中正嶺學報 第三十五卷 第二期 民國 96.05.
JOURNAL OF C.C.I.T., VOL. 35, NO. 2, May, 2007
短文 Short Note

Using Empirical Mode Decomposition for Underwater Acoustic Signals Recognition

Chung-Ling Hung, Chung-Shi Chiang, and Te-Ming Tu

Department of Electrical Engineering, Chung Cheng Institute of Technology, National Defense University

ABSTRACT

Due to the non-linear and time-varying properties, to process and analyze underwater acoustic signals becomes very difficult and complicated. Fourier transform is only suitable for analyzing stationary signals, but not appropriate for the detection of short time and transient signals embedded in the underwater acoustic signals. Based on the property of multi-scale and multi-translation, wavelet transform can be used for analyzing transient signals. Although the scale and translation parameters in the wavelet basis functions can be adjusted for different signals, the basis functions are fixed, same as the basis functions of Fourier transform. Hence, the requirements of analyzing time-vary underwater acoustic signals are not satisfied. The algorithm of empirical mode decomposition provides a useful analysis scheme for non-linear and non-stationary nature signals. In this work, empirical mode decomposition method is design to extract the features from underwater acoustic signals for recognition. In the experiments, we utilize the data set of different ship classes recorded by the hydrophone to test the proposed scheme. Experimental results demonstrate the robustness of the proposed method.

Keywords: underwater acoustic signals, Empirical Mode Decomposition, Intrinsic Mode Function

利用經驗模態分解法識別水下音響信號

洪崇嶺　江中熙　杜德銘

國防大學中正理工學院電機系

摘 要

由於水下音響信號具有非線性與時變的特性，使得它的處理與分析變得困難與複雜。雖然傅立葉轉換能有效地分析穩態信號，但它無法偵測短時與暫態的信號；而小波轉換因具有多重解析的特性，可以應用於短時與暫態信號的分析，但它的基底函數與傅立葉轉換一樣都是固定的，對於不同的暫態信號其多重解析的參數必須做適度的調整，因此小波轉換無法滿足時變信號的分析需求。針對傅立葉與小波轉換的局限性，經驗模態分解法是最近設計出來分析非線性與非穩態信號的新方法；在本文中我們運用經驗模態分解法來萃取水下音響信號的特徵，尋求最佳識別的模態分解階數。在實驗中我們將以真實的船艦水下音響信號來驗證，實驗結果證明經驗模態分解法所萃取水下音響信號的特徵的確能夠提昇後續的判讀作業。

關鍵詞：水下音響信號、經驗模態分解法、本質模態函數

文稿收件日期 95.11.06；文稿修正後接受日期 96. 4.16
Manuscript received November 20, 2006; revised April 16, 2007

I. INTRODUCTION

Recently, the classification problem of underwater targets from the acoustic backscattered signals has attracted a lot of attention. These signals consist of passive sonar signals radiated from various vessels as well as underwater transients such as whale clicks, porpoise whistles and ice crackles, etc. Each type of signals has distinct characteristics, and is conventionally identified by human experts either by listening to or by looking at the spectrograms of the proposed sonar signals. The work [1] on the modeling of the underwater signals radiated from ships has led to the conclusion that the signal sources can be divided into three captions: machinery signals, propeller signals, and hydrodynamic signals. Machinery signals are produced from the diverse parts of a moving ship, such as pumps, pipes, and motor armature, etc. Propeller signals create tonal components in addition to the continuous spectrum of cavitation signals. Hydrodynamic signals consist of Gaussian signals generated from the hull of the vessel and flow signals similar to the ambient signals in the ocean. Hence, the spectrum of the radiated signals consists of two types. First one is the narrow band with a discontinuous spectrum containing line components occurring at discrete frequencies. The other one is the broad band with a continuous spectrum. Radiated signals from ships include a mixture of these two types of signals. The tone is an important feature for ships , so to extract tonal features from mixed spectra is effectively a key task for the recognition of acoustic signals.

Various schemes of signal processing [2-10] have been proposed to extract signatures of submerged targets from the narrow band of sonar data mainly for detection purposes. Fourier transform is used to analyze and detect the sound signals in many approaches. Although this transform is extremely useful and well established, it does have principal difficulties in analyzing short-time transient sound behaviors. Assorted short-time Fourier transforms (STFT), using a variety of "windows" with different relative advantages, have been developed to address this problem [2]. Furthermore, alternatives to the STFT with better time-frequency localization have been also suggested. Cohen [3] devised a time frequency distribution by composing spectrums at different time intervals, which has been studied as the characterization using the Wigner-Ville distribution [4] and its variants. Such distribution describes the energy or intensity of a signal simultaneously in time and frequency, and is also a powerful tool for the analysis of non-stationary signals. Those time-frequency approaches have been extensively reviewed [5,6].

The potential uses of applying the wavelet transform for statistical problems have been discussed and developed [7,8]. They have demonstrated the appropriate threshold values on the coefficients resulting from the wavelet decomposition and a function of unknown smoothness can be recovered from sampled data contaminated with white noise. Subsequently, their idea has been used by various researchers to uncover useful structural information from complex noisy datasets. Such thresholding techniques are essentially concerned with the recovery of "smooth" features against a background that is often assumed to be "white" or "colored" noise. Similar to those methods, the approach [9] is applied to data from an entirely different arena for different purposes. The wavelet transform is used and then the kernel smoothing is applied to the coefficients in each resolution level to produce useful results from the data set. Apart from those approaches using the wavelet transform, Johnstone and Silverman [10] have developed a "level-dependent" thresholding approach for data with correlated noise.

