EXTRACTING ELEMENTS FROM FIELD-BIRD SOUNDS USING SONOGRAM BASED ON WAVELET TRANSFORM

Fang Ren[1,2], Aiwu Zhang[1,2], Shumin Wang[1,2], YihaoDuan[1,2]
1. Capital Normal University, 105, Xi San Huan Bei Lu, Beijing, 100037, China;
2. Key Lab of 3D Information Acquisition and Application of Education Ministry, 105, Xi San Huan Bei Lu, Beijing, 100037, China

KEYWORDS：Field-bird Sound, Filtering Denoising, Features Extraction, Segmentation

ABSTRACT: Ecological environment has many great changes with the intensive human activities. Birds are sensitive to changes of environment, so recently, more and more people start to research bird sounds in order to monitoring ecological environment and geological hazard. Hierarchical levels of each bird sound are phrase, syllable and element. Element is the smallest separable unit, syllables are produced by one or more elements, and series of syllables in particular pattern is a phrase. Here, we choose the element as the best unit for bird population identification. But bird sound is different with human voice, which has a wide range of frequencies from 100Hz to 10000Hz, furthermore, the field-bird sound has always a very complex natural background including wind, water, vehicles and so on, which lead to low signal noise ratio. Therefore, extracting elements from field-bird sounds have many challenges, the element extraction algorithms should be sensitive to noise and range of frequencies. We present a new method of field-bird sound element extraction using sonogram based on wavelet transform. Firstly, we use wavelet decomposition to denoise original field-bird sounds and compare with wiener filter method. Secondly, we extract short term energy and zero cross rate from waveform and frequency band of bird sounds from sonogram as the feature vectors to extract elements automatically. The results show that the proposed algorithm greatly improves the accuracy of bird sound automatically segmentation.

1. INTRODUCTION

Ecological environment has many great changes with the intensive human activities. In order to evaluate this impact of human activities, we need fundamental information on the degree of changes in the living environment. Birds are a good indicator for changes in biodiversity as they are very sensitive to changes of living environment and very easy to detect in comparison to other animals. So recently, more and more people start to research bird sounds in order to monitoring ecological environment and ecological hazard. Bird sound is different with human voice, which has a wide range of frequencies from 100Hz to 10000Hz, and nearly all birds will make different kinds of sounds in communication with different birds, and all sounds have some meaning, for example, flight, alarm, excitement and so on. In order to provide a basis for monitoring the changes of living environment, data sets on the different sounds of birds are of great importance. Meanwhile, human ear, eye and brain constitute a very effective voice recognition system. For human ears they are easy to notice even subtle different in sounds, and for human eyes they are easy to recognition different substances, whereas human decision is always subjective, so, the automatically process of bird sounds would be an important new tool for bird research.

For the case of bird sounds, the usual approaches to identify the species of bird depend on manpower and experience often ornithologist, which is subject. Therefore, the computer signal processing technology will be an important tool for song research. Since the 1970s, many researchers analyzed bird sounds using the filter, acoustic spectrometer and other instruments. In recent years, they are replaced by a variety of sonogram analysis software, such as Wavesurfer, Parat(BECKERS 2003)，Batsound(GENNINGS 2004)，Syrinx(MENNILL 2000), etc. In the sonogram of birds, hierarchical levels are phrase, syllable and element. Element is the smallest separable unit and continuous curve, syllables are fixed together by one or more elements and repeated in a phrase, series of syllables in particular pattern is a phrase, between the phrases is usually separated by a blank pause. Here, we choose the element as the best unit for bird population identification(see figure 1). In this paper, we choose the field-bird who has always a very complex natural background including wind, water, vehicles and so on, that lead to low signal noise ratio. Therefore, extracting elements from field-bird sounds have many challenges. The aim of our study is to develop a new method for field-bird sound element extraction using sonogram based on wavelet transform in real-world. The model of the whole extraction process is presented in figure 2. Firstly, we use wavelet decomposition to denoise original field-bird sounds and compare with wiener filter method. Secondly, we extract short term energy and zero cross rate from waveform and frequency range of bird sounds from sonogram as the feature vectors to extract elements automatically. The results show that the proposed algorithm greatly improves the accuracy of bird sound automatically segmentation.

