IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received April 28, 2019, accepted May 21, 2019, date of publication May 27, 2019, date of current version October 2, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2919321

Whale Vocalization Classification Using Feature Extraction With Resonance Sparse Signal Decomposition and Ridge Extraction

HAILAN CHEN[1,2], **HAIXIN SUN**[1], **NAVEED UR REHMAN JUNEJO**[1], **GUANGSONG YANG**[3], **AND JIE QI**[1]

[1]School of Information Science and Engineering, Xiamen University, Xiamen 361005, China
[2]School of Science, Jimei University, Xiamen 361021, China
[3]School of Information Engineering, Jimei University, Xiamen 361021, China

Corresponding authors: Haixin Sun (hxsun@ xmu.edu.cn) and Jie Qi (qijie@ xmu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61671394, in part by the Young and Middle-Aged Teachers' Educational Scientific Research Project of Fujian Province of China under Grant JAT160268, in part by the Science and Technology Program of Shenzhen, China, under Grant JSGG20170414090428464, in part by the Fundamental Research Funds for the Central Universities under Grant 20720170044, and in part by the National Key Research and Development Program of China under Grant 2018YFC0809200.

ABSTRACT Whales communicate using whistle vocalizations that are essentially underwater acoustic frequency-modulated tones. Inevitable environmental noise decreases recognition accuracy of these sounds during wide range detection. In this paper, we propose a robust time − frequency analysis method that combines resonance sparse signal decomposition (RSSD) and spectrogram ridge extraction. We apply RSSD to extract whistle components from the raw signal, and then we segment the ridge regions of the whistle spectrograms. By applying a partial derivative method, we extract the whistle spectrogram ridge representing an accurate trace of the whistle vocalization. From these results, we extract ridge features and use an SVM or a random forest to identify the whale species. We evaluated our method using experiments with samples for four whale species. Compared with direct ridge extraction directly without RSSD, our proposed method achieved better extraction of frequency characteristics of the vocalizations. Our proposed method achieved an accuracy rate of over 98% for sounds from four species when using five training samples.

INDEX TERMS Classification of whale vocalization, resonance sparse signal decomposition, tunable Q-factor wavelet transform (TQWT), morphological component analysis (MCA), ridge extraction.

I. INTRODUCTION

Whales use narrowband tones (whistles) to communicate with each other. The ability to automatically determine time−frequency tracks corresponding to these vocalizations have numerous applications for describing, identifying, and estimating the density of whale species [1]. However, recordings of whale sounds contain ship-radiated noise and ambient ocean noise among other types that make it challenging for researchers to extract the features from the whales' acoustic signals. More accurate feature extraction better characterizes the tonal calls and improves their classification accuracy. In this study, we present a method using resonance sparse signal decomposition (RSSD) and ridge extraction to extract accurate feature information from noisy recordings.

Many scholars have proposed numerous methods to extract whistle features from underwater whale recordings. Some extraction techniques extracted features directly without removing noise [2], [3]. Others have developed whistle detection and classification programs that search for spectral peaks within a user-specified frequency band [4]–[6]. Oswald proposed a technique for whistle detection and classification that requires manual selection of high-quality whistles suitable for classification [7]. Whistle detectors such as those developed by Johansson and White and Roch et al. employ Bayesian filtering methods to track tonal sounds [8], [9]. An automatic detector described by Gillespie et al. applies a series of noise cancellation techniques to the spectrogram [10]. Recognition of whale communication patterns requires the extraction of whistles from the composite signal. Many methods have been presented to decompose such multi-component

The associate editor coordinating the review of this manuscript and approving it for publication was Li He.

signals, including blind source separation [11]–[13], dual-tree complex wavelet transforms [14], [15], empirical mode decomposition (EMD) [16]–[18], ensemble empirical mode decomposition (EEMD) [19], [20], and independent component analysis (ICA) [21]–[24]. The above methods decompose the target signals in the frequency domain. However, whale sound components may occupy the same frequency band and overlap in the frequency domain. Thus, these methods are unable to extract components of cetacean sounds precisely.

In this study, we propose a new method using RSSD and ridge extraction to separate whistles from background noise extract ridges in whistle spectrograms. Unlike frequency-based methods, the RSSD sparse signal representation method can decompose the multi-component signal according to the oscillatory behavior of each component because whistles, clicks, and other sounds have distinctive oscillatory and transient impulse characteristics. In addition, STFT is a time−frequency representation method, the more significant coefficients modulus more often distributed in several regions, the components of ridge show signal feature in the time−frequency plane. The spectrogram ridges contain the main information of the signal. In this study, we have utilized ridge extraction algorithm based on the partial derivative method. Our proposed method shows very effective extraction of the whistle component from the composite cetacean sounds and ridge extraction from the whistle spectrogram as demonstrated by real-world recordings of cetacean vocalizations.

The rest of the study is organized as follows: The proposed method is explained in Section 2. Section 3 examines the capabilities of our approach using noisy underwater cetacean recordings. Finally, we present our conclusions in Section 4.

II. METHODOLOGY

Our method aims to capture an accurate trace of the whistle signal from the raw signal, which is corrupted by noise and interference. The method consists of five steps: RSSD application, ridge region segmentation, ridge extraction, ridge joining, and feature extraction and classification. We employ RSSD to extract the high-oscillation whistle signal hidden in the raw signal. We then perform ridge region segmentation on the spectrogram of the extracted whistle signal to acquire the time−frequency sub-signatures with energy concentration. Next, we perform ridge extraction of all the sub-signatures to capture an accurate trace of whistle vocalizations. We then join the ridges by polynomial fitting. Finally, we use an SVM to classify the cetacean species. Figure 1 shows the flow chart of the proposed method.

A. RESONANCE SPARSE SIGNAL DECOMPOSITION

The RSSD algorithm adopts the tunable Q-factor wavelet transform (TQWT) and morphological component analysis (MCA) [25]. The TQWT algorithm provides one set of over-complete basis to estimate high- and low-resonance components. MCA, which performs signal decomposition based on sparse representations, decomposes the raw signal into the high, low, and residual components [26].

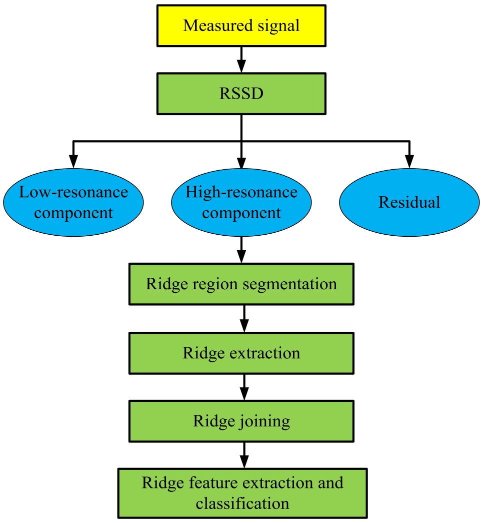


FIGURE 1. Flow chart of cetacean sound classification.

