New similarity scale to recognize bird calls and abnormal sounds of concrete/machine
Development of pattern matching software using multi-CPU

Michihiro Jinnai [*1]   Edward James Pedersen [*2]

[*1] Nagoya Women's University   [*2] Central Queensland University, Australia

A new similarity scale called the Geometric Distance, that numerically evaluates the degree of likeness between the standard pattern and the input pattern is proposed. Traditionally, the similarity scales known as the Euclidean distance and cosine similarity have been widely used to measure likeness. Traditional methods do not perform well in the presence of noise or pattern distortions. In this paper, a mathematical model for similarity is proposed to overcome these limitations of the earlier models, and a new algorithm based on a one-to-many point mapping is proposed to realize the mathematical model. Using the new similarity scale, experiments in bird call recognition were carried out in noisy environments. Furthermore, experiments in abnormal sound recognition of concrete structure were carried out. In all cases a significant improvement in recognition accuracy is demonstrated.

1. Introduction

Human beings, dogs, cats, and other such mammals exhibit a "sense of similarity" in their perception of sounds and sighted objects. To emulate this sense of similarity algorithmically in a "similarity scale" is an important objective for developing computer intelligence.

In an acoustic similarity scale, the degree of likeness between an acoustic standard pattern (control) and an undetermined input pattern is evaluated as a "distance" between the two patterns. This process arbitrarily emulates a human perceiving a sound and comparing that autonomously with 'templated' or remembered sounds. Ostensibly, a software-based similarity scale would return a short 'distance' for two patterns that humans would consider as similar to, or the same as each other, and a long 'distance' for two patterns that humans would consider as dissimilar.

Euclidean distance and cosine similarity are widely used to measure likeness. Conventional similarity scales compare patterns using one-to-one mapping. The result of one-to-one mapping is that the distance metric is highly sensitive to noise, and the distance metric changes in a staircase pattern when a difference occurs between peaks of the standard and input patterns. As an improvement, we have developed a new similarity scale called "Geometric Distance (GD)" [1]. GD is more accurate than the conventional similarity scales in noisy environments.

The GD similarity scale and allied detection software have proven applicable in a broad range of dynamics: from automated call detection of endangered species [2] from within 'big' environmental data sets; to real-time fault detection in concrete structures and operating machinery [3]. In this paper, we describe the underlying mathematical model for similarity; the GD algorithm; and we introduce automatic recognition software for bird vocalisations that uses the GD algorithm.

2. The LPC spectrogram of bird sounds

The lower diagram of Fig. 1 shows the waveform of a 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).
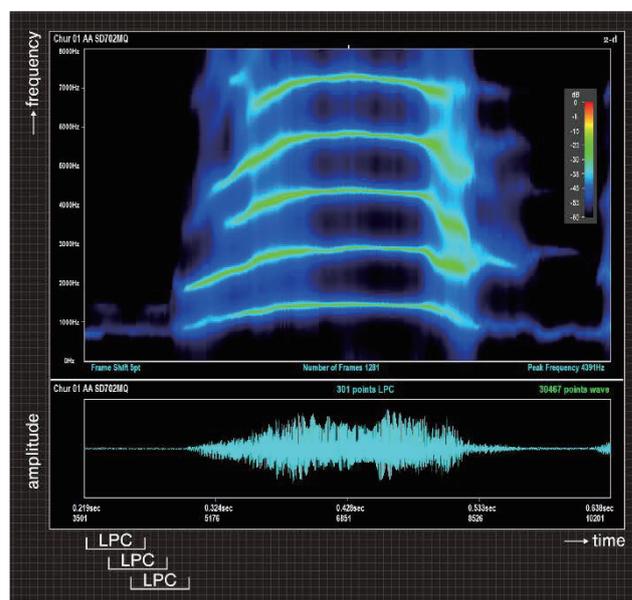
Fig. 1 Spectrogram of bird vocalisation
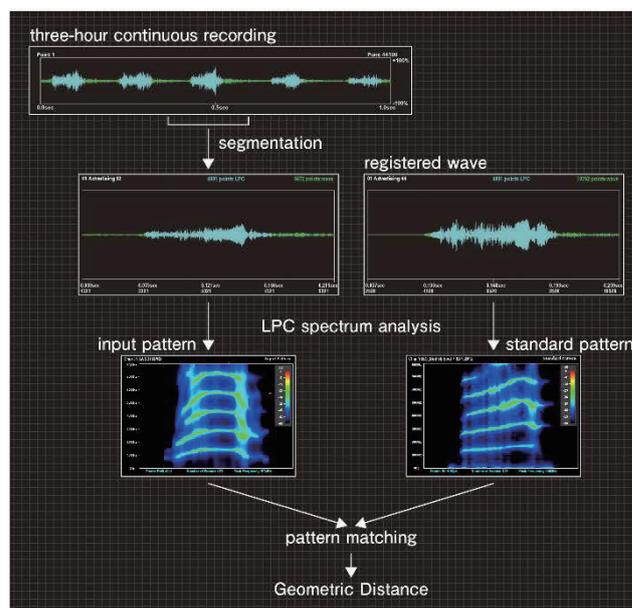
Fig. 2 Processing procedure in recognition system

