Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Recognition method for underwater communication signals that mimic dolphin whistles using phase-shifting modulation*

Qingwang YAO[1], Jiajia JIANG[†‡1], Xiaolong YU[2], Zhuochen LI[1], Xiaozong HOU[1], Xiao FU[1], Fajie DUAN[1]

[1]*State Key Lab of Precision Measuring Technology and Instruments, Tianjin University, Tianjin 300354, China*

[2]*Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China*

[†]E-mail: jiajiajiang@tju.edu.cn

Received July 5, 2024; Revision accepted Mar. 5, 2025; Crosschecked Aug. 6, 2025

Abstract: With the introduction of underwater bionic camouflage covert communication, conventional communication signal recognition methods can no longer meet the needs of current underwater military confrontations. However, the research on bionic communication signal recognition is still not comprehensive. This paper takes underwater communication signals that mimic dolphin whistles through phase-shifting modulation as the research object, and proposes a recognition method based on a convolutional neural network. A time–frequency contour (TFC) masking filtering method is designed, which uses image technology to obtain the TFC mask of whistles and extracts whistles from the obtained mask. Spatial diversity combining is used to suppress the signal fading in multipath channels. The phase derivative spectrum image is obtained by Hilbert transform and continuous wavelet transform, and is then used as the basis for recognition. Finally, the effectiveness of the proposed method is verified by simulations and lake experiments. In the simulations, a recognition accuracy of 90% is achieved at a signal-to-noise ratio (SNR) of 0 dB in multipath channels. In the real underwater communication environment, a recognition accuracy of 81% is achieved at a symbol width of 50 ms and an SNR of 6.36 dB.

Key words: Underwater acoustic signal recognition; Bionic camouflage covert communication; Time–frequency contour masking filtering; Convolutional neural network

https://doi.org/10.1631/FITEE.2400572 **CLC number:** TN911.72

1 Introduction

Underwater acoustic communication signal recognition represents a critical technology in contemporary maritime military confrontations. Conventional underwater acoustic communication signal recognition methods have demonstrated significant limitations and inherent deficiencies with the development of underwater bionic camouflage covert communication (UBCCC). Conventional covert communication is mainly based on two covert schemes: low probability of interception (LPI) (Li Y et al., 2019; Huang et al., 2020) and low probability of detection (LPD) (Schoolcraft, 1991; Diamant et al., 2017). LPI communication approaches achieve covertness by employing artificially constructed signals with varying parameters, such as frequency-hopping and time-hopping. LPD communication approaches achieve covert communication by reducing the signal transmission power and hiding communication signals in noise. Although signals generated based on the latter communication strategy afford a certain degree of concealment, they remain artificially constructed and exhibit characteristics significantly distinct from natural underwater ambient noise in the time, frequency, and time–frequency domains (Jiang et al., 2019, 2020). Consequently, these

‡ Corresponding author

* Project supported by the National Natural Science Foundation of China (No. 62231011) and the Tianjin Outstanding Young Scientists Fund Project (No. 24JCJQJC00240)

ORCID: Qingwang YAO, https://orcid.org/0009-0000-4743-1543; Jiajia JIANG, https://orcid.org/0000-0002-0611-6501

signals can be easily recognized by conventional signal detection methods (Iglesias et al., 2015; Liu F et al., 2016; Lee and Oh, 2019).

Traditional underwater communication relies on regular modulation methods, such as amplitude-shift keying (ASK), frequency-shift keying (FSK), phase-shift keying (PSK), minimum-shift keying (MSK), and quadrature phase-shift keying (QPSK). UBCCC is grounded in the principle of low probability of recognition (LPR), and is characterized by modulation patterns and non-periodic properties. Both of these kinds of signals have essential differences in terms of modulation mode, and traditional communication recognition methods cannot recognize non-standard signals such as UBCCC. UBCCC takes large marine mammal original or modified calls as communication signals directly; therefore, this type of signal is very similar to natural acoustic signals. Traditional communication signal recognition processes typically classify such signals as natural acoustics or background noise, and filter them out. Therefore, UBCCC signals have high concealment and deception properties. UBCCC coding modes can be categorized as time–frequency contour (TFC) coding, symbol coding, and time delay difference (TDD) coding. Marine mammal calls are diverse, and symbol coding leverages this feature for encoding, using different calls to convey different data (Ahn et al., 2020; Bilal et al., 2020; Kaveh and Falahati, 2021). TDD coding converts communication data into the intervals of adjacent call fragments according to certain rules (Qiao et al., 2017; Jiang et al., 2018; Li CY et al., 2021). Unlike these two coding modes, TFC coding constructs communication signals by imitating the TFC of real call signals, and the communication data can be encoded by the TFC slope (Qiao et al., 2021), TFC jitter (Liu SZ et al., 2017), and phase shift (Ahn et al., 2019).

With the diversification of UBCCC coding forms, bionic communication recognition has gradually attracted widespread attention from researchers. Jiang et al. (2022) studied underwater bionic signal recognition, and proposed a recognition method based on the distribution characteristics of inter-click intervals for a TDD-encoded bionic click communication train. They found that the sorted inter-click intervals of bionic signals exhibit a ladder-like distribution. Casari et al. (2023) revealed the discernible differences between the acoustic properties of authentic dolphin whistles and their transducer-generated playback under specific conditions. Through field experiments, they identified transitive entropy as the most effective metric for distinguishing between the original and artificially generated signals. Jiang et al. (2023) studied TFC-encoded continuous-phase multiple FSK signals. The normalized power spectrum and Wiener entropy were extracted to construct feature vectors based on the time and frequency characteristics of real and bionic signals, and support vector machine (SVM) was used to recognize real and bionic signals. Yao et al. (2023) proposed a recognition method for TFC-encoded bionic binary orthogonal keying modulated signals. They found that the time–frequency curve slopes are mainly distributed around two positive and negative symmetric values, which can serve as the recognition characteristics. The method achieved a recognition accuracy of more than 90% at a signal-to-noise ratio (SNR) of −5 dB.