1) TUNABLE Q-FACTORS WAVELET TRANSFORM (TQWT)

In TQWT, the Q-factor reflects the oscillatory properties of one signal and is defined as [27]

$$Q = f_c/BW, \tag{1}$$

where f_c is the center frequency, and BW is the bandwidth. TQWT is implemented by iteratively applying the two-channel bandpass filter banks on its low-pass channel. The center frequency f_c of the level j is derived from the input signal sampling rate f_s given by [28]

$$f_c = \alpha^j \frac{2-\beta}{4\alpha} f_s. \tag{2}$$

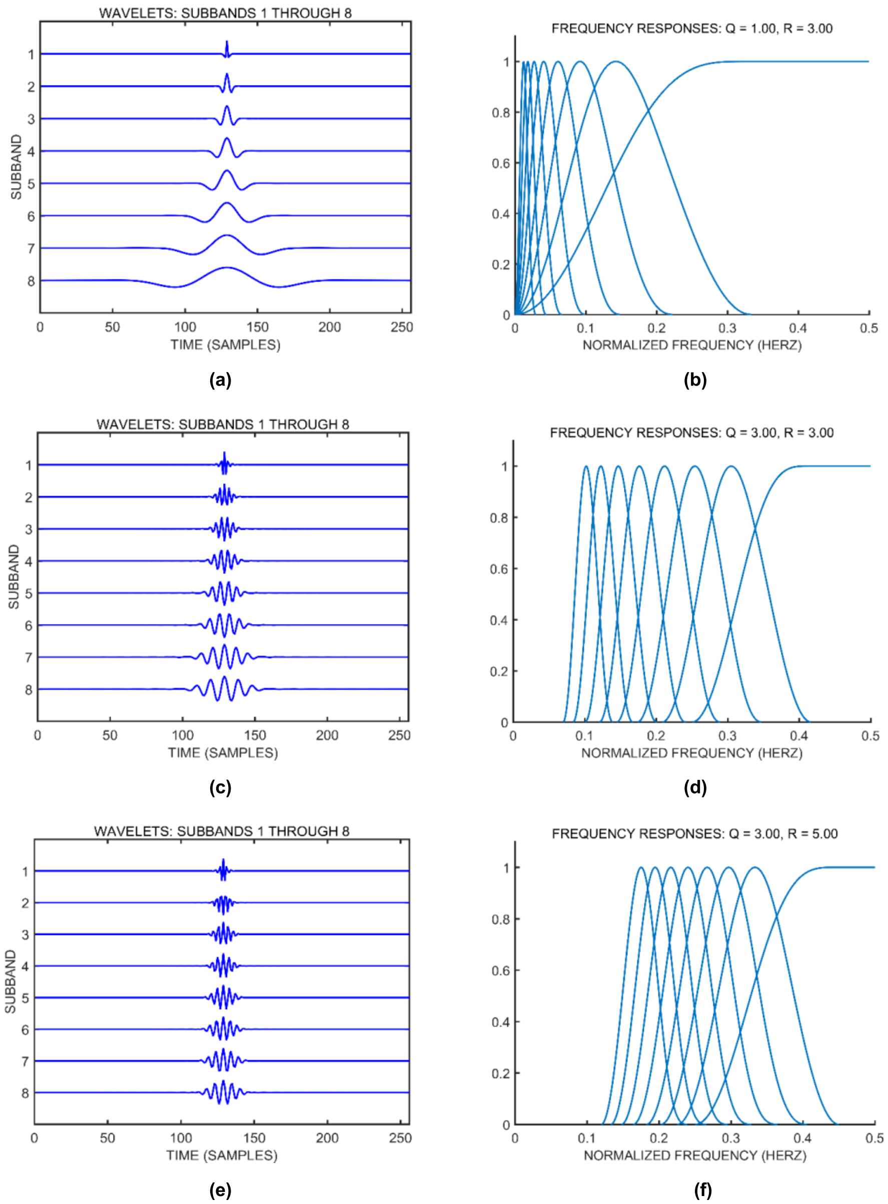
The corresponding bandwidth BW is expressed as [25]

$$BW = \frac{1}{2}\beta\alpha^{j-1}\pi, \tag{3}$$

where α and β are the low-pass and high-pass scaling parameters, respectively. Using (1), (2) and (3), the Q-factor can be formulated in terms of α and β as [29]

$$Q = \frac{2-\beta}{\beta} \tag{4}$$

The signal oscillatory property can be described with the Q-factor. As shown in Figure 2, higher Q values result in higher oscillatory intensity in the time domain and better frequency aggregation in the frequency domain at the same time, and vice versa [24]. Hence, the difference between the low-Q and high-Q wavelet functions represents the oscillation of the signal, which we use to solve the component extraction problem. Figure 3 presents the flow chart of applying TQWT to decompose an N-point discrete-time signal into J-level sub-bands. The TQWT structure is based on the discrete dyadic DWT, which employs two-channel analysis and synthesis filter banks. The analysis filter banks are repeatedly



(a)

(b)

(c)

(d)

(e)

(f)

FIGURE 2. Flow chart of cetacean sound classification. The graphs show the wavelet waveform and the corresponding frequency responses spectrum for different parameters: (a) wavelet waveform with $Q = 1, r = 3$, (b) frequency response spectrum with $Q = 1, r = 3$, (c) wavelet waveform with $Q = 3, r = 3$, (d) frequency response spectrum with $Q = 3, r = 3$, (e) wavelet waveform with $Q = 3, r = 5$, and (f) frequency response spectrum with $Q = 3, r = 5$.

applied on its low-pass channel and then further processed by low- and high-pass scaling operations with α and β as the corresponding scaling parameters [25]. The synthesis filter banks execute the same steps. Each level's two-channel filters are composed of low- and high-pass filter, with $H_l(\omega)$ and $H_h(\omega)$ as the corresponding frequency responses defined by the equations

$$H_l(\omega) = \begin{cases} 1, & |\omega| \leq (1-\beta)\pi \\ \theta\left(\dfrac{\omega + (\beta - 1)\pi}{\alpha + \beta - 1}\right), & (1-\beta)\pi < |\omega| < \alpha\pi \\ 0, & \alpha\pi \leq |\omega| \leq \pi \end{cases} \tag{5}$$

and

$$H_h(\omega) = \begin{cases} 0, & |\omega| \leq (1-\beta)\pi \\ \theta\left(\dfrac{\alpha\pi - \omega}{\alpha + \beta - 1}\right), & (1-\beta) < |\omega| < \alpha\pi \\ 1, & \alpha\pi \leq |\omega| \leq \pi, \end{cases} \tag{6}$$

where the parameters must satisfy $0 < \alpha < 1, 0 < \beta \leq 1, \alpha + \beta > 1$, and $\theta(\omega) = 0.5(1 + \cos\omega)\sqrt{2 - \cos\omega}, |\omega| \leq \pi$. The outputs of the filters are re-scaled by a low-pass scaling with parameter α (HPS α) and a high-pass scaling with parameter β (LPS β) expressed by