To recognize passive tonal signals radiated from ships and propagated through an ocean environment at various speeds is crucial for the processing stages. Traditionally, basis decomposition techniques such as Fourier decomposition or wavelet decomposition are selected to analyze real world signals as mentioned in the previous paragraph. Also, Fourier and wavelet descriptors have long been used as powerful tools for feature extraction. However, the main drawback of those approaches is that the basis functions are fixed and do not necessarily match the varying nature of signals. The empirical mode decomposition (EMD) was firstly proposed by Huang et al. [11], with which any complicated data set can be decomposed into a finite and often small number of intrinsic mode function (IMF) components, which become the basis representing the data. Those extracted components can match the signals themselves very well. Motivated by that EMD provides a decomposition method to analyze the signals locally and separate the component holding locally the highest frequency from the rest into a separate IMF, in this paper, we adopt the EMD technique to extract the feature of the underwater targets. The first advantage of this technique is that EMD is a fully data driven method, and does not use any predetermined filter as Fourier basis or wavelet function does. The second one is that it can be easily implemented and the speed of matching time can be improved. Therefore, here the EMD approach is design to extract residual components of the underwater targets as the discriminative features for recognition.

II. PREPROCESSING OF ACOUSTIC SIGNALS

The selection of a suitable database is critical to verify the robustness of a classification methodology. It was recognized that the acoustic database, as it stands, is limited in its applications for developing and comparing specific classification algorithms. However, as will be discussed next, these limitations can be mitigated by 1. removing certain artifacts; 2. appropriate signals scaling; and 3. normalization. These modifications create an effective database that can be used to investigate challenging classification problems. The entire preprocessing procedure consists of the following steps that are implemented in order:

1. Artifact removing: The length of the data set is recorded for 60 seconds and the sampling rate is 11025 Hz. Hence, the signal size is 661500 points. Artifact removing is done by extracting the middle part of 40 seconds from data.
2. Resampling: The original sampling rate is significantly higher than the signal bandwidth. This step is to reduce the sampling rate to 1000 Hz and make the signal size down to 40000 points. The data is denoted by X for further uses.
3. Normalization: the linear rescaling [12] is applied to each vector to adjust the average of each data set to zero and to normalize the standard deviation to unity before using the acoustic signal vector. By calculating the mean $\bar{x}$ and variance σ_x^2 with respect to the spatial template, the linear rescaled version of X is represented by X^N given by

$$X^N = \frac{X-\bar{x}}{\sigma_x} = \{x_1^N, x_2^N, \cdots x_j^N, \cdots, x_n^N\} \qquad (1)$$

where the mean $\bar{x} = \sum_{j=1}^{n} x_j \Big/ n$ and

the variance $\sigma_x^2 = \sum_{j=1}^{n} (x_j - \bar{x})^2 \Big/ n-1$

Figure 1 shows the results of the preprocessing procedure using the same sample from one class. Figure 2 displays the preprocessing results for different samples of the same class, and Figure 3 reveals the preprocessing results of different classes.

III. FEATURE EXTRACTION USING EMPIRICAL MODE DECOMPOSITION

Joint space-spatial frequency representations have received special attention in the fields of image and signal processing, feature extraction and pattern recognition. Huang et al. [11] introduced a multi-resolution decomposition technique: the empirical mode decomposition (EMD), which is adaptive and appears to be suitable for non-linear, non-stationary data analysis.

Different applications as medical and seismic signals have shown the effectiveness of this method. The most attractive one among the facts is that EMD acts as dyadic filter banks [13]. The principle is to adaptively decompose a given signal into components called intrinsic mode functions (IMFs) by satisfying two properties. One is that the numbers of extrema and zero-crossing must equal. Another property of IMF is that the mean value of the envelope defined by the local maxima and the local minima, respectively, is locally symmetric around the envelope mean . Those IMF components are

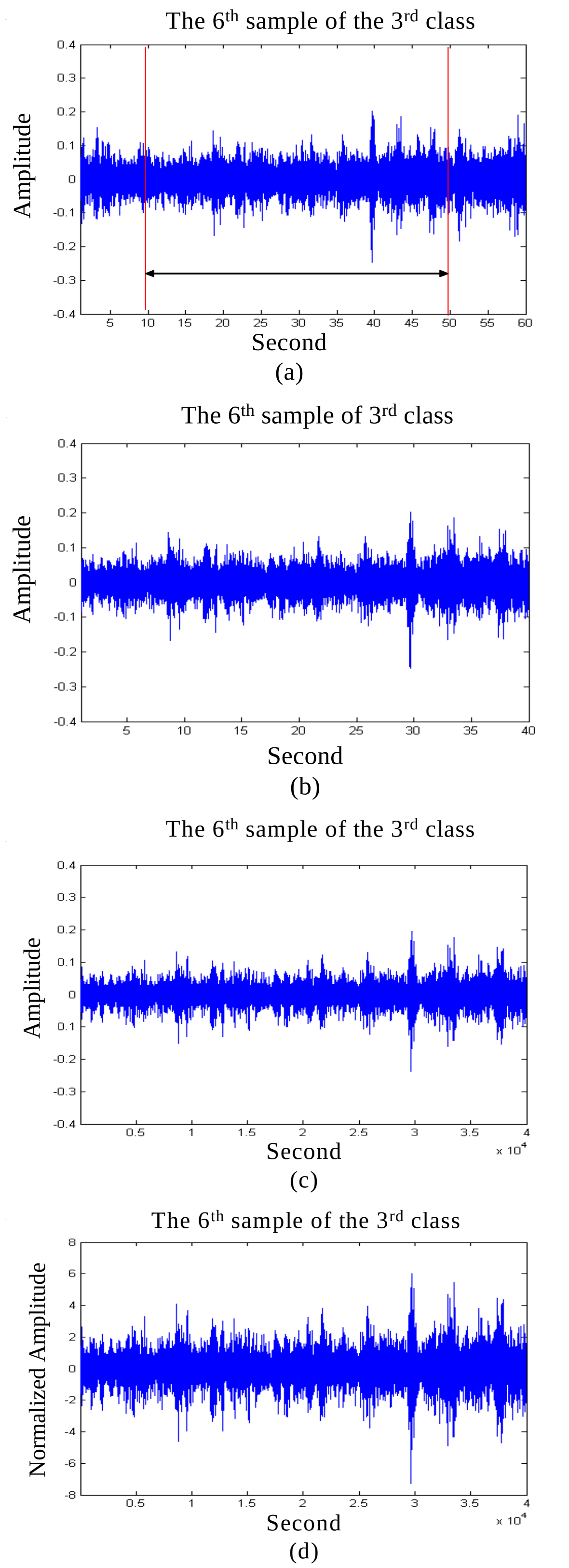


Fig.1. The results of the preprocessing procedure (a) Original signal, (b) Artifact removing, (c) Resampling, (d) Normalization.

