UNDERWATER ACOUSTIC SENSING WITH RATIONAL ORTHOGONAL WAVELET PULSE AND AUDITORY FREQUENCY CEPSTRAL COEFFICIENT-BASED FEATURE EXTRACTION

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

[1]Dept. of Communications and Networking, School of Advanced Technology, Xi'an Jiaotong-Liverpool University (XJTLU), Suzhou, China
[2]Dept. of Electrical Engineering and Electronics, University of Liverpool, Liverpool, Merseyside, UK
[3]ARIES Research Centre, Antonio de Nebrija University, 28040 Madrid, Spain
[4]Dept. 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:

Active pulse design, target detection and classification play an essential role in underwater acoustic sensing. This paper addresses the system design with three kinds of pulse signals, including continuous wave (CW) linear frequency modulation (LFM) signal and rational orthogonal wavelet (ROW) signal. The detector design has an architecture of feature extraction and convolutional neural network (CNN) based classification. A geometric underwater channel model is adopted to facilitate the generation of training datasets with designated geometric underwater environment parameters. The simulated received pulse signals are converted into feature maps as the input of the classifier. This paper applies the acoustic features, Short Time Fourier Transform (STFT), Mel Frequency Cepstral Coefficients (MFCC) and Gammatone Frequency Cepstral Coefficients (GFCC) to construct different feature maps. A lightweight CNN model is used as the classifier. Experiments demonstrate the superiority of the ROW wavelet pulse signals and the proposed algorithm in target localization and underwater signal classification.

Keywords:

Tracking; Underwater communication, CNN, Mel frequency cepstral coefficient (MFCC), Gammatone frequency cepstral coefficient (GFCC), Rational orthogonal wavelet (ROW)

1. Introduction

Ocean noises, sound velocity characteristics, subsea acoustic properties, and other distorting factors, such as multipath attenuation, influence acoustic sensing in the marine environment. In recent years, feature extraction of underwater acoustic signals has grown fast, the most prominent of which are speech recognition-based auditory feature extraction methods. However, the performance of traditional methods could not be well suited to a variety of complex environments [1].

Steven B. Davis first proposed the theory of Mel Frequency Spectral Coefficients (MFCC) by studying the characteristics of human hearing and finding that the human ear has different sensitivities to different frequency bands [2]. The gammatone filter has a simple time-domain impulse response [3].

Underwater acoustic signals have similar characteristics to speech signals. The auditory feature extraction methods used in speech signal processing can also be applied to underwater acoustic signal processing. The feasibility of using auditory for underwater signal analysis was theoretically demonstrated by Brown et al. [4]. Gammatone frequency cepstral coefficient (GFCC) algorithm avoids the errors caused by the spectral estimation in the MFCC algorithm because the signal is filtered directly in the time domain [5].

The wavelet transform provides a more accurate localization both in the time and frequency domains. The multi-resolution of the wavelet transform facilitates the extraction of different features at each resolution. Research has shown that wavelets have been applied to speech signal feature extraction with good results. The non-smooth nature of underwater signals is well suited to applying the wavelet transform.

The rest of this paper is organized as follows: Section 2 introduces the broadband underwater ray-tracing model, the database construction, feature extraction methods and our CNN architecture. In Section 3, classification results and

analysis are presented in bar charts. Section 4 is the conclusion and future work.

2. Methodology

Classical ray tracing (geometric acoustics) was used to synthesize the echo generated when the target is impacted in a simple horizontal layered three-layer marine acoustic environment. We only consider plane waves processed by their associated rays, which are vectors perpendicular to the wavefront. Next, three different signals will be considered as the transmitted signals.

2.1. Underwater channel model

A ray-tracing model was used to simulate broadband underwater acoustic channels. Figure 1 shows the underwater environment simulation. Target moving directions are illustrated as 270°, 90° and 0/180°.

