View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Elsevier - Publisher Connector

Available online at www.sciencedirect.com

ScienceDirect

Defence Technology 9 (2013) 115−120

www.elsevier.com/locate/dt

# Bark-wavelet Analysis and Hilbert−Huang Transform for Underwater Target Recognition

Xiang-yang ZENG*, Shu-guang WANG

*College of Marine Engineering, Northwestern Polytechnical University, Xi'an 710072, China*

Received 3 January 2012; revised 5 June 2012; accepted 15 September 2012
Available online 2 November 2013

**Abstract**

Recognizing the underwater targets by the radiated noise information is one of the most significant subjects in the area of underwater acoustics. A novel recognition approach which consists of the algorithms of Bark-wavelet analysis, Hilbert−Huang transform and support vector machine is proposed based on the theory of auditory perception. The performance of the proposed method is validated by comparing with traditional method and evaluated by the recognition experiments for SNRs of 0 dB, 5 dB, 10 dB, 15 dB and 20 dB. The results show that the average recognition rate of the method is above 88% and can be increased by 0.75%−6.25% under various SNR conditions compared to the baseline system.
Copyright © 2013, China Ordnance Society. Production and hosting by Elsevier B.V. All rights reserved.

*Keywords:* Acoustics; Underwater target; Recognition; Bark-wavelet; Hilbert−Huang transform

## 1. Introduction

Recognition of underwater targets is significant in the military and economic fields. The acoustic feature extraction and classification of the signals collected by sonars are two important procedures in underwater noise target recognition. At present, numerous techniques have been brought forward for these subjects, including spectrogram correlation [1], time-frequency analysis [2,3], hidden Markov models [4−6], wavelet transformation [7−9], and other approaches [10,11]. Since the noise signals radiated by underwater targets, such as torpedo and vessels, consist of deterministic mechanical sound, propeller noise and hydrodynamic noise, and are often time-variant and non-stationary, the typical feature extraction methods are not suitable for its classification. It is necessary to find new feature extraction approaches. Auditory perception has been proposed recently, and the study shows that it is appropriate for underwater signal processing [12−14].

In this paper, we propose an underwater noise target classification algorithm which consists of Bark- wavelet analysis (BWA), Hilbert−Huang transform (HHT) and support vector machine (SVM), in which BWA is used for auditory denoising and enhanced SNR of the signal, HHT is used for the feature extraction, and SVM is used for the classification. Three types of underwater noise targets with different SNRs are used for the recognition experiment. The results show that the proposed method is effective.

## 2. Using Bark-wavelet to enhance SNR

Noise masking is a well-known characteristic of the human auditory system. It means the fact that the auditory system is incapable of distinguishing two signals close in the time or frequency domains. This is manifested by an elevation of the minimum threshold of audibility due to a masking signal, which motivates its use in the enhancement process to mask

* Corresponding author.
*E-mail address:* zenggxy@nwpu.edu.cn (X.Y. ZENG).
Peer review under responsibility of China Ordnance Society

Production and hosting by Elsevier

2214-9147/$ - see front matter Copyright © 2013, China Ordnance Society. Production and hosting by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.dt.2012.09.001

the residual noise and signal distortion. Bark-wavelet combines the advantages of Bark scale and wavelet analysis. It is appropriate for the analysis of signals in noisy environment.

### 2.1. Bark-wavelet analysis

The perception of the human auditory system is nonlinear to actual frequency, but linear to Bark frequency. Eq. (1) presents the function between the linear frequency and Bark frequency

$$b = 13\arctan(0.76f) + 3.5\arctan(f/7.5)^2 \tag{1}$$

where $b$ is the Bark frequency, and $f$ is the linear frequency. The basic thought of constructing Bark wavelet is as follows. First, for the minimization of time and bandwidth product, Gaussian function is selected as mother function of Bark wavelet. Second, for being consistent with conception of the frequency group, the mother wavelet should have the equal bandwidth in the Bark domain. Furthermore, their bandwidths are the unit bandwidths, namely 1 Bark. This is consistent with the frequency group. According to the above analysis, the formation of wavelet function in the Bark domain is

$$W(b) = e^{-c_1 b^2} \tag{2}$$

When the bandwidth is 3 dB, the constant $c_1$ is selected as 4ln2. Supposing that the linear frequency bandwidth of the signal satisfies $|f| \in [f_l, f_k]$, then the corresponding Bark frequency bandwidth is $[b_l, b_k]$. Thus, the wavelet function in Bark domain can be defined as