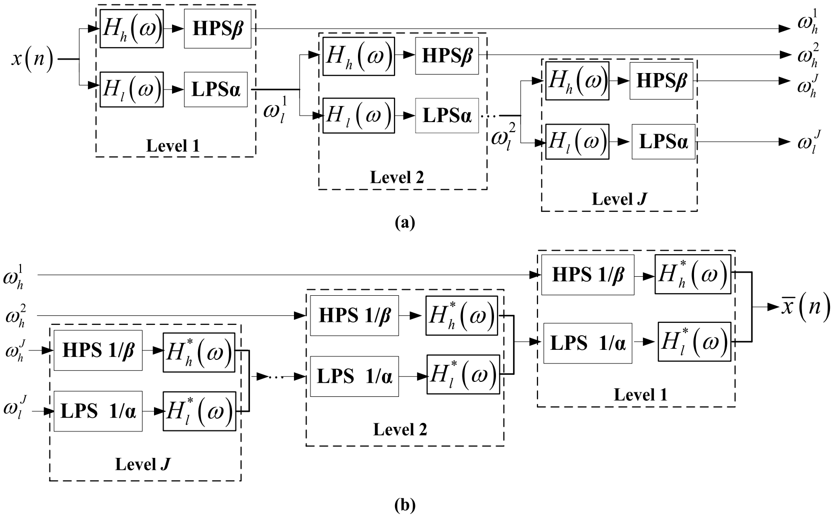


FIGURE 3. The TQWT filter banks. (a) The analysis filter banks, (b) the synthetic filter bank.

Eqs. (7) and (8):

$$Y(\omega) = X(\alpha\omega), \quad |\omega| \leq \pi \tag{7}$$

$$Y(\omega) = \begin{cases} X(\beta\omega + (1-\beta)\pi), & 0 < \omega < \pi \\ X(\beta\omega - (1-\beta)\pi), & -\pi < \omega < 0. \end{cases} \tag{8}$$

Note that the outputs of the high-pass filters are wavelet coefficients.

The most significant parameters of the TQWT algorithm are the Q-factor, redundancy factor r, and decomposition level J. Q-factor is described and set by the degree of oscillation of the objected signal. To decompose the original signal into high- and low-oscillatory components, the Q-factor should be chosen to match the oscillatory levels of the extracted components. A high Q-factor results in wavelets having more intense oscillatory cycles that are suitable for extracting high-oscillatory components. Low Q-factor values produce wavelets consisting of non-oscillatory elements that are suitable for abstracting the low-oscillatory component. The redundancy factor r controls the overlapping rate between the frequency responses of adjacent wavelets. As shown in Figure 2(e) and (f), an increase in r for a fixed Q-factor leads to a higher overlapping rate of frequency responses. Note that r must be greater than 1, with a value of 3 or greater recommended for perfect reconstruction and sparsity. The decomposition level J adjusts the frequency coverage of the wavelets. Higher values of J lead to cover the wider frequency band and become much closer to $0Hz$. The value of J should be as large as possible to include as many lower frequencies as possible. We set the maximum number of decomposing levels $J_{\max}$ to

$$J_{\max} = \left[\frac{\log(\beta N/8)}{\log(1/\alpha)} \right] = \left[\frac{\log\left(\frac{N}{4(Q+1)}\right)}{\log\left(\frac{(Q+1)r}{(Q+1)r-2}\right)} \right], \tag{9}$$

where N is the length of the input signal $x(n)$.

2) MCA

For the morphological component analysis (MCA), we consider the problem of writing an observed noisy signal x as the sum of an oscillatory signal x_1, a non-oscillatory signal x_2, and noise n:

$$x = x_1 + x_2 + n. \tag{10}$$

To estimate x_1 and x_2 with sparse representation from the observed signal x, one approach is to model x_1 and x_2 as having sparse representations using both high Q-factor and low Q-factor TQWT jointly. Our approach uses MCA, so we decompose the signal x into two components x_1 and x_2 by constructing the objective function

$$\begin{aligned} \arg\min_{w_1,w_2} \|x - \Phi_1 w_1 - \Phi_2 w_2\|_2^2 &+ \sum_{j=1}^{J_1+1} \lambda_{1,j} \left\| w_{1,j} \right\|_1 \\ &+ \sum_{j=1}^{J_2+1} \lambda_{2,j} \left\| w_{2,j} \right\|_1, \end{aligned} \tag{11}$$

where $\|\cdot\|_1$ and $\|\cdot\|_2$ denote l_1 and l_2 norms, respectively; Φ_1 and Φ_2 represent the inverse TQWT having high and low Q-factors, respectively; ω_1 and ω_2 denote the transform coefficients of signals x_1 and x_2 within the framework of Φ_1 and Φ_2 [18]. ω_1 and ω_2 can be written as $\omega_1 = \left[\omega_{1,1}, \omega_{1,2}, \ldots, \omega_{1,J_1}, \omega_{1,J_1+1}\right]$ and $\omega_2 = \left[\omega_{2,1}, \omega_{2,2}, \ldots, \omega_{2,J_2}, \omega_{2,J_2+1}\right]$, where $\omega_{1,j}$ and $\omega_{2,j}$ denote sub-bands' j wavelet coefficient of ω_1 and ω_2. J_1 and J_2 are the number of filter banks in the high and low Q-factor TQWT, respectively. For the radix-2 TQWT, the synthesis functions (wavelets) do not have the same energy which in the mathematical expression form of l_1-norm squared. The energy, in particular, differs for each sub-band [30]. Therefore, each wavelet coefficient of $\omega_{1,j}$ and $\omega_{2,j}$ is compensated for by using the energy of the corresponding wavelet function through the regularization parameters vectors $\lambda_{1,j}$ and $\lambda_{2,j}$ [27]:

$$\begin{cases} \lambda_{1,j} = \theta \left\| \psi_{1,j} \right\|_2 \\ \lambda_{2,j} = (1-\theta) \left\| \psi_{2,j} \right\|_2, \end{cases} \tag{12}$$

where $\psi_{1,j}$ and $\psi_{2,j}$ denote the discrete mother wavelet functions of the sub-band j. The parameter θ affects the energy distribution between the high- and low-resonance components. We set θ to 0.5 to ensure an equal balance of the energy distribution.