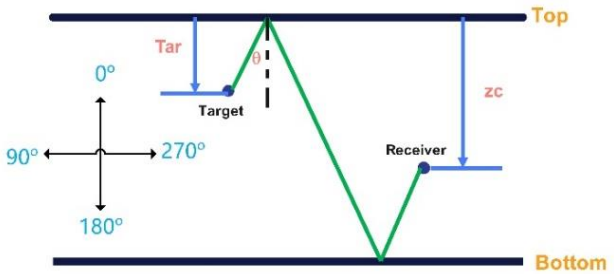

Figure 1: Underwater environment simulation

2.2. Transmitted pulses

The available pulse types are CW (Continuous Wave), LFM (Linear frequency modulation) and ROW (Rational Orthogonal Wavelet). CW is a narrow-band signal, and LFM and ROW are broadband signals.

The CW pulse is of the form:

$$\omega(t) * exp(-i * 2\pi f_c t) \quad (1)$$

Where f_c is the carrier frequency, and the LFM is

$$\omega(t) * exp(-i * 2\pi(f_a t + (f_b - f_a) * t^2/(2 * T))) \quad (2)$$

Function $\omega(t)$ is a "window", being rectangular in our experiment,

$$\omega(t) = 1 \quad (3)$$

Where f_a is the low frequency and f_b is the high frequency, T is the pulse duration.

ROW signals are designed wavelet signals in [6], the formulation is

$$\Psi(\omega) = \begin{cases} (2\pi)^{-\frac{1}{2}} e^{j\frac{\omega}{2}} \sin(\frac{\pi}{2}\beta(\frac{q}{\omega_1}|\omega| - q)), \omega_1 \le |\omega| \le \omega_2 \\ (2\pi)^{-\frac{1}{2}} e^{j\frac{\omega}{2}} \cos(\frac{\pi}{2}\beta(\frac{q}{\omega_2}|\omega| - q)), \omega_2 \le |\omega| \le \omega_3 \\ 0, \quad |\omega| \notin [\omega_1,\ \omega_3] \end{cases} \quad (4)$$

Where q is the dilation factor of ROW, and

$$\omega_1 = a_0 \cdot \left(q - \frac{q}{2q+1}\right)\pi,\ \omega_2 = a\omega_1,\ \omega_3 = a\omega_2 = a^2\omega_1 \quad (5)$$

$\beta(x)$ is the construction function and has the form in (6). It is not unique.

$$\beta(x) = x^4(35 - 84x + 70x^2 - 20x^3) \quad (6)$$

Figure 2 illustrates the time responses of three different transmit pulses. For ROW pulses, we will also consider various dilation factors of rational orthogonal wavelets.

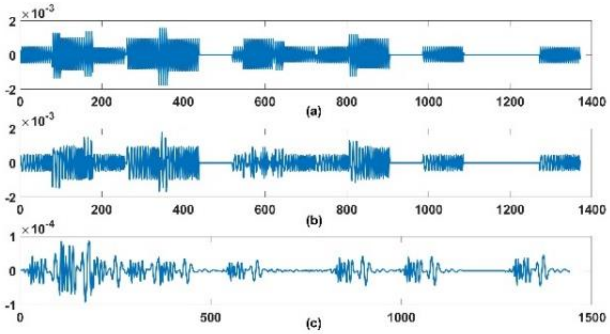

Figure 2: Three different transmit pulses: (a) CW; (b) LFM; (c)ROW.

2.3. Database construction

2.3.1. Noise type

All the noise data are from the underwater sound recording database [7]. We selected two ambient noises, two whale calls, two sea fish noises, two boat noises and AWGN, nine noises with different SNRs to generate the database. The complexity of the channel strongly influences the performance of underwater acoustic communication systems. Noise interfering with underwater acoustic communication includes underwater dynamics, activity noise generated by aquatic organisms, and natural noise from sea surface waves and storms. These noises can seriously affect the SNR of the signal. Therefore, our SNR setting is from -45dB to 0dB with a step of 5dB.

