Journal of Computer Science and Applications.
ISSN 2231-1270 Volume 5, Number 1 (2013), pp. 39-50

http://www.irphouse.com

A New Wavelet Improvement Method for Wild Bird Species Identification

[1]Rong Sun, Yihenew Wondie Marye and [*] Hua-An Zhao

[1, First Author], 102d9308@st.kumamoto-u.ac.jp
*[*Corresponding Author] cho@cs.kuamoto-u.ac.jp*
yihenew@st.cs.kumamoto-u.ac.jp
Kumamoto University, Japan

Abstract

Syllable segmentation, feature extraction and classification are three important aspects of an automatic bird song recognition system. To improve the recognition rate, in this study, a new method of signal denoising based on wavelet transform for nocturnal wild bird species identification has been developed and researched; and nocturnal wild bird species identification by sound information using wavelet is investigated. As a method, firstly, frequency conversion of a bird call is performed. Then, the mean value that expresses the strength of the frequency ingredient is obtained; further, the modulation spectrum expresses the frequency of the birdcall. Feature quantities are then input into the neural network to simplify classification.

Keywords: Bird song recognition, Wavelet transform, Denoising, Thresholding

1. Introduction

Recently, environmental problems are becoming more and more serious, such problems include earthquakes, storms, desertification, El Nino phenomenon, global warming and so on. These environmental influences are making survival of living things on earth difficult. New scientific environmental protection and evaluation mechanisms should be in place to mitigate such influences.

Generally, birds have the following characteristics: they are sensitive to the environment; they are high-level consumers in the ecosystem and they are easy to observe, with the result that the kinds and numbers of different birds become good indices in the study of species diversity. So, the investigation of the diversity of bird species is a key in monitoring the environment and ecosystem recovery. In practice, the investigation of bird species requires long-term monitoring and recording. The complex nature and the wide area covered by an ecosystem, however, raise the difficulty of the study because of manpower, material resources, and time restrictions. So, the automatic recognition of bird vocalization has become an invaluable technique for the long-term investigation of bird species.

The vocalization of bird species includes bird song and bird call. Bird song, being complicated, varied, agreeable and pleasant to listen to, is usually generated by a male bird and is used to declare his turf or attract a mate. Bird call, on the other hand, is monotonous, brief, repeated, fixed and sexless and is used to contact or alert companions. Bird calls are usually short and acoustically simple while a birdsong is longer and is composed of a succession of musical notes [1].

Many studies have been done on human voice recognition, and some of them have been applied to birdsong recognition. These methods include the linear prediction coefficients, the dynamic time warping (DTW) algorithm, the hidden Markov model (HMM), the idea of maximum power (energy) frequency, the usage of wavelet packet decomposition, and the application of neural network classifier [2]. However, specific sound features are required for bird recognition because of the vocalization diversity of birds. Birdsongs are typically divided into four hierarchical levels: note, syllable, phrase, and song, of which syllable plays an important role in bird species recognition.

The wild birds which are active at night are very difficult to be watched, and it is very dangerous for the investigators. So, using sound as the method of investigation is the most convenient and the safest. It is necessary to set the system for the investigation of the nocturnal birds [3].

Currently, technology for sound-based identification of birds would be a significant addition to the research methodology in ornithology, and biology in general. There is also significant commercial potential for such systems because bird watching is a popular hobby in many countries. Extensive international programs are also boosting the activity in the area of bioacoustics signal processing and pattern recognition. Sounds have been widely used in the recognition of bird and insect species, and the estimation of acoustical environment health [4].

Speech denoising technology is an important branch of speech signal processing. It plays an important role in removing the noise pollution; improve voice quality and so on. The wavelet transform is a multifunction signal analysis method, and useful tool for the analysis of non-stationary signals. It can analyze both the overview and details of the signal.