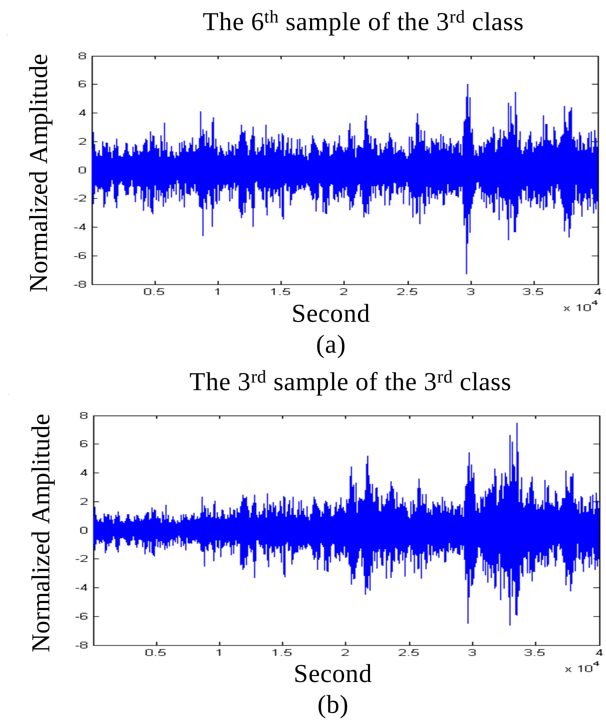


Fig.2. The preprocessing results of the same class (a) The 6th sample of the 3rd class, (b) The 3rd sample of the 3rd class.

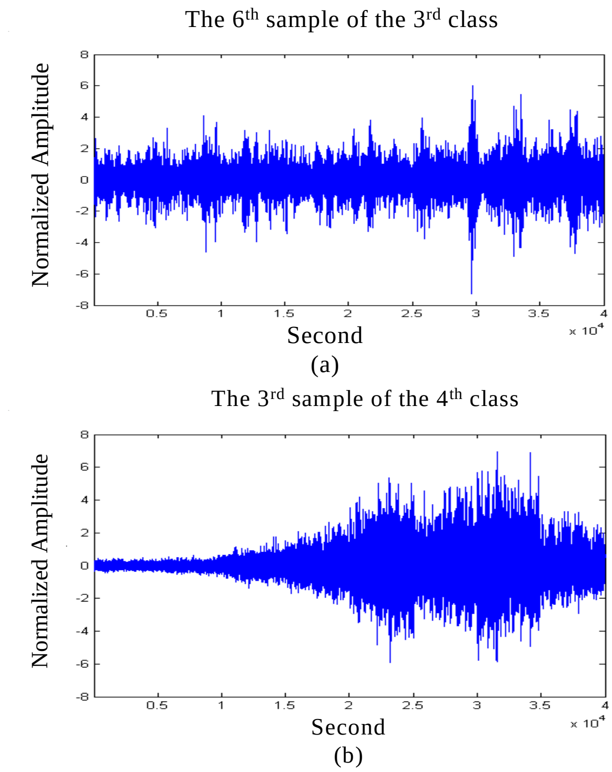


Fig.3. The preprocessing results of different classes (a) The 6th sample of the 3rd class, (b) The 3rd sample of the 4th class.

obtained from the signals by means of an algorithm called sifting process. This algorithm extracts locally for each mode the highest frequency oscillations out of the original signals.

3.1 Sifting Procedure

Given a one-dimensional signal X , the sifting process to find the IMFs is summarized as follows [11]:

(1) Initialization:
Set $r_0 = X$ (the residue) and set $i = 1$ (index number of IMF).

(2) Extracting the ith IMF component:

(a) Initialize $h_{i0} = r_{i-1}$ and $j = 1$.

(b) Find all the points of local maxima, and all the points of local minima in h_{ij-1}.

(c) From the input signal h_{ij-1}, create the upper envelope of local maxima, denoted by $\max_{ij-1}$, and the lower envelope of local minima, denoted by $\min_{ij-1}$, by interpolation.

(d) Calculate the mean of the upper and the lower envelopes by

$$m_{ij-1} = (\max_{ij-1} + \min_{ij-1})/2 \tag{2}$$

(e) Update: $h_{ij} = h_{ij-1} - m_{ij-1}$, then $j = j+1$,

(f) Calculate the stopping criterion (standard deviation SD_{ij} as defined hereinafter)

(g) Repeat step (b) to (f) until $\mathrm{SD}_{ij} \le \mathrm{SD}_{\mathrm{MAX}}$, where the $\mathrm{SD}_{\mathrm{MAX}}$ is usually set between 0.2 and 0.3, and then set $c_i = h_{ij}$ (c_i means the ith IMF)

(3) Update the residue $r_i = r_{i-1} - c_i$,

(4) Repeat steps (2) to (3) with $i = i+1$ until the number of extrema in r_i is less than 2.

Note that the interpolation method usually used is the cubic spline interpolation [11, 13].