The problem in equation (11) is a convex problem solvable with a fast iterative algorithm known as the "split variable augmented Lagrangian shrinkage algorithm" (SALSA) [31]. We apply the SALSA to the iteration update and reevaluate the wavelet coefficients until we obtain the optimal wavelet coefficients ω_1^* and ω_2^* to minimize the objective function value. Then the extracted high- and low-resonance components $\hat{x}_1$ and $\hat{x}_2$ are given by

$$\hat{x}_1 = \Phi_1 \omega_1^*, \quad \hat{x}_2 = \Phi_2 \omega_2^*. \tag{13}$$

The energy in signal spectrograms is often centered near a series of curves with similar "mountain ridges." The ridge distribution describes important features of the raw signal. Thus, signal ridge extraction becomes an important method for analyzing signal features and is useful for signal compression, reconstruction, and de-noising.

B. RIDGE REGION SEGMENTATION

Rather than applying ridge extraction to all pixels in the spectrogram, we apply dual threshold processing to limit processing to those regions of the spectrogram with a relatively high signal to noise ratio, that is, where there is high spectral power in the local region. This operation is often called "ridge region segmentation," which aims to remove those regions of an image where a signal is unlikely to exist. First, we generated a mask to exclude all pixels of the spectrogram image with an absolute value of amplitude less than the threshold Q_1. Second, to remove those pixels scattering in the spectrogram which are unlikely belong to the signal, we excluded those points where the number of points belonging to the same connected domain is less than the threshold Q_2. After these two operations, we had the "Ridge region" where the ridges exist.

C. RIDGE EXTRACTION

As the whistle component is non-stationary and frequency-modulated, the ridge concept is suitable for characterizing signal components. The extraction of the ridges was implemented through the first and second order derivatives maps of the spectral image. The ridge extraction method is briefly explained as follows:

First, we calculate the two-dimensional STFT transform coefficients of the extracted whistle signal in the ridge region. We denote the first- and second-order derivative of the two-dimensional STFT transform coefficients as ∂t, ∂f, ∂tt, ∂ff, and ∂tf (where t and f indicate the time and frequency, respectively). The gradient of the two-dimensional STFT transform coefficients can be shown as a first-order partial derivative $(\partial t\ \partial f)_{t,f}$. We denote the Hessian matrix of the second-order partial derivative of the two-dimensional STFT transform coefficients is represented as $H_{t,f}$ and the largest magnitude eigenvector of the Hessian matrix as $E_{t,f}$:

$$H_{t,f} = \begin{bmatrix} \partial tt_{t,f} & \partial tf_{t,f} \\ \partial ft_{t,f} & \partial ff_{t,f} \end{bmatrix}. \tag{14}$$

A ridge point in a spectrogram occurs for any pixel (t, f) where the gradient vector is perpendicular to the eigenvector $E_{t,f}$ of the Hessian matrix $H_{t,f}$ [27]. The dot product of the eigenvector and the gradient will be equal to zero for the two points on either side of a ridge [7]:

$$E_{t,f} \cdot \begin{pmatrix} \partial t \\ \partial f \end{pmatrix}_{t,f} = 0. \tag{15}$$

To find the ridge points (t^*, f^*), we calculate the dot product between $H_{t,f}$ and $E_{t,f}$ for every pixel in the ridge regions and track the zero-crossing of this function by searching the 3×3 pixel neighborhood $(t \pm 1, f \pm 1)$ for a sign change. Tracking ceases when all pixels in the ridge regions have been used up.

Various types of noise (ambient, thermal/instrument, aliasing) can cause discontinuities in the ridges in the spectrogram image. To form longer, smoother ridges, we employ a polynomial fitting method.

D. FEATURES EXTRACTION AND CLASSIFICATION

The extracted ridges in whistle spectrogram are the fundamental structure of the tonal vocalizations and characterize the frequency modulation of the tonal calls for classification purposes. In this study, feature set F_{ridge} for each ridge includes the starting and ending frequencies f_{start} and f_{end}, the time index t_1 of the lowest frequency , the lowest frequency f_l, the time index t_2 of the highest frequency, the highest frequency f_h , and the time duration Δt. After feature extraction, we represent each acoustic event as

$$F_{ridge} = \left\{ f_{start}, f_{end}, t_1, f_l, t_{2,}, f_h, \Delta t \right\}. \tag{16}$$

Support vector machine (SVM) are widely used for classification, especially with small samples. We use the extracted prominent ridge feature sets F_{ridge} as the training data. Because there are few data samples, we adopt the round-by-round method.

III. RESULTS AND DISCUSSION

We selected four globally widespread cetacean species for our experiments as listed in Table 1 using recordings downloaded from http://www.mobysound.org/ and http://www.aigei.com/. Figure 4(a) shows the time−frequency signature representation of the raw signals with whistle components visualized with the background noise in the spectrogram. We applied the RSSD algorithm to extract whistle

TABLE 1. Summary of the whale data used in experiments.

No.	Scientific name	Total syllables
1	Pilot whale	12
2	Blue whale	52
3	Bottlenose dolphin	10
4	Orca whale	66