2.3.2. Preprocessing method

In our experiments, three feature extraction methods were considered to transfer one-dimension signals to two-dimensions feature maps as the input of the Neural Network. Short-time Fourier transform (STFT) is the comparison traditional signal processing technique benchmark. In

addition, two auditory-based feature extraction methods, MFCC and GFCC, are also investigated in our experiment. Their derivative features are also introduced in feature extraction to incorporate dynamic features. Figure 3 shows examples of MFCC and GFCC feature maps. Our investigation set 13 MFCC and GFCC features and their first and second derivation, totalling 39 features in one frame. So feature map size is 39 ∗ F, where F is the number of frames.

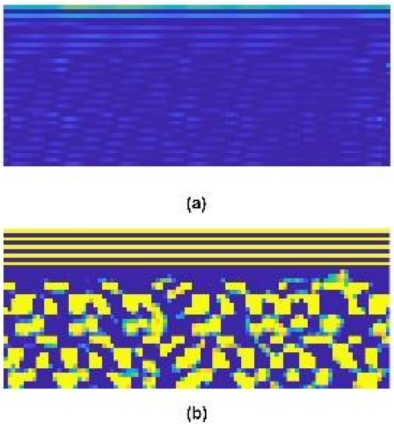

Figure 3: Feature map examples: (a) GFCC; (b) MFCC

2.3.3. Classification setting.

In our experiment, we consider a 3-convolution-layer CNN structure. Figure 4 is the designed structure of CNN. Conv denotes the convolution layer, and Fc represents the fully connected layer. The input is a 39 ∗ F ∗ 1 feature map. The last layer gives a classification score to five categories for each underwater signal set.

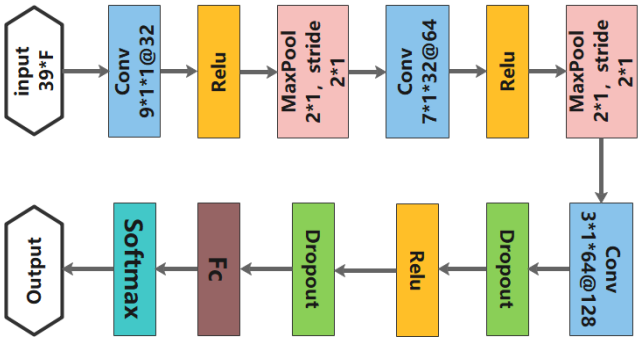

Figure 4: CNN architecture

3. Underwater classification

In our experiments, we choose target depth, speed, and distance to classify for underwater target localization and tracking. Each situation has five classes. We will introduce five pulses (CW, LFM, ROW q=2, ROW q=8, ROW q=50) and three preprocessing methods (STFT, MFCC, GFCC). 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 degrees because the target moves in a 90-degree direction with no change in depth, as shown in Figure 1. Other parameters, such as target speed and distance between receiver and target, are fixed. Figure 5 shows the classification results of 5 pulses under three different preprocessing methods for target depth.

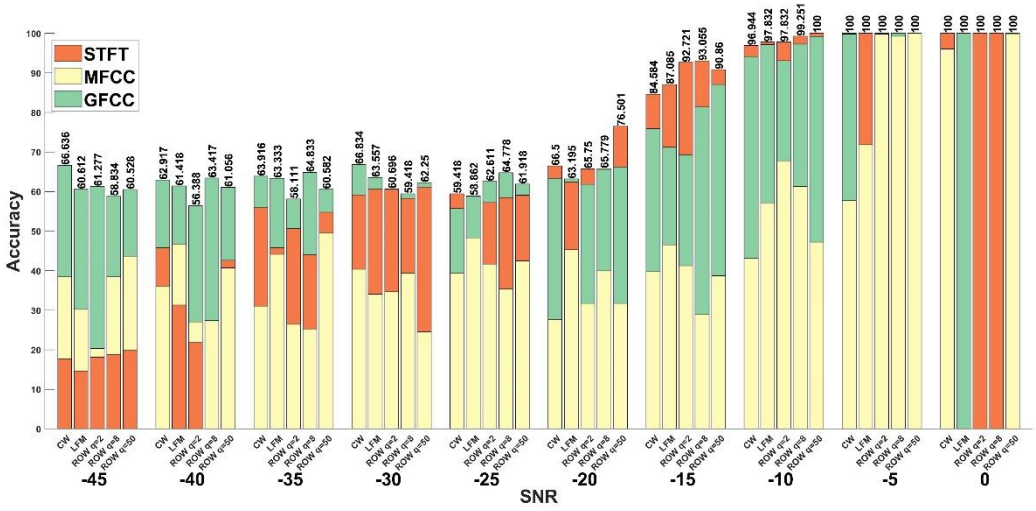

Figure 5: Target depth classification results.

