RATIONAL ORTHOGONAL WAVELET PULSE AND FEATURE EXTRACTION METHOD BASED ON AUDITORY FREQUENCY CEPSTRAL COEFFICIENT FOR UNDERWATER TARGET LOCALIZATION

TIANTIAN GUO[1,2], ENG GEE LIM[1], MIGUEL LÓPEZ-BENÍTEZ[2,3], MA FEI[4], YU LIMIN[1]

[1]Department of Communications and Networking, School of Advanced Technology, Xi'an Jiaotong-Liverpool University (XJTLU), Suzhou, China
[2]Department of Electrical Engineering and Electronics, University of Liverpool, Liverpool, Merseyside, UK
[3]ARIES Research Centre, Antonio de Nebrija University, 28040 Madrid, Spain
[4]Department Applied Mathematics, School of Science, Xi'an Jiaotong-Liverpool University (XJTLU), Suzhou, China
E-MAIL: Tiantian.guo19@student.xjtlu.edu.cn, Enggee.lim@xjtlu.edu.cn, M.Lopez-Benitez@liverpool.ac.uk, Fei.ma@xjtlu.edu.cn, Limin.yu@xjtlu.edu.cn

Abstract:

Underwater target localization has always been a challenging and important research topic in underwater acoustic sensing, especially when the underwater target is moving. This paper uses three pulse signals for system design: continuous wave, linear frequency modulated signal and rational orthogonal wavelet signal. We use a geometric underwater channel model to generate a database of underwater signals with specified geometric parameters of ocean environments to simulate the ocean. The received pulses signal are converted into feature maps as the classifier's input. In this paper, Short-time Fourier transform, Mel-frequency cepstral coefficient, Gammatone frequency cepstral coefficient and Perceptual linear prediction coefficient are applied to construct different feature maps. The classifier uses a lightweight CNN model. Experiments demonstrate the superiority of wavelet pulse signals in underwater target localization. The multipath effect will also contribute to underwater acoustic sensing.

Keywords:

Underwater communication; CNN; Mel frequency cepstral coefficient; Gammatone frequency cepstral coefficient; Rational orthogonal wavelet

1. Introduction

Due to the complexity of the underwater environment and the limitation of underwater acoustic equipment, strong noise will interfere with the received target signal [1]. In recent years, the localization of shallow sea sound sources, especially underwater low-frequency broadband sound sources, has received much attention from researchers. Similar to electromagnetic positioning on land, wireless transmission of information underwater is often carried out using acoustic waves as the carrier.

Underwater propagation has high time delays and limited available bandwidth [2]. The propagation rate of sound waves underwater is low and varies with the water's temperature, salinity and pressure. The acoustic bandwidth that can be used is much smaller due to the hydroacoustic channel. At the same time, underwater propagation multipath is a serious phenomenon. Signal transmission energy attenuation is high in the underwater channel due to absorption, scattering, and reflection losses. Reflections from the surface and bottom of the water distort the amplitude and phase of the received acoustic signal. Therefore, underwater target localization is complicated.

Similar to speech signals, underwater acoustic signals are non-stationary signals. Brown et al. [3] theoretically demonstrated the feasibility of using auditory for underwater signal analysis. Commonly used auditory-based features are the Mel frequency cepstral coefficients (MFCC), Perceptual linear predictive coefficients (PLPC), and Gammatone frequency cepstral coefficients (GFCC). GFCC algorithm is filtered directly in the time domain, which avoids the errors caused by the spectral estimation in the MFCC algorithm [4]. The gammatone filter has a simple time-domain impulse response [5].

Wavelets are suitable for the analysis of non-stationary signals, such as speech signals, because they have good dynamic properties and correspond to the auditory properties of the human ear. The orthogonality and good time-frequency localization properties of some wavelets provide great flexibility in pulse design and form the basis for good noise suppression, Doppler robustness and detection performance [6].