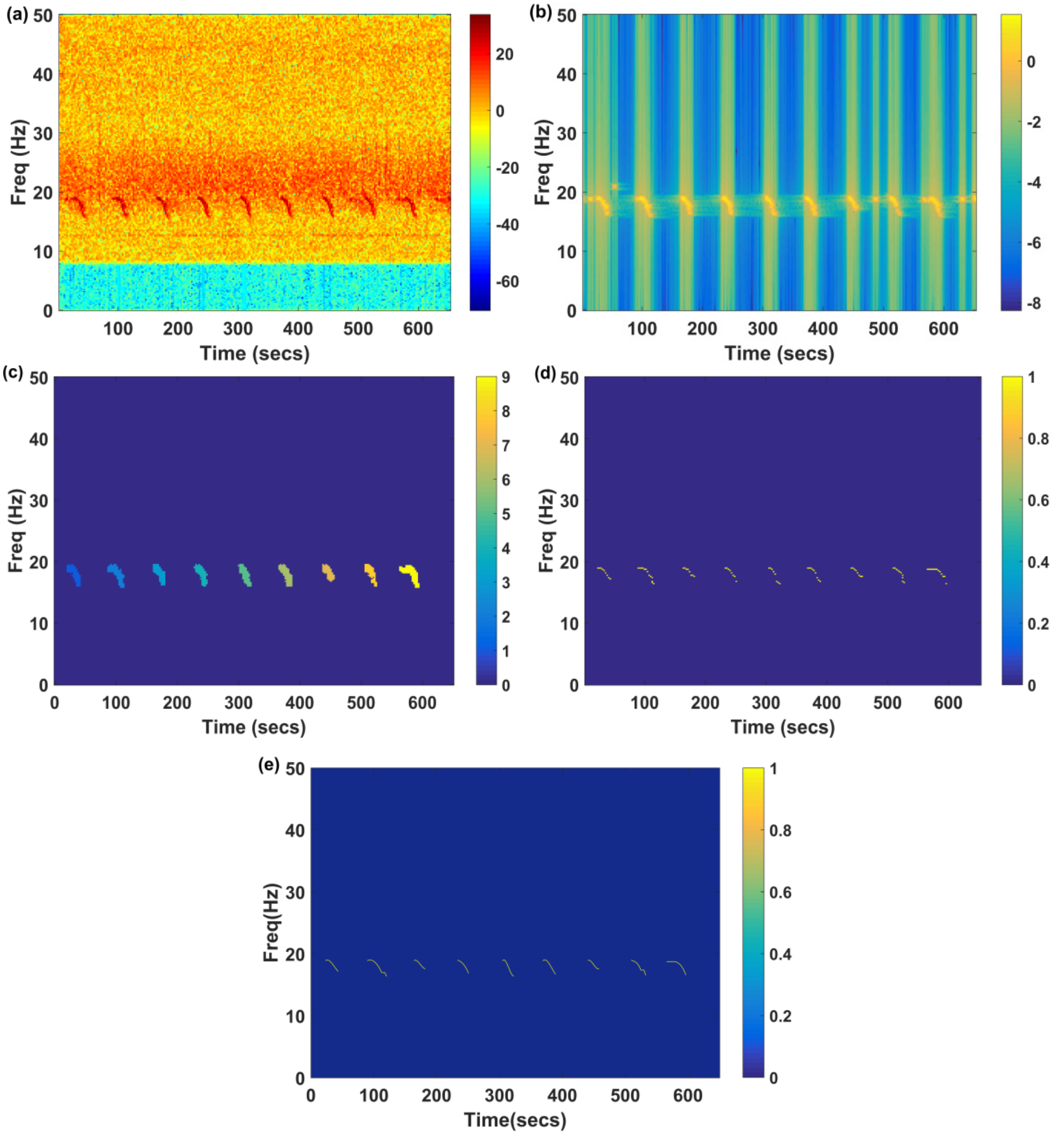


FIGURE 4. Ridge extraction in the spectrogram for the blue whale: (a) the original spectrogram, (b) spectrogram of the extracted whistle after RRSD method, (c) ridge region segmentation, (d) ridge extraction, and (e) ridge was joining.

components from the noisy raw signal to enable accurate and effective extraction of the features in the whale vocalizations.

We also analyzed the sounds by exploring the ridges due to their ability to mark the leading edge in the spectrogram. We note that some whale sounds are characterized as having a single harmonic while others have multiple harmonics. For example, a blue whale's single harmonic sound is depicted in Figure 4(a), while the pilot whale's multiple harmonic sound is shown in Figure 5(a). Whales, as with all mammalian animals, have different minimum frequency ridges according to species. We selected the minimum frequency ridge from all ridges to represent each sound.

Analyzing the sounds of a blue whale before ridge extraction, we performed two pre-processing steps.

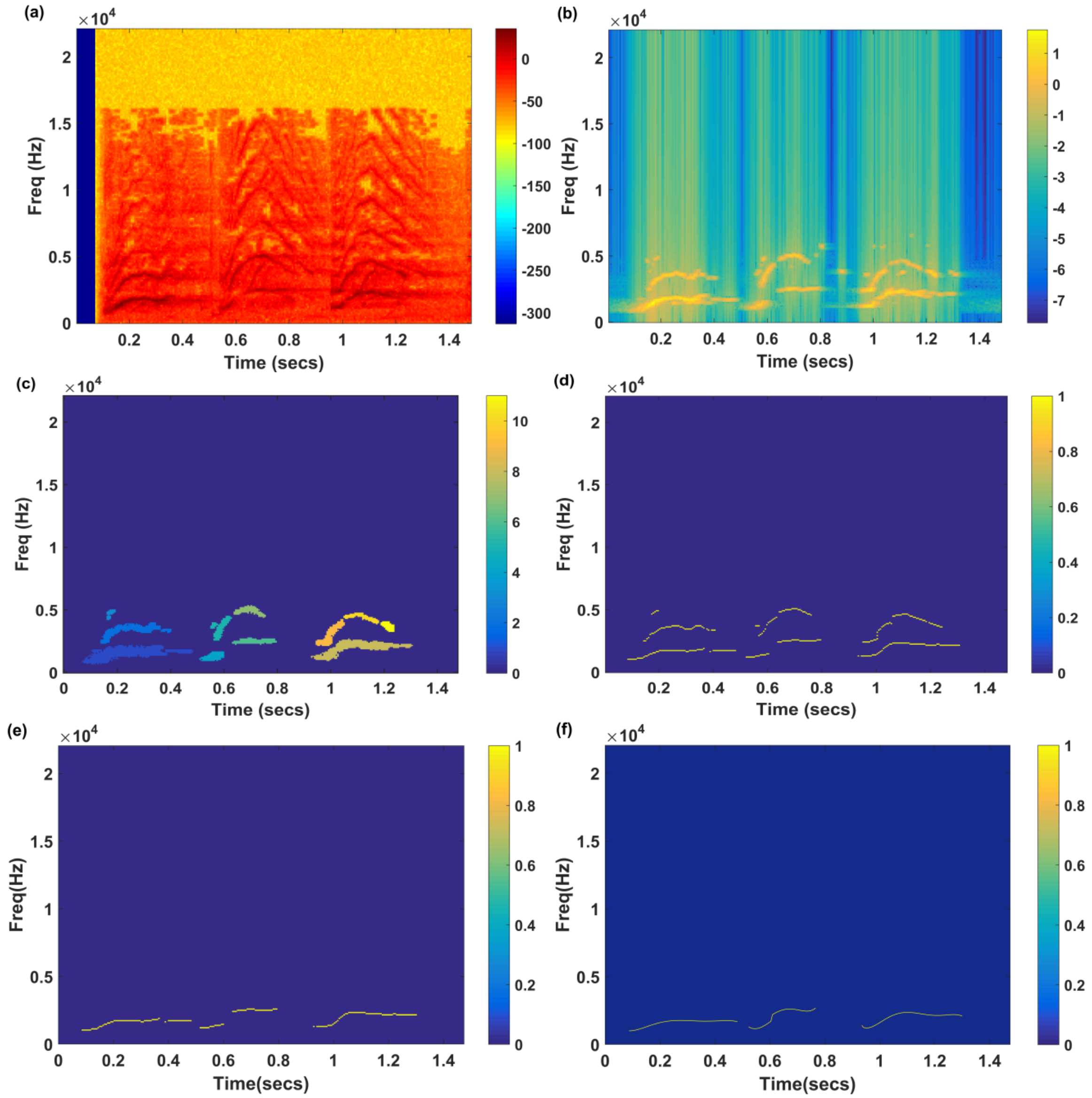


FIGURE 5. Ridge extraction in the spectrogram for the pilot whale: (a) the original spectrogram, (b) the spectrogram of the extracted whistle after the RSSD method, (c) ridge region segmentation, (d) ridge extraction, (e) the prominent ridge extraction, and (f) ridge was joining.

