行政院國家科學委員會專題研究計畫 成果報告

以鳥鳴聲為基礎之鳥種辨識系統架構之研究與開發(III)
研究成果報告(精簡版)

計 畫 類 別 ： 個別型
計 畫 編 號 ： NSC 98-2221-E-216-033-
執 行 期 間 ： 98 年 08 月 01 日至 99 年 07 月 31 日
執 行 單 位 ： 中華大學資訊工程學系

計 畫 主 持 人 ： 周智勳
共 同 主 持 人 ： 陳建宏

報 告 附 件 ： 出席國際會議研究心得報告及發表論文

處 理 方 式 ： 本計畫涉及專利或其他智慧財產權，2 年後可公開查詢

中 華 民 國 99 年 10 月 30 日

目錄

以鳥鳴聲為基礎之鳥種辨識系統架構之研究與開發(III)

主持人： 周智勳

執行機構： 中華大學資訊工程學系

執行期間： 民國98年08月01日至99年07月31日

國科會計畫編號：NSC 98-2221-E-216-033

Abstract

In this study, an automatic birdsong recognition system based on syllable features was developed. In this system, after the syllables have been segmented, three syllable features, namely mean, QI and QE, were computed from the MFCCs of each syllable. The first feature has been applied in many studies, however, QI and QE are novel features. Adding the advantages of the fuzzy c-mean (FCM) clustering algorithm and the linear discriminant analysis (LDA), the presented feature vector was used to construct an automatic birdsong recognition system. In the experiment, the proposed system was applied to a birdsong database with 420 bird species and achieved an average recognition rate of 83.3%.

Keywords: Birdsong, MFCC, syllable, linear discriminant analysis, transition matrix

1. Introduction

The investigation of bird species diversity is the key in monitoring environment and ecosystem recovery, and automatic bird species recognition by recognizing their birdsongs has become an invaluable study method in the long-term investigation of bird species. The vocalization types of bird species include birdsong and birdcall. Birdsong 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. Birdcall, on the other hand, is monotonous, brief, repeated, fixed and sexless and is used to contact or alert companions. The time duration and acoustic structure of a birdcall are usually short and simple while the duration of a birdsong is longer and is composed of a succession of melodious musical notes.

Although MFCCs have been well-applied in bird species recognition, further study on this feature is necessary to increase the recognition rate. In (Lee et al. 2003; Lee et al., 2001; Skowronski and Harris, 2002, 2003; Bou-Ghazale and Hansen, 2000) optimal theories were used to obtain the center frequencies and bandwidths of the triangular filters. The discrete cosine transform (DCT) was replaced with the wavelet transform in (Ricotti, 2005). Filter weighting was applied in (Hung and Wang, 2001) to assign a weight for each order of MFCCs. In (Kwan et al., 2006) the MFCCs as well as their first-order and second-order differences were used to form the feature vector. Combination of MFCCs with a lot of low-level descriptive parameters such as zero-crossing rate, short time energy, syllable length, spectrum centroid, bandwidth and so on was applied in (Somervuo, 2006) for recognizing 14 bird species.

In this study, neither modifying the steps for computing the MFCCs nor combining the MFCCs with other types of features, three features were computed from the MFCCs of a syllable to form the syllable feature vector. The proposed method aimed at easy computation and small time complexity. Integrating with the advantages of FCM clustering algorithm and the LDA, the proposed system achieved an average recognition rate over 83% when recognizing the birdsongs of 420 bird species. The remaining of this paper is as follows: Section 2 describes the structure of the proposed system. Experimental results are shown in Section 3. Section 4 is the conclusion.

2. The Proposed System

The block diagram of the proposed system containing the training part and the testing (recognition) part is shown in Fig. 2.1. Three terms named mean, QI and QE were computed to form the feature vectors of each syllable, and the recognition was achieved by comparing the matching degrees between the feature vector of the test syllable and the template syllables. Each step in the diagram is described in detail in the following.

