A new segmentation method of bird vocalisations recorded in the distance

Development of recognition software using multi-CPU

○ Michihiro JINNAI（Nagoya Women's University）

1　Introduction

The introduction of the ornithological spectrograph in the 1950s changed the way birdsong was measured, quantified, and interpreted by biologists. The spectrograph was correlated with objective knowledge and quickly superseded subjective aural assessment. Over the past 15 years, we have used advanced LPC (Linear Predictive Coefficient) spectrography to develop and update automated detection and recognition software for bird vocalisations [1][2].

The same system has been diversified and is used successfully to detect sounds associated with anomalous equipment behavior and machinery failures such as steam leaks. Another iteration is used to detect abnormalities in pre-stressed concrete structures. A technician taps the structures with a hammer and the software quickly identifies off-sounds associated with structural abnormalities. Over the past 20 years, we have advanced the automatic recognition and detection system for abnormal anthropogenic sounds [3][4]. The same core recognition software is used in both the bioacoustic and anthropogenic applications.

Procedurally the detection and recognition software: (1) detects and segments a waveform of potential signals automatically from up to a three-hour continuous recording, (2) extracts the spectrum patterns from the segmented waveforms, and (3) matches the detected spectrum patterns in the segmented waveforms with the exemplar signal using a similarity scale. To expedite the process, the software executes parallel processing using multi-CPU. Using a bird vocalisation (*Manorina melanocephala*, Noisy Miner), we demonstrate a new segmentation method for bioacoustic signals recorded at variable distances in noisy environmental backgrounds. This method provides increased detection and extraction accuracy where amplitude varies from call to call within a bout, such as when a bird is calling in flight; or where, due to shifts in head or body orientation, call axis and hence amplitude varies within a call or during a call bout.

2　The spectrogram of bird sounds

The upper diagram of Fig. 1 shows a spectrogram (time-frequency-power) extracted from the vocalisation of the Noisy Miner Chur call (Paul G. McDonald; University of New England, Australia 2012, Cooperative bird differentiates between the calls of different individuals, even when vocalisations were from completely unfamiliar individuals. *Biology Letters* 8: 365-368). In Fig. 1, the waveform has been segmented with 18.8 msec frame width and 0.31 msec frame period, and the LPC spectrum has been