3.2 Finding all the IMFs

Once the first set of 'sifting' results in an IMF, defined by $c_1 = h_{1j}$, this first IMF component contains the finest spatial scale in the signal. Then the first residue, r_1, of the signal can be generated by subtracting out c_1, $r_1 = r_0 - c_1$. The residue now contains information about large scales. The other IMFs can be computed by performing the resifting to find additional components $r_2 = r_1 - c_2$,…, $r_n = r_{n-1} - c_n$. The original signal can then be reconstructed, using the following equation

$$X = \sum_{i=1}^{n} (c_i) + r_n \tag{3}$$

To stop the sifting process, a criterion needs to be determined. This can be accomplished by limiting the standard deviation (SD), computed from two consecutive sifting results given by

$$\mathrm{SD}_{ij} = \sum_{k=1}^{K} \left[\frac{\left| h_{i(j-1)}(k) - h_{ij}(k) \right|^2}{h_{i(j-1)}^2(k)} \right] \tag{4}$$

Figure 4 shows a simulated example of an iteration of sifting process by EMD decomposition, where the analyzed signal on Figure 4(c) is composed of two sinusoidal waves with two frequencies. Low frequency is shown in Figure 4(a) and high frequency is shown in Figure 4(b). Figure 4(d) displays the lower envelope and Figure 4(e) shows the upper envelope. The mean envelope is demonstrated in Figure 4(f) which is computed from Equation (2). Figure 4(g) shows the integration of the upper envelope, the lower envelope and the mean envelope. Finally, the first IMF is obtained and shown in Figure 4(h) and the first residue is shown in Figure 4(i).

From Figure 4, the EMD algorithm extracts the oscillatory mode that exhibits the highest local information from the data ("detail" in the wavelet context), leaving the remainder as a "residue" ("approximation" in wavelet analysis). According to the major advantage of EMD that the process of deriving the basis functions is empirical, the basis functions are derived dynamically from the signal itself. As shown in Figure 4, the EMD approach is used to denoise the noisy acoustic backscattered

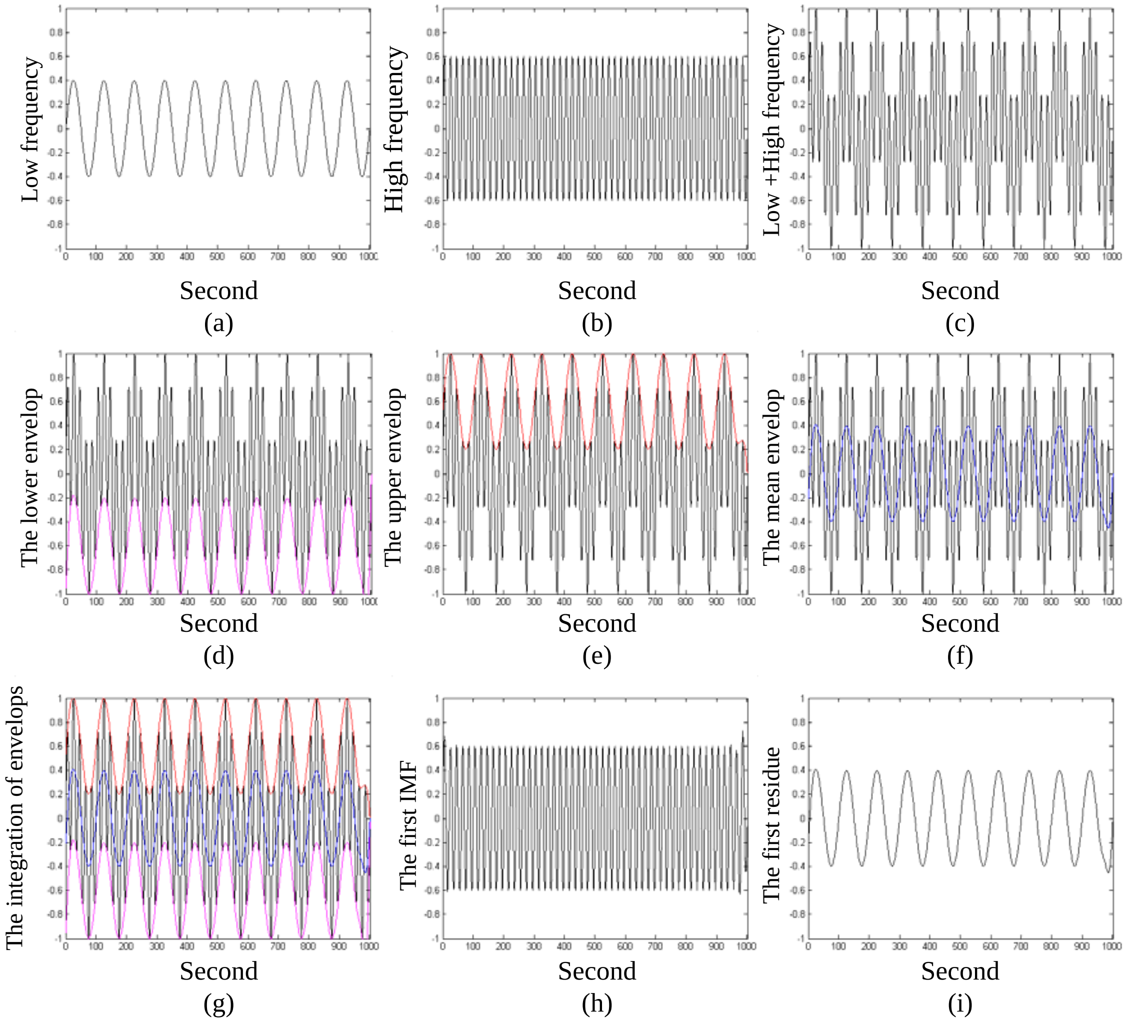


Fig.4. A simulated example for an iteration of the sifting process by EMD decomposition (a) Low frequency of the analyzed signal on (c), (b) High frequency of the analyzed signal on (c), (c) The analyzed signal simulated, (d) The lower envelope, (e) The upper envelope, (f) The mean envelope, (g) The integration of the upper envelope, the lower envelope and the mean envelope, (h) The first IMF, (i) The first residue.

signals and greatly improve the SNR. Therefore, it is reasonable to consider that the residue presents the basic characteristics of the signal and the detail denotes the variation of the noise represented by the highest local information. That is, we use the EMD as a low-pass filter and only the distinct signal characteristics (the residue of EMD) are utilized as discriminating features for accurate signals recognition.