The upper diagram of Fig. 1 shows a spectrogram (time-frequency-power) of the exact same signal. In Fig. 1, the waveform has been segmented with 18.8msec frame width and 0.31msec frame period, and the LPC spectrum has been calculated for each frame. Next, the spectrogram has been coloured according to logarithmic power of the LPC spectrum. We have set the software analysis parameters for this bird vocalisation recorded at 16kHz sampling frequency, 16bit quantization, to an LPC order of 12; restricted the spectral frequency range to 0Hz to 8000Hz, with an 11.5Hz frequency resolution; and set a dB threshold filter of 0dB to -60dB logarithmic power spectrum. To analyze transient signals such as bird vocalisations, we set a fractional frame width (with respect to total signal period) as shown at the bottom of Fig. 1. LPC spectrum analysis is suitable for spectral modeling of 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 the GD similarity scale. To expedite the process, the software executes parallel processing using multiple CPUs [4][5].

4. Mathematical model for similarity

For a functional similarity scale, we need first to develop a mathematical model for similarity, that can perform numerical processing by computation. In the GD process, a mathematical model incorporating the following two characteristics is used:
< 1 > A distance metric which shows good immunity to noise.
< 2 > A distance metric which increases monotonically when a difference increases between peaks of the standard and input patterns.

Figs. 3 and 4 graphically demonstrate the underlying computational and algorithmetric processes. The upper diagram of Fig. 3 shows an example of the "difference" where the standard pattern has two peaks in the spectrogram, and input patterns 1, 2, and 3 have a different position on the first peak. Note that both the standard and input patterns have the same volume. Fig. 4 shows an example of a "wobble" where the standard pattern has a flat spectrogram. Input patterns 4 and 5 have a "wobble" on the flat spectrogram, and input pattern 6 has a single peak. Each pattern however, is assumed to have variable α in the relationship shown in Fig. 4. Therefore, the standard and input patterns always have the same volume.

Bar graphs at the bottom right of Figs. 3 and 4 express the characteristics < 1 > and < 2 > of the mathematical model diagrammatically.