First, considering the oscillatory nature of the whistle signal, we set $Q = 60, r = 4, J = 511$. Figure 4(b) shows the time-frequency signature of the whistle components extracted by our method. This step effectively extracted whistle components from the raw signal. We also noted that the RSSD method is ineffective for the extraction of weak whistle components. The differences in Figure 4(a) and (b) make this weakness clear. Two harmonic components with a lower average frequency were perfectly extracted, but harmonic components with a higher average frequency were not.

Second, we applied two thresholds to remove those spectrogram regions, where whistle components were unlikely to exist. As shown in Figure 4(c), this processing effectively segmented the regions with concentrated whistle components.

Figure 4(d) shows the results of ridge extraction using the derivative method. A comparison of Figure 4(d) and Figure 4(a) shows that the ridges were successfully extracted through this approach. We then employed polynomial fitting with order 5 to form a continuous ridge, as shown in Figure 4(e). The ridges after polynomial fitting took on better shapes.

The processing procedure for cetacean species with multiple harmonic sounds, such as the pilot whale, was similar to that for the blue whale despite some differences in details. Figure 5(d) shows the relevant results after pre-processing as explained previously, we selected only the ridge with the lowest frequency from each sound to accommodate the differences between species. Figure 5(e) shows

TABLE 2. Accuracy rates using an svm classifier with 3 standard samples.

Common name	Accuracy rate			
	Pilot whale	Blue whale	Bottlenose dolphin	Orca whale
Pilot whale	**77.35%**	0%	22.65%	0%
Blue whale	0%	**100%**	0%	0%
Bottlenose dolphin	27.57%	0%	**72.43%**	0%
Orca whale	0%	0%	0%	**100%**

TABLE 3. Accuracy rates using an svm classifier with 4 standard samples.

Common name	Accuracy rate			
	Pilot whale	Blue whale	Bottlenose dolphin	Orca whale
Pilot whale	**76.4%**	0%	23.6%	0%
Blue whale	0%	**100%**	0%	0%
Bottlenose dolphin	16.27%	0%	**83.73%**	0%
Orca whale	0%	0%	0%	**100%**

TABLE 4. Accuracy rates using an svm classifier with 5 standard samples.

Common name	Accuracy rate			
	Pilot whale	Blue whale	Bottlenose dolphin	Orca whale
Pilot whale	**98.82%**	0%	1.18%	0%
Blue whale	0%	**100%**	0%	0%
Bottlenose dolphin	1.16%	0%	**98.84%**	0%
Orca whale	0%	0%	0%	**100%**

TABLE 5. Accuracy rates using a random forest classifier with 3 standard samples.

Common name	Accuracy rate			
	Pilot whale	Blue whale	Bottlenose dolphin	Orca whale
Pilot whale	**98.96%**	0%	1.04%	0%
Blue whale	0%	**99.65%**	0%	0.35%
Bottlenose dolphin	0.34%	0%	**99.66%**	0%
Orca whale	0%	0%	0%	**100%**

TABLE 6. Accuracy rates using a random forest classifier with 4 standard samples.

Common name	Accuracy rate			
	Pilot whale	Blue whale	Bottlenose dolphin	Orca whale
Pilot whale	**98.9%**	0%	1.1%	0%
Blue whale	0%	**99.73%**	0%	0.27%
Bottlenose dolphin	0.13%	0%	**99.87%**	0%
Orca whale	0%	0%	0%	**100%**

prominent ridge extracted, with a discontinuous second sound ridge. In response, we employed polynomial fitting with order 5 to form the continuous ridge shown in Figure 5(g). The results show that our proposed methods work best with multi-harmonic cetacean species as it does for single harmonic.

Finally, we extracted the ridge feature parameters (the start frequency, the end frequency, the time corresponding to the lowest frequency, the lowest frequency, the time corresponding to the highest frequency, the highest frequency, and the time duration) from every ridge. We then classified the ridges using the SVM. Because the number of pilot whale and bottlenose dolphin samples was all small, we adopted the round-by-round method in the classification process. Our proposed classification method obtained good results when extracting ridges with the derivative method. Table 2, Table 3 and Table 4 show the classification results using an SVM classifier with 3, 4, and 5 standard samples, respectively. Table 2 shows that when only three samples were used as standard samples, the average classification

TABLE 7. Accuracy rates using a random forest classifier with 5 standard samples.

Common name	Accuracy rate			
	Pilot whale	Blue whale	Bottlenose dolphin	Orca whale
Pilot whale	**99.06%**	0%	0.94%	0%
Blue whale	0%	**99.86%**	0%	0.14%
Bottlenose dolphin	0.17%	0%	**99.83%**	0%
Orca whale	0%	0%	0%	**100%**