In this paper, we will focus on increasing the performance of underwater target localization with the auditory feature extraction method. We will apply rational orthogonal wavelet (ROW) as transmit pulses. The rest of this paper is organized as follows: Section 2 presents our designed broadband underwater environment with the ray-tracing model, the construction of the database, and feature extraction methods. In Section 3, classification results are presented in tables and bar charts. We will discuss the findings. Section 4 is the conclusion and future work.

2. Methodology

Classical ray tracing (geometric acoustics) is used to simulate the received echoes generated in a simple horizontally stratified three-layer acoustic environment. We will only consider plane waves. Next, three different signals are regarded as emitting pulse signals for comparison.

2.1. Underwater channel simulation

We use a ray-tracing model to simulate broadband underwater acoustic channels. Our channel model refers to the channel model in [7], which also considers multipath fading. We use the number of quartets nq to measure the extent of multipath. A larger value of nq means a larger number of eigenpaths are included. Figure 1 shows the simulation of underwater environment with receiver and target parameters. For target moving directions, we consider 270°, 90° and 0° to illustrate West, East and North.

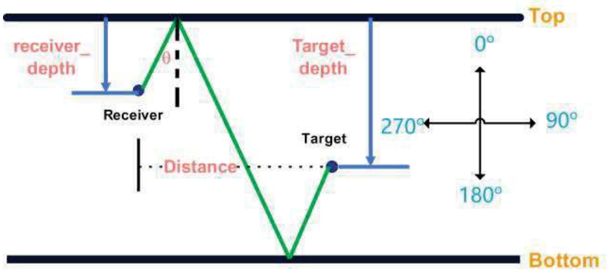

FIGURE 1. Underwater environment simulation.

2.2. Transmitted pulses

The available transmitted pulse types are CW (Continuous Wave), LFM (Linear frequency modulation) and ROW pulses. CW is a narrow-band signal with a carrier frequency of 20 kHz. LFM and ROW are broadband signals. The pulse duration is set at 0.001s for each pulse type. For LFM signals, the frequency bandwidth is from 15 kHz to 20 kHz. All sampling rate is 100 kHz. Detailed formulations for CW and LFM are illustrated in [8].

The ROW pulse we will consider is the superposition of four sub-pulses, forming the ROW pulse with frequency shifts and designed in the frequency domain [9]. The pulse duration is also set as 0.001s. For ROW pulses, various dilation factors of rational orthogonal wavelets are investigated in our experiment. In this paper, we consider dilation factors q=2, 4, 8, and a much higher number q=50. For dilation factor q=50, we increased the sampling rate to 20 kHz. Figure 2 illustrates the time response and frequency spectrum of three different pulses.

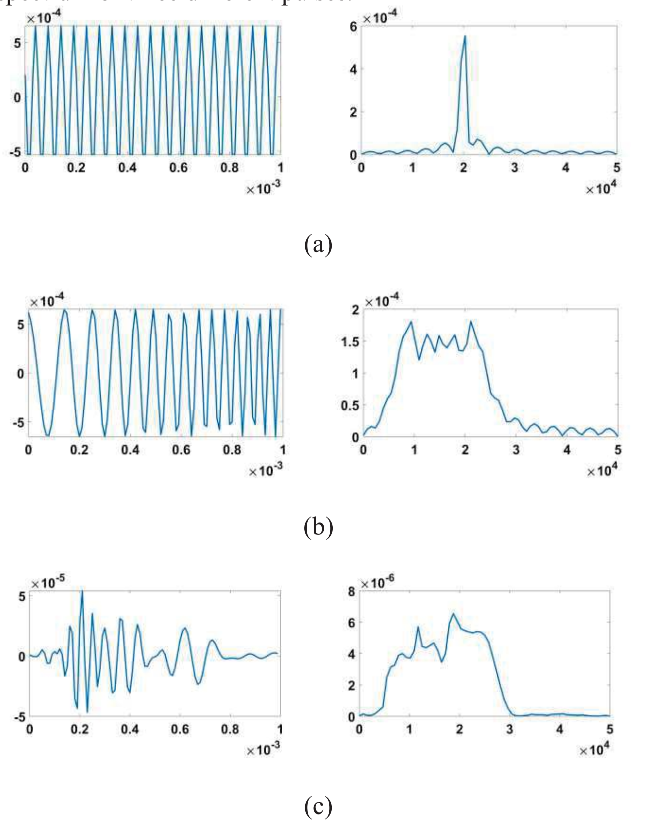

FIGURE 2. Three different pulses and their frequency spectrum: (a) CW; (b) LFM; (c) ROW.