Compared with existing UBCCC coding forms, the underwater bionic signal recognition methods proposed in the above research are still not comprehensive. Therefore, we select the continuously varying carrier frequency-modulated signal (CVCFMS) (Ahn et al., 2019) as the object of study. CVCFMS is generated by mimicking the dolphin whistle signal (DWS) and encoding the communication data via phase shifts. The TFCs of the CVCFMSs are diverse and consistent with those of the DWSs, making it difficult to recognize the CVCFMSs using auditory and conventional detection methods. While DWSs exhibit continuous phase changes, CVCFMSs exhibit a phase shift phenomenon at different code symbol connections because of the need for data transmission; this represents an inherent coding characteristic that can be used to recognize CVCFMSs. Our main contributions are as follows:

1. In view of CVCFMS recognition, we propose a recognition method based on convolutional neural networks (CNNs), which takes the scalogram image of the phase derivative obtained by the Hilbert transform and continuous wavelet transform (CWT) as the input of the CNNs.

2. Aiming at whistle extraction, time–frequency spectrum (TFS) mask filtering is proposed to accurately extract whistle signals from the noise environment, where the TFS mask is obtained using the trend search and extension method. The phase of the whistle is obtained using the Hilbert transform.

3. To address the problems of multipath interference and signal fading in underwater acoustic channels, spatial diversity combining (SDC) is first used to suppress signal fading, and then TFS mask filtering is used to extract whistle signals accurately and remove multipath interference signals.

2 Signal characteristic analysis

DWSs are frequency-modulated signals characterized by time-varying frequencies, with durations typically ranging from hundreds of milliseconds to several seconds. DWSs are primarily used for communication between the individual dolphins or groups, emotional expression, and information transfer. The process for generating CVCFMS is depicted in Fig. 1, and the DWS and imitation object are digitized at 44.1 kHz (16 bits). First, the time–frequency curve of DWS in Fig. 1c is obtained, and then the symbol width τ is determined according to the coding requirements. Finally, the CVCFMS in Fig. 1b is generated according to Eq. (1):

$$C(t) = \cos\left[2\pi F(t)t + \pi S_k\right], k = 1, 2, \cdots, K, \quad (1)$$

where $F(t)$ denotes the carrier frequency, t denotes the time, S_k denotes the k^{th} communication value (0 or 1), and $C(t)$ denotes the CVCFMS waveform.

It can be observed that the TFC of CVCFMS in Fig. 1b exhibits a high degree of consistency with that of DWS in Fig. 1a. Meanwhile, the TFCs of imitation signals exhibit time-varying and diverse patterns, and the communication frequency and frequency band are not fixed, posing significant challenges for the conventional communication recognition methods in capturing and recognizing CVCFMS. However, Figs. 1e and 1d show that there is a clear difference between the phase derivative curve of the CVCFMS and that of DWS. The phase derivative curve of DWS shows continuous changes, but there are mutation points at the connections of different symbols on the CVCFMS phase derivative curve. Therefore, the mutation points on the phase derivative curve are used as a key distinguishing feature for recognizing CVCFMSs.

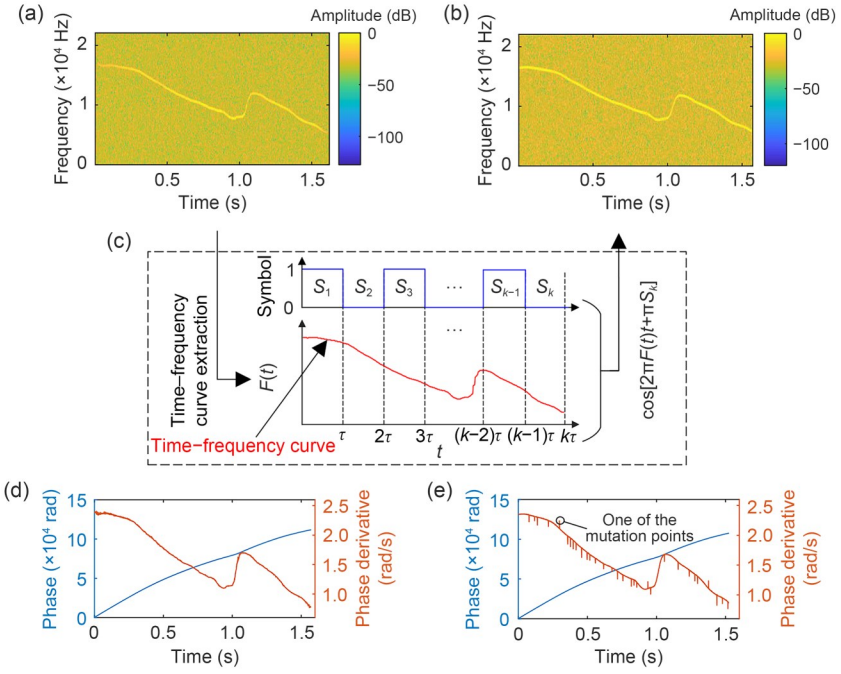

Fig. 1 TFC of DWS (a), TFC of CVCFMS (b), CVCFMS coding scheme (c), DWS phase and its derivative (d), and CVCFMS phase and its derivative (e)

3 The CVCFMS recognition process

Fig. 2 illustrates the CVCFMS recognition process. First, spatial diversity combining is used to combine multiple hydrophone signals. Second, TFS mask filtering is applied to extract the whistle from the combined signals. Third, the phase is obtained by the Hilbert transform for the above whistle signal, and then the phase derivative is calculated. Finally, the scalogram is calculated by the CWT, and the obtained scalogram image is taken as the input for CNN to recognize the CVCFMS.

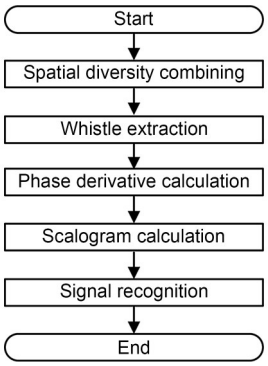

Fig. 2 The CVCFMS recognition process

3.1 Whistle extraction

The extraction of whistle signals is the primary step in the process of recognizing DWSs and CVCFMSs. The signal extraction directly affects the later phase calculation and thus affects the signal recognition accuracy. For the time-varying characteristics of whistle signal frequency, a TFS mask filtering method is proposed for whistle extraction, where the TFS mask is obtained using the image processing method. The TFS of a DWS is obtained using the short-time Fourier transform (STFT), and is shown in Fig. 3. The STFT can be expressed as

$$Z(r,c)=\frac{1}{N}\sum_{n=0}^{N-1}x(n+Sc)\omega(n)\exp\left(-\mathrm{i}\frac{2\pi r}{N}n\right), \quad (2)$$

where $Z(r, c)$ denotes the result of the STFT with window length N of 512 and step length S of 128, r denotes the frequency point index, c denotes the time window index, n denotes the sampling point index of the discrete digital signal, $\omega(n)$ represents the hamming window function, and $x(\,)$ is the input signal discrete sequence.