3.3 Feature Vector

For the normalized acoustic signal, the sequences are formed to the 1-D vector represented by

$$V = \{v_1,\ v_2,\ \cdots v_j,\ \cdots,\ v_n\} \tag{5}$$

where v_j defines the position of the vector V, and n is the number of total components, herein, $n = 40000$. Features from the EMD residue of the 1-D vector V^N can be obtained by

$$\mathbf{R}^m = \{R_1^m,\ R_2^m,\ \cdots R_j^m,\ \cdots,\ R_n^m\} \tag{6}$$

where $\mathbf{R}^m$ represents the mth residue of the EMD results and R_j^m denotes the feature from the jth position of the R^m. In our experiments, the feature vector consists of 40000 components.

IV. MATCHING AND RECOGNITION

The main goal of acoustic signals recognition is to match the unknown acoustic feature with those known acoustic feature classes in the database and determine whether the unknown feature comes from the authentic one or the imposter. The matching process is to be made with the unknown feature, which will be calculated depending on different metrics. The different similarity measures used as the matching criterion are

1. The mean of the Euclidean distance (MED) measure:

$$d_1(p,\ q) = 1 - \sqrt{\frac{1}{M}\sum_{i=1}^{M}(p_i - q_i)^2} \qquad (7)$$

where M is the dimension of the feature vector, p_i is the ith component of sample feature vector, and q_i is the ith component of unknown sample feature vector.

2. The cosine similarity measure:

$$d_2(p,\ q) = \frac{p}{\|p\|} \bullet \frac{q}{\|q\|} \qquad (8)$$

where p and q are two different feature vector, and $\|\bullet\|$ indicates the Euclidean norm. The range of $\frac{p}{\|p\|} \bullet \frac{q}{\|q\|}$ is $[0,\ 1]$. The more similar the two vectors are, the bigger the $d_2(p,\ q)$ value is.

3. The correlation coefficient (CC) measure:

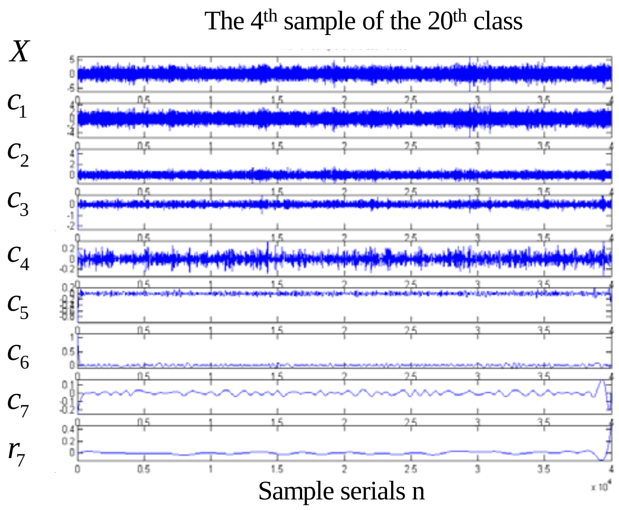
$$\mathbf{CC} = \frac{\sum(p_i - \bar{p})(q_i - \bar{q})}{\sqrt{\sum(p_i - \bar{p})^2}\sqrt{\sum(q_i - \bar{q})^2}} \qquad (9)$$

where $\bar{p}$ and $\bar{q}$ are the mean values of the database sample feature vector and the unknown sample feature vector, respectively. The peak correlation coefficient value equals one when the two signals are completely identical.

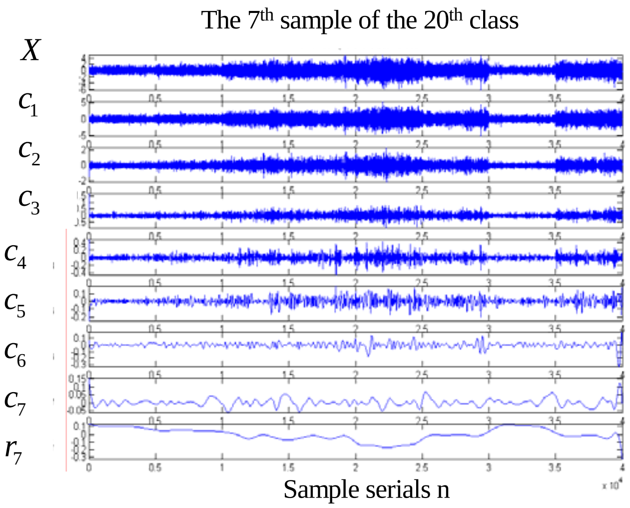
V. EXPERIMENTAL RESULTS

This section describes the experimental results obtained from the experiments performed at the proposed approach. Before the results are presented, we will introduce the adopted database and the verification (identification) scheme for an acoustic signal recognition system. Our available data relate to underwater sounds of the different speed of the different vessels. They consist of selected extracts from lengthy recordings taken in real life conditions in the ocean, using the hydrophone placed several meters beneath sea level. The recording apparatus sampled the sound at the 11025 Hz with 16-bit resolution, and thus the volume of raw data is considerable. The sampling rate is over twice the highest frequency of the signals that we potentially wish to identify, so there are few aliasing problems. The experiments conducted below are running on the computing environment of 1.8 GHz PC with 736 MB RAM using Matlab 6.5.