This paper is organized as follows. Section 2 describes the identification system, wavelet transform, wavelet thresholding and denoising, method of identification and some recognition results in detail. The conclusion and acknowledgement are shown in section 3 and section 4 respectively.

In addition, while using wavelet approach [4], the number of bird species that could be identified is only eight, and it uses only information of the height of the song after wavelet analysis. In this paper, 13 species can be identified. And it uses not only the height part of bird song, but also the modulation spectrum.

2. Identification System

The block diagram of the proposed system, shown in Figure 1, entails preprocessing, feature extraction and recognition (species decision) as described in detail in the following.

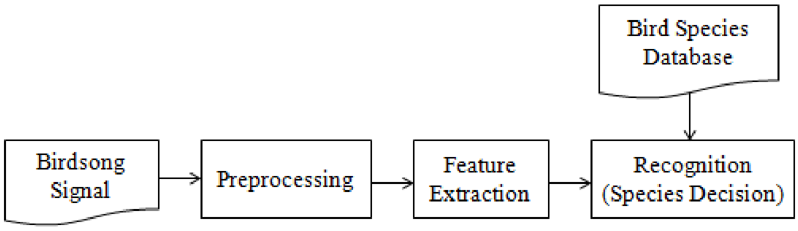

Figure 1. Block diagram of the proposed system

Preprocessing filters the signal and properly segments the syllables for feature extraction. Both time-domain and frequency-domain information are used in the proposed approach.

2.1. Wavelet Transform

In this process, the time resolution doesn't change even when the frequency is raised since the width of the window function is constant in the Fourier transform for a short time. The time width is shortened in the wavelet transform at a high frequency, and on the other hand, local frequency information is obtained because the base that expands the time width is used and efficient time- frequency analysis can be done at low frequency values [4].

A wavelet is a wave-like oscillation with amplitude that starts out at zero, increases, and then decreases back to zero. The “-let" refers to its smallness as it is decaying while the "wave" refers to its volatility, its alternating amplitude positive and negative shock forms. Compared with the Fourier transform, wavelet transform is the time -space (frequency- localization) analysis in which signal through the telescopic translation operations (functions) gets a step by step multi-scale refinement; the final breakdown of time to reach from high frequencies to low frequencies, the frequency that can automatically adapted to the requirements of time-frequency signal analysis which can be focused to any details of the signal’s Fourier transform to solve difficult problems. This is called the basic wavelet or mother wavelet, and is expressed as $\psi(t)$.

The wavelet function uses two real number parameters *a* and *b*,

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right). \tag{1}$$

The wavelet function is basic wavelet $\psi(t)$ which is scaled in the horizontal direction by *a* fold, and translated *b* units to the right [5][6]. The factor $1/\sqrt{a}$ is a normalizing factor to keep the energy constant. So a is called the scale, and is $1/a$ corresponds to the frequency which shows the width of the time of wavelet. The inner product of an arbitrary signal $f(x)$ and the wavelet function $\psi_{a,b}(x)$ as

$$\left\langle f(x), \psi_{a,b}(x) \right\rangle = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(x) \psi^{*}\left(\frac{x-b}{a}\right) dx \tag{2}$$

is defined as wavelet transform or continuous wavelet transform where $\psi^{*}(\cdot)$ represents operation of complex conjugate of $\psi(\cdot)$.

2.2. Wavelet Thresholding and Denoising

2.2.1. Theory

The wavelet transform has a concentrating ability that makes signal energies in wavelet domain to be expressed in a few coefficients. These coefficients of wavelet transform must be greater than whole domain transform as energy is dispersed in wide spectrum in small wave coefficients in the latter case. So it means that using transform coefficients to cut signals, and applying threshold processing method to remove noise as is shown in Figure 2.