2.1 Syllable Segmentation

Due to the resistance to signal fading and echoing, the frequency domain analysis of the birdsong signal is better than in the time domain approach. So the frequency domain approach was utilized in this study. The segmentation process applied in this study is described in the following.

Step 1 Compute the short time Fourier transform of $x(t)$ with frame size N = 512, and form the spectrogram of the signal. The Hamming window for short time analysis has the form of

$$w[n] = \begin{cases} 0.54 - 0.46\cos\left(\dfrac{2\pi n}{N-1}\right), 0 \le n \le N-1 \\ 0 \qquad\qquad\qquad\qquad\qquad , \text{otherwise} \end{cases} \quad (1)$$

Step 2 For each frame m, find the frequency Bin bin_m with the greatest magnitude.

Step 3 Initialize the syllable index j, j = 1.

Step 4 Compute the frame t at which the maximum magnitude occurs

$$t = \arg\max_{1 \le m \le M}(|X[bin_m]|), \quad (2)$$

and set the amplitude of syllable j as

$$A_j = 20 \cdot \log_{10}|X[bin_t]|(\text{dB}), \quad (3)$$

in which M is the number of frames of $x(t)$, and $X[\cdot]$ denotes the spectrum of $x(t)$.

Step 5 Start from frame t and move backward and forward up to frames h_j and t_j such that both $20 \cdot \log_{10}|X[bin_{h_j}]|$ and $20 \cdot \log_{10}|X[bin_{t_j}]|$ are smaller than $(A_j - 20)$ (dB).

Step 6 Start from frames h_j and t_j, find frames $h_j - \alpha$ and $t_j + \beta$ (α, β > 0) such that both $20 \cdot \log_{10}\left|X[bin_{h_j-\alpha-1}]\right|$ and $20 \cdot \log_{10}\left|X[bin_{t_j+\beta+1}]\right|$ are greater than $(A_j - 20)$. Then $h_j - \alpha$ and $t_j + \beta$ are called the head frame and tail frame of syllable j.

Step 7 Set

$$|X[bin_m]| = 0, m = h_j - \alpha, h_j - \alpha + 1, \cdots, t_j + \beta - 1, t_j + \beta \cdot \quad (4)$$

Step 8 Let $j = j + 1$.

Step 9 Repeat Step 4 to Step 8 until $A_j < A_1 - 20$.

2.2. Feature Extraction

After syllable segmentation, three features named mean, QI and QE were computed to form the feature vector of the syllable as described in the following.

2.2.1. Compute the MFCCs of each frame

The steps for computing the MFCCs of each frame are as follows:

Step 1 Compute the fast Fourier transform (FFT) of each framed signal.

$$X[k] = \sum_{n=0}^{N-1} x[n]w[n]e^{-j2\pi nk/N}, 0 \le k < N, \quad (5)$$

Step 2 Compute the energy of each triangular filter band

$$E_j = \sum_{k=0}^{N/2-1} \phi_j[k]|X[k]|^2, 0 \le j < J, \quad (6)$$

where $\phi_j[k]$ denotes the amplitude(weight) of the

j^{th} triangular filter at frequency bin k as shown in Fig. 2.2, E_j denotes the energy of j^{th} filter band, and J is the number of triangular filters.

Step 3 Compute the MFCCs by Cosine transformation

$$c_i(m) = \sum_{j=0}^{J-1} \cos\left(m\frac{\pi}{J}(j+0.5) \right) \log_{10}(E_j),\ 0 \le m < 15, \tag{7}$$

where $c_i(m)$ denotes the m^{th} order MFCC of the i^{th} frame.

In the following, three features named mean, QI and QE computed from the MFCCs were used to form the feature vector of a syllable.

2.2.2. Computing the mean, QI, and QE

In this study, three features, namely mean, QI, and QE were used to form the feature vector.