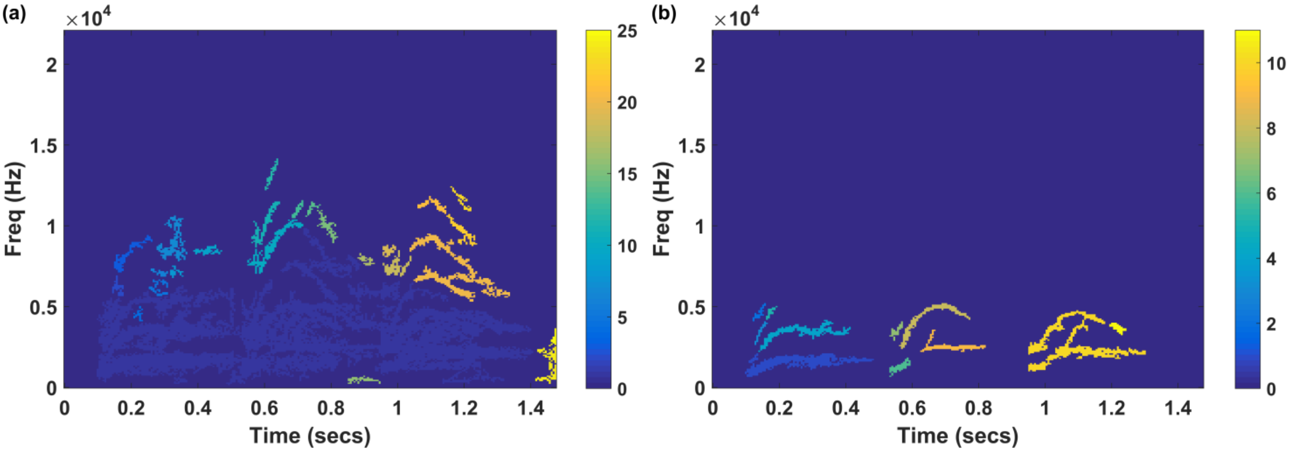


FIGURE 6. Ridge extraction without RSSD in the spectrogram of the pilot whale: (a) low threshold and (b) high threshold.

accuracy of the Pilot whale, the blue whale, and the Orca whale was 77.35%, 100%, 72.43%, and 100%, respectively. Table 3 shows the result when four standard samples were used at a time. Increasing the number of standard samples to five as shown in Table 4, our method achieved a 98.82% average classification accuracy of the pilot whale and 98.84% of the bottlenose dolphin, as we have expected earlier.

Table 5, Table 6, and Table 7 show the classification results using a random forest classifier with 3, 4, and 5 standard samples, respectively. Comparisons of Table 2, Table 3, Table 4, Table 5, Table 6, and Table 7 show that random forest classifier obtains better classification performance than the SVM classifier when using 3 or 4 standard samples, but achieves almost the same classification accuracy when using 5 standard samples. In comparison with the random forest classifier, the SVM classifier has relatively worse performance when using less standard samples, however, the two classifiers can achieve almost the same good performance when using more standard samples.

We compared our proposed method using RSSD to extract whistle components with the results of direct ridge extraction without RSSD. Figure 6 shows this comparison for the pilot whale. The figure shows the ineffective extraction from the non-RSSD method using the same threshold values from our own method; the whistle components remained embedded in noise at the same frequencies. To remove the noise we increased the threshold value to obtain the results in Figure 6(b). A comparison of Figures 5(c) with 6(b) shows that the higher threshold value removed significantly more noise, but, without RSSD, still missed parts of the whistle information. Thus, we conclude that whistle features cannot be extracted effectively with ridge region segmentation and ridge extraction alone, when the signal-to-noise ratio is low.

IV. CONCLUSION

In this study, we have explored a new method for classifying cetacean sounds using RSSD and ridge extraction. Considering the disparate oscillatory intensity of whistle signals, we use an RSSD algorithm with high-Q factor to extract whistle components from raw noisy whale signals. After ridge region segmentation, we extract ridges using partial derivatives and ridge joining (i.e., smoothing) from the whistle spectrogram. Finally, we classify the different whale species vocalizations using an SVM or a random forest. We compared our proposed method and the method without RSSD to the same pilot whale vocalization. Our proposed method offered significantly better extraction results from the whistles. By increasing the number of standard samples used for training the SVM or the random forest. Our method can achieve an accuracy rate in excess of 98% for all the tested species. Our proposed method offers excellent performance in isolating whale whistles and removing noise and, consequently, excellent performance in capturing whale whistle frequency information and in species classification.

REFERENCES

[1] R. Miralles, G. Lara, J. A. Esteban, and A. Rodriguez, "The pulsed to tonal strength parameter and its importance in characterizing and classifying Beluga whale sounds," *J. Acoust. Soc. Amer.*, vol. 131, no. 3, pp. 2173–2179, 2012.

[2] H. Ou, W. W. L. Au, L. M. Zurk, and M. O. Lammers, "Automated extraction and classification of time-frequency contours in humpback vocalizations," *J. Acoust. Soc. Amer.*, vol. 133, no. 1, pp. 301–310, 2013.

[3] A. Kershenbaum and M. A. Roch, "An image processing based paradigm for the extraction of tonal sounds in cetacean communications," *J. Acoust. Soc. Amer.*, vol. 134, no. 6, pp. 4435–4445, 2013.

[4] F. Erbs, S. H. Elwen, and T. Gridley, "Automatic classification of whistles from coastal dolphins of the Southern African subregion," *J. Acoust. Soc. Amer.*, vol. 141, no. 4, pp. 2489–2500, 2017.

[5] D. K. Mellinger, S. W. Martin, R. P. Morrissey, L. Thomas, and J. J. Yosco, "A method for detecting whistles, moans, and other frequency contour sounds," *J. Acoust. Soc. Amer.*, vol. 129, no. 6, pp. 4055–4061, 2011.

[6] S. Datta and C. Sturtivant, "Dolphin whistle classification for determining group identities," *Signal Process.*, vol. 82, no. 2, pp. 251–258, 2002.

[7] J. N. Oswald, S. Rankin, J. Barlow, and M. O. Lammers, "A tool for real-time acoustic species identification of delphinid whistles," *J. Acoust. Soc. Amer.*, vol. 122, no. 1, pp. 587–595, 2007.

[8] A. T. Johansson and P. R. White, "An adaptive filter-based method for robust, automatic detection and frequency estimation of whistles," *J. Acoust. Soc. Amer.*, vol. 130, no. 2, pp. 893–903, 2011.

[9] M. A. Roch, T. S. Brandes, B. Patel, Y. Barkley, S. Baumann-Pickering, and M. S. Soldevilla, "Automated extraction of odontocete whistle contours," *J. Acoust. Soc. Amer.*, vol. 130, no. 4, pp. 2212–2223, 2011.

[10] D. Gillespie, M. Caillat, J. Gordon, and P. White, "Automatic detection and classification of odontocete whistles," *J. Acoust. Soc. Amer.*, vol. 134, no. 3, pp. 2427–2437, 2013.

[11] J. Jing and G. Meng, "A novel method for multi-fault diagnosis of rotor system," *Mech. Mach. Theory*, vol. 44, no. 4, pp. 697–709, 2009.

[12] L. Tong, R. Liu, V. C. Soon, and Y.-F. Huang, "Indeterminacy and identifiability of blind identification," *IEEE Trans. Circuits Syst.*, vol. 38, no. 5, pp. 499–509, May 1991.

[13] P. W. Tse, J. Y. Zhang, and X. J. Wang, "Blind source separation and blind equalization algorithms for mechanical signal separation and identification," *J. Vib. Control*, vol. 12, no. 4, pp. 395–423, 2006.

