates effective real-time performance. For a typical dysarthric speech input of 5–10 seconds, the system delivers the converted, clear speech output in approximately 2–3 seconds. This latency is sufficiently low for practical use in daily conversations, which often consist of shorter utterances. Qualitatively, the system successfully removes the unnatural pauses and disfluencies characteristic of dysarthric speech, producing a smooth, rhythmically natural utterance. This transformation allows for clearer and more efficient communication.

4. CONCLUSIONS

This study successfully developed and validated a web-based, real-time communication aid that converts sentence-level Korean dysarthric speech into clear, non-dysarthric speech. By leveraging a high-performance conformer-based ASR model and a high-quality JETS-based TTS model, our system achieves both high accuracy and natural-sounding output. The significant improvements in ASR performance (over 38% WER reduction) and the superior subjective quality of the synthesized speech (4.64 MOS) confirm the effectiveness of our chosen architectures.

The web-based implementation ensures the system is accessible and easy to use across various devices without requiring specialized hardware. While future work will focus on reducing latency for very long utterances and conducting extensive usability testing with the target user group, the proposed system represents a significant step toward providing a practical and powerful communication tool for individuals with dysarthria.

ACKNOWLEDGEMENTS

This research was supported by the Academic Promotion Program for Basic and Protected Academic Fields at Changwon National University in 2025.

REFERENCES

1. Enderby P. Disorders of communication: Dysarthria. Handb. Clin. Neurol. 2013;110:273–81.
2. Hawley MS, Cunningham SP, Green PD, Enderby P, Palmer R, Sehgal S, et al. A voice-input voice-output communication aid for people with severe speech impairment. IEEE Trans. Neural Syst. Rehabil. Eng. 2012;21:23–31.
3. Qian Z, Xiao K. A survey of automatic speech recognition for dysarthric speech. Electronics. 2023;12:4278.
4. Guo P, Watanabe S, Kawahara T, Takeda K. Recent developments on ESPnet toolkit boosted by conformer. Proc ICASSP 2021; 6-11 June 2021; Toronto, ON, Canada. 2021. p. 5874-8.
5. Hannun A, Case C, Casper J, Catanzaro B, Cong Y, Fougner C, et al. Differentiable weighted finite-state transducers. arXiv. 2020;arXiv:2010.01003.
6. Lim D, Jung S, Kim E. JETS: Jointly training FastSpeech2 and HiFi-GAN for end-to-end text-to-speech. arXiv. 2022;arXiv:2203.16852.

7. AI-Hub. Dysarthric speech recognition data. Available from: https://aihub.or.kr/aihubdata/data/view.do?currMenu=115&topMenu=100&dataSetSn=608
8. KSS Dataset: Korean Single Speaker Speech Dataset. Available from: https://www.kaggle.com/datasets/bryanpark/korean-single-speaker-speech-dataset
9. Watanabe S, Hori T, Karita S, Hayashi T, Nishitoba S, Unno Y, et al. ESPnet: End-to-end speech processing toolkit. arXiv. 2018;arXiv:1804.00015.
10. Shen J, Pang R, Weiss RJ, Schuster M, Jaitly N, Yang Z, et al. Natural TTS synthesis by conditioning WaveNet on mel spectrogram predictions. Proc 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 15-20 April 2018; Calgary, AB, Canada. 2018. p. 4779-83.
11. Li N, Liu S, Liu Y, Zhao M, Liu M, Zhou L. Neural speech synthesis with transformer network. Proc AAAI Conference on Artificial Intelligence; 27 January–1 February 2019; Honolulu, HI, USA. 2019. p. 6706-13.

NOTES

NOTES

NOTES

BOSTECH Co., Ltd.

Total IT Infrastructure Integration & Maintenance Services

BOSTECH is a specialized Korean IT company providing fully integrated infrastructure solutions across defense, public, and enterprise sectors. From custom domestic server manufacturing (QTX) to virtualization, SDN, and encrypted data solutions, we connect your systems with trusted performance.

