ICSV 21
13-17 July, 2014, Beijing/China

The 21st International Congress on Sound and Vibration

13-17 July, 2014, Beijing/China

UNDERWATER SOUND CLASSIFICATION BASED ON GAMMATONE FILTER BANK AND HILBERT-HUANG TRANSFORM

Xiangyang Zeng, Shuguang Wang

School of Marine Science and Technology, Northwestern Polytechnical University, 710072 Xi'an, China

e-mail: zenggxy@nwpu.edu.cn

The variable acoustic environment makes it harder for the application of underwater sound recognition system. However, human auditory system has remarkable ability on dealing with complex acoustic conditions. A robust underwater noise target classification system is expected if this ability can be simulated. Aimed at this purpose, a robust underwater sound classification algorithm which employs Gammatone filter bank and Hilbert-Huang transform is studied in this paper. Gammatone filter bank is used for the simulation of nonlinear dividing of human ears. Then the wavelet denoising procedure is applied on the divided sub-bands. At last, Hilbert-Huang transform is used as the time-frequency analysis tool for the feature extraction. With the help of Hilbert-Huang transform, instantaneous features are extracted, and then used for the built of feature vector. Experimental results indicated the expected efficiency of the proposed algorithm.
Key words: Gammatone filter bank; Hilbert-Huang transform; Time-frequency analysis.

1. Introduction

Underwater sound detection and recognition systems have applications on various fields, such as commercial usage [1-2], scientific research [3-4] and national defence [5-6]. Attentions have been attracted on this study and numerous methods have been proposed and improved in the past decades [7-9]. Using various time-frequency representations, time-frequency analysis can study a signal in both the time and frequency domains simultaneously. This characteristic makes it suitable for the processing of underwater sounds [10-11].

Due to the complex underwater conditions, inevitable noises will appear in the application and then decrease the accuracy of the recognition system. Thus, the built of noise robust recognition system has become the key solution.

Knowing as "cocktail party effect", human auditory has obvious advantages on robust signal processing. If this remarkable ability can be introduced and combined in the existing method, a robust underwater sound recognition system can be expected [12-13].

Aimed at this purpose, a robust time-frequency analysis algorithm which combines Gammatone filter bank and Hilbert-Huang transform (HHT) has been proposed. Gammatone filter bank is used for the simulation of auditory perception. On which the denoising processing is applied subsequently for the enhancement of the signal-to-noise ratio (SNR). HHT is used as the analysis tool for the feature extraction. Support Vector Machine (SVM) is employed as the classifier for the classification under different SNR conditions.

2. Methodology

The diagram of the presented algorithm is showing in Figure 1. Using Gammatone filter bank, input signal is divided into different sub-bands, which simulates the nonlinear characteristic of auditory perception system. Wavelet threshold denoising is applied on each sub-band. Denoised signal can be obtained by combining each sub-band's results. By using HHT as the analysis tool, instantaneous amplitudes and frequencies are extracted and then used for the building of feature vector. SVM is employed as the classifier for the classification experiments.

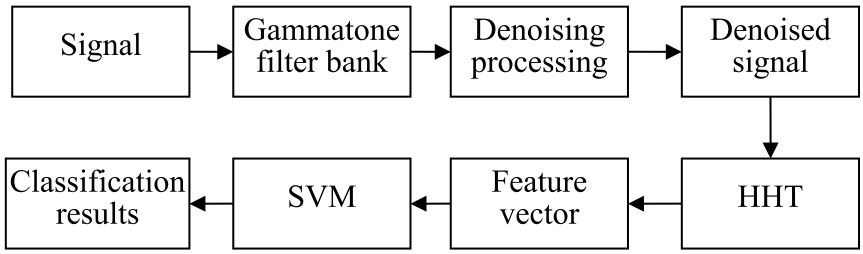

Figure 1. Diagram of the presented algorithm.