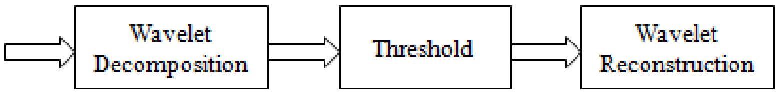

Figure 2. Wavelet threshold denoising principle

2.2.2. Wavelet threshold

The data acquired during non-destructive testing of actual structures in service usually contains noise. To ensure a better detection of damage without losing the property of the original data, the noise needs to be reduced or, if possible, totally removed. There have been several investigations on additive noise suppression in signals using wavelet transforms, such as the work of Donoho and Johnstone [5] and Donoho [6], which is based on thresholding the discrete wavelet transform of signals. The method of Donoho and Johnstone thresholds the wavelet coefficients to zero if their values are below a threshold. This is called hard threshold method. An alternative to such hard thresholding is soft threshold method.

A signal *s(t)* from noisy data *x(t)* with a Gaussian white noise *n(t)* is

$$x(t) = s(t) + n(t). \tag{3}$$

Then use the discrete wavelet transform of *x(t)*,

$$\omega_x(j,k) = \omega_s(j,k) + \omega_n(j,k),\ j = 0,1,2,\cdots,J,\ k = 0,1,2,\cdots,N. \tag{4}$$

Where $\omega_x(j,k), \omega_s(j,k), \omega_n(j,k)$ is the j-layer's wavelet coefficient; j is the maximum decomposition layer of wavelet transform. N is the length of the signal *s(t)*.

In here, soft threshold and hard threshold method is proposed.
The soft threshold function is

$$\hat{w}_{j,k} = \begin{cases} \operatorname{sgn}(w_{j,k})(|w_{j,k}| - \lambda) & if\,|w_{j,k}| \ge \lambda \\ 0 & if\,|w_{j,k}| < \lambda \end{cases}. \tag{5}$$

The hard threshold function is

$$\hat{w}_{j,k} = \begin{cases} w_{j,k} & if\,|w_{j,k}| \ge \lambda \\ 0 & if\,|w_{j,k}| < \lambda \end{cases}, \tag{6}$$

where $\hat{w}_{j,k}$ represents the wavelet coefficients,

$$\lambda = \sigma\sqrt{2\log(N)}, \tag{7}$$

where σ is variance of noise, N is length of signal [7] [8].

Figure 3 and Figure 4 show hard and soft the threshold method respectively. Usually, the hard threshold method preserves the sharp features of original signal, but the smoothness is worse. The soft threshold has better visual smoothness.

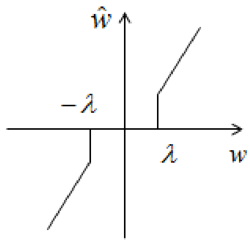

Figure 3. Hard threshold method

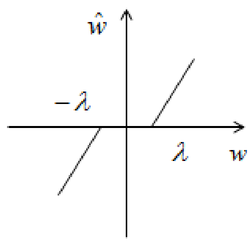

Figure 4. Soft threshold method

2.3. Denoising using Matlab

Matlab is a very powerful toolbox to compute a lots of signal processing, especially using the wavelet toolbox, the threshold method has been programmed which can thus be chosen at will. For instance, there are some samples of bird song denoising using soft and hard threshold method. In this paper, for the recorded data of 13 species of wild birds, it shows that using the hard threshold is better than using soft threshold. Fig. 5 is the sample for the White's Thrush, with bit rate of 705 kbps, for duration of 20 seconds. The picture with red shows the original signal, the black part of the picture shows the signal after denoising [3].