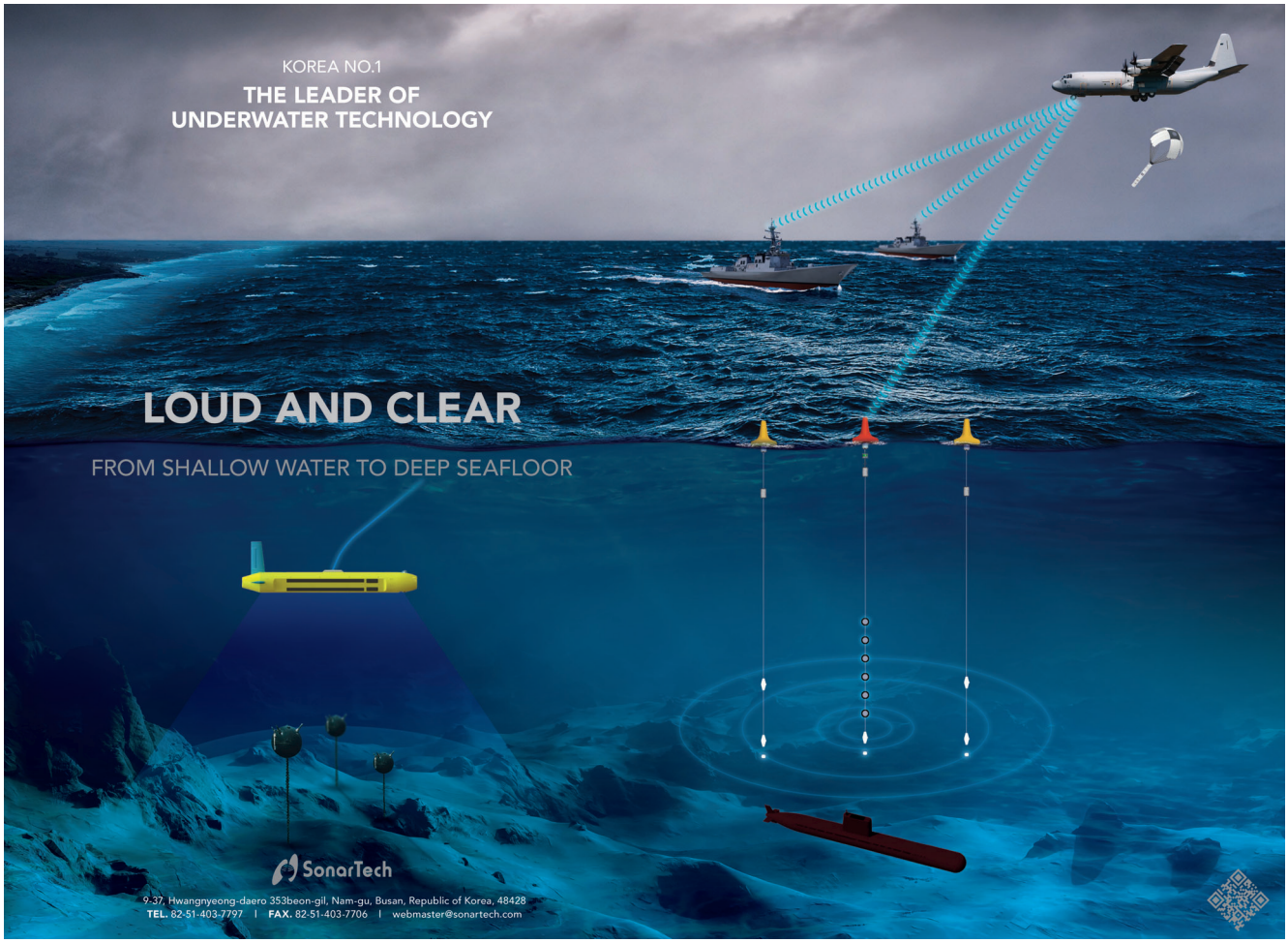
Our expertise spans:

- Integrated maintenance services across server, storage, and network
- Cloud & HCI solutions using VMWare, Arista, Dell, and SUSE
- DB encryption support for mission-critical government systems
- 10+ years of experience with Korea's defense infrastructure (MOD, Cyber Command, DIDC)

We build secure, scalable, and high-availability systems — from planning to operation — through a proven, customer-first process.