Feature 1: mean of MFCCs

After computing the first 15 (order) MFCCs of each frame, the coefficients of the same order of all frames were averaged. The average of m^{th} order MFCCs $a(m)$ was obtained by the following equation:

$$a(m) = \frac{1}{W}\sum_{i=1}^{W} c_i(m),\ 0 \le m < L, \tag{8}$$

where W is the number of frames and L = 15 is the order of MFCCs applied in this study. Due to the scale diversity between different orders of MFCCs, a normalization process for $a(m)$, $\hat{a}(m)$, is required.

Feature 2: QI of MFCCs

For saving on computation complexity, consecutive frames were used as a time unit to compute QI. The process for computing QI is described in the following.

Step 1 Quantize the MFCCs of each order in all frames ($c_1(m), c_2(m), ..., c_W(m)$) into Q levels (from level 0 to level Q-1).

$$v(m) = \frac{c_{\max}(m) - c_{\min}(m)}{Q}, \tag{9}$$

where $\max = \arg\max_{i=1,2,...,W} c_i(m)$, $\min = \arg\min_{i=1,2,...,W} c_i(m)$ and $v(m)$ is the quantization interval of the m^{th} order MFCCs.

Step 2 Segment the W frames into S equal sections, then compute the mean of each order of MFCCs in every section.

$$\tilde{a}_s(m) = \frac{S}{W}\sum_{k=(s-1)\cdot W/S+1}^{s\cdot W/S} c_k(m),\ 0 \le m < L,\ 1 \le s \le S, \tag{10}$$

where s is the section index.

Step 3 Find the level $I_s(m)$ at which the value $\tilde{a}_s(m)$ locates, where

$$I_s(m)\cdot v(m) \le \tilde{a}_s(m) - a_{\min}(m) < (I_s(m)+1)\cdot v(m),\quad 0 \le m < L,\ 1 \le s \le S. \tag{11}$$

Step 4 Form the sequence $I_1(m), I_2(m), ..., I_S(m)$ for each order of MFCCs.

Step 5 Those sequences obtained in Step 4 for all the 15 orders of MFCCs form the second feature QI.

$$\mathrm{QI} = I_1(0),..., I_S(0), I_1(1),..., I_S(1), ..., I_1(L-1),..., I_S(L-1) \tag{12}$$

Feature 3: QE of MFCCs

The process for obtaining QE is described in the following.

Step 1 Perform the same quantization process (Step 1) used in computing feature 2.

Step 2 Find the level $I_i(m)$ at which the value $c_i(m)$ locates, where

$$I_i(m)\cdot v(m) \le c_i(m) - c_{\min}(m) < (I_i(m)+1)\cdot v(m),\quad 0 \le m < L,\ 1 \le i \le W. \tag{13}$$

Step 3 For each order of MFCCs, record the frames that transit from level x to level y, $0 \le x, y \le Q-1$, and denote it as $G_m(x,y)$. That is

$$G_m(x,y)=\{i|x\cdot v(m)\le c_i(m)-c_{\min}(m)<(x+1)\cdot v(m),\\ y\cdot v(m)\le c_{i+1}(m)-c_{\min}(m)<(y+1)\cdot v(m), 1\le i<W\}. \quad (14)$$

Step 4 Compute the level transition matrix $T_m(X,Y)$ for each order of MFCC by using $G_m(x,y)$

$$T_m(x,y)=\frac{|G_m(x,y)|}{\sum_{x=0}^{Q-1}\sum_{y=0}^{Q-1}|G_m(x,y)|},\quad 0\le m<L. \quad (15)$$

Step 5 Compute and sort the eigenvalues of $T_m(X,Y)$, $\lambda_m^1\ge\lambda_m^2\ge...\ge\lambda_m^Q$, $0\le m<L$.

Step 6 Form the feature vector QE by using all the eigenvalues

$$\mathrm{QE}=\lambda_0^1,...,\lambda_0^Q,\lambda_1^1,...,\lambda_1^Q,...,\lambda_{L-1}^1,...,\lambda_{L-1}^Q \quad (16)$$

2.2.3. Construct the feature vector by using mean, QI and QE

Combining the three features forms a 15+$S\cdot L$+$Q\cdot L$ dimensional feature vector. The Linear Discriminant Analysis (LDA) was applied to the feature vector form by QI and QE.