2.1 Gammatone filter bank

Gammatone filter bank [14] is a cochlear filtering model which decomposes an input signal into the time–frequency domain using a set of Gammatone filters. The impulse response of i-th Gammatone filter centred at frequency f_i is mathematically expressed as

$$g_i(t) = t^{n-1}\exp(-2\pi b_i t)\cos(2\pi f_i + \varphi_i) \qquad 1 \le i \le N \tag{1}$$

where N is the number of filter, n is filter order, φ_i is the phase of the i-th filter. b_i defines the duration of the impulse response and thus the bandwidth.

For filter orders of $n = 3,...,5$, the Gammatone filter is reported to give a good approximation of the human auditory filter. Thus, 4th order Gammatone filters are used and implemented as infinite impulse response filters in our algorithm [15-17]. As the phase of acoustic signal has little effect on hearing, φ_i was equal to 0 in this paper.

The band of each filter is related to human ear's critical band, which can be measured by equivalent rectangular bandwidth in auditory psychology. The relation is showed below.

$$ERB(f) = 24.7 * (4.37 f / 1000 + 1) \tag{2}$$

where f is frequency in Hz. Each subband filter can be defined as

$$b_i = 1.019 ERB(f_i) \tag{3}$$

2.2 Wavelet threshold denoising

Wavelet threshold de-noising is based on the multi-resolution analysis of wavelet transform. After wavelet transform, analyzed data can be divided into approximation sequence and detail sequence at each resolution. The variances and amplitudes of the details of noise at the various levels decrease regularly as the level increases. On the other hand, the amplitudes and variances of wavelet transform of the available signal are not related to the change of scale [18-19]. According to this property, by selecting an appropriate threshold, the noise can be eliminated. Then denoised signal can be obtained by inverse wavelet transform.

Soft threshold is used as the thresholding method. Its equation is showed below

$$\hat{x} = \begin{cases} \mathrm{sgn}(x) * |x - th| & |x| \geq th \\ 0 & |x| < th \end{cases} \tag{4}$$

where x is the input signal, $\hat{x}$ stands for the thresholding result, and th represents the threshold.

2.3 Hilbert-Huang transform

HHT contains Empirical Mode Decomposition (EMD) and Hilbert transform. EMD is a pre-processing step to Hilbert transform. It adopts a nonlinear procedure to decompose an input signal into several intrinsic mode functions (IMF). Each IMF satisfies two conditions: 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.

For an input signal $x(t)$, finding its local maxima then the upper envelope $x_U(t)$ can be constructed by cubic spline fitting. So is the same to its lower envelope $x_L(t)$. The EMD process is to find the mean function $m(t)$ of $x_U(t)$ and $x_L(t)$. An IMF candidate, denoted by $h(t)$ can be acquired via $x(t) - m(t)$. If $h(t)$ meets the two criteria of IMF, the first IMF is confirmed $c_1(t) = h(t)$. Otherwise, the residual $r(t) = x(t) - h(t)$ is used as the new input and same processing is applied until the first IMF is found.

After the extraction of first IMF, the residual component can be represented as

$$r_1(t) = x(t) - c_1(t) \tag{5}$$

$r_1(t)$ is considered as a new signal and subjected to the same extracting process as described above

$$r_n(t) = x(t) - \sum_{i=1}^{n} c_i(t) \tag{6}$$

where $r_n(t)$ is the residue of the signal and $c_i(t)$ stands for the *i-th* IMF.

Finally, $x(t)$ can be expressed as

$$x(t) = \sum_{i=1}^{n} c_i(t) + r_n(t) \tag{7}$$

For any of the IMF $c_i(t)$, the corresponding $\hat{c}_i(t)$ is computed by Hilbert transform. Then an analytic signal $z_i(t)$ is composed

$$z_i(t) = c_i(t) + j\hat{c}_i(t) = a_i(t)e^{j\theta(t)} \tag{8}$$

$$a_i(t) = \sqrt{c_i(t)^2 + \hat{c}_i(t)^2} \tag{9}$$

where $a_i(t)$ is its instantaneous amplitude. Then the phase is readily obtained and the instantaneous frequency $f(t)$ can be defined by the derivative of the phase