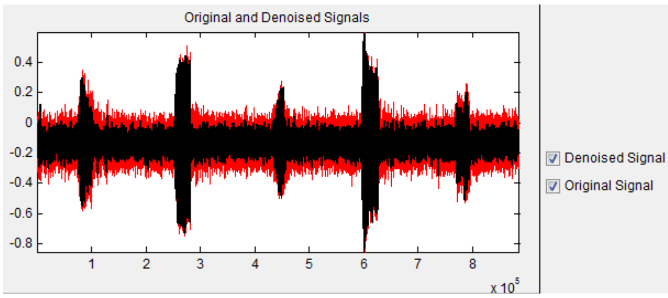

Figure 5. (a) Hard threshold of White's Thrush

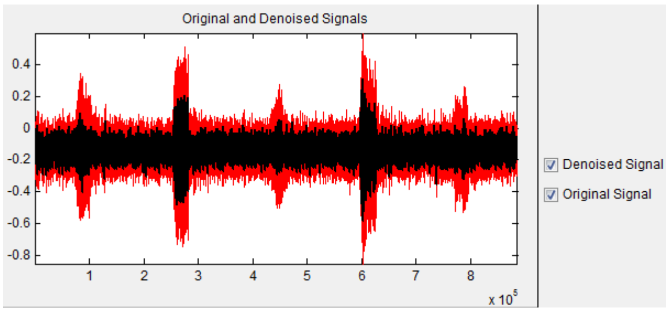

Figure 5. (b) Soft threshold of White's Thrush

From this example, we can find that by using hard threshold method, the noise is obviously reduced, and the signal is mostly reserved. Using the soft threshold method, not only the noise is reduced, but also the signal as well which would represent a very bad recognition for the wild bird. Retaining the original signal is an important aspect of this research. Lots of experiments show that hard threshold method is better than soft threshold method in this particular case. And it has been used for the identification of various wild birdsongs.

2.4. Speech Waveform

Wild bird's sound is saved as WAV format. WAV format is Microsoft developed sound file format. Also known as wave sound file, it is the first digital audio format in Windows platform and its applications have been widely supported. WAV format supports many compression algorithms and multiple audio bits. It uses a sampling frequency of 44.1 kHz with 16-bit quantization, which means WAV and CD sound quality is almost the same.

2.5. Relation between scale and frequency

Continuous wavelet transform (CWT) is an implementation of the wavelet transform using arbitrary scales and almost arbitrary wavelets. The wavelets used are not orthogonal and the data obtained by this transform are highly correlated. For the discrete time series we can use this transform as well, with the limitation that the smallest wavelet translations must be equal to the data sampling.

In Figure 6, four cycles are assumed to be 512*pt*, and the sound data of one second is 44100pt for scale 1. The frequency can be requested with this relation. When the scale is *a*, the points can be calculated to be 512 times *a*.

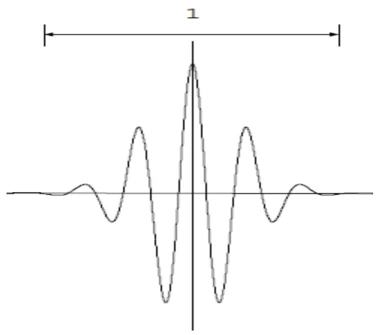

Figure 6. Basic wavelet

The scale used in the present study is assumed to be 11 pieces. The relation between the scale and the center frequency is shown in Table 1. The sampling frequency is 44.1 kHz. When it happens to be 48 kHz, to keep the center frequency unchanged, the scale becomes 48/44.1 times.

Table 1 Scale and center frequency

No.	**Scale**	**Center frequency [Hz]**
1	2	172.25
2	1.414	243
3	1	345
4	0.707	487
5	0.5	690
6	0.354	974
7	0.25	1380
8	0.177	1948
9	0.125	2760
10	0.088	3896
11	0.0625	5520

Using the scale of Table 1, the relation of the sine wave with different frequency to each scale can be shown in Figure 7.