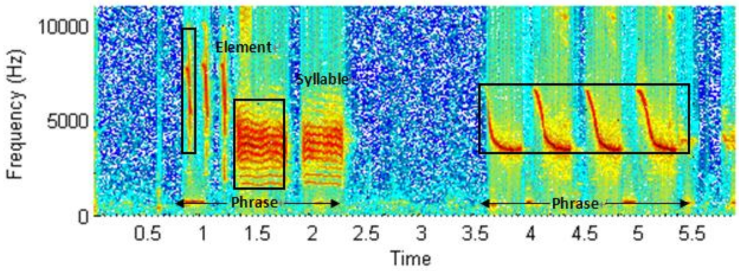


Figure 1: Hierarchical structure of tit (after wavelet denosing)

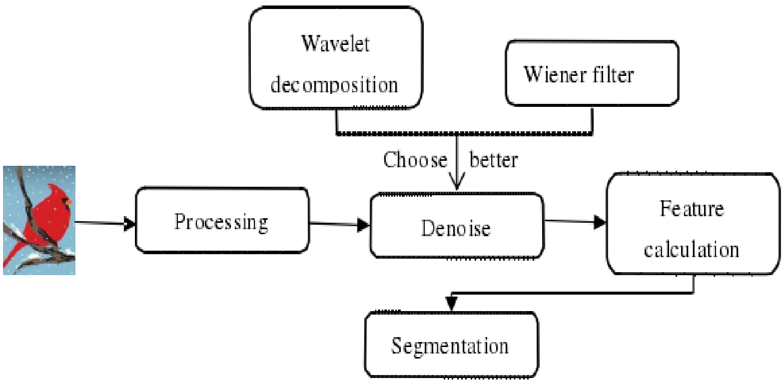


Figure 2: The extraction process

2. WAVELET DECOMPOSITION AND DENOISE

The aim of our study is to develop a new method for field-bird sound element extraction in real-world. Nevertheless, field-bird sound is more complicated and different with the song collected in door due to the complexity of the wild environment, so it's inevitably to collect a variety of noise, such as wind, water, vehicles, mechanical and so on. According to the signal's quality which is directly related to the accuracy of song's feature parameters, before the segmentation, it's very necessary to denoise to make songs as clean as possible.

The wavelet analysis has gained a great deal of attention in the field of digital signal processing (Rioul1991). It's a very efficient tool to deal with the temporal-frequency signal. For example, compared with the short-time Fourier transform, it can find out both temporal and frequency information, and to analyze signals which contain discontinuities and sharp spikes(Aria Selin 2007). In the low frequency has higher frequency resolution and lower time resolution in contrast to the high frequency (Le yang 2008), these properties are appropriate for bird songs. In the wavelet transform the original signal is converted into wavelet coefficients. In this paper, the song signal is mainly divided into the following two parts: high frequency signal and low frequency signal(see figure 3). According to the two parts, we choose different threshold methods to remove the environment noise and compare with the wiener filter method. There are several wavelet families that have proved to be particularly(Mathworks 2004). The Daubechies wavelet family was selected because in it both scaling and wavelet function are compactly supported and they are orthogonal. The figure 4 shows the denoise results of wavelet and wiener filter.

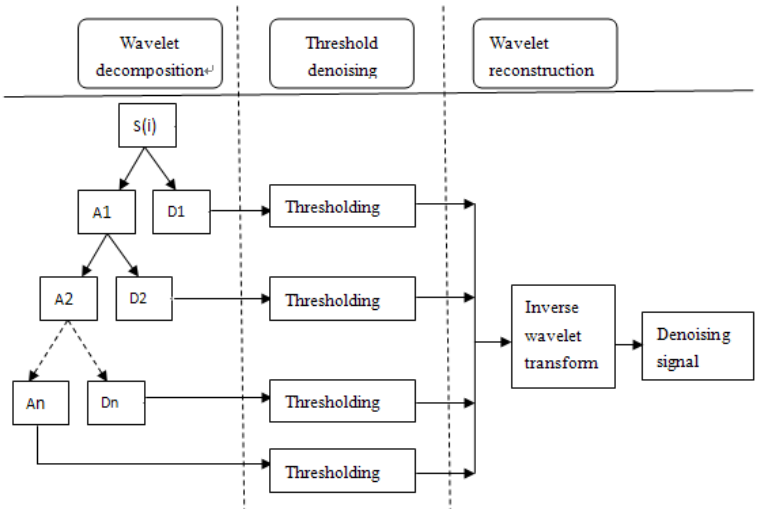
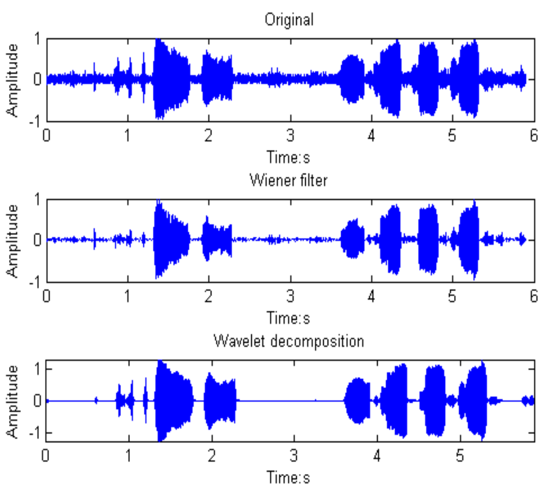