$$W_k(k) = W(b - b_l - k\Delta b),$$

$$k = (0, 1, 2 \ldots K-1), \tag{3}$$

where $\Delta b$ is the translation step-length of $W_k(b)$, and $k$ is the scale parameter. According to the principle of equal bandwidth in the Bark domain, the following equation can be obtained

$$\Delta b = \frac{(b_k - b_l)}{K-1}, \tag{4}$$

where $K$ is the total number of sub-bands, $b_k$ is the highest Bark frequency number of signal, and $b_l$ is the lowest Bark frequency number of signal. Considering $W(b) = e^{-c_1 b^2}$, we can obtain

$$W_k(b) = e^{-4\ln 2(b-b_l-k\Delta b)^2} = 2^{-4(b-b_l-k\Delta b)^2},$$

$$k = (0, 1, 2 \ldots K-1). \tag{5}$$

By substituting Eq. (1) into Eq. (5), in the linear frequency, the Bark wavelet function can be described as

$$W_k(f) = c_2 \cdot e^{-4\left[13\arctan(0.76f)+3.5\arctan(f/7.5)^2-(b_l+k\Delta b)\right]^2} \tag{6}$$

In Eq. (6), $c_2$ is a normalized factor and can be obtained through Eq. (7).

$$c_2 \sum_{K=0}^{K-1} W_K(b) = 1, \quad 0 < b_l \le b \le b_k. \tag{7}$$

We can define Bark wavelet transform in linear frequency from Eq. (8)

$$s_k(t) = \int_{-\infty}^{\infty} S(f) W_K(f) e^{j2\pi ft} df, \tag{8}$$

where $S(f)$ is the frequency spectrum of signal $s(t)$, and $s_K(t)$ is the $K^{th}$ sub-band signal which is translated by Bark wavelet.

It should be noted that the relation between the linear frequency and Bark frequency has no influence on Eq. (7), so we can obtain

$$\begin{aligned} \sum_{K=0}^{K-1} s(t) &= \sum_{K=0}^{K-1} \int_{-\infty}^{+\infty} W_k(f) S(f) e^{j2\pi ft} df \\ &= \int_{-\infty}^{+\infty} \sum_{K=0}^{K=1} W_k(f) S(f) e^{j2\pi ft} df \end{aligned} \tag{9}$$

In Eq. (7), let $c_2 = 1$, we can obtain Eq. (10) from Eq. (9)

$$\int_{-\infty}^{+\infty} S(f) e^{j2\pi ft} df = s(t). \tag{10}$$

Therefore, Eq. (7) is called as the engineering perfect reconstruction condition of the Bark wavelet.

From the above analysis we can find that Bark wavelet transform is a nonorthogonal but self- contained, reversible and self-reversal transform. This wavelet still provides good spectrum resolution when the analytic frames are very small, and the frequency bandwidth derived from the Bark domain is very close to the characteristic frequency of the basilar membrane of human ear.

### 2.2. Bark-wavelet analysis of underwater noise signals

After Bark wavelet transform, the underwater signal can be decomposed into various frequency bands which are consistent with human auditory perception. Wavelet denoising is applied to each sub-band signal. Finally, the denoised sub-band signal is merged into the denoised signal. Bark wavelet transform takes the advantages of Bark scale and wavelet denoising. The results are suitable for human auditory perception and have better robustness in noise condition. After wavelet decomposition, the remained noise is in the form of most of wavelet coefficients with smaller value, and at the same time, the underwater signals exist with bigger coefficient values. Soft thresholding is performed on the wavelet coefficients for denoising. The corresponding mathematical expression is

$$f(i) = \begin{cases} \mathrm{sgn}(f(i))(|f(i)| - th), & |f(i)| \ge th \\ 0, & |f(i)| < th \end{cases} \tag{11}$$

where sgn($i$) is the Signum function. The adaptive threshold $th$ is selected using principle of Stein's unbiased risk estimate.

Fig. 1 shows the result of a paddle splashing sound denoised by the Bark wavelet. The sound was recorded underwater at the sample rate of 44,100 Hz. Fig. 1(a) shows the original signal, Fig. 1(b) shows the noise signal with white Gaussian noise at SNR 10 dB, and Fig. 1(c) shows the denoised signal.