5. New algorithm for similarity scale

A new algorithm based on one-to-many point mapping is proposed to realize the mathematical model. In the GD algorithm, when a "difference" occurs between peaks of the standard and input patterns with a "wobble" due to noise, the "wobble" is
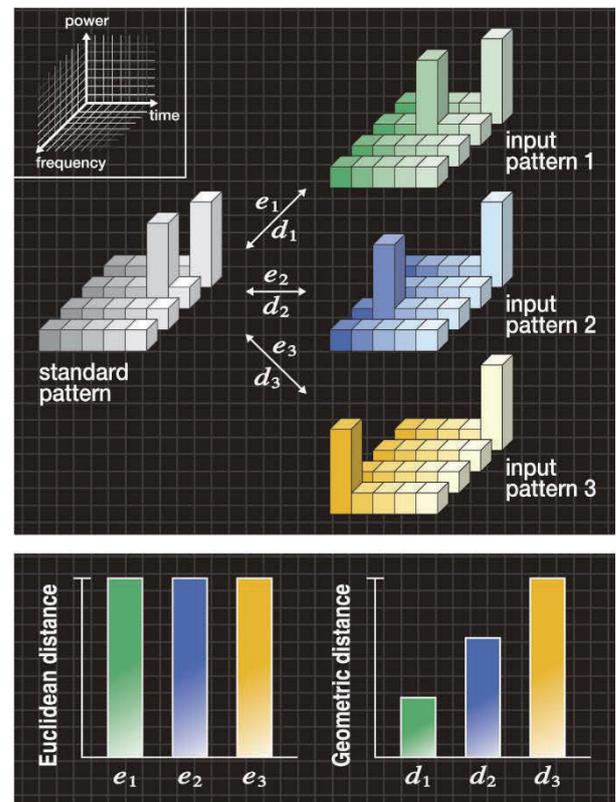
Fig. 3   Typical example of "difference"
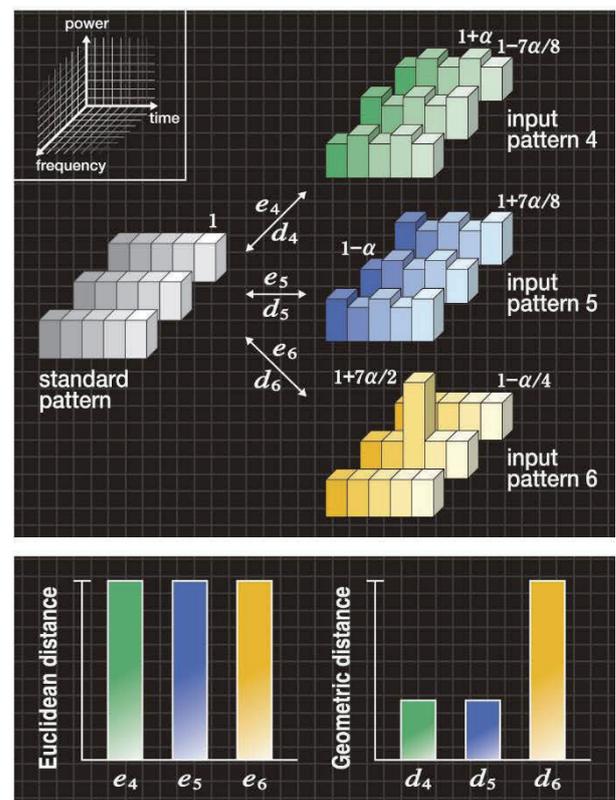
Fig. 4   Typical example of "wobble"

absorbed and the distance metric increases monotonically according to the increase of the "difference".

In statistical analysis, a normal distribution is usually used as a model for a phenomenon. Then, a "kurtosis" and a "skewness" are used to verify whether the phenomenon obeys the normal distribution or not. Here, the kurtosis '*a*' and the skewness '*b*' are statistics, and we explain them using Figs. 5 and 6. If a probability distribution of the phenomenon follows the normal distribution, then $a = 3$ (Fig. 5(b)). If it has flatness relative to the normal distribution, then $a < 3$ (Fig. 5(a)). Conversely, if it has peakedness relative to the normal distribution, then $a > 3$ (Fig. 5(c)). Also, if a probability distribution of the phenomenon is symmetrical about the mean μ, then $b = 0$ (Fig. 6(b)). If the tail on the left side of the probability distribution is longer than the right side, then $b < 0$ (Fig. 6(a)). Conversely, if the tail on the right side of the probability distribution is longer than the left side, then $b > 0$ (Fig. 6(c)).