Figure 3: The flow of noise reduction



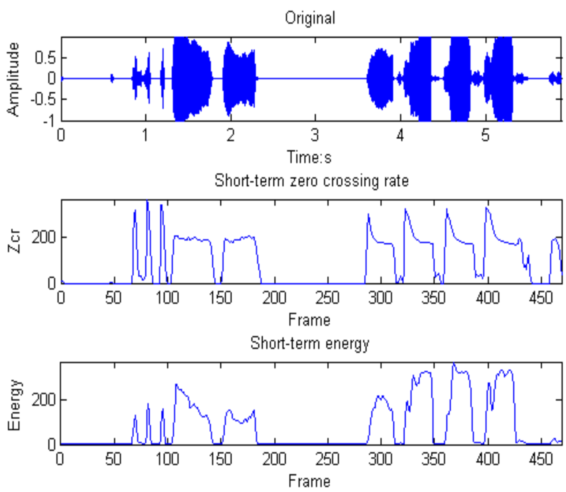


Figure 4: The results of wavelet and wiener filter
The figure shows that wavelet is much better than wiener filter. It can efficiently reduce the noise and keep the useful signal simultaneously.

Figure 5: Time-domain characteristics
Top: the tit signal after wavelet denosing.
Middle: short-term zero crossing rate.
Bottom: short-term energy.

3. EXTRACTING FEATURES

Time-domain characteristics contain short-term energy and zero crossing rate. The main difference of efficient signal and inefficient signal (noise) is always expressed as the energy distribution; obviously, the energy of song is higher than noise. From the short-term energy graph, it does not only describe the time changing characteristic between elements of bird song, but also changing of energy and difference. The signal of each frame is denoted as

$x_n(m)$:

$$x_n(m) = w(m)x(n+m)\ , \qquad 0 \le m \le n-1$$
$$w(m) = \begin{cases} \frac{1}{2n},\ m = 0\sim n-1 \\ 0,\ m = else \end{cases} \tag{1}$$

The short-term energy is defined as:

$$E_n = \sum_{m=0}^{n-1} {x_n}^2(m) \cdot h(n-m) \tag{2}$$

Where n is the frame length, h(n-m) is the window function. In this paper, we choose n as 25ms, frame shift as half of n as the suitable distance to avoid too much amplitude change.

Short-term zero-crossing rate refers to the change number of algebraic sign when the frame signal cross zero-level axis. It's an efficient and simple method for signal analysis without multiplication, division and square root. The more useful information we extract, the greater the effect of real-time signal monitoring. The figure 5 shows the two time-domain characteristics: short-term energy and zero crossing rate.

The signal of each frame is denoted as $x_n(m)$, the short-term zero crossing rate is defined as:

$$Z_n = \frac{1}{2}\sum_{m=0}^{n-1} |sgn[x_n(m)] - sgn[x_n(m-1)]| \tag{3}$$

Where sgn(x) is the sign function: $sgn(x) = \begin{cases} 1,\ x \ge 0 \\ -1,\ x < 0 \end{cases}$

After time-domain characteristic extraction, we also need to obtain the distinct frequency bands, covering most of the frequencies of bird, due to the reference value of frequency band in set the threshold of elements segmentation. In this case, we only need a band located index, defining the bands minimum and maximum frequencies. These variables are set according to power spectral density (PSD). Figure 6 below shows the main frequency band of tit is between 3000-6500Hz.