The LDA (Duda et al., 2000) transforms data from the original space to a new space which is better for classification. To find such a transformation matrix W, it requires maximizing the Fisher criterion

$$\max_W J(W)=\max_W\frac{tr(W^TS_bW)}{tr(W^TS_wW)}. \quad (17)$$

The matrices S_w and S_b, called within-class scatter matrix and between-class scatter matrix, are computed by the following equations:

$$S_w=\sum_{j=1}^{C}\sum_{i=1}^{N_j}(\mathbf{x}_i^j-\boldsymbol{\mu}_j)(\mathbf{x}_i^j-\boldsymbol{\mu}_j)^T, \quad (18)$$

$$S_b=\sum_{j=1}^{C}(\boldsymbol{\mu}_j-\boldsymbol{\mu})(\boldsymbol{\mu}_j-\boldsymbol{\mu})^T, \quad (19)$$

in which $\mathbf{x}_i^j$ denotes the i^{th} vector in class j, $\boldsymbol{\mu}_j$ is the mean vector of class j, C is the number of classes, N_j is the number of vectors in class j and $\boldsymbol{\mu}$ is the mean of all data vectors. It was found that the optimal matrix W_{opt} solved by Eq. (17) is composed of the principal eigenvectors of the matrix $S_w^{-1}S_b$. The principal eigenvectors of a matrix are defined by the corresponding eigenvalues. The eigenvectors whose corresponding eigenvalues are the largest d eigenvalues of a matrix form the d principal eigenvectors of the matrix. Determination of d can be accomplished by the following equation

$$d=\min_t\sum_{i=1}^{t}\lambda_i\ge\theta\cdot\sum_{i=1}^{m}\lambda_i, \quad (20)$$

where λ_i is the i^{th} largest eigenvalue, m is the number of eigenvalues and θ is a parameter to be set.

After the LDA of QI and QE, the dimension of the feature vector formed by the three features was reduced. To obtain representative feature vectors for a birdsong requires the clustering of the syllable feature vectors. In this study, the clustering process was accomplished by using the fuzzy c-mean (FCM) clustering method. The FCM, proposed by Dunn in 1973 and enhanced by Bezdek in 1981, is an un-supervised clustering algorithm iterative tuning the cluster centers and the cluster memberships of data vectors. The clustering process is described in the following.

Step 1: Select the cluster number c.

Step 2: Set the initial fuzzy pseudopartition at t = 0 satisfying

$$\sum_{i=1}^{c}\mu_{ij}^{(t)}=1,\quad j=1,2,...,J, \quad (21a)$$

$$0<\sum_{j=1}^{J}\mu_{ij}^{(t)}<J,\quad i=1,2,...,c. \quad (21b)$$

In these two equations, $\mu_{ij}^{(t)}$ denotes the membership grade of feature vector $\mathbf{v}_{s_j}$ belonging to cluster i at time t, and J is the number of feature vectors to be clustered.

Step 3: Set the initial performance index $J_m^{(t)}$, t = 0, as 0.

Step 4: Calculate the c cluster centers $\mathbf{v}_1^{(t)},\ldots,\mathbf{v}_c^{(t)}$ by

$$\mathbf{v}_i^{(t)} = \frac{\sum_{j=1}^{J} (\mu_{ij}^{(t)})^m \cdot \mathbf{v}_{s_j}}{\sum_{j=1}^{J} (\mu_{ij}^{(t)})^m}, \quad i = 1,\ldots,c, \quad 1 < m < 2. \tag{22}$$

Step 5: Update the membership grade for each feature vector $\mathbf{v}_{s_j}$,

$$\mu_{ij}^{(t+1)} = \left[\sum_{k=1}^{c} \left(\frac{\left\| \mathbf{v}_{s_j} - \mathbf{v}_i^{(t)} \right\|^2}{\left\| \mathbf{v}_{s_j} - \mathbf{v}_k^{(t)} \right\|^2} \right)^{\frac{1}{m-1}} \right]^{-1}. \tag{23}$$

Step 6: Compute the performance index

$$J_m^{(t+1)} = \sum_{j=1}^{J} \sum_{i=1}^{c} \left[(\mu_{ij}^{(t+1)})^m \cdot \left\| \mathbf{v}_{s_j} - \mathbf{v}_i \right\|^2 \right]. \tag{24}$$

Step 7: If $\left| J_m^{(t+1)} - J_m^{(t)} \right| \geq \varepsilon$ (a threshold), then t = t + 1, go to step 4.