2.3. Underwater database construction

All noise data were obtained from the underwater sound recording database [10]. We selected various ambient noises, whale calls, fish noises, boat noises and AWGN to simulate the ocean noise. Different signal-to-noise ratios are used to generate the database. Our SNR setting is from -45 dB to -5 dB with a step of 5 dB, indicating the extent of signals distorted by noise.

In our experiments, we divided received signals into some frames and then extracted features of each frame with different feature extraction methods. Our investigation chooses 13 features and their first and second derivation so that there are 39 features in one frame. The derivation features are used to demonstrate the dynamic characteristics of signals. Feature map size is 39*F, where F is the number of frames. Therefore, we converted one-dimension signals into two-dimension images as input of the classifier. Firstly, we consider Short-time Fourier transform (STFT), MFCC, GFCC and PLPC. Figure 3 demonstrates the flowcharts of each of the four feature extraction methods, which makes it easy to see the differences. For target localization, we will consider three criteria: target depth and distance between the target and receiver. For the multipath measure, we choose nq=4, a medium value.

Secondly, we will investigate the effect of multipath on target localization by varying the number of quartets. We will study 6 values of nq, nq=1, 2, 3, 4, 6 and 8. In addition, we will use different dilation factors for the ROW pulse signals. For the pre-processing method, we use the classical STFT. Therefore, a database of different orientations is also created.

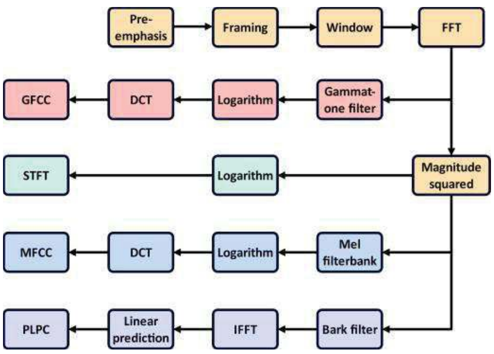

FIGURE 3. Underwater environment simulation.

3. Underwater target classification

Each situation has five classes to classify. We will introduce six pulses (CW, LFM, ROW q=2, ROW q=4, ROW q=8, ROW q=50) and four pre-processing methods (STFT, MFCC, GFCC, PLPC). We consider a 3-convolution-layer CNN; the structure is shown in [8]. The input of CNN is a 39*F*1 feature map. For each classification situation, we selected the average of the results of 10 CNN classifications as the final result for comparison.

3.1. Target depth

For the target depth, we chose the direction of motion to be 90^{o} because when the target is moved in a 90^{o} direction, its depth does not change, as shown in Figure 1. The target depth is from 10 m to 50 m with a step of 10m. Other parameters are fixed, such as the velocity of the target and the distance between the receiver and the target. Since all pre-processing methods achieve 100% accuracy after the SNR is raised above 0 dB, Table 1 demonstrates the case of SNR = -45 dB to -25 dB. Figure 4 shows the classification results for six pulses under four pre-processing methods with SNR = -25 dB to -5 dB.

In Table 1, the conventional STFT method is far inferior to other feature extraction methods when the signal is heavily affected by noise, i.e., when the SNR is less than -25 dB. In addition, ROW q=50 pulse gives the best results when compared within the same SNR value. As the SNR increases, the STFT method shows its superiority, as shown in Figure 4. On the contrary, the MFCC method is the least effective, as the results under the MFCC method are not stable among the 10 CNN results, which is related to the value of the features extracted by MFCC. This is because CNN is a numerical calculation of the values in feature maps. The GFCC and PLPC methods give good results when the noise is very influential. In this case, GFCC results are the most stable; although they are affected by different pulses, they can always maintain an accuracy rate of about 60%.