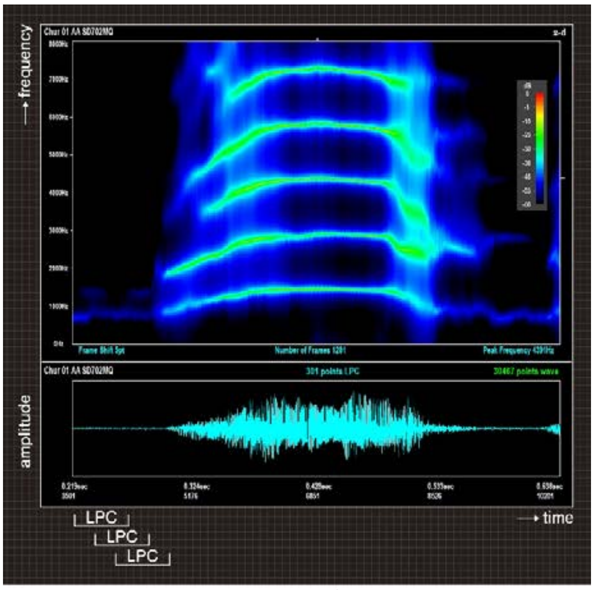

Fig. 1　Spectrogram of bird vocalisation

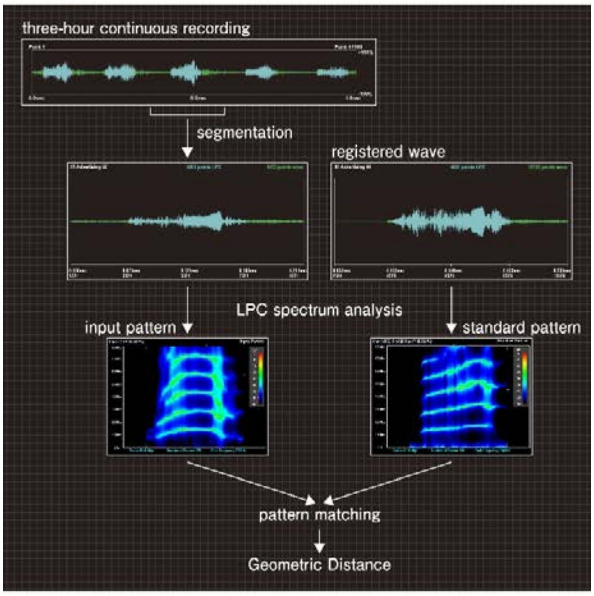

Fig. 2　Processing procedure in recognition system

calculated in each frame. Next, the spectrogram has been coloured according to logarithmic power of the LPC spectrum. We have set the analysis conditions of the bird vocalisation with a 16 kHz sampling frequency, 16 bit quantization, 12 order LPC, 0 Hz to 8000 Hz frequency range, 11.5 Hz frequency resolution, and 0 dB to -60 dB logarithmic power spectrum. If we analyze transient signals such as bird vocalisations, we then need to set the short frame width as shown at the bottom of Fig. 1. The LPC spectrum analysis is suitable for such transient signals.

3 Automatic recognition procedure

Fig. 2 shows the software procedural stages for automated signal detection and recognition. First, the software differentiates potential signals from background noise (segmentation). Second, the software extracts the segmented signals and establishes the segmented signal's spectral characteristics (time-frequency-power) using LPC spectral analysis. Third, the software compares the spectral characteristics of the extracted signal (the input pattern) with a previously registered standard pattern of the focal signal (the signal to be automatically detected). Comparison is effected using a new similarity scale called the GD (Geometric Distance [5][6]). Sections 4 and 5 describe the automatic segmentation method of the bird vocalisation.

4 Segmentation

The amplitude of received signals in field recordings of bird vocalisations can be highly variable. The method presented here, periodically normalizes the otherwise variable energy curve, providing advanced signal discrimination. This compensates for amplitude shifts and yet remains sensitive to low power signals. Fig. 3 shows a four-syllable noisy miner 'chur' call in which amplitude varies from syllable to syllable. This is a common scenario where calls are produced during flight and the receiver is stationary.

Eq. (1) and Fig. 3 show the method for calculating an energy curve E_k. We suppose that w_i (i =1, 2, … , N) is the amplitude of a continuous recording waveform at a receiver, where N is the number of the recorded sample points and M is the range of energy calculation. In Fig. 3, the calcu-