Step 8: Stop

Applying the FCM algorithm requires the determination of the optimal cluster number, that is, to treat the cluster validity problem. In this study, the *WB* index proposed in (Tan, 2000) was applied to solve it. The *WB* index has the purpose of finding cluster number c that minimizes the intra-group variance $W(\mu, \mathbf{v})$ and maximizes the inter-group variance $B(\mu, \mathbf{v})$. That is, to find c_{opt} such that

$$c_{opt} = \arg\max_{c} WB = \arg\max_{c} \frac{B(\mu, \mathbf{v})}{W(\mu, \mathbf{v})}. \tag{25}$$

The two terms $W(\mu, \mathbf{v})$ and B(μ,v) are defined as

$$W(\mu, \mathbf{v}) = \sum_{i=1}^{c} \sum_{j=1}^{J} (C_i)^{-1} \left\| \mathbf{v}_{s_j} - \mathbf{v}_i \right\|^2, \tag{26}$$

$$B(\mu, \mathbf{v}) = \frac{1}{C_2^c} \sum_{\lambda=1}^{c-1} \sum_{m=\lambda+1}^{c} \left(\frac{\sum_{j \in S_\lambda \cup S_m} (\mu_{\lambda j} \cdot \mu_{mj})}{|S_\lambda| + |S_m|} \right)^{-1} \left\| \mathbf{v}_\lambda - \mathbf{v}_m \right\|^2, \tag{27}$$

in which

$$C_i = \frac{\sum_{j=1}^{J} \mu_{ij}^2}{|S_i|}, \quad i = 1,2,\ldots,c, \tag{28}$$

where

$$S_i = \{ j \mid I_{ij} = 1 \}, \tag{29}$$

and

$$I_{ij} = \begin{cases} 1, & \text{if } \mu_{ij} = \max_{1 \le k \le c} \mu_{kj} \\ 0, & \text{otherwise} \end{cases}, \tag{30}$$

and C_2^c is a combination computation.

After the FCM clustering, several mean vectors were obtained as the feature vectors of each bird species. Before applying them, the LDA was applied again to extract the principle components of the feature vector and improve the recognition rate. In the following, the song of Pallas's Leaf Warbler was used as an example for the feature extraction process.

2.3. Recognition

In the recognition process, after the same feature extraction procedure (without the clustering process), as shown in Fig. 1.1, the feature vector of a testing syllable was matched to those of the template bird species. A template bird species usually has several syllable feature vectors and so do the matching degrees defined by the inverse of the Euclidean distance between the feature vectors of the testing syllable and the template syllables. The recognition of the testing syllable was accomplished by finding the template bird species that had syllable with the largest matching degree.

3. Experimental Results

The bird species vocalization database used in this study was obtained from a commercial CD (Kabaya and Matsuda, 2001) containing both birdcall and birdsong files of 420 bird species recorded in the field in Japan. Each file contains vocalizations of the same bird species. The sampling rate of these vocalization signals was 44.1 kHz with 16-bit resolution and a monotone type PCM format.