The signal is decomposed into 25 Bark scales due to its sampling rate. Each scale is decomposed with 3 level wavelet decompositions and denoised. In order to evaluate the efficiency of the denoising method, the cross correlation coefficient between the original signal and the denoised signal is calculated. Its value is 0.91, which means the denoised signal is very close to the original signal.

After Bark-wavelet analysis, the extracted signal approaches to human hearing. Due to the wavelet denoising, it is also robust in noise conditions. Then Hilbert–Hwang transform is applied to the extracted signals to obtain the feature vector. As an adaptive time-frequency method, HHT is suitable to the non-stationary signals. It is sensitive to the small change of the signal and makes the classification performance better.

## 3. Using HHT to extract the features of underwater noises

HHT comprises the Empirical Mode Decomposition (EMD) and Hilbert transform (HT). The key part of HHT is the EMD method with which any data set can be decomposed into a finite and often small number of intrinsic mode functions (IMF) (fission process). Each IMF satisfies two conditions [15]:1) in the whole data set, the number of extrema and the number of zero crossings must either equal or differ at most by one; 2) at any point, the mean value of the envelopes defined by the local maxima and the local minima is zero.

### 3.1. Empirical model decomposition

EMD is a method which uses the envelopes defined by the local maximum and minimum separately. Once the extrema are identified, all the local maxima are connected by a cubic spline function, as the upper envelope. The procedure of the local minimum to produce the lower envelope is repeated. For a signal designated as $x(t)$, the upper and lower envelopes should cover all the data between them, their mean is designated as $m_1$, and the difference between the data and $m_1$ is the first basic component $h_1$. The EMD procedure is repeated for $k$ times until $h_{1k}$ is a basic component, that is the first IMF designated as $\mathrm{IMF}_1$. Then $\mathrm{IMF}_1$ is separated from $x(t)$ to get the rest of the signal designated as $r_1$ by

$$r_1 = x(t) - \mathrm{IMF}_1 \tag{12}$$

The first process is repeated to get $\mathrm{IMF}_2, r_2, \cdots, r_n$, where

$$r_2 = r_1 - \mathrm{IMF}_2$$

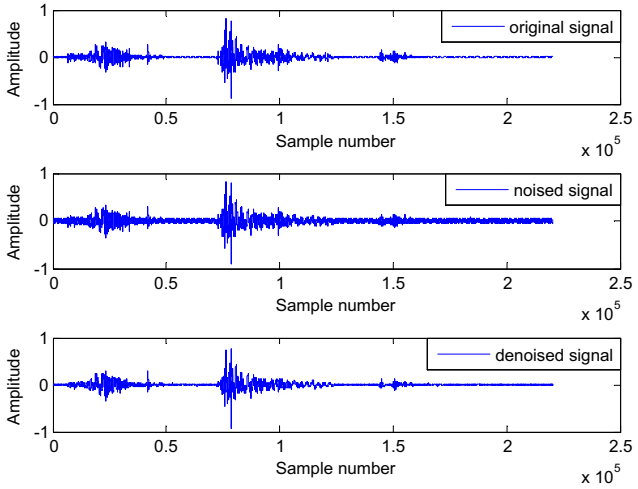


Fig. 1. Original signal and denoised signals.

$$r_3 = r_2 - \mathrm{IMF}_3$$

$$r_n = r_{n-1} - \mathrm{IMF}_n \tag{13}$$

Then summing Eqs. (12) and (13), we can finally obtain

$$x(t) = \sum_{i=1}^{n} \mathrm{IMF}_i + r_n \tag{14}$$

Because EMD merely uses the information about local extrema of signal, it's suitable to analyze the nonlinear and irregular series [16]. EMD can be used to extract the distinct modes of oscillations from the highly nonlinear and nonstationary exchange rate time series so as to clearly understand the change trend of different components.

### 3.2. Hilbert transform

HT can be applied to each IMF so that IF can be acquired. The analysis is shown as follows.