$$E_k = \sum_{i=k-M}^{k+M} w_i^2 \qquad (k=1+M, 2+M, \cdots, N-M) \tag{1}$$

$$E_k = E_{k-1} - w_{k-M-1}^2 + w_{k+M}^2 \tag{2}$$

$$(k=2+M, 3+M, \cdots, N-M) \qquad E_{1+M} = \sum_{i=1}^{1+2M} w_i^2$$

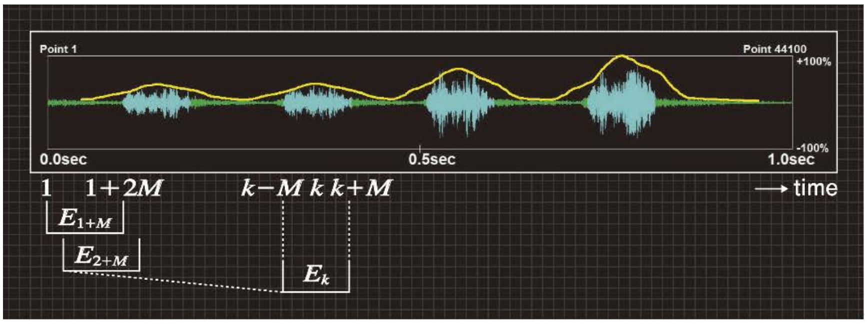

Fig. 3 Calculation of energy curve

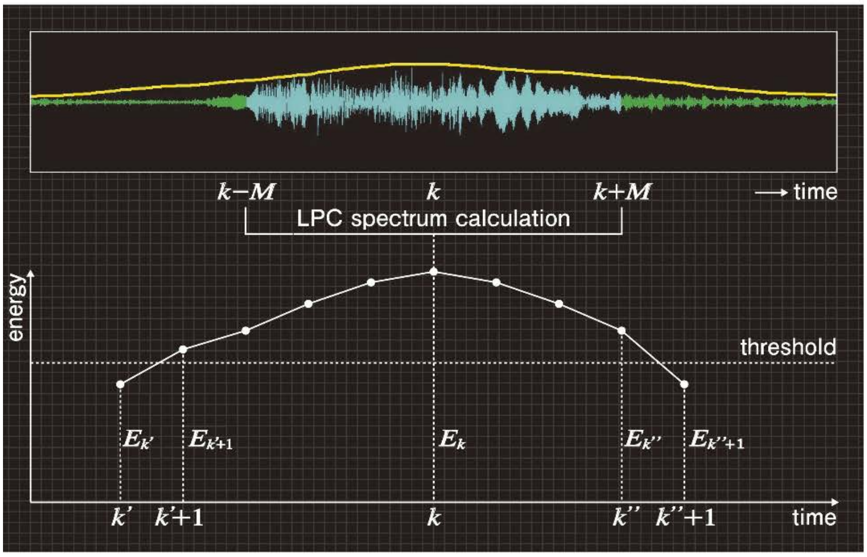

Fig. 4 Comparison of energy curve with threshold

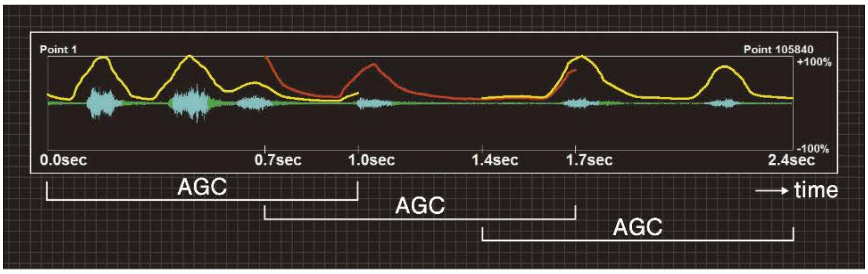

Fig. 5 Auto gain control of energy curve

lated energy curve is shown by the yellow line. In the actual software, the energy curve E_k can be calculated by the recurrence formula Eq. (2) using long integers E_k and $w_i{}^2$ in order to reduce the processing overhead.

Fig. 4 shows the method for segmenting the bird vocalization from a three-hour continuous recording using the energy curve E_k. As shown in Fig. 4, we set the detection threshold value arbitrarily in advance. Here, if ($E_{k'} \leq$ threshold and threshold $< E_{k'+1}$) and ($E_{k''} >$ threshold and threshold$\geq E_{k''+1}$), then we find the position k that corresponds to the maximum value E_k within the range of $k'+1$ to k''. Next, we calculate the LPC spectrum of the extracted waveform segment of the range of $k-M$ to $k+M$ (the blue segmented waveforms in Fig. 4). While effective, this simple method can be insensitive to signals of varying strength. To improve detection and segmentation efficiency we have added an Automatic Gain Control (AGC) and AGC shift function.

Fig. 5 shows the combined AGC and AGC shift functions act to normalize the energy curve of signals of disparate strength (yellow and red lines). In Fig. 5, we have set the AGC with a 1.0 second frame width