In this section, we explain the GD algorithm using Figs. 7 and 8. Fig. 7 shows the spectra (frequency-power) extracted from a Macleay's Fig Parrot (*Cyclopsitta diophthalma macleayana*) vocalisation. Fig. 7 shows standard and input patterns that have been created using the momentary power spectrum (frequency-power) of standard and input sounds. Figs. 8(a)-(e) respectively show typical examples of the standard and input patterns. Note that the power spectrum is generated from the output of a filter bank with *m* frequency bands. The *i*-th power spectrum values (where, $i = 1, 2, \ldots, m$) of the standard and input sounds are divided by their total energy, so that normalized power spectra s_i and x_i have been calculated, respectively. At this moment, the standard and input patterns have the same area size. Moreover, Figs. 8(a)-(e) respectively show reference patterns that have the initial shape r_i of a normal distribution.

With the GD algorithm, a difference in shapes between standard and input patterns is replaced by the shape change of the reference pattern using the following equation.

$$r_i \longleftarrow r_i + (x_i - s_i) \qquad (i = 1, 2, 3, \cdots, m) \qquad (1)$$
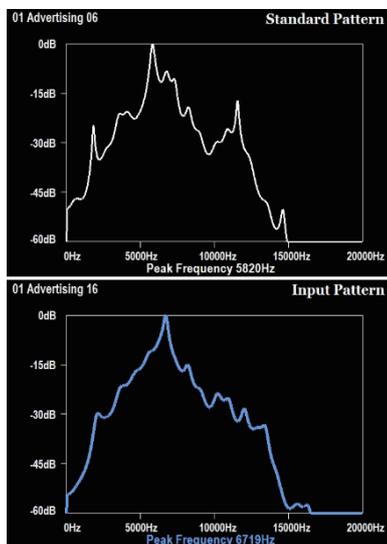
Next, we explain Eq. (1) using Figs. 8(a)-(e).
- Fig. 8(a) gives an example of the case where the standard and input patterns have the same shape. Because values r_i of Eq. (1) do not change during this time, the reference pattern shown in Fig. 8(a) does not change in the shape from the normal distribution.
- Figs. 8(b)-(d) respectively show examples exhibiting a small, medium, and large "difference" of peaks between the standard and input patterns. If Eq. (1) is represented by the shapes, as shown in Figs. 8(b)-(d), value r_i decreases at peak position '*i*' of each standard pattern. At the same time, value r_i increases at peak position *i* of each input pattern.
- Fig. 8(e) typically shows the standard pattern having a flat shape and the input pattern where a "wobble" occurs in the flat shape. Because values r_i increase and decrease alternatively in Eq. (1)
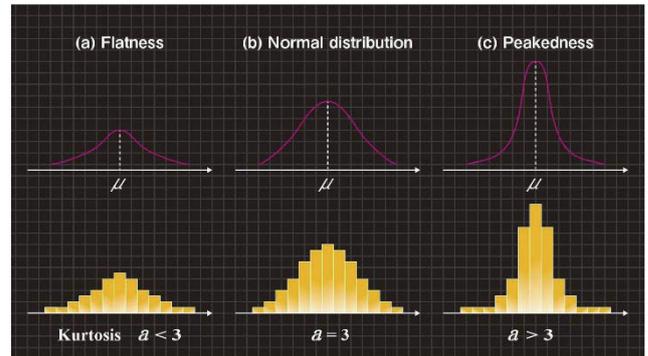
Fig. 5   Shape change and kurtosis value '*a*'
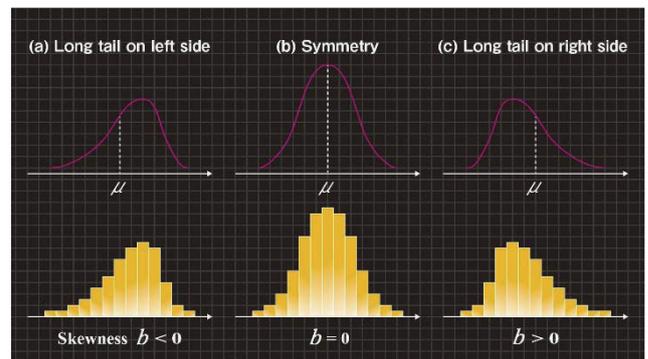
Fig. 6   Shape change and skewness value '*b*'