$$\theta_i(t) = \arctan\frac{\hat{c}_i(t)}{c_i(t)} \tag{10}$$

$$f_i(t) = \frac{1}{2\pi}\frac{d\theta_i(t)}{dt} \tag{11}$$

3. Experiments

Four types of underwater noise signals (A, B, C and D) which were recorded under real conditions are applied to the classification experiments. Its sampling rate is 8000 Hz. Each type of signals contains 15 different individuals. Each individual has 48000 sample points, which lasts for about 6 seconds. Fig.2 shows the first 4 seconds waveform of the random selected individuals from each type of signals.

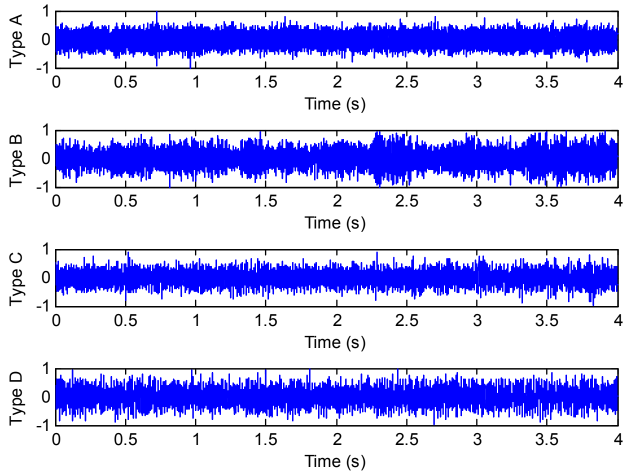

Figure 2. Waveform of four types signals.

The filter numbers of the designed Gammatone filter bank is 32. Figure 3 shows the spectrum of the designed Gammatone filter bank.

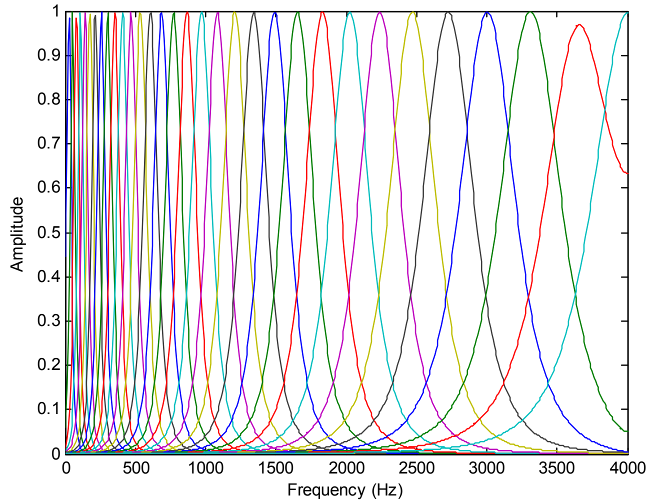

Figure 3. Spectrum of the designed Gammatone filter bank.

Using Gammatone filter bank, the input signals are decomposed into 32 different sub-bands. Then 2-level wavelet denoising is applied. Daubechies wavelet 'db8' is selected as the wavelet function. The adaptive threshold *th* in equation (4) is calculated via the principle of stein's unbiased risk estimate [19-20]. Finally, denoised signal can be obtained through the combination of denoised sub-band signals.

After denoising, EMD is applied to the denoised signal for the extraction of IMFs. The resulting IMF numbers for each signal varies from 11 to 13. Based on our former research, first 8 IMFs can obtain at least 90% of the energy of the denoised signal [21]. Thus, the first 8 IMFs are selected and used for the feature extraction. Weighted mean instantaneous frequency is extracted as the feature:

$$wMIF = \frac{\sum_t f_i(t) a_i^2(t)}{\sum_t a_i^2(t)} \tag{12}$$

SVM is used as the classifier. Leave-one-out cross-validation (LOOCV) method is used in the recognition experiments. A single individual from each type signals is selected as the testing data, and the remaining observations are used as the training data. This process repeated until all the observations are tested. Then the results are averaged as the recognition results.