$\mathrm{ZA}(r, c)$ and $\mathrm{ZP}(r, c)$ denote the amplitude and phase of $Z(r, c)$, respectively, and can be calculated by

$$\begin{cases}\mathrm{ZA}(r,c)=\sqrt{a^2(r,c)+b^2(r,c)},\\ \mathrm{ZP}(r,c)=\arctan\left(b(r,c)/a(r,c)\right),\end{cases} \quad (3)$$

where $a(r, c)$ denotes the real $Z(r, c)$, and $b(r, c)$ denotes the image of $Z(r, c)$.

ZA can be treated as a two-dimensional image matrix, and the corresponding row and column index coordinate system of the two-dimensional image matrix coordinate *R-O-C* is labeled in Fig. 3a. First, mean filtering and median filtering are applied to suppress the noise in **ZA**, and then the image segmentation based on the dynamic threshold is used to extract the TFCs initially. The binary image of TFC $\mathbf{ZA}_{\mathrm{bin}}$ can be obtained by

$$\mathrm{ZA}_{\mathrm{bin}}(r,c)=\begin{cases}1, & \mathrm{ZA}(r,c)\geq \mathrm{Er}(r,c)\ \text{or}\\ & \mathrm{ZA}(r,c)\geq \mathrm{Ec}(r,c),\\ 0, & \text{otherwise},\end{cases} \quad (4)$$

where $\mathrm{Er}(r, c)$ and $\mathrm{Ec}(r, c)$ denote the row and column thresholds, respectively, and can be calculated by

$$\begin{cases}\mathrm{Er}(r,c)=\mathrm{mean}\left[\mathrm{ZA}(r-\Delta R{:}r+\Delta R,c)\right]\times 1.5,\\ \mathrm{Ec}(r,c)=\mathrm{mean}\left[\mathrm{ZA}(r,c-\Delta C{:}c+\Delta C)\right]\times 1.5.\end{cases} \quad (5)$$

Herein, ΔR is the row interval set to 10, and ΔC is the column interval set to 10.

Based on the 8-connected domain principle, the different TFCs of $\mathbf{ZA}_{\mathrm{bin}}$ are labeled with different tag values. The labeled result **ZAL** is shown in Fig. 3b, where different colors correspond to different tag values. From the labeled result **ZAL**, a large number of noise TFCs can be found, other than the TFCs of the whistle in $\mathbf{ZA}_{\mathrm{bin}}$. Therefore, the characteristic variable CS is constructed to characterize the TFCs of whistle and noise. The CS of the TFC with tag value L ($\mathrm{CS}(L)$) can be expressed by

$$\begin{gathered}\mathrm{CS}(L)=\mathrm{ColLength}(L)+\mathrm{AverageGray}(L),\\ L=1,2,\cdots,N_L,\end{gathered} \quad (6)$$

where N_L denotes the maximum tag value, and the $\mathrm{ColLength}(L)$ and $\mathrm{AverageGray}(L)$ represent the column

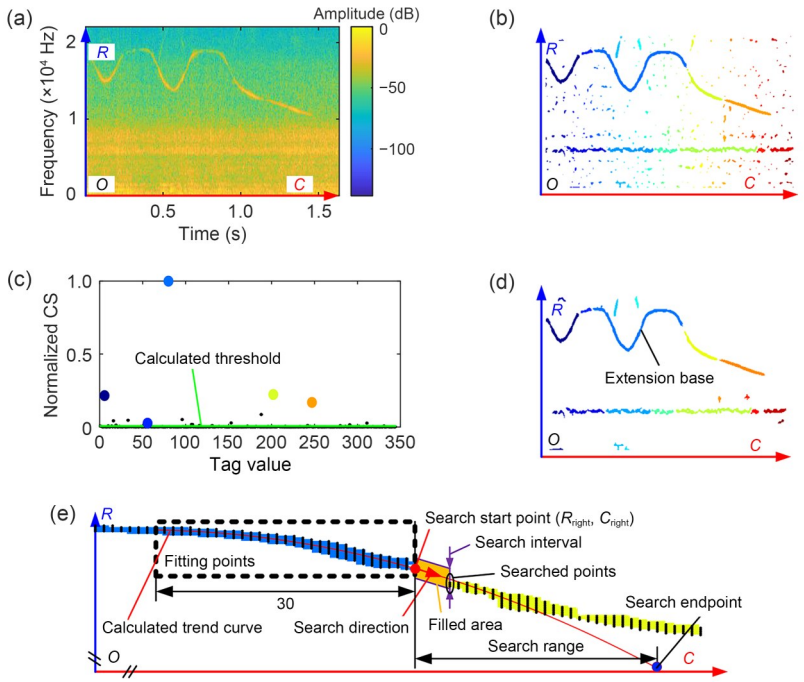

Fig. 3 TFCs of instance DWS (a), labeled result of the extracted TFCs (b), characteristic variable result of TFCs (c), TFC filtering result by means of characteristic variable CS (d), and schematic of TFC expansion connection (e). References to color refer to the online version of this figure

span and grayscale average of the marked area, respectively, and can be calculated by

$$\begin{cases} \mathrm{ColLength}(L) = \max(c) - \min(c), \\ \mathrm{AverageGray}(L) = \dfrac{1}{K_\mathrm{p}} \sum \mathrm{ZA}(r, c). \end{cases} \tag{7}$$

Herein, L=ZAL (r, c) and K_p denotes the pixel number of the TFC with tag value L.