Fig. 7   Standard and input patterns
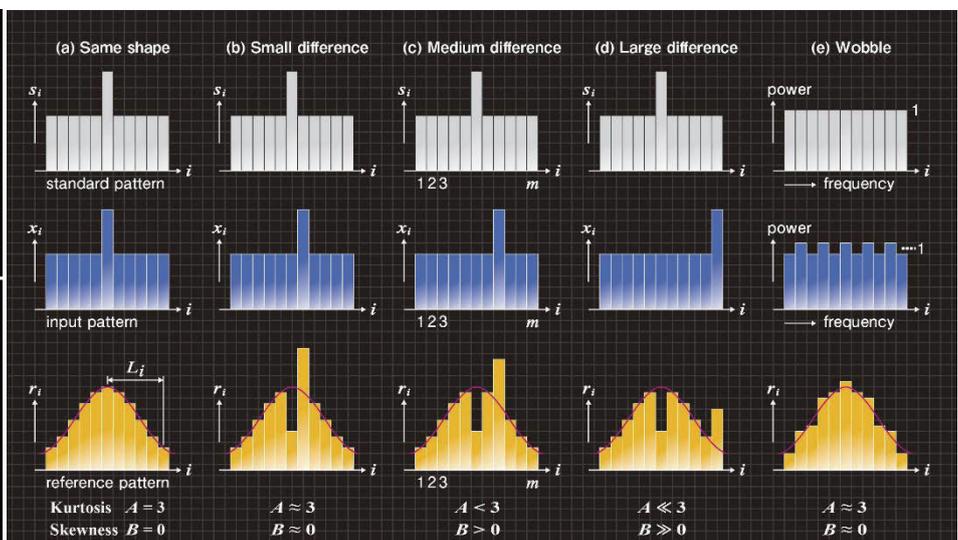
Fig. 8   Shape change of reference patterns

during this time, the reference pattern shown in Fig. 8(e) has a small shape change from the normal distribution.

With the GD algorithm, we replace the mean μ shown in Figs. 5 and 6 with the centre axis of the normal distribution (reference pattern) shown in Fig. 8(a). Then, we replace the kurtosis 'a' and the skewness 'b' with a kurtosis 'A' and a skewness 'B' shown in the following equations.

$$A = \frac{\left\{\sum_{i=1}^{m} r_i\right\} \cdot \left\{\sum_{i=1}^{m} (L_i)^4 \cdot r_i\right\}}{\left\{\sum_{i=1}^{m} (L_i)^2 \cdot r_i\right\}^2} \quad B = \frac{\left\{\sqrt{\sum_{i=1}^{m} r_i}\right\} \cdot \left\{\sum_{i=1}^{m} (L_i)^3 \cdot r_i\right\}}{\left\{\sum_{i=1}^{m} (L_i)^2 \cdot r_i\right\}^{\frac{3}{2}}} \quad (2)$$

Where, L_i ($i = 1, 2, \ldots , m$) is a deviation from the centre axis of the normal distribution as shown in the reference pattern of Fig. 8(a). Then, numerical experiments were carried out to study the relationships between the kurtosis 'a' and the kurtosis 'A' or between the skewness 'b' and the skewness 'B'. As a result of the experiments, we have confirmed that they have the same characteristics [1][3].

For the reference pattern whose shape has changed by Eq. (1), the magnitude of shape change is numerically evaluated as the variable of kurtosis A and skewness B. The kurtosis and the skewness of the reference pattern can be calculated using Eq. (2). Figs. 8(a)-(e) show how A and B vary with r_i.

● In Fig. 8(a), the values r_i do not change. The kurtosis becomes $A = 3$ and the skewness becomes $B = 0$.