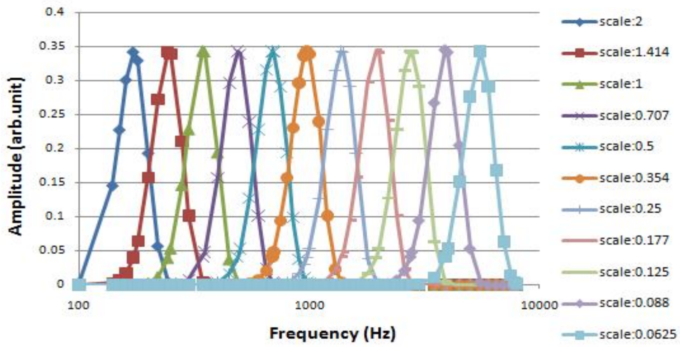

Figure 7. Log plot of amplitude versus frequency

Based on the result, a frequency of bird sound of 8 kHz is enough. The values of the wavelet coefficients do not get small. In order to cover all the frequency of 8 kHz, we choose the different scales. To show the changes clearly as the scale varies, Figure 7 shows the log plot of amplitude against frequency.

2.6. Mean value of wavelet coefficient

To determine the frequency characteristic of bird's sound, the mean value of wavelet coefficient is calculated. The procedure to determine the mean value of wavelet coefficient is as follows.

① Prepare the bird's sound data for one second.

② Set the scale to 1.

③ Compute the wavelet coefficient.

④ Change the scale to $2, \sqrt{2}, 1, 1/\sqrt{2}, 1/2, 1/2\sqrt{2}, 1/4, 1/4\sqrt{2}, 1/8, 1/8\sqrt{2}, 1/16$

⑤ For each scale, compute the mean value of wavelet coefficients.

⑥ For all the 11 scales, compute the mean values.

⑦ Normalize the value such that maximum is 1 and minimum is 0.

2.7. Method of identification

Neural network is used as the method of identification. Neural network is well used to simulate, research, develop a wide array of adaptive systems. It is suitable for processing a lot of data because it is possible to identify easily while studying at the same time. For this reason, this section will explain the neural network used for wild bird identification [9][10].

For identification, the neural network of three layer structure shown in Figure 8 was used. It was studied by 145 teaching data using the back propagation. At first, the sound was Fourier transformed. Next, the result was divided into 11 bands, and average is calculated for each band. Then 20 values of modulation spectra were used to assess the time change of bandwidth power by using the mean values of the 11 bands. So, total number of input data for the neural network is 31. The number of output data is 13, the same as the number of species to be identified.

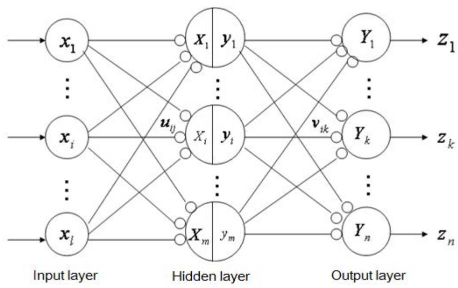

Figure 8. 3-layer neural network

2.8. Recognition condition

There are 13 species of forest nocturnal wild birds for identification. Table 2 shows the name of nocturnal wild bird used in the experiment. Sound data is taken from "The Songs & Calls of 333 Birds in Japan", "283 Wild Birdsongs of Japan", and "98 Wild Birdsongs of Japan", along with local recorded data.

Table 2. The name of nocturnal wild bird and number.

No.	Wild bird Name
1	Brown Hawk Owl
2	Brown Thrush
3	Collared Scops Owl
4	Oriental Scops Owl
5	White's Thrush
6	Long-eared Owl
7	Ural Owl
8	Little Cuckoo
9	Siberian Ground Thrush
10	Japanese Night Heron
11	Woodcock
12	Horsfield's Hawk Cuckoo
13	Jungle Nightjar

The identification procedure by the neural network is explained as follows.
Input layer: 1; Hidden layer: 1; Output layer: 1
Input units: 31 (mean value 11, modulation spectra 20)
Hidden units: 20, 25, 30, 35
Output units: 13 (wild bird species)
Teaching data: 145

Learning times: 20000, 30000
Recognition rate of the wild bird is calculated using the following equation.

$$Recognition\ Rate = \frac{Identified\ number\ of\ wild\ bird\ sound}{Input\ number\ of\ wild\ bird\ sound} \times 100\%$$

The wild birds and the number of each species used in this experiment are shown in Table 3.