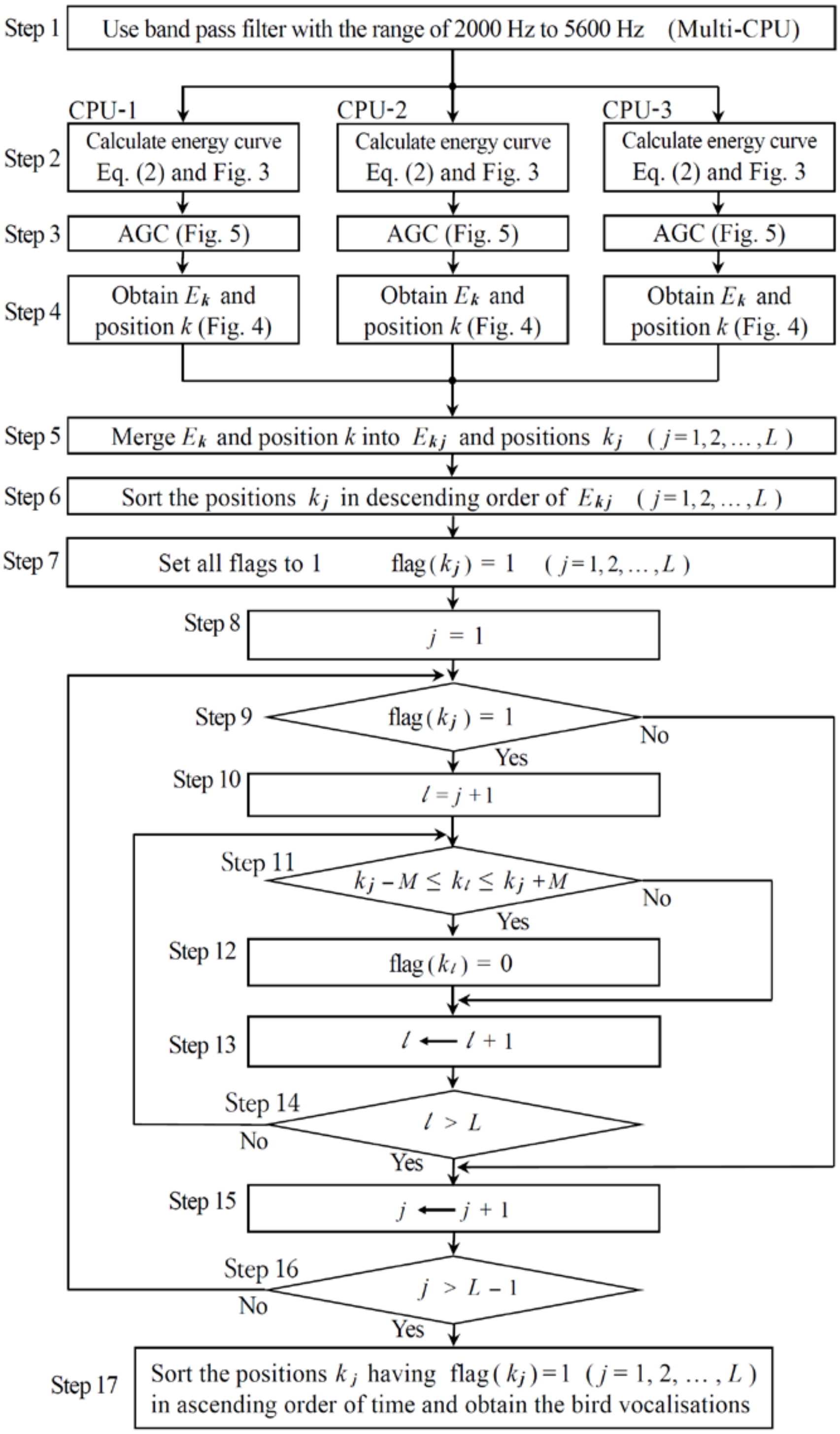

Fig. 6 Flowchart for segmentation

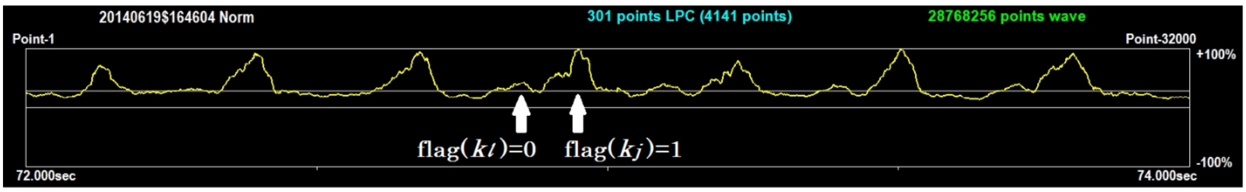

Fig. 7 Setting flags and resetting flags

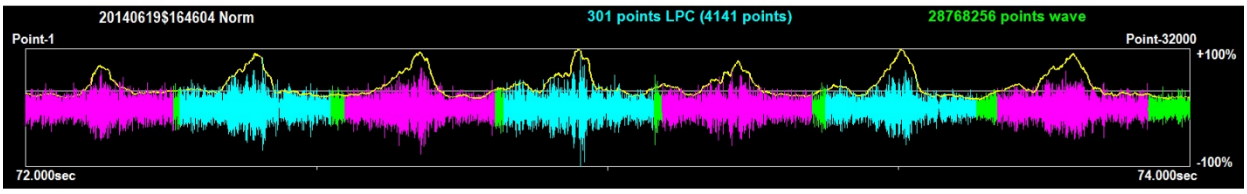

Fig. 8 Results of segmentation

and a 0.7 second frame period. These settings are established with respect to the inter-syllable and call intervals of the target species. Considering the blue segmented waveform sections shown in Fig. 5, we find that the bird vocalisations can be segmented accurately even if the amplitude of the received signal is reduced or variable.

5 Segmentation for low sound pressure level signals and signals in noisy environments

While the base segmentation algorithm described in the previous section performs well on good quality signals clear of noise, it performs less consistently on low sound pressure level signals and signals in noisy environments. Here we describe an advance on the previous segmentation algorithm that performs well, even on low strength signals in comparatively noisy environments.