First, the analytic expression can be acquired by HT

$$z(t) = \mathrm{IMF}(t) + j * H(\mathrm{IMF}(t)) = a(t)e^{j\phi(t)} \tag{15}$$

where

$$H(\mathrm{IMF}(t)) = \mathrm{IMF}(t) * \frac{1}{\pi t} = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\mathrm{IMF}(\tau)}{t-\tau} d\tau \tag{16}$$

$$A(t) = \left[\mathrm{IMF}(t)^2 + H(\mathrm{IMF}(t))^2\right]^{1/2} \tag{17}$$

$$\varphi(t) = \arctan[H(\mathrm{IMF}(t))/\mathrm{IMF}(t)] \tag{18}$$

Then the IF can be acquired

$$f(t) = \frac{1}{2\pi}\frac{d[\varphi(t)]}{dt} = \frac{1}{2\pi}\frac{d[\arctan[H(\mathrm{IMF}(t))/\mathrm{IMF}(t)]]}{dt} \tag{19}$$

The root mean square, maximum amplitude of IMF, mean instantaneous frequency, weighted mean instantaneous

Table 1
Derived and selected features for classification.

| Set no. | Feature extraction | Feature selection decision |
|---|---|---|
| 1 | $\mathrm{rms}(\mathrm{IMF}_i)$ | First 2 selected |
| 2 | $\max(A_i)$ | First 2 selected |
| 3 | $\mathrm{mean}(IF_i)$ | First 3 selected |
| 4 | $\text{weighted\_MIF} = \frac{\sum_{k=1}^{m} \mathrm{IF}_i(k)A_i^2(k)}{\sum_{k=1}^{m} A_i^2(k)}$ | First 3 selected |
| 5 | $\mathrm{autocorrelation}(\mathrm{IF}_i)$ | Al discarded |
| 6 | $\mathrm{crosscorrelation}(\mathrm{IF}_i)$ | All discarded |

frequency of each IMF, autocorrelation of each IMF and crosscorrelation between different IMFs can be calculated and are selected experimentally through the recognition rate. Table 1 lists these derived and selected features.

As time-frequency analysis tool, Bark-wavelet and Hilbert–Huang transform are not only suitable to the traditional stationary data but also the non-linear and non-stationary data. This means that the proposed method has the various wide applications.

## 4. Experiments and analysis

Three types of underwater targets (A, B and C) are used for the recognition experiments. Each type of signals has 448 samples, with a sample rate of 4 kHz. Fig. 2 shows the waveform of the three types of signals.

The difference between the different signals cannot be seen clearly from their waveform. Thus, the spectra of the three types of signals are calculated and shown in Fig. 3. The spectra are obtained through the FFT from a 20 ms segment. From Fig. 3, it can be seen that there are small differences among their frequency spectra. The frequency component of Type A is about 420 Hz, while the frequency components of Type B and Type C are about 300 Hz and 180 Hz, respectively.

Fig. 4 shows the corresponding Bark-wavelet waveform. The signal is decomposed into 13 Bark scales according to its sample rate. Each signal is denoised, and then the features are extracted from them. Finally the support vector machine

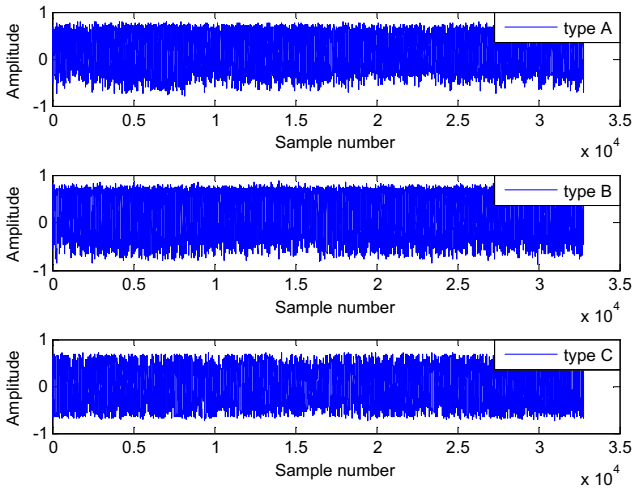


Fig. 2. Waveforms of 3 types of signals.

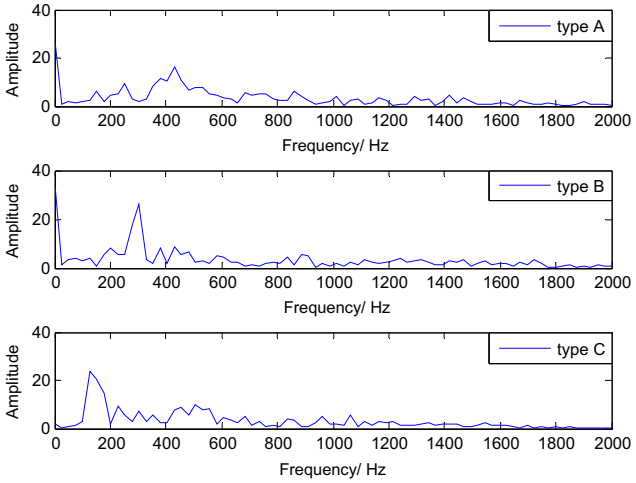


Fig. 3. Spectra of 3 types of signals.