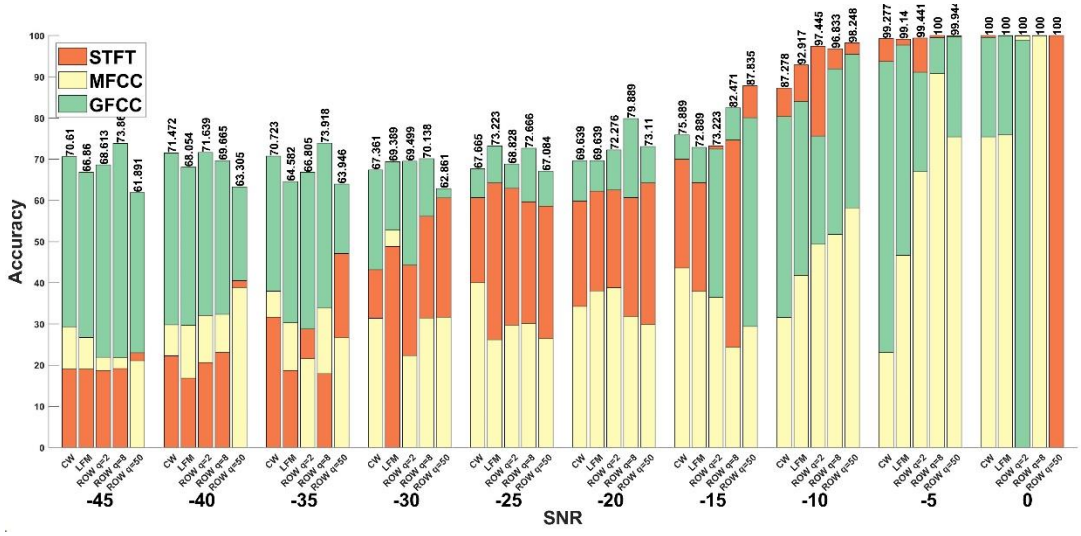

Figure 6: Distance classification results.

We can find from the figures that the GFCC method generally outperforms the other two methods when SNR=-45dB to -25dB. When SNR is very low, the MFCC method also has better performance than the STFT method. However, as SNR increases, the MFCC method, on the other hand, produced the worst results due to the fact that the results of the MFCC were erratic and intermittently high and low over the ten classifications of each case, resulting in a lower mean value than the other two methods. Therefore, for target depth classification, GFCC is the best choice when the SNR is very low. However, the traditional STFT method has no advantages as the SNR increases, especially with ROW pulse signals.

3.2. Distance between target and receiver

For the distance between the target and receiver, we chose the direction of motion to be 0 degrees because the target moves in a 0-degree direction with no change in initial distance, as shown in Figure 1. Other parameters, such as speed and depth, are fixed. Figure 6 shows the classification results of 5 pulses under three different preprocessing methods for distance.

The five pulses of GFCC are clearly superior at SNRs below -20dB, and MFCC has a slight advantage over STFT when SNR is very low. ROW pulse signals do not have an absolute advantage over other pulses. At low SNRs, the five pulses are similar due to the effectiveness of GFCC, and at higher SNRs, ROW pulse signals are more effective for GFCC and MFCC. This phenomenon also verifies the superiority of the GFCC method in underwater target localization.