TABLE 1. Accuracy results for different pulses under lower SNR

	SNR	**-45**	**-40**	**-35**	**-30**	**-25**
STFT	CW	22.806	30.306	57.029	61.223	85.666
	LFM	18.166	22.472	35.056	56.667	66.694
	q=2	18.722	20.472	29.138	36.972	59.168
	q=4	18.194	22.862	17.222	48.61	62.111
	q=8	19.583	19.944	26.25	57.806	65.642
	q=50	34.917	40.002	56.805	64.305	81.111
MFCC	CW	48.861	47.055	56.668	59.36	57.666
	LFM	56.138	53.501	57.361	58.055	58.195
	q=2	36.278	47.194	47.751	57.584	40.695
	q=4	54.668	51.585	55.362	58.417	52.64
	q=8	60.14	62.415	57.941	59.723	58.001
	q=50	58.335	60.501	59.612	57.612	58.861
GFCC	CW	61.639	63.416	59.194	64.527	66.917
	LFM	58.861	59.11	61.972	63.22	62.361
	q=2	61.389	62.111	63.999	60.75	63.972
	q=4	59.084	63.722	59.167	65.971	71.89
	q=8	58.499	60.306	62.054	58.446	64.112
	q=50	58.029	61.695	64.389	62.251	60.028
PLPC	CW	61.027	55.389	60.694	59.499	67.472
	LFM	62.305	56.748	59.695	61.833	64.695
	q=2	61.778	58.53	57.332	59.973	59.222
	q=4	61.054	60.333	59.638	58.389	70.584
	q=8	60.083	57.943	57.75	58.583	58.833
	q=50	59.973	61.498	61.138	61.667	71.527

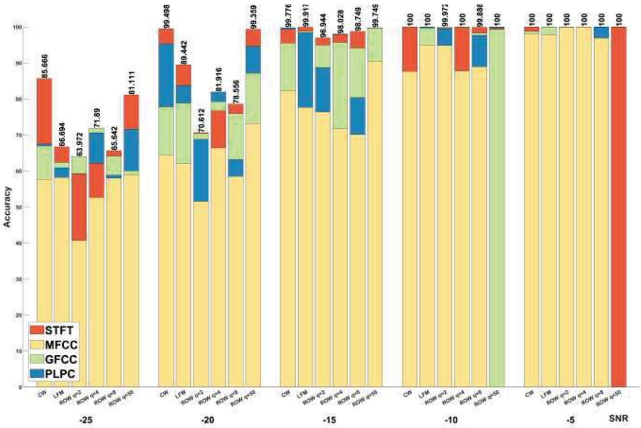

FIGURE 4. Target depth classification results under higher SNR.

3.2. Distance between receiver and target

For the horizontal distance between the target and the receiver, we also choose the target's direction of motion to be 0º, as the target moves in the 0º direction and the initial distance to the receiver does not change, as shown in Figure 1. The distance setting is from 1000 m to 1400 m with a step of 100 m. Other parameters, such as velocity and depth, are fixed. Figure 5 shows the classification results for six pulses under four different distance pre-processing methods. Since all pre-processing methods achieve 100% accuracy after the SNR is raised above 0 dB, we also draw the case of SNR = -45 dB to -5 dB.