As shown in Fig. 3c, the CS values of the whistle's TFCs are generally higher than those of noise's TFCs. However, there are still some CS values of noise's TFCs that are higher than those of whistles' TFCs, such as those regions with tag values of 188, 96, 33, which makes it impossible to completely filter out the TFCs of the noisy areas using a fixed threshold. Meanwhile, due to the influence of noise, the extracted TFCs are discontinuous. Therefore, the mean of CS is used as the filtering threshold to filter out part of the noise's TFCs, and the discontinuous TFCs of the whistle are connected by the trend search and extension method.

First, the TFC with the maximum CS value is selected as the extension base. As shown in Fig. 3e, the pixel coordinates of the extension base in the interval [C_{right}–30, C_{right}] are extracted as fitting points, and the trend curve of the extension base is calculated based on the polynomial fitting. Taking (R_{right}, C_{right}) as the search start point, gradually search for the TFC to the right based on the calculated trend curve, where the search interval is set to 10 and the search range is set to 30. If another TFC is searched, the break between the two TFCs is filled based on the calculated trend curve. Repeat the above process, and if there is no TFC to be searched, the current search for start point is the right endpoint of the whistle. In the same way, follow the above process to search for the TFC to the left, and fill the break between TFCs until the left endpoint of the whistle. Fig. 4a shows the finally obtained TFC of the whistle ($\mathbf{ZA}_{\mathrm{TFC}}$) after removing the TFCs not connected to the extension base. It is evident that the resulted TFC is irregular and cannot be directly used as a mask. Therefore, the dilation operation is first used to obtain the expanded TFC ($\mathbf{ZAE}_{\mathrm{TFC}}$), and then the TFS mask is generated by dilating the time–frequency curve CUR. The CUR can be obtained by

$$\begin{aligned} &\mathrm{CUR}(m) \\ &= \mathrm{maxindex}\left[\mathrm{ZAE}_{\mathrm{TFC}}(:, m) \cdot \mathrm{ZA}(:, m)\right], \end{aligned} \tag{8}$$

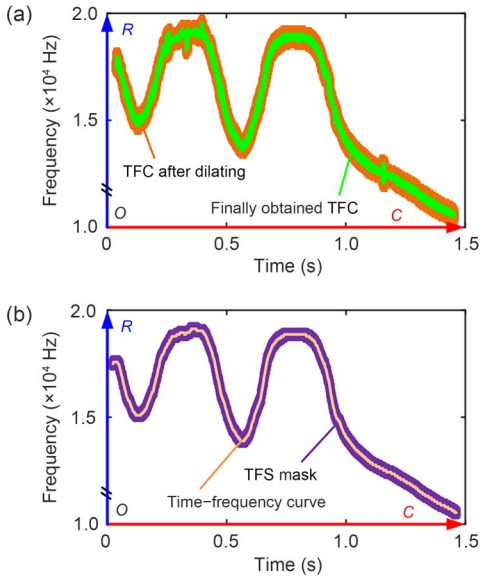

Fig. 4 Finally obtained TFC (a) and time−frequency curve and TFS mask (b)

where m denotes the column coordinates, "$\cdot$" denotes the dot product, and max index[ ] denotes the operation of calculating the row coordinate of the maximum value point.

The CUR and TFC mask ($\mathbf{ZAM}_{\mathrm{TFC}}$) are shown in Fig. 4b. The whistle signal can be calculated using the inverse Fourier transform. The whistle signal (WS) obtained by TFS mask filtering can be expressed by Eqs. (9) and (10):

$$\mathrm{WS}_c(n)=\begin{cases}\dfrac{1}{N}\displaystyle\sum_{r=1}^{N}\mathrm{ZAM}_{\mathrm{TFC}}(r,c)\,\mathrm{ZA}(r,c)\exp(\beta), n<N,\\ 0, \text{otherwise},\end{cases} \tag{9}$$

$$\begin{aligned}\mathrm{WS}=\ &\mathrm{WS}_0(n-S)+\mathrm{WS}_1(n-S)+\cdots\\ &+\mathrm{WS}_{c-1}\left(n-\left(C_{\mathrm{size}}-1\right)S\right),\end{aligned} \tag{10}$$

where $\beta=2\pi\mathrm{i}\dfrac{rn}{N}+\mathrm{iZP}(r,c)$, $n=0,1,\cdots,N-1$, $\mathrm{WS}_c(n)$ is the c^{th} short window signal, and C_{size} is the column size of the two-dimensional matrix $\mathbf{ZA}$.

3.2 Characteristic extraction and recognition

Fig. 5a illustrates the TFS of the whistle signal WS obtained by TFS mask filtering, showing that TFS mask filtering can effectively extract the whistle signal. After completing the filtering operation, the phase of the whistle $\varphi(t)$ can be calculated by

$$\varphi(t)=\arctan\left(\mathrm{WS}(t)/\mathrm{WS}_{\mathrm{j}}(t)\right), \tag{11}$$

where $\mathrm{WS}(t)$ denotes the whistle signal at time t, and $\mathrm{WS}_{\mathrm{j}}(t)$ denotes the Hilbert transform of $\mathrm{WS}(t)$, which can be expressed as

$$\mathrm{WS}_{\mathrm{j}}(t)=\mathrm{WS}(t)\otimes\frac{1}{\pi t}. \tag{12}$$

Herein, $\otimes$ denotes the convolution operation.

$\varphi(t)$ of the whistle and phase derivative $D\varphi(t)$ are shown in Fig. 5b. Due to the influence of noise, the phase derivative curve of the DWS has mutation points. To analyze and differentiate the non-stationary phase, the spectrum of $D\varphi(t)$ is obtained based on CWT, which is widely used in local signal mutation problems (Martinez-Ríos et al., 2023).

Fig. 5c demonstrates that the manually coded mutation points exhibit a short-time broadband characteristic in the frequency domain, and appear as obvious vertical lines on the CWT spectrum image. At the same time, Fig. 5d shows that the mutation points caused by noise appear as vertical lines on the CWT spectrum image, but there are fewer vertical lines in the DWS CWT spectrum image than that of the CVCFMS. As shown in Fig. 5e, there are no mutation points on the phase derivative curve of the DWS without noise interference, so there are no obvious vertical lines on the corresponding CWT spectrum image.