4. ELEMENTS SEGMENTATION

In the sonogram of birds, hierarchical levels are phrase, syllable and element. Element is the smallest separable unit and continuous curve, and, we choose the element as the best unit for bird population identification, because element segmentation can provide the template data for follow-up song identified. So, we through the computer digital process technology to find the element's start point, end point and construction. According to the unique characteristics of bird song, we choose the short-term energy and zero crossing rate to make double insurance. First, divide the whole song into two parts: mute and voice, the end of mute is the start of voice. Then set a high threshold and a low threshold respectively for short-term energy and zero crossing rate, denoted by EH, EL, ZH, ZL before testing. If one of the short-term energy and zero crossing rate exceed the high threshold in the mute, then it's considered to enter the voice. And if both of the short-term energy and zero crossing rate reduce to the low threshold, then it's the end. Furthermore, during this step, the most important thing is to set the detective threshold which will directly affect the endpoint detection results. In the past years, there is no one method general and reliability but experience. Figure 7 shows the result of segmentation, where red line is the star, green line is the end.

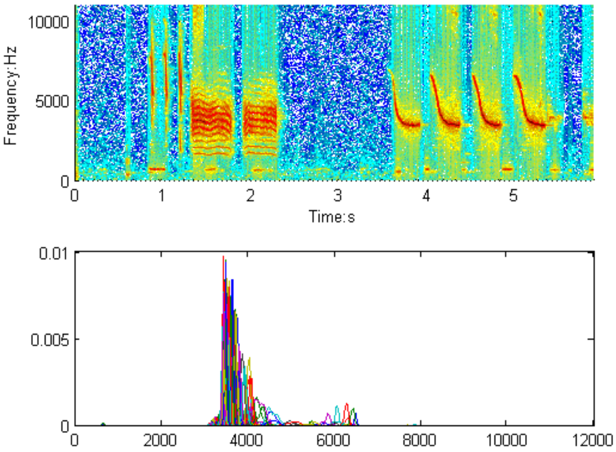


Figure 6: Power spectral density

Top: the sonogram of tit signal after wavelet.

Bottom: the power spectral density of tit which Shows the main frequency band (3000-6500Hz).

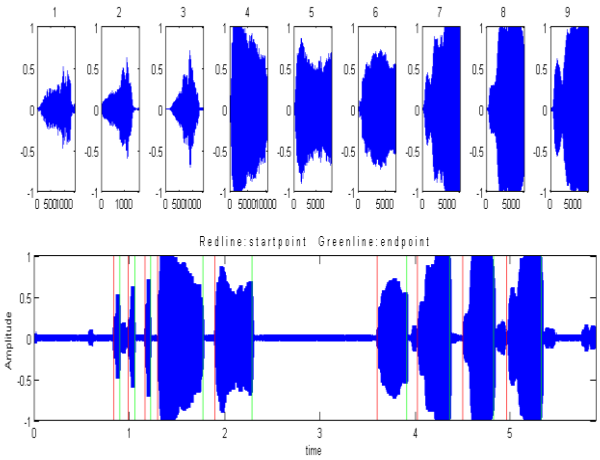


Figure 7: The result of segmentation

Top: lists of segmentation signal.

Bottom: red line is the star point, green line is the end point.

5. CONCLUSION

Element is the smallest separable unit and continuous curve, we choose the element as the best unit for bird population identification. Nevertheless, field-bird sound is more complicated and different with the song collected in door due to the complexity of the wild environment. According to the different characteristics between bird song and noise, we developed a new method for field-bird sound element extraction in real-world. First, we use wavelet decomposition to denoise original field-bird sounds and compare with wiener filter method. Then, through the computer digital process technology to find the element's start point, end point and construction. The results show that the proposed algorithm greatly improves the accuracy of bird sound automatically segmentation.

REFERENCES:

[1] Aria Selin, Jari Turunen, Juha T. Tanttu. 2007. Wavelets in recognition of bird sounds. Journal on Applied Signal Processing. pp. 141:150.

[2] BECKERS G J L, SUTHERS R A, CATE C T. 2003. Mechanisms of frequency and amplitude modulation in ring dove song. J ExpBio, pp. 1833-1843.

[3] GENNINGS N V, PARSONS S, BARLOW K E, eta. 2004. Echolocation calls and wing morphology of bats from the West Indies. Acta Chiropterologica, 6(1): 75-90.

[4] Le Yang, Li Xie. 2008. The research and analysis of song singal based on wavelet translate. Software Guide. vol.7, No.4. pp. 89-90.

[5] MENNILL D J, RATCLIFFE L M. 2000. A field test of syrinx sound analysis software in interactive playback. Bioacoustics, 11:77-86.

[6] O. Rioul and M. Vetter li, 1991. Wavelets and signal processing. IEEE Signal Processing Magazine, vol.8, no.4, pp.14–38.

[7] Wavelet Toolbox User's Guide. 2004. Mathworks. Inc.