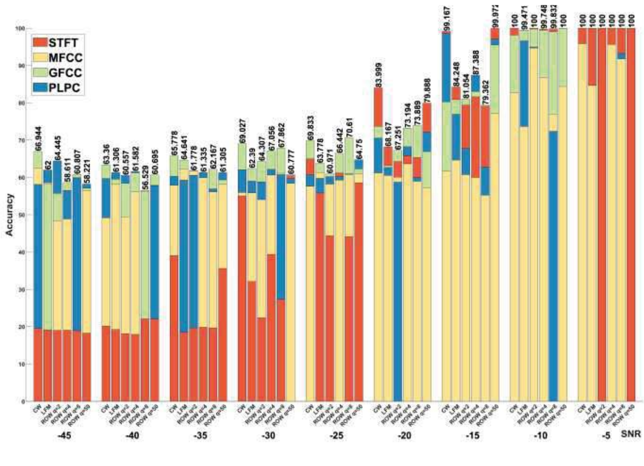

FIGURE 5. Target distance classification

From Figure 5, the GFCC has an advantage in the classification of distances, especially when the signals are heavily disturbed by noise. MFCC is only better at lower SNRs and always worse at higher SNRs. For distance classification, however, the PLPC results are even slightly better than the GFCC results. Compared with traditional MFCC, GFCC has higher robustness to low SNR situations. This characteristic lead GFCC to have good noise immunity. The ROW pulse signal with a higher dilation factor has a slight advantage when SNR increases. However, CW pulses can also achieve such results under other pre-processing methods. Again, ROW q=8 is slightly better than the other dilation factors.

3.3. Different nq values for target depth

For the effect of multipath, we considered different values of nq, from a small multipath effect nq=1 to a severe multipath effect nq=8 so that there are a total of 6 nq values (nq=1, 2, 3, 4, 6, 8). The signals emitted were selected from the ROW pulse signals with different dilation factors. The settings for the target depth are the same as in Section 3.1. Figure 6 illustrates the accuracies of varying nq values with four dilation factors ROW pulse signals. For the convenience of comparison, we selected STFT as the pre-processing method. Because under the STFT method, the result of the signal affected by the noise is very different, it is convenient for us to study the influence caused by different nq values and dilation factors.

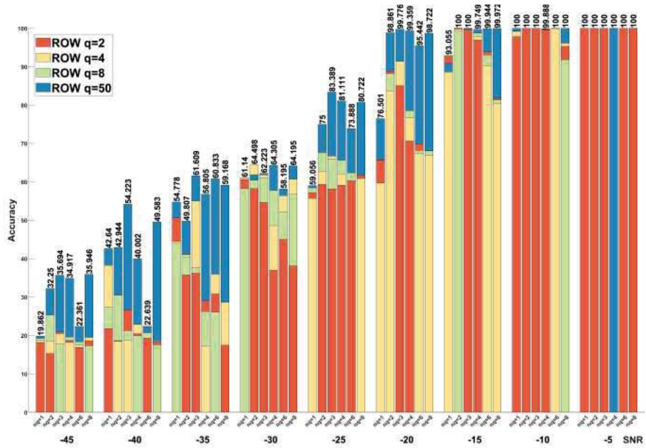

FIGURE 6. Target depth classification with different q and nq values

The ROW pulse signal with dilation factor q=50 undoubtedly occupies a great advantage, especially when noise seriously affects the signal. The blue bar representing q=50 almost exceeds the other colours by a large margin. The reason is that when q=50 is selected, we increase the sampling frequency, and the extracted frequency features are better. Theoretically, self-interference may occur between the various paths of signal propagation. The increase in the signal propagation path will significantly impact the received signal, which is an important cause of fading. The orthogonal rational wavelet we adopt can resist the interference produced by multipath. In this way, multipath can even increase the amount of signal information, helping us better locate underwater targets. This can also be seen in Figure 6. It is not that the lower the nq value, the higher the classification accuracy. When the signal is affected by noise in different degrees, the result of nq=1 is almost the worst. However, the results of nq=8 are not the best, indicating that the negative impact of multipath is also great. When the SNR is very low, the classification accuracy of nq=3 is the highest. This demonstrates that multipath will contribute to target depth classification.

3.4. Different nq values for target distance

For the distance between the target and the receiver, the parameter settings are consistent with Section 3.3. The SNR range was also selected from -45 dB to -5 dB. Table 2 contains the accuracy results of different nq values and dilation factors when SNR = -25 dB to -5 dB. Figure 7 illustrates the accuracies of 6 nq values with four dilation factors ROW pulse signals under lower SNR.