Based on the above characteristics, a CNN is used to recognize CVCFMSs, using GoogLeNet and the cross-entropy loss function. The value of loss function L_{v} can be calculated by

$$L_{\mathrm{v}}=-\frac{1}{N_{\mathrm{s}}}\sum_{i=1}^{N_{\mathrm{s}}}\left[y_i\ln(p_i)+(1-p_i)\ln(1-p_i)\right], \tag{13}$$

where N_{s} denotes the number of samples, y_i denotes the label value of the i^{th} sample (0 or 1), p_i denotes the probability of predicting that the i^{th} sample belongs to the positive class, and ln( ) is the natural logarithm function with a base of e.

The sound dataset includes 700 DWSs sourced from a whale sound database, and 600 CVCFMSs generated with symbol widths of 20, 30, 40, and 50 ms (50 signals per symbol width). To simulate realistic underwater conditions, ocean background noise is added to the CVCFMSs at SNRs of −5, 0, and 5 dB for each

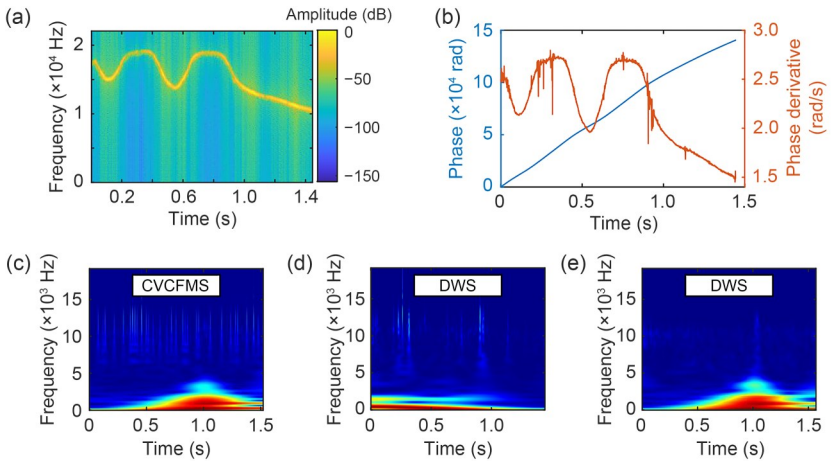

Fig. 5 TFS of DWS obtained by TFS mask filtering (a), the obtained DWS phase and its derivative (b), phase derivative spectrum of Fig. 1e (c), phase derivative spectrum of Fig. 5b (d), and phase derivative spectrum of Fig. 1d (e)

symbol width. The training dataset comprises CWT spectral images (150×300×3) derived from the phase information of the processed sound signals from the aforementioned dataset. The model training process is shown in Fig. 6.

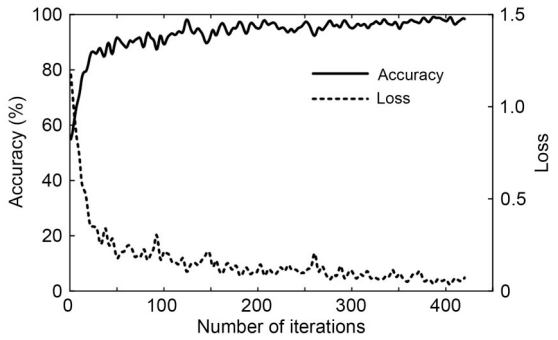

Fig. 6 The model training process

3.3 Multipath channel process

The multipath effect of underwater acoustic channels introduces multipath interference signals into the received signals, causing signal fading, seriously affecting the extraction of phase characteristics. As shown in Fig. 7c, the channel impact responses (CIRs) are obtained using Bellhop according to Fig. 7a.

Fig. 7b illustrates the CVCFMS with a symbol width of 20 ms, sent by the transmitter, and Fig. 7d1 depicts the signal received by receiver 1 with an SNR of 3 dB. It can be seen that the received signal exhibits significant multipath interference and signal fading, particularly in the interference regions where overlapping signals occur. As shown in Figs. 7d2 and 7e2, TFS mask filtering can effectively remove the multipath interference signals that do not overlap with the main path signal. Due to the problem of signal fading, the phase of the interference area cannot be accurately extracted, which causes many interference mutation points on the calculated phase derivative curve and seriously affects the later signal recognition. Therefore, the SDC (Shin and Chan, 2002; Ikki and Ahmed, 2009; Rohilla et al., 2013) is used to solve the signal fading problem. The SDC process is shown in Fig. 8.

The combining of two received signals S_{a_b} can be calculated by

$$S_{a\_b} = S_a(t) + S_b(t + \tau_{\max}), \tag{14}$$

where $\tau_{\max}$ denotes the alignment time, calculated by

$$\begin{cases} \tau_{\max} = \max\_\text{time}\big(R(\tau)\big), \\ R(\tau) = \int S_a(t) S_b(t+\tau)\,\mathrm{d}t. \end{cases} \tag{15}$$

Herein, max_time( ) denotes calculating the value of τ when $R(\tau)$ reaches the maximum value.

It is known that increasing the number of receivers (N_r) in the signal-receiving array enhances signal fading suppression, but also escalates computational demands. Therefore, the number of receivers is set to three. Fig. 7e1 shows that the SDC has a strong focusing gain effect on the main path signal. As shown in Figs. 7e2 and 7e3, after TFS mask filtering, the phase of whistles can be accurately calculated, and the manual coding characteristic can be effectively characterized.