● In Fig. 8(b), the position i of the decreased r_i and that of the increased r_i are close. Because the effect of an increase and a decrease is cancelled out, the kurtosis becomes $A \approx 3$ and the skewness becomes $B \approx 0$.

● In Fig. 8(d), because the shape of the reference pattern is flattened relative to the normal distribution and the shape of the reference pattern has a long tail to the right side, the kurtosis becomes $A \ll 3$ and the skewness becomes $B \gg 0$.

● In Fig. 8(c), because the shape of the reference pattern is an intermediate state between (b) and (d), the kurtosis becomes $A < 3$ and the skewness becomes $B > 0$.

● In Fig. 8(e), the reference pattern has a small shape change from the normal distribution, and the kurtosis becomes $A \approx 3$ and the skewness becomes $B \approx 0$.

From Figs. 8(a)-(d), we can understand that the values $|A|$ and $|B|$ respectively increase monotonically according to the increase of the "difference" between peaks of the standard and input patterns. Also, from Fig. 8(e), it is clear that $A \approx 3$ and $B \approx 0$ for the "wobble". In this method, when a "difference" occurs between peaks of the standard and input patterns with a "wobble" due to noise, the "wobble" is absorbed and the distance metric increases monotonically purely in accord with the increase of the "difference". On this basis, we verify that the GD algorithm matches the characteristics ＜1＞ and ＜2＞ of the mathematical model. GD is defined using both the kurtosis A and the skewness B [3]. We have both one-dimensional GD and two-dimensional GD. In addition, we have a fast calculation GD algorithm [1][3].

6.　Evaluation experiments

To authenticate the effectiveness of the GD algorithm described in Section 5, we performed evaluation experiments for the vocalisations of Macleay's Fig-Parrot. Fig. 9 shows that, using the GD algorithm, pattern matching even in a noisy environment is accurate. The same GD algorithm has been successfully used to locate cavities in concrete structures by comparing the acoustic response to controlled surface tapping above integral concrete and concrete compromised by erosion cavities. Recognition accuracy comparing taps arising from integral and cavity-compromised concrete is $17/20$ [5]. These applied experiments verify the effectiveness of the GD algorithm.

7.　Conclusions and future work

We have described the GD algorithm and introduced associated automatic recognition software for bird vocalisations. The software executes parallel processing using multiple CPUs. In our future work, we will continue to improve the recognition software.

References

[1] 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

[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, 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)

[4] M. Jinnai, "A new segmentation method of bird vocalisations recorded in the distance", *2018 Autumn Meeting, Acoustical Society of Japan*, 3-1-5

[5] M. Jinnai, K. Hashimoto, S. Hayashi and E. J. Pedersen, "Recognition of abnormal vibrational responses of concrete structures", *2019 Spring Meeting, Acoustical Society of Japan*, 1-6-15
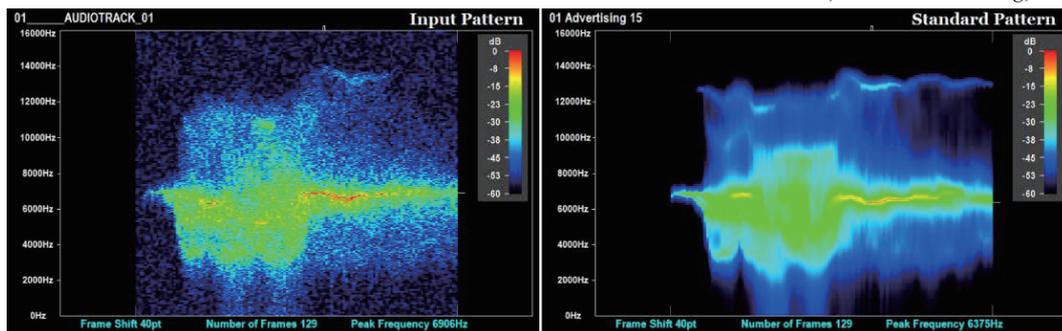
Fig. 9　Result of pattern matching for bird call recognition in a noisy environment