TABLE 2. Accuracy results for different nq values and dilation factors under higher SNR

	SNR	**-25**	**-20**	**-15**	**-10**	**-5**
nq=1	q=2	62.473	59.39	76.444	98.638	100
	q=4	59.805	62.194	73.028	98.055	100
	q=8	58.723	65.944	82.722	99.441	100
	q=50	60.918	67.054	91.667	99.972	100
nq=2	q=2	59.667	70.14	95.722	100	100
	q=4	61.612	66.305	93.668	99.972	100
	q=8	60.917	63.554	91.973	100	100
	q=50	63.695	88.249	100	100	100
nq=3	q=2	58.028	70.417	98.084	100	100
	q=4	60.333	71.083	97.196	100	100
	q=8	61.972	67.194	94.611	100	100
	q=50	64.638	89.888	99.944	100	100
nq=4	q=2	44.361	64.251	79.39	100	100
	q=4	61.195	65.777	81.61	99.72	100
	q=8	44.14	65.361	79.111	99.832	100
	q=50	58.555	79.888	99.972	100	100
nq=6	q=2	42.695	59.499	71.083	95.778	100
	q=4	61.778	61.166	75.749	93.999	99.972
	q=8	58.89	62.583	72.446	92.916	100
	q=50	61.304	68.029	93.64	100	100
nq=8	q=2	22.667	54.111	63.249	77.696	98.861
	q=4	29.166	52.473	61.334	75.556	99.028
	q=8	21.89	59.501	58.333	75.804	98.749
	q=50	60.972	64.056	92.25	99.748	100

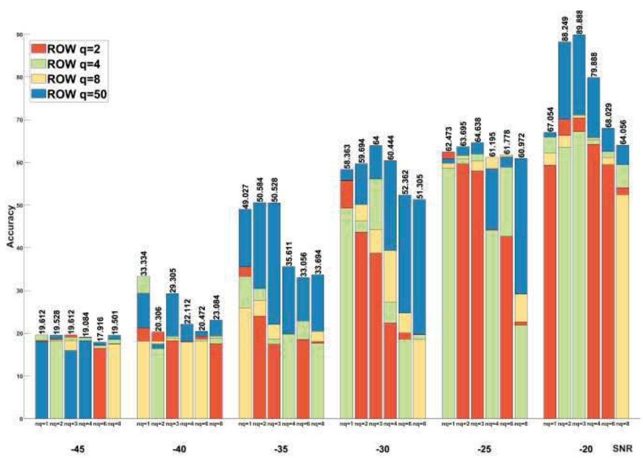

FIGURE 7. Target distance classification with different q and nq values under lower SNR.

Since this experiment is for classifying target and receiver distances, we need to consider various distance lengths, so the influence of multipath effect will become more prominent. It can be seen from Figure 7 that when the signal is greatly affected by noise, the situations are similar. However, as the SNR increases, the features extracted from the signal become more apparent. In most SNR situations, the result is the best when nq=3 with dilation factor q = 50. The multipath effect can also contribute to underwater acoustic sensing, but severe multipath still costs channel loss. A higher dilation factor will increase the classification results for each degree of the multipath effect. From Table 2, the ROW pulse signal with q=50 has the best outcome, especially when the multipath effect is burdensome (nq=8). It shows that in the case of a sizeable multipath effect, the dilation factor of the pulse signal should be increased. There are still some drawbacks for q = 50, such as the higher computation complexity and computation time cost.

4. Conclusions

In this paper, we design underwater pulse signals using rational orthogonal wavelets. We also consider four different dilation factors, ranging from a smaller value of q = 2 to a larger value of q = 50. Based on the underwater channel model, we have established a database of underwater acoustic received echo signals with different features of the target depth and the distance between the target and the receiver, including simple CW pulse signals, traditional LFM pulse signals, and four ROW pulse signals with different dilation factors. Four feature extraction methods, STFT, MFCC, GFCC and PLPC, are used for pre-processing received signals. Experimental results show that, in most cases, the results of ROW pulse signals are superior to other pulse signals. However, increasing dilation factors do not always lead to higher accuracy. When the SNR is very low, MFCC, GFCC, and PLPC can get better classification results. However, the effect of MFCC improvement is insignificant when SNR increases. GFCC has excellent advantages in distance classification between targets and receivers, and is more suitable in low SNR situations, with an accuracy of 66% when SNR = -45 dB.