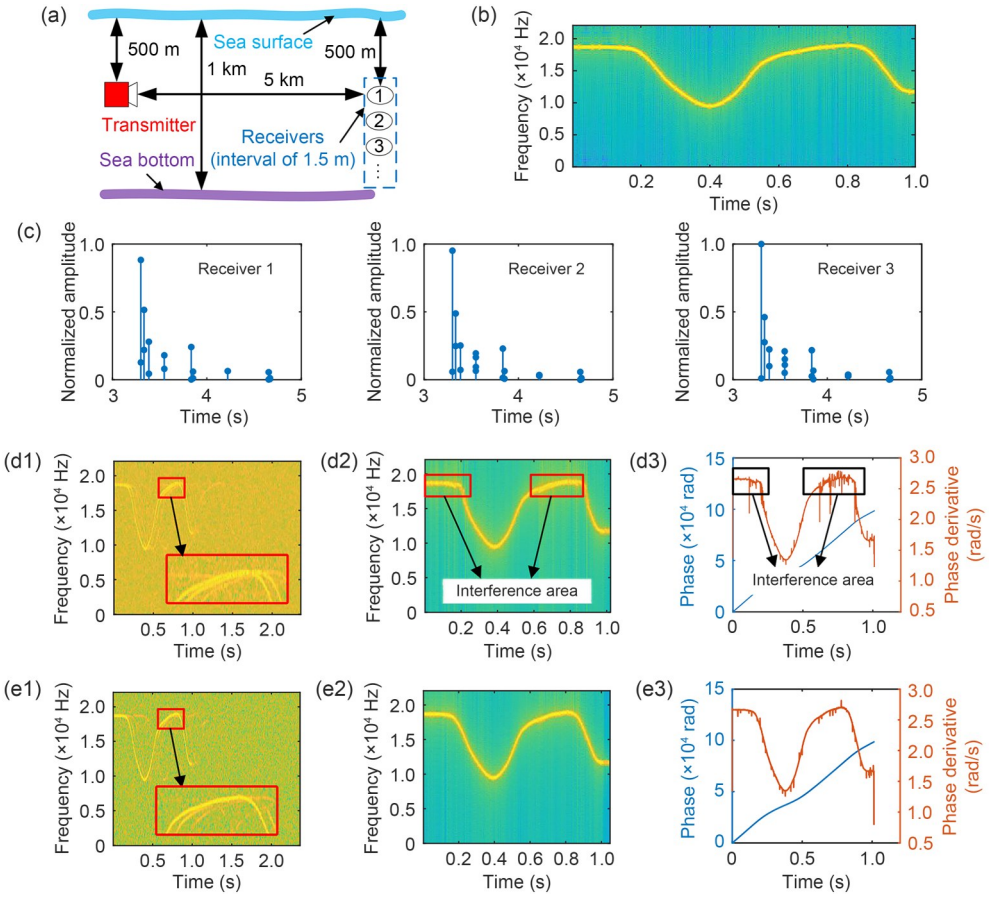

Fig. 7 Channel construction parameters (a), transmitted CVCFMS with a symbol width of 20 ms (b), CIRs shown in Fig. 7a (c), TFS of the signal received by receiver 1 (d1), TFS of the signal obtained by TFS mask filtering shown in Fig. 7d1 (d2), phase and its derivative of the signal shown in Fig. 7d2 (d3), TFS of the signal obtained by SDC (e1), TFS of the signal obtained by TFS mask filtering shown in Fig. 7e1 (e2), and phase and its derivative of the signal shown in Fig. 7e2 (e3)

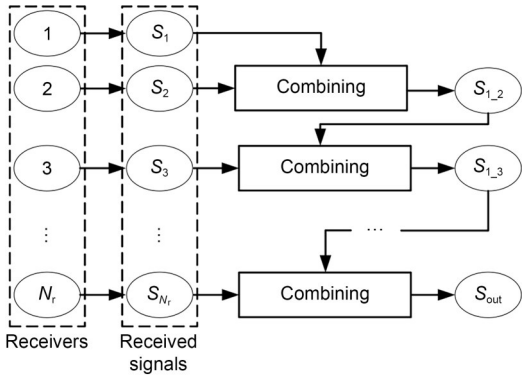

Fig. 8 The process of SDC

4 Numerical simulations and lake experiments

4.1 Numerical simulations

Numerical simulations were conducted to evaluate the performance of the proposed recognition method under different SNRs, channel conditions, and coding parameters. The results are presented in Figs. 9a and 9b. The number of phase mutation points per unit time decreases with increasing symbol width, thereby weakening the distinctiveness of the phase coding features. Consequently, as shown in Fig. 9a, the recognition accuracy under the same SNR gradually declines with increasing coding width due to the diminished distinctiveness of the phase coding features. Fig. 9b shows the recognition accuracy of the recognition method proposed in this paper under different channels constructed by Bellhop according to the parameters presented in Table 1 and the sound speed profile depicted in Figs. 9c and 9d. It can be observed that the different channel conditions have different effects on recognition accuracy.

In shallow water channels, the reflection and refraction of signals on the water surface and bottom lead

Table 1 Channel construction parameters

Channel type	Water depth (m)	Communication distance (km)	Transmitter depth (m)	Receiver depth (m)	Bottom acousto-elastic parameters		
					P-wave speed (m/s)	Density (g/cm^3)	P-wave attenuation (dB)
Channel 1	100	2	50	50, 51.5, 53	1759	1.962	0.439
Channel 2	100	20	50	50, 51.5, 53	1759	1.962	0.439
Channel 3	1000	2	500	500, 501.5, 503	1528	1.454	0.382
Channel 4	1000	20	500	500, 501.5, 503	1528	1.454	0.382

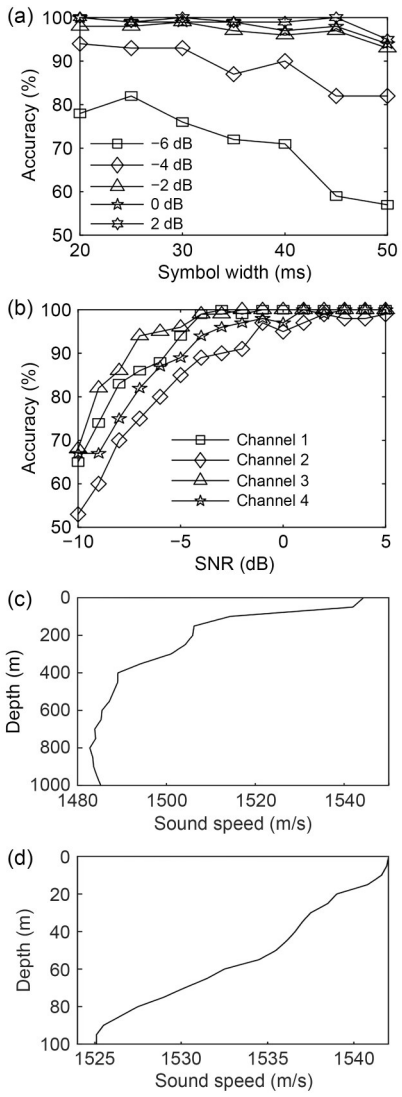

Fig. 9 Numerical simulations: (a) accuracy with different symbol widths; (b) accuracy under different channels; (c) sound speed profile of 1000 m; (d) sound speed profile of 100 m

to serious multipath effects. As the water depth increases, the chance of the signal reaching the bottom decreases, and it cannot be effectively reflected back to the receivers. Therefore, under the same conditions, the multipath interference of deep water channels is relatively small, which can be verified by the fact that the recognition accuracy under channel 3 is higher than that under channel 1, and that the recognition accuracy under channel 4 is higher than that under channel 2.