Table 3. Wild bird species and number of teach data (number in () is the number of individuals)

Brown hawk owl	14(6)	Brown thrush	18(6)	Collared scops owl	9(4)
Scops owl	11(4)	White’s thrush	12(4)	Long-eared owl	8(4)
Ural Owl	9(4)	Little cuckoo	13(3)	Siberian thrush	12(4)
Japanese night heron	11(2)	Woodcock	4(2)	Horsfield's hawk cuckoo	14(4)
Jungle Nightjar	10(4)				

2.9. Recognition results

Figure 9 shows the results of signals using hard threshold method. The recognition rate has improved a little bit after using hard threshold method. Especially the No. 8 and No. 9 have better result of using hard threshold method after than before.

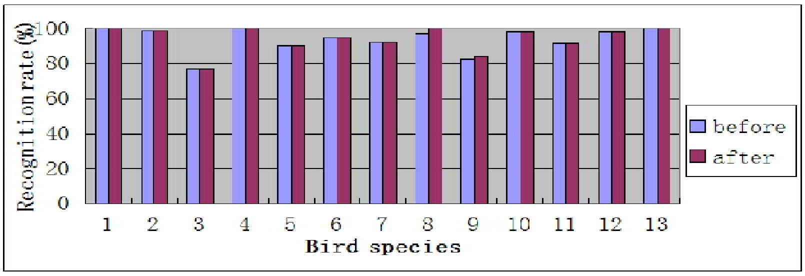

Figure 9. Recognition rate

Figure 10 shows recognition rate for different learning times. Higher recognition rate is obtained when the learning time is 20000 than when it is 30000. Figure 11 shows that when the number of the hidden units changes, different recognition rate is obtained. The highest recognition rate shows up when the number of the hidden units is 20.