[14] Y. Wang, Z. He, and Y. Zi, "Enhancement of signal denoising and multiple fault signatures detecting in rotating machinery using dual-tree complex wavelet transform," *Mech. Syst. Signal Process.*, vol. 24, no. 1, pp. 119–137, 2010.

[15] I. W. Selesnick, R. G. Baraniuk, and N. C. Kingsbury, "The dual-tree complex wavelet transform," *IEEE Signal Process. Mag.*, vol. 22, no. 6, pp. 123–151, Nov. 2005.

[16] R. Shao, W. Hu, and J. Li, "Multi-fault feature extraction and diagnosis of gear transmission system using time-frequency analysis and wavelet threshold de-noising based on EMD," *Shock Vib.*, vol. 20, no. 4, pp. 763–780, 2013.

[17] M. Kedadouche, M. Thomas, and A. Tahan, "A comparative study between empirical wavelet transforms and empirical decomposition methods: Application to bearing defect diagnosis," *Mech. Syst. Signal Process.*, vol. 81, pp. 88–107, Dec. 2016.

[18] A. B. Ali, N. Fnaiech, L. Saidi, B. Chebel-Morello, and F. Fnaiech, "Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration signals," *Appl. Acoust.*, vol. 89, no. 3, pp. 16–27, Mar. 2015.

[19] H. Jiang, C. Li, and H. Li, "An improved EEMD with multiwavelet packet for rotating machinery multi-fault diagnosis," *Mech. Syst. Signal Process.*, vol. 36, no. 2, pp. 225–239, 2013.

[20] X. Zhang, Y. Liang, and J. Zhou, "A novel bearing fault diagnosis model integrated permutation entropy, ensemble empirical mode decomposition and optimized SVM," *Measurement*, vol. 69, pp. 164–179, Jun. 2015.

[21] B. Patil, R. Shastri, and A. Das, "Wavelet denoising with ICA for the segmentation of bio-acoustic sources in a noisy underwater environment," in *Proc. 3rd Int. Conf. Commun. Signal Process. (ICCSP)*, Apr. 2014, pp. 472–475.

[22] S. Seramani, E. A. Taylor, P. J. Seekings, and K. P. Yeo, "Wavelet denoising with independent component analysis for segmentation of dolphin whistles in a noisy underwater environment," in *Proc. IEEE Oceans Asia–Pacific Conf.*, May 2006, pp. 1–7.

[23] J. R. Wessel, "Testing multiple psychological processes for common neural mechanisms using EEG and independent component analysis," *Brain Topogr.*, vol. 31, no. 1, pp. 90–100, 2018.

[24] B. Liu, W. Dai, and N. Liu, "Extracting seasonal deformations of the Nepal Himalaya region from vertical GPS position time series using independent component analysis," *Adv. Space Res.*, vol. 60, no. 12, pp. 2910–2917, 2017.

[25] D. Zhang and D. Yu, "Multi-fault diagnosis of gearbox based on resonance-based signal sparse decomposition and comb filter," *Measurement*, vol. 103, pp. 361–369, Jun. 2017.

[26] H. Wang, J. Chen, and G. Dong, "Feature extraction of rolling bearing's early weak fault based on EEMD and tunable Q-factor wavelet transform," *Mech. Syst. Signal Process.*, vol. 48, nos. 1–2, pp. 103–119, Oct. 2014.

[27] I. W. Selesnick, "Wavelet transform with tunable Q-factor," *IEEE Trans. Signal Process.*, vol. 59, no. 8, pp. 3560–3575, Aug. 2011.

[28] I. W. Selesnick, "TQWT toolbox guide," Dept. Elect. Comput. Eng., Polytech. Inst. New York Univ., New York, NY, USA, Tech. Rep., 2011, pp. 1–36. [Online]. Available: http://eeweb.poly.edu/iselesni/TQWT/index.html

[29] M. V. Afonso, J.-M. Bioucas-Dias, and M. A. T. Figueiredo, "Fast image recovery using variable splitting and constrained optimization," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2345–2356, Sep. 2010.

[30] L. Helene and M. Catherine, "Ridge extraction from the scalogram of the uterine electromyogram," in *Proc. IEEE-SP Int. Symp. Time-Frequency Time-Scale Anal.*, Oct. 1998, pp. 245–248.

[31] J. H. Wang and Y. X. Yang, "A wavelet ridge extraction method employing a novel cost function in two-dimensional wavelet transform profilometry," *AIP Adv.*, vol. 8, no. 5, 2018, Art. no. 055020.

HAILAN CHEN received the B.S. degree in electronic devices and technology from Fuzhou University, in 1999, and the M.S. degree in radio physics from Xiamen University, Xiamen, China, in 2008, where she is currently pursuing the Ph.D. degree in communications engineering with the School of Information Science and Engineering. She is currently a Lecturer with the School of Science, Jimei University. Her research interests include underwater acoustic communication and networks, underwater acoustic signal processing, speech signal processing, and sparse representation.

HAIXIN SUN received the B.S. and M.S. degrees in electronic engineering from the Shandong University of Science and Technology, Shandong, China, in 1999 and 2003, respectively, and the Ph.D. degree in communication engineering from the Institute of Acoustic, Chinese Academy of Science, Shanghai, China, in 2006. He is currently an Associate Professor and a Doctorial Tutor with the School of Information Science and Engineering, Xiamen University. His research interests include high speed underwater acoustic communication, and the application of decision feedback equalization and coding in underwater acoustic communication has been put forward.

NAVEED UR REHMAN JUNEJO received the B.S. degree in telecommunication engineering from the National University of Computer and Emerging Sciences, Pakistan, in 2011, and the M.S. degree in information and communication engineering from Harbin Engineering University, China, in 2014. He is currently pursuing the Ph.D. degree in information and communication engineering with Xiamen University, China. His research interests include digital signal processing, compressive sensing for wireless, and underwater acoustic communications.

GUANGSONG YANG received the Ph.D. degree from the School of Information Technology, Xiamen University. He visited UC Davis, USA, from 2009 to 2010, and Griffith University, Australia, from 2017 to 2018, as a Visiting Scholar. He is currently a Professor with the School of Information Engineering, Jimei University, Xiamen, China. His current research interests include underwater sensor networks and intelligent information processing.

JIE QI received the B.S. degree in automatic control and the M.S. degree in electronic engineering from Northwestern Polytechnical University, China, in 1995 and 1999, respectively, and the Ph.D. degree from the School of Aeronautics and Astronautics, Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2011. She is currently a Lecturer with the School of Information Science and Engineering, Xiamen University. Her research interests include optical sensing and optical communication.