We also demonstrated that multipath brings not only the negative impact of signal interference but also may contribute to the primary performance of the target location. It is demonstrated that ROW pulses, as the pulse signal transmitted by the target, can resolve the interference of multipath and preserve the amount of information contained in multipath signals. Therefore, the estimation of the multipath effect for hydroacoustic channels is necessary because we need more information for target localization. However, we also need to balance the loss of multipath effects with its contribution. After estimating the multipath effect, an appropriate dilation factor for ROW pulse signals needs to be chosen. A higher dilation factor can perform better but will have more computational complexity and require more computational time.

Acknowledgements

This research was partially funded by Research Enhancement Fund of XJTLU (REF-19-01-04), National Natural Science Foundation of China (NSFC) (Grant No. 61501380), and by AI University Research Center (AI-URC) and XJTLU Laboratory for Intelligent Computation and Financial Technology through XJTLU Key Programme Special Fund (KSFP-02), Jiangsu Data Science and Cognitive Computational Engineering Research Centre, ARIES Research Centre, and Suzhou Municipal Key Laboratory Broadband Wireless Access Technology.

References

[1] Z. Lian, K. Xu, J. Wan, G. Li, and Y. Chen, "Underwater acoustic target recognition based on gammatone filterbank and instantaneous frequency," in 2017 IEEE 9th International Conference on Communication Software and Networks (ICCSN), 2017, pp. 1207–1211.

[2] M. Stojanovic, "Underwater Wireless Communications: Current Achievements and Research Challenges," *IEEE Oceanic Engineering Society Newsletter,* Vol. XXXXI, No. 2, pp. 1-5, 2006.

[3] G. J. Brown, R. W. Mill, and S. Tucker, "Auditory-motivated techniques for detection and classification of passive sonar signals," *The Journal of the Acoustical Society of America,* vol. 123, no. 5, p. 3344, 2008.

[4] X. Wang, A. Liu, Y. Zhang, and F. Xue, "Underwater acoustic target recognition: A combination of multi-dimensional fusion features and modified deep neural network," *Remote. Sens.,* vol. 11, p. 1888, 2019

[5] R. F. Lyon, A. G. Katsiamis, and E. M. Drakakis, "History and future of auditory filter models," in Proceedings of 2010 IEEE International Symposium on Circuits and Systems, 2010, pp. 3809–3812.

[6] L. Yu, F. Ma, E. Lim, E. Cheng and L. B. White, "Rational-Orthogonal-Wavelet-Based Active Sonar Pulse and Detector Design," *IEEE Journal of Oceanic Engineering,* vol. 44, no. 1, pp. 167-178, Jan. 2019, doi: 10.1109/JOE.2018.2801158.

[7] J. Li and Y. V. Zakharov, "Efficient Use of Space-Time Clustering for Underwater Acoustic Communications," *IEEE Journal of Oceanic Engineering,* vol. 43, no. 1, pp. 173-183, Jan. 2018, doi: 10.1109/JOE.2017.2688558.

[8] T. Guo, E. G. Lim, M. López-Benítez, F. Ma and L. Yu, "Underwater Acoustic Sensing with Rational Orthogonal Wavelet Pulse and Auditory Frequency Cepstral Coefficient-Based Feature Extraction," 2022 19th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP), Chengdu, China, 2022, pp. 1-6, doi: 10.1109/ICCWAMTIP56608.2022.10016489.

[9] L. Yu and L. B. White, "Complex rational orthogonal wavelet and its application in communications," *IEEE Signal Processing Letters,* vol. 13, no. 8, pp. 477-480, 2006.

[10] "Sound in the Sea." Historic Naval Ships Association, https://www.hnsa.org/manuals-documents/historic-naval-sound-and-video/sound-in-the-sea/.