Figure 5 shows the results of seven times EMD decomposition, computed from Equation (3), for two samples of the same acoustic signal. Figure 6 shows the same for two samples of the different acoustic signals. The top upper part, X, is the normalized original signal, and the bottom part, r_7, is the seventh residue. In this work, we choose the three different residues as feature sequences. Herein, two sets of feature sequences with their original length that are the first residue displayed in Figure 7(a) and 7(b). For easy comparison, Figure 7(c) and 7(d) show only the first 1024 components of Figure 7(a) and 7(b), respectively. Figure 7(e) and 7(f) show only the first 128 components of their original feature sequences. Figure 8 demonstrates the EMD results of two samples



(a)



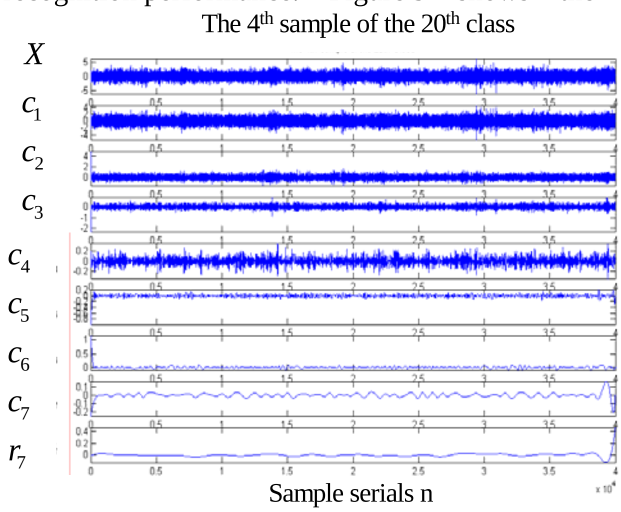
(b)

Fig.5. EMD decomposition results, obtaining IMFs and residue, of two samples from the same class, (a) and (b) are two sets of 7 times decomposition sequences with their original length where X is the normalized original signal and r_7 is the 7th residue.

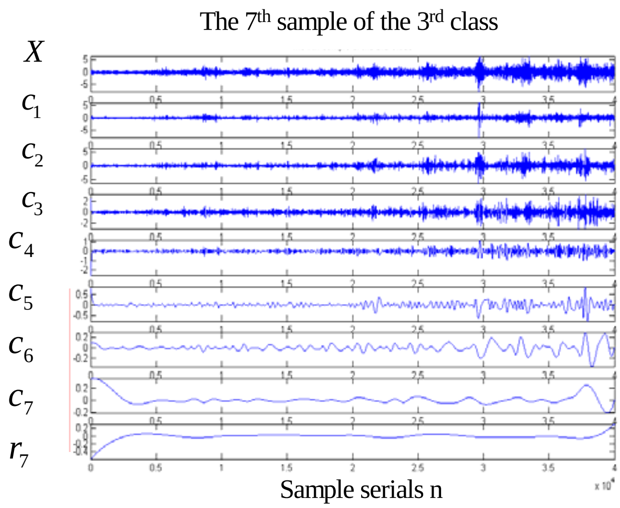
corresponding EMD results for the first 1024 and 128 components of Figure 8(a) and 8(b), respectively. To demonstrate the similarity of two acoustic signals from the same class captured at different time, it is easily proved by checking those corresponding circles marked in Figure 7(e) and 7(f). Also, those circles marked in Figure 8(e) from two different acoustic signals. Same as Figure 7 , Figure 8(c) - 8(f) present the

and 8(f) point out the differences of two samples from two classes.

Furthermore, we use the data to test the recognition performance. Figure 9 shows the



(a)



(b)

Fig.6. EMD decomposition results, obtaining IMFs and residue, of two samples from the different classes, (a) and (b) are two sets of 7 times decomposition sequences with their original length where X is the normalized original signal and r_7 is the 7th residue.

recognition results using the first residue as the feature, Figure 10 displays the recognition results using the second residue as the feature, and Figure 11 demonstrates the results using the third residue as the feature. Note that they all use three similarity measures from Equation (7) to (9). As shown in Figure 9, Figure 10, and Figure 11, the same class demonstrates the highest peak, and the different class shows the lowest values. However, as shown in Figure 11, while using the third residue as the feature and the MED metric as the similarity measure , the achieved recognition