In the case of the same water depth, the reflection and refraction of the signal in short-distance communication are relatively small. As the communication distance increases, the number of signal paths will increase, so the multipath effect will be aggravated, which can be verified by the fact that the recognition accuracy under channel 1 is higher than that under channel 2, and that the recognition accuracy under channel 3 is higher than that under channel 4.

4.2 Lake experiments

To verify the proposed recognition method's performance in the actual underwater acoustic channel, the experiments were carried out in the Qingnian Lake at Tianjin University. As shown in Fig. 10, the communication distance was about 150 m. The lake depth was 6 m, and the transmitter depth was 3 m. The depths of the three receivers were 1.0, 2.5, and 4.0 m, separately. Simultaneously, a Bellhop simulation was performed based on the lake test scenario. The sound speed in the lake was set to 1480 m/s, while the bottom acousto-elastic parameters included a P-wave speed of 1520 m/s, a density of 1.421 g/cm^3, and a P-wave attenuation of 0.38 dB per wavelength. The CIRs of the actual and simulated lakes are illustrated in Fig. 11. After calculation, the SNR of the received signals during the lake test was 6.36 dB, so the SNR of the Bellhop simulation was also set to 6.36 dB.

There were seven test CVCFMS signal groups, and the experimental results are presented in Table 2. It can be observed that the recognition accuracy gradually decreases as the symbol width increases, which is

consistent with the simulation results shown in Fig. 9a. The recognition accuracies of the simulations and the actual lake test are generally comparable under the same condition. Due to simulations' limitations in replicating the complexities of actual underwater environments, the recognition accuracy in the simulations is higher than that observed in the actual lake test.

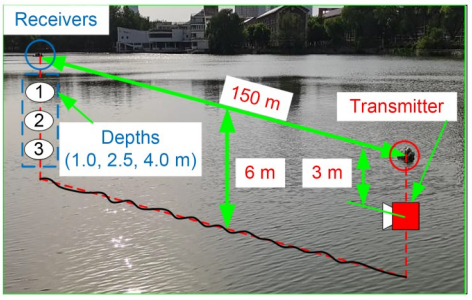

Fig. 10 Scene photo of the lake experiments

Table 2 Experimental results

Group number	Symbol width (ms)	RA of the actual lake (%)	RA of the simulations (%)
1	20	100	98
2	25	95	98
3	30	96	96
4	35	95	97
5	40	93	95
6	45	91	94
7	50	81	85

RA: recognition accuracy

5 Conclusions

This paper proposes a method for recognizing underwater communication signals that mimic dolphin whistles with shifting phase modulation. Aiming to solve the whistle extraction problem, a TFS mask filtering-based signal extraction method is proposed, which takes the spectrum as a two-dimensional matrix image and obtains the whole TFC using a trend search and extension method. Additionally, spatial diversity combining is employed to mitigate signal fading caused by multipath propagation. The phase derivative spectrum image is obtained using the Hilbert transform and CWT, and is used as the CNN input. Both simulations and lake experiments confirm the recognition method's effectiveness, as detailed in the results. In the simulations, a recognition accuracy of 90% at an SNR of 0 dB can be achieved in multipath channels, and in the actual underwater communication environment, a recognition accuracy of 81% at a symbol width of 50 ms and an SNR of 6.36 dB can be achieved, proving the effectiveness of the recognition method.

Contributors

Qingwang YAO, Jiajia JIANG, Xiaolong YU, and Zhuochen LI proposed the main idea. All the authors designed the research. Qingwang YAO and Xiaozong HOU performed the simulations and processed the data. Jiajia JIANG, Xiao FU, and Fajie DUAN participated in the theoretical analysis, and revised and finalized the paper.

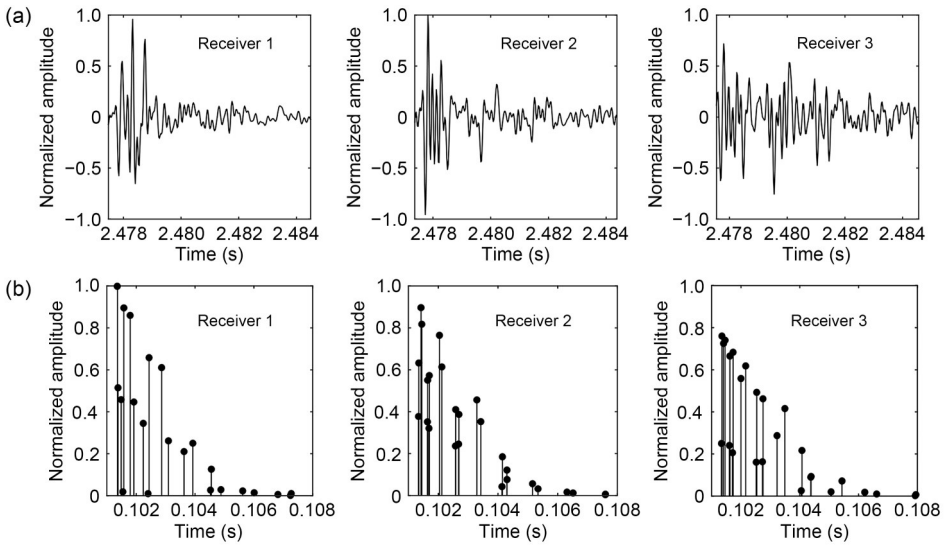

Fig. 11 CIRs of the actual lake (a) and CIRs of the simulated lake (b)

Conflict of interest

All the authors declare that they have no conflict of interest.

Data availability

Due to the nature of this research, participants of this study did not agree for their data to be shared publicly, so supporting data are not available.