(SVM) is used for the classification. A half of the signals are used for training and the other half are used for recognition test.

In order to evaluate the classification accuracy of the proposed method, a baseline system is designed for the classification experiment. The baseline system uses the noise loudness as a feature and SVM as a decision tool, of which the performance was proved in our previous work and is compared with the novel method proposed in this paper. Both methods are used for the classification of the three underwater targets, and the results are shown in Table 2. It can be seen from Table 2 that the mean recognition rate of the proposed method is higher than that of the baseline system. The recognition rates of Type A and C are approximate. For Type B which is difficult to classify, the proposed method is better than the baseline system.

In order to test the efficiency of the proposed method, it is also compared with a published method [17], which consists of wavelet packet transform (WPT) and support vector machine (SVM). Its performance has been proved through the classification experiment. As its description, after 5 level wavelet packet transform, the normalized energy of sub-band signal is used as the feature, and SVM is used for the classification. We tested the method with the same data. The results are also shown in Table 2.

It can be seen from Table 2 that the average classification rate of the method we proposed is about 4.4% higher than that of WPT method. Its performance is especially better for Type A and B.

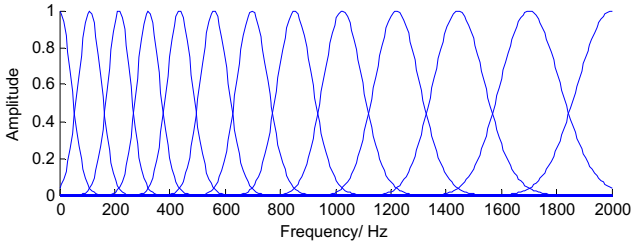


Fig. 4. Waveform of Bark wavelet.

Table 2
Classification results of different methods and baseline system.

| Method | Recognition rate | | | |
|---|---|---|---|---|
| | A | B | C | Mean |
| WPT/% | 79.31 | 76.57 | 96.91 | 84.27 |
| Baseline/% | 85.39 | 75.65 | 100 | 87.01 |
| Proposed/% | 84.82 | 82.14 | 99.10 | 88.70 |

The recognition rates of WPT method and the baseline method are lower than that of the proposed method for Type B signal. This may be because the frequency deviation of the two methods is relatively constant and the divided sub-frequency bands are close to each other. Also, the main frequencies of three types of signals are relatively approximated. When the main frequency of Type B signal fluctuates, it is highly possible to be mis-divided into the nearby sub- frequency bands, which would be confused with Type A or C. Compared with the two methods, the frequency deviation of the proposed method is adaptive, and the first 3 IMFs are used for the feature extraction, which increases the robustness of the classification performance.

To test the performance of the proposed method under various noise conditions, the white Gaussian noises are added to the signals to achieve different SNRs. The noise signals and the denoised signals are tested under each SNR condition. The recognition rates are shown in Table 3.

From Table 3, it can be seen that the recognition rate of denoised signal is higher than that of noise signal. The results are plotted and shown in Fig. 5.

Although the recognition rates of both signals decrease with the decrease in SNR, the recognition rate of denoised signal is always higher than that of the noise signal. It also can be seen that the decrease in the recognition rate of denoised signal is slower than that in the noised signal. While SNR decreases, the gap of recognition rate becomes larger. The recognition rate of denoised signal is 6.25% higher than that of the noise signal.

## 5. Conclusions

An underwater noise target recognition method was proposed by combining the Bark-wavelet analysis, Hilbert−Huang transform and support vector machine. It is an attempt of applying auditory perception theory into underwater noise target classification. The experimental results have validated the performance of the method. In future work, the efforts will be made on feature selection and robustness enhancement.