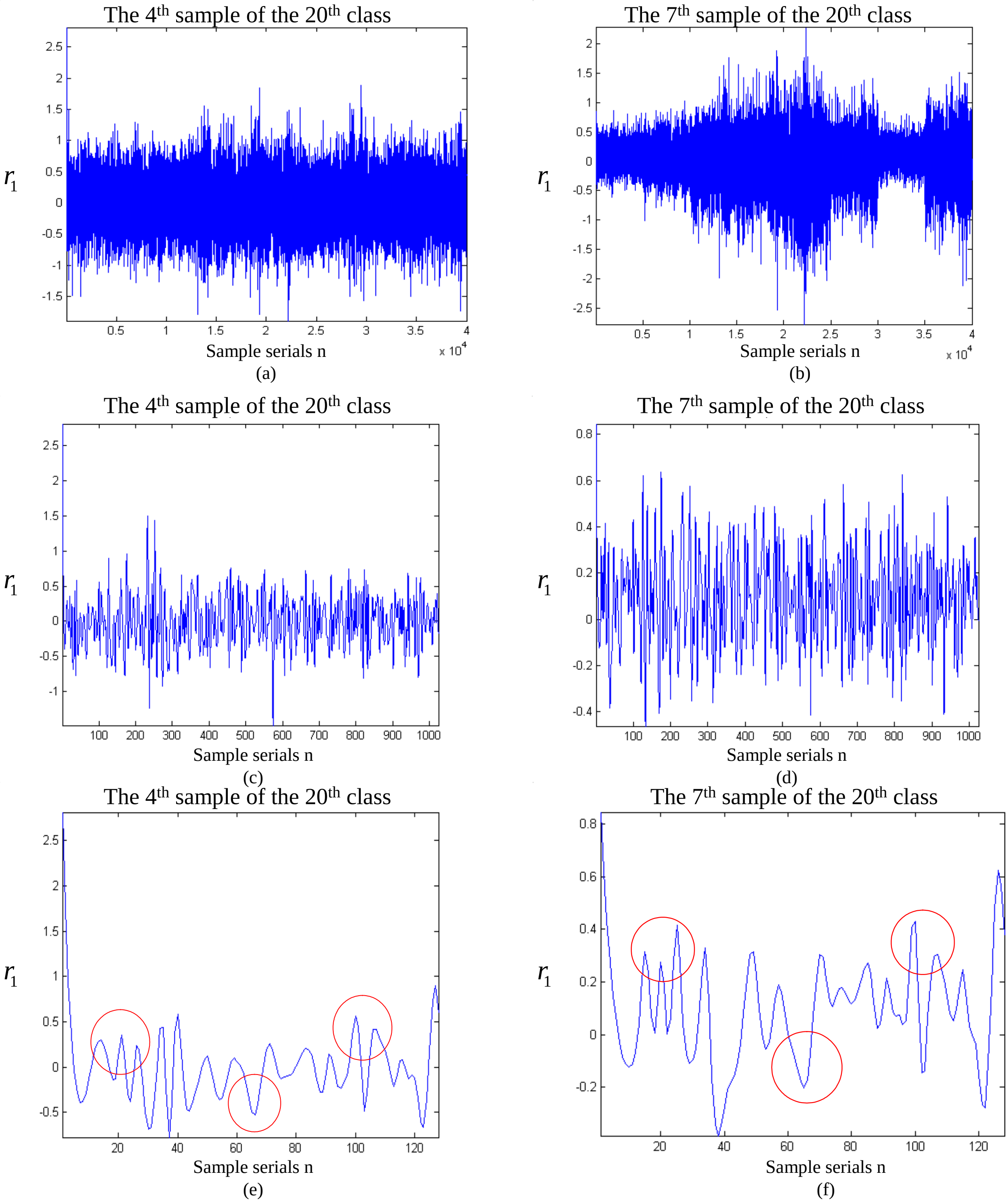


Fig.7.EMD decomposition results of two samples from the same class, (a) and (b) are two sets of feature sequences with their original length, (c) and (d) show the first 1024 components of the original features, (e) and (f) show the first 128 components of the original features.

performance is not good. Hence, the MED metric is not suitable for measuring the similarity of the EMD feature. The other metrics demonstrate promising performance in the recognition results.

VI. CONCLUSIONS

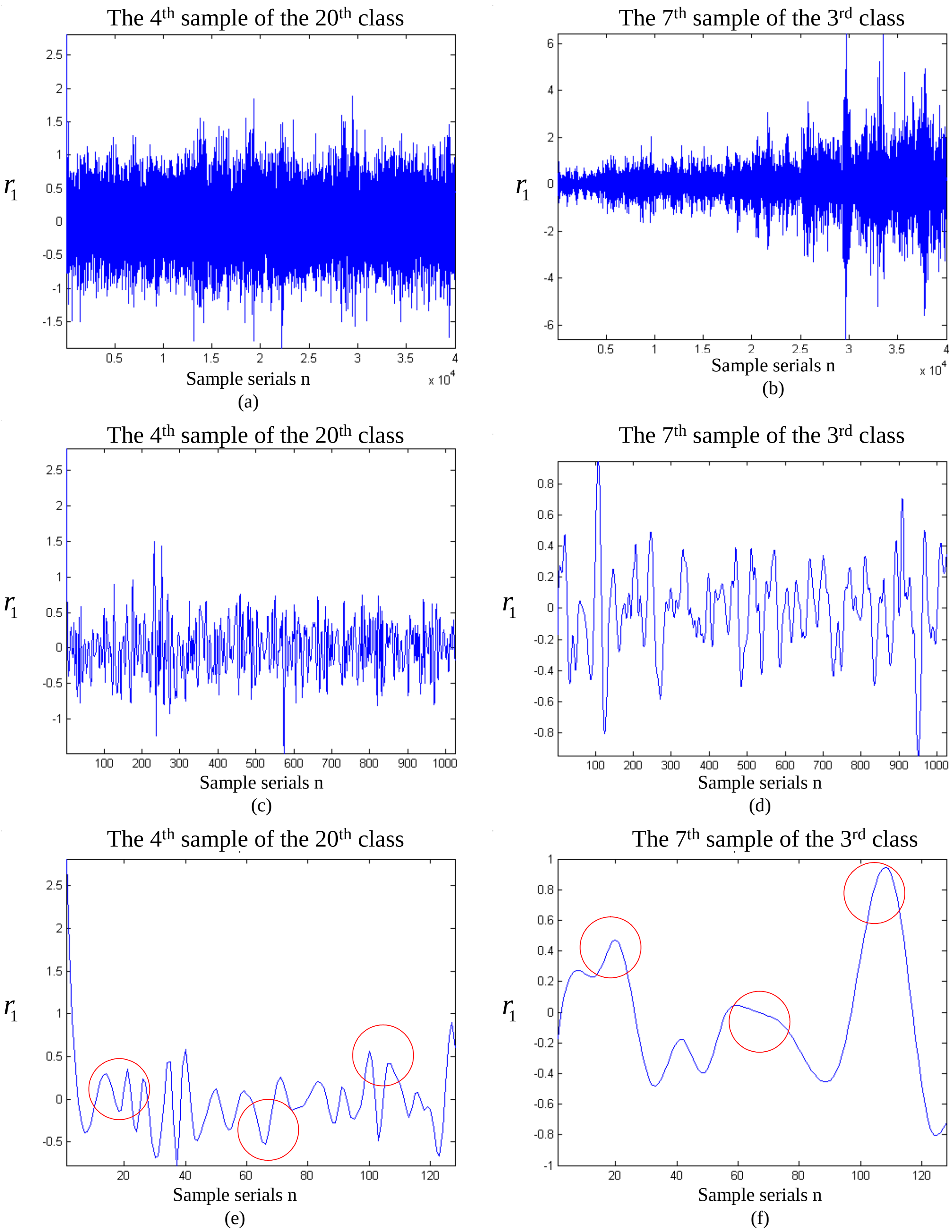


Fig.8.EMD decomposition results of two samples from two classes, (a) and (b) are two sets of feature sequences with their original length, (c) and (d) show the first 1024 components of the original features, (e) and (f) show the first 128 components of the original features.