Seoul, Republic of Korea | www.bostech.co.kr
kdson@bostech.co.kr | +82-2-3667-5803

BOSTECH
BUILDERS OF OUTSTANDING SOLUTION



ORUN Technologies Co., Ltd.

Pioneering Advanced Simulation and Underwater Acoustic Solutions

Since 2001, ORUN Technologies has delivered cutting-edge simulation and underwater acoustic solutions for naval defense applications. Based in Daejeon, Korea, we collaborate with LIG Nex1, Hanwha, and ADD to develop mission-critical systems.

Core Capabilities:

• Sonar & Combat System Simulators
Customized simulators for sonar signal generation, target detection, and tactical training—applied to KSS-I/III, KDDX, and harbor surveillance systems.

• 3D Tactical Visualization with Unreal Engine
Immersive, real-time 3D combat environments enabling realistic scenario-based training and operational analysis.

• OriRay: High-Speed Acoustic Propagation Model
Our proprietary OriRay model rapidly simulates broadband acoustic transmission loss and detection probability using hybrid Gaussian–Geometric beams and frequency-dependent acoustic physics. Optimized for AI-driven sonar performance evaluation.

Technical Highlights:

• Real-time sonar signal simulation with Doppler effect and multipath propagation
• Generation of acoustic visibility maps for AI-based training and analysis

With over 20 years of expertise and more than 50 national defense projects successfully completed, ORUN is a trusted leader in sonar simulation and underwater acoustic innovation.

Daejeon, Korea | www.orun.co.kr
njseoung@orun.co.kr | +82-42-336-4600

Business field

- Acoustic targeting systems for sonar/torpedoes.
- Sonar transmitters/receivers and signal processors.
- High power amplifiers and power supplies for active sonar.
- Underwater acoustic transducers.
- Marine underwater acoustic signal acquisition/processing systems.
- Underwater navigation sonar systems.

TOV
TECHNOLOGY

직원이 행복한 기업, 나라를 생각하는 기업

www.tovtek.com

TOV tech. Co., Ltd.

주식회사 토브텍

대표이사	허을회
설립년도	2012년
임 직 원	15명 : 연구원 12명 (석/박사 개발자 5명)
소 재 지	대전광역시 유성구 테크노4로 17, B동 314~315호 (관평동, 대덕비즈센터)

방위산업의 미래기술을 개척하고 더 안전한 내일을 설계합니다

2012년 설립된 주식회사 토브텍은 수중무기체계 관련 다양한 시뮬레이터, M&S(Modeling & Simulation) 및 시험 장비를 개발하고 있습니다. 또한 신호처리 기술을 바탕으로 해양산업을 비롯하여 항공 등 다양한 분야로 비즈니스 영역을 확장하고 있습니다.

최근 급변하는 세계 안보의 중심에서 한국 방위산업은 그 우수성을 꾸준히 인정받고 있습니다.
10여 년간 대한민국의 주요 방위산업의 파트너로 함께 성장해 온 토브텍은 도전과 혁신으로 미래기술을 개척하고 더 안전한 우리의 내일을 설계해 나갈 것입니다.

토브텍의 최고의 가치는 우리 모두가 행복한 동행입니다.
기술혁신과 인재양성으로 고객과 함께 더 행복한 내일을 꿈꾸고 지속적으로 성장하는 토브텍이 되겠습니다.

주요 사업분야

- 수중무기체계 시뮬레이터
- 시험/분석 장비 해상 시험 장비
- 솔루션 개발