Fig. 6 and Fig. 7 show the processing procedure to segment a waveform of the sample bird vocalisation from a three-hour continuous recording. In Fig. 7, the calculated energy curve E_k is shown by the yellow line. As shown by a white arrow flag(k_l) in Fig. 7, the energy curve has a lot of small peaks due to noise. Fig. 6 is a flowchart used to remove these small peaks.

● In Step 1 of Fig. 6, a band pass filter with the range of 2000 Hz to 5600 Hz is used to remove the noise of traffic. The band pass filter software executes parallel processing using multi-CPU.

● In Step 2, as shown in Fig. 3, the energy curves are calculated by using Eq. (2).

● In Step 3, as shown in Fig. 5, the energy curves are periodically normalized for each period of time (Auto Gain Control).

● In Step 4, as shown in Fig. 4, we obtain the position k that corresponds to the maximum value E_k within the range of k'+1 to k''.

○ Note that each of the CPUs executes parallel processing for the process of Steps 2-4. The software detects the number of CPUs automatically.

● In Step 5, a single CPU merges E_k and position k into E_{kj} and positions k_j (j = 1, 2, … , L).

● In Step 6, the CPU sorts the positions k_j in descending order of E_{kj} (j = 1, 2, … , L).

● In Step 7, all flags are set to 1. Namely, flag(k_j) = 1 (j = 1, 2, … , L) .

● In Steps 8-16, if ($E_{kl} \leq E_{kj}$) and ($k_j - M \leq k_l \leq k_j + M$), then flag($k_l$) = 0. This is, if the smaller peak E_{kl} of the energy curve is close to the greater peak E_{kj}, then we determine that the smaller peak E_{kl} is not the bird vocalisation. Fig. 7 shows that flag(k_l) is reset to 0 because the smaller peak E_{kl} is close to the greater peak E_{kj}.

● In Step 17, the CPU sorts the positions k_j having flag(k_j) = 1 (j = 1, 2, … , L) in ascending order of time and we obtain the bird vocalisations.

6 Evaluation experiments

To check the effectiveness of the segmentation method described in Section 5, we have performed an evaluation experiments for the Noisy Miner Chur call (Fig. 1) using the algorithm shown in Fig. 6. Following blue and magenta lines of the waveform shown in Fig. 8, we find that the bird vocalisations can be segmented accurately even if they are poor quality signals recorded within a noisy environment. Moreover, as a result of listening to these bird vocalisations, we have verified the effectiveness of the proposed method.

7 Conclusions and future work

We have introduced automatic recognition software that executes parallel processing using multi-CPU and have proposed here a new segmentation method, particularly relevant for bird vocalisations recorded at distance or in noisy environments. In our future work, we will continue to develop the recognition software using the Two-dimensional Geometric Distance.

References

[1] M. Jinnai, N. Boucher, J. Robertson and S. Kleindorfer, "Design considerations in an automatic classification system for bird vocalisations using the Two-dimensional Geometric Distance and cluster analysis", *20th International Congress on Acoustics, ICA2010*, 130, (August 2010, Sydney) http://www.soundid.net/

[2] M. Jinnai, N. Boucher, M. Fukumi and H. Taylor, "A new optimization method of the geometric distance in an automatic recognition system for bird vocalisations", *International Congress on Acoustics*, 105, (April 2012, Nantes, France) http://www.soundid.net/

[3] M. Jinnai, H. Yamaguchi, Y. Ishihara, J. Ohshima and Y. Kidu, "Apparatus for detecting abnormal sound and method for judging wrong in machine", *Patent No.* US6170333(2001), JP3426905(2003)

[4] M. Jinnai, Y. Akashi, K. Hashimoto and S. Hayashi, "Method for detecting abnormal sound and method for judging abnormality in structure by use of detected value thereof", *Patent No.* US9552831(2017), AU2016200487(2017), CA2918533(2018), JP5956624(2016)

[5] M. Jinnai, S. Tsuge, S. Kuroiwa, F. Ren and M. Fukumi, "New Similarity Scale to Measure the Difference in Like Patterns with Noise", *International Journal of Advanced Intelligence*, Volume 1, Number 1, pp. 59-88 (2009) http://aia-i.com/ijai/index.html

[6] M. Jinnai, S. Tsuge, S. Kuroiwa and M. Fukumi, "A New Geometric Distance Method to Remove Pseudo Difference in Shapes", *International Journal of Advanced Intelligence*, Volume 2, Number 1, pp. 119-144 (2010) http://aia-i.com/ijai/index.html