White Gaussian noise is added to the original signal for the simulation of different SNRs. For convenient comparison, a former proposed algorithm which contains bark-wavelet and 3-level wavelet denoising is used as the baseline system [21]. Table 1 shows the recognition results of the classification experiments.

Table 1. Classification results under different SNRs.

SNRs	Proposed algorithm					Baseline system				
	Recognition rate (%)				Averaged rate (%)	Recognition rate (%)				Averaged rate (%)
	A	B	C	D		A	B	C	D	
original	100.0	100.0	86.7	100.0	96.7	100.0	100.0	93.3	100.0	98.3
25	100.0	100.0	86.7	100.0	96.7	100.0	100.0	93.3	80.0	93.3
20	100.0	93.3	86.7	100.0	95.0	100.0	93.3	93.3	93.3	95.0
15	93.3	100.0	93.3	100.0	96.7	93.3	93.3	86.7	100.0	93.3
10	100.0	100.0	93.3	93.3	96.7	93.3	53.3	93.3	100.0	85.0
5	100.0	100.0	86.7	93.3	95.0	100.0	6.7	86.7	93.3	71.7

The two algorithms get similar recognition results when the SNRs are higher than 10 dB. After SNR drops to 10 dB or lower, the proposed algorithm get higher recognition rates than the that of the baseline system. Furthermore, the recognition rates on each signal types of the proposed algorithm are more stable to the SNR changes.

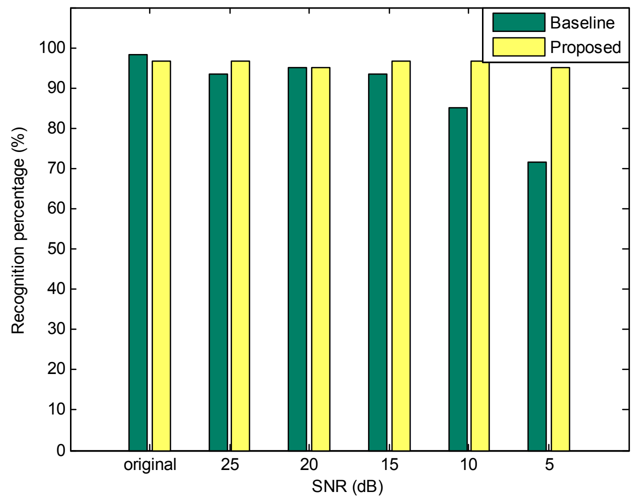

Figure 4. Recognition results of four type signals.

For a direct understanding of the performances between the two algorithms, figure 4 gives the bar plot of the final recognition rates. From the figure it can be seen that the proposed algorithm maintains a final recognition rate of above 95% through different SNR conditions. This verifies the robustness and efficiency of the proposed algorithm we expected.

4. Conclusion

Combined Gammatone filter bank, wavelet denoising and Hilbert-Huang transform, a robust underwater sound classification algorithm has been proposed in this paper. Gammatone filter bank is used for the simulation of auditory perception, based on which the wavelet denoising is applied. Hilbert-Huang transform is used as the time-frequency analysis tool on the denoised signal. Feature vectors are then built with the extracted instantaneous parameters. The performance of the proposed algorithm has been tested by the classification experiments under different SNR conditions. Experimental results verify the robust performance of the proposed algorithm.

REFERENCES

1 Langkau, M. C., Balk, H., Schmidt, M. B. and Borcherding, J. Can acoustic shadows identify fish species? A novel application of imaging sonar data, *Fisheries Management and Ecology*, **19**(4), 313–322, (2012).

2 Korneliussen, R. J., Heggelund, Y., Eliassen, I. K. and Johansen, G. O. Acoustic species identification of schooling fish, *ICES Journal of Marine Science*, **66**(6), 1111–1118, (2009).