Table 3
Classification results at various SNRs.

| SNR/dB | Recognition rate | |
|---|---|---|
| | Noise signal/% | Denoised signal/% |
| 20 | 84.52 | 85.27 |
| 15 | 83.33 | 85.12 |
| 10 | 80.21 | 84.23 |
| 5 | 73.51 | 79.76 |
| 0 | 66.52 | 71.28 |

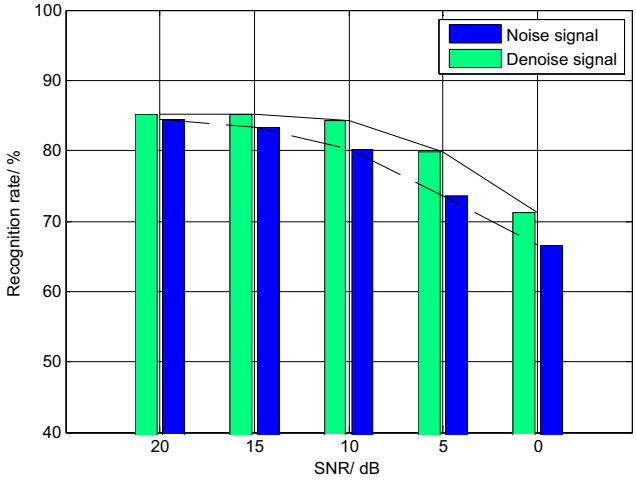


Fig. 5. Recognition rate of noise and denoised signals.

## Acknowledgment

This research was support by the program for New Century Excellent Talents in University for the support.

## References

[1] Ali Pezeshki, Azimi-Sadjadi Mahmood R, Scharf Louis L. Undersea target classification using canonical correlation analysis. IEEE J Ocean Eng 2007;32(4):948−55.
[2] Cornel Ioana, André Quinquis, Yann Stephan. Feature extraction from underwater signals using time-frequency warping operators. IEEE J Ocean Eng 2006;31(3):628−45.
[3] Huynh Quyen Q, Cooper Leon N, Nathan Intrator, Shouavl H. Classification of underwater mammals using feature extraction based on time−frequency analysis and BCM theory. IEEE Trans Signal Process. 1998;46(5):1202−7.
[4] Robinson M, Azimi-Sadjadi MR, Salazar J. Multiaspect target discrimination using hidden Markov models and neural networks. IEEE Trans Neural Network 2005;16(2):447−59.
[5] Dasgupta N, Runkle P, Couchman L, Carin L. Dual hidden Markov model for characterizing wavelet coefficients from multi-aspect scattering data. Signal Process. 2001;81:1303−16.
[6] Runkle P, Bharadwaj P, Carin L. Hidden Markov model multiaspect target classification. IEEE Trans Signal Process. 1999;47(7):2035−40.
[7] Shi GZ, Hu JC. Extraction and fusion of frequency domain features from ship radiated-noise based on wavelet packet and 1(1/2)-dimensional spectrum. Tech Acoust 2004;23(1):4−7.
[8] Li XY, Peng Y, Lin LJ, Lin ZQ. Study on classification of underwater targets based on modulation spectrum by wavelet transforms and cubic spline technique. Acta Acustica 2004;29(1):64−6.
[9] Azimi-Sadjadi MR, Yao D, Huang Q, Dobeck GJ. Underwater target classification using wavelet packet and neural networks. IEEE Trans Neural Network 2000;11(3):784−94.
[10] Ji SL, Cong FY, Jia P, Shi ZX. Principal component analysis on LOFAR spectrum features of underwater target signals. J Data Acquis Process 2003;18(2):123−6.
[11] Runkle P, Carin L, Couchman L, Yoder T, Bucaro J. Multiaspect identification of submerged elastic targets via wave-based on matching

pursuits and hidden Markov models. IEEE Trans Pattern Anal Mach Intell 1999;21(12):1371−8.

[12] Lu ZB, Zhang XH, Hu HB. Auditory feature extraction of the underwater target. J Syst Eng Electron 2004;26(12):1801−3.

[13] Peng Y, Wang S, Wang KJ, Li XY, Lin LJ, Lin ZQ, et al. A study on underwater target classification applying perception linear prediction method. Acta Acustica 2006;31(2):146−50.

[14] Wang Y, Sun JC, Chen KA, Fu LL. Feature extraction of underwater targets based on psychoacoustic parameters. J Data Acquis Process 2006;21(3):313−7.

[15] Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng Q, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proc Royal Soc A Math Phys Eng Sci 1998;454:903−95.

[16] Huang NE, Wu Z. A review on Hilbert-Huang transform: method and its applications to geophysical studies. Rev Geophys 2008;46:1−23.

[17] Liu J, Liu Z, Xiong Y. Underwater target recognition based on wavelet packet energy spectrum and support vector machine. J Wuhan Univ Technol (Transport Sci Eng) 2012;36(2):361−5.