In this paper, an effective method for acoustic signals recognition is presented, which operates the Empirical Mode Decomposition (EMD) technique. The performance of signals recognition achieved by the EMD approach associated with three different similarity measures has been evaluated. Experimental results have shown eminent performance. The cosine similarity measure and the correlation metric have achieved similar performance. Therefore, the proposed method has demonstrated to be promising for acoustic signals recognition and EMD is suitable for feature extraction. In the future, we will conduct more experiments on intraclass and interclass comparison to evaluate the proposed algorithm. We are also working at increasing the database in order to further verify the performance.

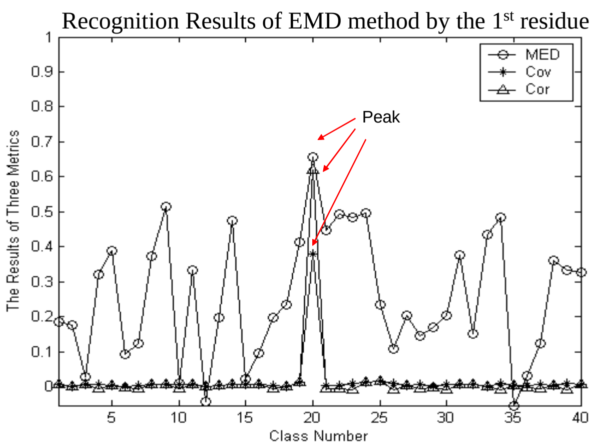


Fig.9. Experimental results using the first residue of EMD as the feature and three similarity metrics.

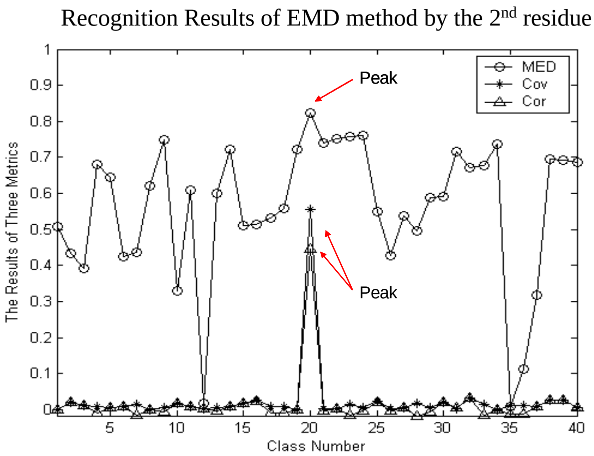


Fig.10. Experimental results using the second residue of EMD as the feature and three similarity metrics.

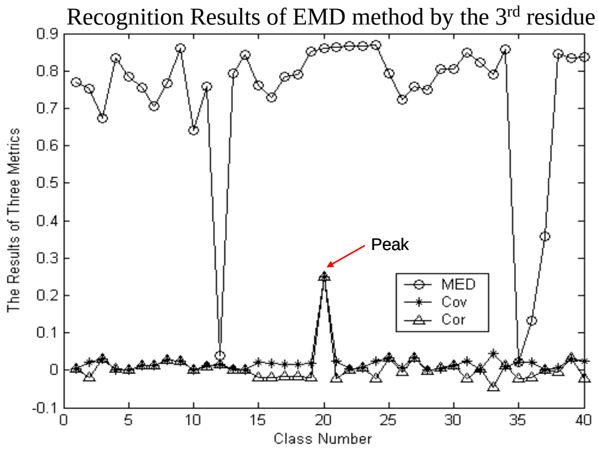


Fig.11. Experimental results using the third residue of EMD as the feature and three similarity metrics.

REFERENCES

[1] Urick, R. J., “Principles of Underwater Sound,” McGraw-Hill, 1983.

[2] DeBilly, M., “Determination of the resonance spectrum of elastic bodies via the use of short pulses and Fourier transform theory,” Journal of the Acoustical Society of America, Vol. 79, No. 2, pp.219-221, Feb, 1986.

[3] Cohen, L., “Time-frequency distribution—a review,” Proceedings of the IEEE, Vol. 77, No. 7, pp.941-981, July, 1989.

[4] Rao, P., and Taylor, F. J., “Detection and localization of narrow-band transient signals using the Wigner distribution,” Journal of the Acoustical Society of America, Vol. 90, pp. 1423-1434, 1991.

[5] Boashash, B., and O’Shea, P., “A methodology for detection and classification of some underwater acoustic signals using time-frequency analysis techniques,” IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. 38, pp. 1829-1841,

1990.

[6] Jones, D. L., and Parks, T. W., "A resolution comparison of several time-frequency representations," IEEE Transactions on Signal Processing, Vol. 40, pp. 413-420, 1992.

[7] Donoho, D. L., and Johnstone, I. M., "Ideal spatial adaption by wavelet shrinkage," Biometrika, Vol. 81, pp. 425-455, 1994.

[8] Donoho, D. L., "Adapting to unknown smoothness via wavelet shrinkage," Journal of the American Statistical Association, Vol. 90, pp.1200-1224, 1995.

[9] Nason, G. P., and Silverman, B. W., "The discrete wavelet transform in S," Journal of Computational and Graphical Statistics, Vol. 3, pp. 163-191, 1994.

[10] Johnstone, I. M., and Silverman, B. W., "Wavelet threshold estimator for data with correlated noise," Journal of the Royal Statistical Society, Series B, Vol. 59, pp.319-351, 1997.

[11] Huang, N., Shen, Z., Long, S., Wu, M., Shih, H., Zheng, Q., Yen, N., Tung, C., and Liu, H., "The Empirical Mode Decomposition and Hilbert Spectrum for Nonlinear and Non-stationary Time Series Analysis,"

[13] Flandrin, P., Rilling, G., and Goncalves, P., "Empirical Mode Decomposition as a Filter Bank," IEEE Signal Processing Letters, Vol. 11, No. 2, pp.112-114 , 2004.

Proceedings of the Royal Society of London, Vol. 454, pp. 903–995, 1998.

[12] Bishop, C. M., "Neural Networks for Pattern Recognition," Oxford University Press, 1996.