3 Adam, O. Advantages of the Hilbert Huang transform for marine mammals signals analysis, *Journal of Acoustical Society of America*, **120**(5), 2965–2973, (2006).

4 Miller, B. and Dawson, S. A large-aperture low-cost hydrophone array for tracking whales from small boats, *Journal of Acoustical Society of America*, **126**(5), 2248–2256, (2009).

5 Tucker, J. D. and Azimi-Sadjadi, M. R. Coherence-based underwater target detection from multiple disparate sonar platforms, *IEEE Journal of Oceanic Engineering*, **36**(1), 37–51, (2011).

6 Stack, J. Automation for underwater mine recognition: current trends and future strategy, *Proceedings of SPIE 8017, Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XVI*, Orlando, Florida, USA, April 25, (T2011T).

7 Azimi-Sadjadi, M. R., Yao, D., Huang, Q. and Dobeck, G. J. Underwater target classification using wavelet packet and neural networks. *IEEE Transactions on Neural Networks,* **11**(3), 784–794, (2000).

8 Li, D., Azimi-Sadjadi, M. R. and Robinson, M. Comparison of different classification algorithms for underwater target discrimination, *IEEE Transactions on Neural Networks*, **15**(1), 189–194, (2004).

9 Zhang, Y., Baggeroer, A. B. and Bellingham, J. G. Spectral-feature classification of oceanographic processes using an autonomous underwater vehicle. *IEEE Journal of Oceanic Engineering*, **26**(4), 726–741, (2001).

10 Peng, Z. K. and Chu, F. L. Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography. *Mechanical systems and signal processing*, **18**(2), 199–221, (2004).

11 Huang, N. E., Shen, Z. Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N. C., Tung, C. C., and Liu, H. H. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences,* **454**(1971), 903–995, (1998).

12 Young, V. W. and Hines, P. C. Perception-based automatic classification of impulsive-source active sonar echoes. *Journal of Acoustical Society of America,* **122**(3), 1502–1517, (2007).

13 Chen, K., Liang, Y., Du, L., Chen, J., Wu, Y. and Wang, H. Two kinds of timbre representations and its application into acoustic target classification. *Journal of Acoustical Society of America,* **131**(4), 3313–3313, (2012).

14 Hu, G. and Wang, D. L. Monaural speech segregation based on pitch tracking and amplitude modulation. *IEEE Transactions on Neural Networks*, **15**(5),1135–1150, (2004).

15 Lopez-Poveda, E. A. and Meddis, R. A human nonlinear cochlear filterbank. *Journal of Acoustical Society of America*, **110**(6), 3107–3118, (2001).

[16] Schluter, R., Zolnay, A. and Ney, H. Feature Combination using Linear Discriminant Analysis and its Pitfalls. *Proceedings of the International Conference on Spoken Language Processing (ICSLP/Interspeech)*, Pittsburgh, PA, September, (2006).

[17] Slaney, M. An efficient implementation of the Patterson-Holdsworth auditory filter bank. Technical Report, Apple Computer, Perception Group, (1993).

[18] Mallat, S. and Hwang, W. L. Singularity detection and processing with wavelets. *IEEE Transactions on Information Theory*, **38**(2), 617–643, (1992).

[19] Rosas-Orea, M. C. E., Hernandez-Diaz, M., Alarcon-Aquino, V. and Guerrero-Ojeda, L. G. A comparative simulation study of wavelet based denoising algorithms. *Proceedings of 15th IEEE International Conference on.Electronics, Communications and Computers*, TPuebla, Mexico,T 28–02 Feb., (2005).

[20] Donoho, L. Denoising by Soft Thresholding. *IEEE Trans. on Information Theory*, 41(3), 965–1001, (1995).

[21] Wang, SG. and Zeng, XY. Robust underwater noise targets classification using auditory inspired time–frequency analysis. *Applied Acoustics*, 78, 68–76, (2014).