References

Ahn J, Lee H, Kim Y, et al., 2019. Mimicking dolphin whistles with continuously varying carrier frequency modulation for covert underwater acoustic communication. *Jpn J Appl Phys*, 58(SG):SGGF05. https://doi.org/10.7567/1347-4065/ab14d2

Ahn J, Lee H, Kim Y, et al., 2020. Machine learning based biomimetic underwater covert acoustic communication method using dolphin whistle contours. *Sensors*, 20(21): 6166. https://doi.org/10.3390/s20216166

Bilal M, Liu SZ, Qiao G, et al., 2020. Bionic Morse coding mimicking humpback whale song for covert underwater communication. *Appl Sci*, 10(1):186. https://doi.org/10.3390/app10010186

Casari P, Neasham J, Gubnitsky G, et al., 2023. Acoustic projectors make covert bioacoustic chirplet signals discoverable. *Sci Rep*, 13(1):2591. https://doi.org/10.1038/s41598-023-29413-2

Diamant R, Lampe L, Gamroth E, 2017. Bounds for low probability of detection for underwater acoustic communication. *IEEE J Oceanic Eng*, 42(1):143-155. https://doi.org/10.1109/JOE.2016.2550278

Huang SH, Hou XG, Liu WW, et al., 2020. Mimicking ship-radiated noise with chaos signal for covert underwater acoustic communication. *IEEE Access*, 8:180341-180351. https://doi.org/10.1109/ACCESS.2020.3027022

Iglesias V, Grajal J, Royer P, et al., 2015. Real-time low-complexity automatic modulation classifier for pulsed radar signals. *IEEE Trans Aerosp Electron Syst*, 51(1):108-126. https://doi.org/10.1109/TAES.2014.130183

Ikki SS, Ahmed MH, 2009. Performance of cooperative diversity using equal gain combining (EGC) over Nakagami-*m* fading channels. *IEEE Trans Wirel Commun*, 8(2):557-562. https://doi.org/10.1109/TWC.2009.070966

Jiang JJ, Wang XQ, Duan FJ, et al., 2018. Bio-inspired steganography for secure underwater acoustic communications. *IEEE Commun Mag*, 56(10):156-162. https://doi.org/10.1109/MCOM.2018.1601228

Jiang JJ, Wang XQ, Duan FJ, et al., 2019. Study of the relationship between pilot whale (*Globicephala melas*) behaviour and the ambiguity function of its sounds. *Appl Acoust*, 146:31-37. https://doi.org/10.1016/j.apacoust.2018.10.032

Jiang JJ, Sun ZB, Duan FJ, et al., 2020. Synthesis and modification of cetacean tonal sounds for underwater bionic covert detection and communication. *IEEE Access*, 8:119980-119994. https://doi.org/10.1109/ACCESS.2020.3004282

Jiang JJ, Qiao F, Li Y, et al., 2022. Recognition method for the bionic camouflage click communication trains modulated by time delay difference. *J Acoust Soc Am*, 152(1):491-500. https://doi.org/10.1121/10.0012693

Jiang JJ, Yao ZG, Li ZC, et al., 2023. Recognition method for the bionic camouflage cetacean whistle modulated by CPMFSK signals. *Appl Acoust*, 207:109326. https://doi.org/10.1016/j.apacoust.2023.109326

Kaveh M, Falahati A, 2021. An improved Merkle hash tree based secure scheme for bionic underwater acoustic communication. *Front Inform Technol Electronic Eng*, 22(7): 1010-1019. https://doi.org/10.1631/FITEE.2000043

Lee KG, Oh SJ, 2019. Detection of fast frequency-hopping signals using dirty template in the frequency domain. *IEEE Wirel Commun Lett*, 8(1):281-284. https://doi.org/10.1109/LWC.2018.2870275

Li CY, Jiang JJ, Wang XQ, et al., 2021. Bionic covert underwater communication focusing on the overlapping of whistles and clicks generated by different cetacean individuals. *Appl Acoust*, 183:108279. https://doi.org/10.1016/j.apacoust.2021.108279

Li Y, Huo K, Li Q, et al., 2019. A novel method of wireless power transfer identification and resonance decoupling based on frequency hopping communication. *IEEE Access*, 7:161201-161210. https://doi.org/10.1109/ACCESS.2019.2950084

Liu F, Marcellin MW, Goodman NA, et al., 2016. Compressive sampling for detection of frequency-hopping spread spectrum signals. *IEEE Trans Signal Process*, 64(21):5513-5524. https://doi.org/10.1109/TSP.2016.2597122

Liu SZ, Ma TL, Qiao G, et al., 2017. Biologically inspired covert underwater acoustic communication by mimicking dolphin whistles. *Appl Acoust*, 120:120-128. https://doi.org/10.1016/j.apacoust.2017.01.018

Martinez-Ríos EA, Bustamante-Bello R, Navarro-Tuch S, et al., 2023. Applications of the generalized Morse wavelets: a review. *IEEE Access*, 11:667-688. https://doi.org/10.1109/ACCESS.2022.3232729

Qiao G, Zhao YJ, Liu SZ, et al., 2017. Dolphin sounds-inspired covert underwater acoustic communication and micromodem. *Sensors*, 17(11):2447. https://doi.org/10.3390/s17112447

Qiao G, Ma TL, Liu SZ, et al., 2021. A frequency hopping pattern inspired bionic underwater acoustic communication. *Phys Commun*, 46:101288. https://doi.org/10.1016/j.phycom.2021.101288

Rohilla S, Patidar DK, Soni NK, 2013. Comparative analysis of maximum ratio combining and equal gain combining diversity technique for WCDMA: a survey. *Int J Eng Invent*, 3(1):72-77.

Schoolcraft R, 1991. Low probability of detection communications-LPD waveform design and detection techniques. MILCOM 91-Conf Record, p.832-840. https://doi.org/10.1109/MILCOM.1991.258378

Shin EJ, Chan VWS, 2002. Optical communication over the turbulent atmospheric channel using spatial diversity. Global Telecommunications Conf, p.2055-2060. https://doi.org/10.1109/GLOCOM.2002.1188992

Yao QW, Jiang JJ, Chen GC, et al., 2023. Recognition method for underwater imitation whistle communication signals by slope distribution. *Appl Acoust*, 211:109531. https://doi.org/10.1016/j.apacoust.2023.109531