3.3. Target speed

For the target speed, we chose the direction of motion to be 0 degrees because when the target moves at 0 degrees, the distance to the receiver does not change, as shown in Figure 1. Therefore, the Doppler effect due to the relative motion of the target does not exist. Since the Doppler effect is related to the velocity of the movement, other directions of movement can be classified by the Doppler effect. Without the Doppler effect, it would be difficult to classify the velocity of the target's motion, which in turn would validate the advantages of ROW pulses. Other parameters, such as speed and distance, are fixed. Figure 7 shows the classification results of 5 pulses under three different preprocessing methods for target speed.

The GFCC method generally outperforms the other two methods in most SNR conditions. MFCC method performs poorly in almost all SNR conditions. The figures also illustrated that ROW pulse signals have a significant advantage over traditional pulses CW and LFM, especially when dilation factor q equals 8 and 50. When SNR increases, the STFT method achieves better performance than the GFCC method. Therefore, for a target moving speed classification, since we choose the motion degree 0 so that there is no Doppler effect to help classification, the GFCC method with ROW pulse signals is the optimal choice when the SNR is not very high. However, as the SNR increases, the traditional method of STFT still has advantages, especially combined with ROW pulse signals.

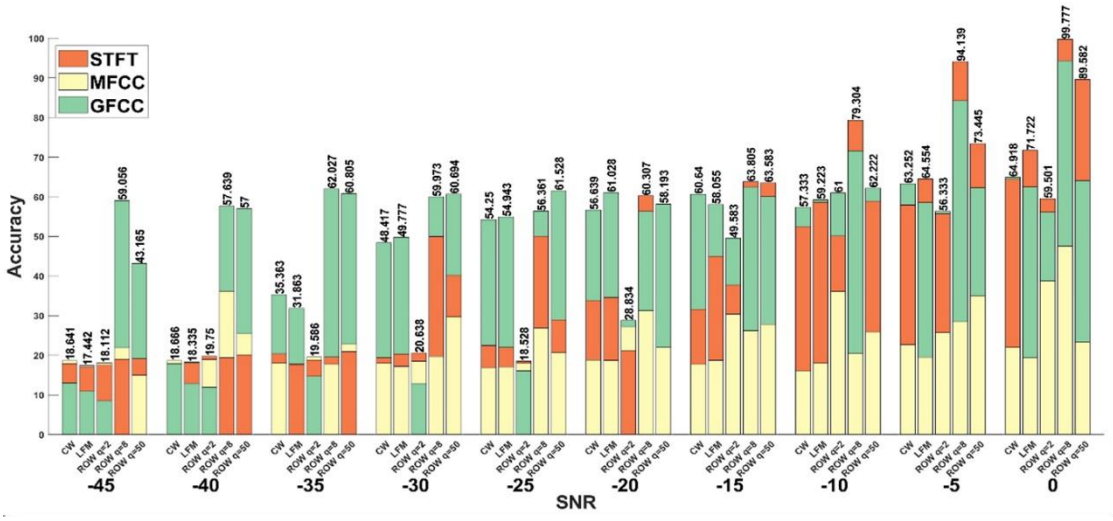

Figure 7: Target speed classification results.

4. Conclusions

The paper applied several feature extraction methods covering STFT, MFCC and GFCC. The extracted feature maps serve as the input of CNN. Traditional pulse signals of CW and LFM signals are compared with the ROW signals of different dilation factors to find more effective pulse signals for target localization in the underwater environment. The echo signals were simulated by a geometric acoustic ray tracing channel model under designated scenarios. Experiments show that the GFCC method has superiority over the other two feature extraction methods in most situations, especially when the SNR is very low and the underwater acoustic signals are severely distorted. For target speed classification, ROW pulse signals show their significant superiority over traditional pulses when there is no Doppler effect. A higher dilation factor leads to better performance. In future work, more experiments should be conducted in the direction of motion, and the classification of the target direction of movement can also be investigated under the same framework. The fusion of different features may achieve better system performance.

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, and ARIES Research Centre.

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] Davis and B. Steven, "Evaluation of acoustic parameters for monosyllabic word identification," *Journal of the Acoustical Society of America*, vol. 64, no. S1, pp. S180–S181, 1978.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] https://www.hnsa.org/manuals-documents/historic-naval-sound-and-video/sound-in-the-sea/.