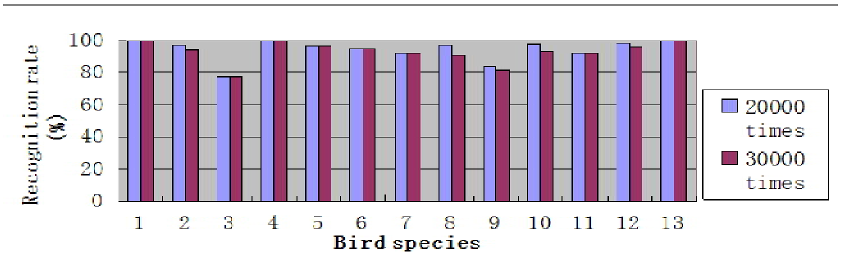

Figure 10. Recognition rate with learning time of 20000 and 30000

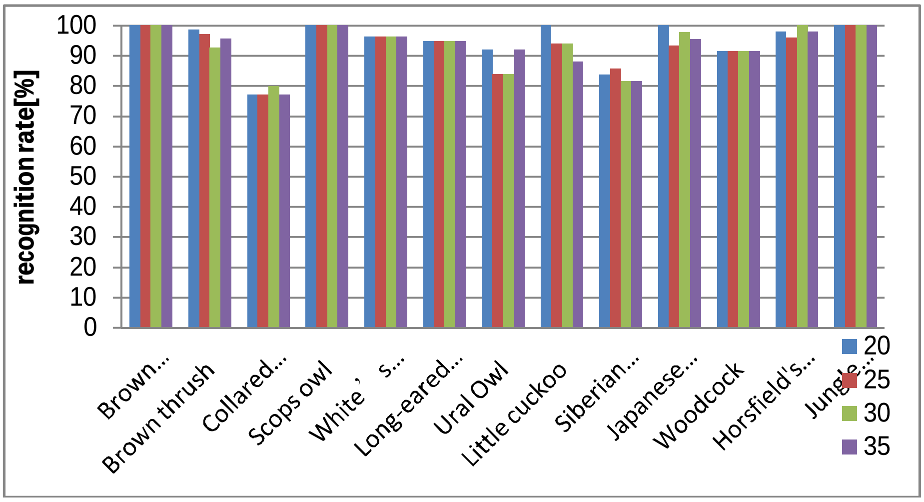

Figure 11. Recognition rate with hidden units of 20, 25, 30, 35

3. Conclusions

In this research, thirteen species of wild bird sounds were identified. In the approach, we used wavelet transform, computed the wavelet coefficient and their mean values, extracted the characteristics of birds' sounds, and put the different characteristics to the neural network for identification.

For the evaluation of this system, various data including the voices of forest nocturnal birds actually recorded during night time as outdoor experiment were used. Identification rate of 100% was achieved for at least two wild birds while the average identification rate for most of the experiments is above 90% before using hard threshold. The highest recognition rate was obtained when the learning times are 20000 and the number of hidden units is 20.

The recognition result after using hard threshold is a little bit better than without it. For example, for the Long-eared Owl, the recognition result without hard threshold is 0.976 while it is 0.981 after.

The average identification rate has shown increase of 1 to 2%. Although the increase achieved is very small, it shows this method can be used as an available alternative for other approaches.

For the future work, in order to get better recognition rate, we will add more teaching data base and record more sound data in various areas.

4. Acknowledgments

The authors would like to thank Prof. Nagahisa MITA for his advice, support and encouragement that made this research possible.

Also, the authors would like to express our gratitude to Dr. Mutsuyuki Ueta for collecting songs of wild birds.

5. References

[1] Hojatollah Yeganeh, Seyed Mohammad Ahadi, Ali Ziaei, “A New MFCC Improvement Method for Robust ASR,” Signal Processing, pp.643-646, ICSP2008.

[2] H. Kim and H. Melhem, “Fourier and Wavelet Analyses for Fatigue Assessment of Concrete Beams,” Journal of Experimental Mechanics, vol.43, Issue 2, pp. 131-140, 2003.

[3] Rong Sun, Nagahisa Mita, “Nocturnal Wild Bird Species Identification By sound information Using Wavelet,” Proc. of Wavelet Active Media Technology and Information Processing (ICWAMTIP), pp.1-3, 2012.

[4] A. Selin, J. Turunen and J.T. Tanttu, “Wavelets in recognition of bird sounds,” EURASIP Journal on Advances in Signal Processing, vol. 2007, Article ID 51806, 9 pages.

[5] Donoho, D.L. and Johnstone, LM., "Ideal Spatial Adaptation by Wavelet Shrinkage," Oxford Journals Biometrika, vol. 81, Issue 3, pp. 425-455, 1994.

[6] Donoho, D.L., "De-Noising by Soft -thresholding,", Information Theory, IEEE Transaction on vol. 41, Issue 3, pp. 613-627, 1995.

[7] MT Johnson, X Yuan, Y Ren, “Speech signal enhancement through adaptive wavelet thresholding,” Speech Communication, vol. 49, Issue 2, pp. 123-133, 2007.

[8] Xiang Jin, Zhai chengrui, Yang wei, Meng lingjun, Zhang wendong, “Sound signal denoising based on wavelet transform,” Journal on Control and Automation Publication Group in Embedded System and SOC, vol. 35, pp. 85-86, 2007.

[9] Alex L. Mcllraith and Card C. Howard, “Birdsong Recognition with DSP and Neural Networks,” Proc. of the IEEE Communications, Power, and Computing conference, vol. 2, pp. 409-414, May 1995.

[10] A.L. McIlraith and H.C. Card, “Bird song identification using artificial neural networks and statistical analysis,” Proc. of the Canadian Conference on Electrical and Computer Engineering, vol. 1, pp. 63-66, 1997.