In the experiment, the frame size was set as 512 samples with one-half frame overlapping. Half the syllables of each birdsong file were randomly selected for training and the remaining for testing. The recognition rate *RR* was defined as

$$RR(\%) = \frac{\text{number of syllables recognized correctly}}{\text{number of all syllables}} \cdot 100\% \qquad (31)$$

The proposed two-stage structure shown in Fig. 1.1 performed LDA of QI and QE before the FCM clustering. Usually the threshold θ used in the LDA is set as 0.95. In this experiment various values from 0.6 to 0.95 were tested to examine the *RRs*. The *RRs* and corresponding feature dimensions using the proposed structure are shown in Table 2.1. It can be seen that when θ was 0.95, the *RR* of the feature mean was increased from 79.52% to about 82% if QI, QE or both was added. In addition, a *RR* of 83.3% was achieved and the feature dimension was reduced to 31 when θ was 0.75. For objectivity, this structure with θ equaling 0.75 was performed 20 times, and the statistics of the resulting *RRs* are shown in Table 2.2. Table 2.2 shows that a maximum RR of 84.34% can be achieved under a relatively low standard deviation of *RRs*. Meanwhile, feature vector with dimension of 31 is more practical for real application.

4. Conclusions

The investigation of bird species diversity is the key in monitoring environment and ecosystem recovery, and automatic bird species recognition based on their songs has become an invaluable study method in the long-term investigation of bird species. In the design of a voice recognition system, a well-known feature that has been widely applied is the MFCC. Nevertheless, designing a MFCC-based birdsong recognition system requires advanced feature extraction processes for obtaining a satisfactory recognition rate because birdsongs are usually recorded in a noise environment, are incomplete or interrupted. In this study, two novel features based on the MFCCs were presented. Adding the techniques of LDA and the FCM algorithm, the mean, QI and QE were applied to develop a birdsong recognition system. The proposed system was applied for birdsong recognition with 420 bird species.

References

Anderson, S.E., Dave, A.S., Margoliash, D., 1996. Template-based automatic recognition of birdsong syllables from continuous recordings. Journal of the Acoustical Society of America. 100 (2) 1209-1219.

Bou-Ghazale, S.E., Assaleh, K., 2002. A robust endpoint detection of speech for noisy environments with application to automatic speech recognition. In: IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 4. pp. 3808-3811.

Bou-Ghazale, S.E., Hansen, J.H.L., 2000. A comparative study of traditional and newly proposed features for recognition of speech under stress. IEEE Transactions on Speech and Audio Processing. 8, 429-442.

Catchpole, C.K., Slater, P.J.B., 1995. Bird Song: Biological Themes and Variations. Cambridge University Press.

Duda, R., Hart, P., Stork, D., 2000. Pattern Classification. New York: Wiley.

Fagerlund, S., 2007. Bird species recognition using support vector machines. EURASIP Journal on Advances in Signal Processing. 2007, Article ID 38637, 8 pages.

Härmä, A., 2003. Automatic identification of bird species based on sinusoidal modeling of syllables. In: IEEE International Conference on Acoustics, Speech, Signal Processing, vol. 5. pp. 545-548.

Härmä, A., Somervuo, P., 2004. Classification of the harmonic structure in bird vocalization. In: IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 5. pp. V701-V704.

He, S.N., Yu, J.B., 2002. A novel Chinese continuous speech endpoint detection method based on time domain features of the word structure. In: IEEE International Conference on Communications, Circuits and Systems and West Sino Expositions, vol. 2. pp. 992-996.

Hung, J.W., Tsai, W.Y., 2008. Constructing Modulation Frequency Domain-Based Features for Robust Speech Recognition. IEEE Transactions on Audio, Speech, and Language Processing. 16 (3), 563-577.

Hung, W.W., Wang, H.C., 2001. On the use of weighted filter bank analysis for the derivation of robust MFCCs. IEEE Signal Processing Letters. 8, 70-73.

Kabaya, T., Matsuda, M., 2001. The Songs & Calls of 420 Birds in Japan. SHOGAKUKAN Inc., Tokyo.

Kwan, C., et al., 2006. An automated acoustic system to monitor and classify birds. EURASIP Journal on Applied Signal Processing. 2006, Article ID 96706, 1-19.

(*)Lee, C.H., **Chou, C.H.**, Han, C.C., Hunag, R. Z., 2006. Automatic recognition of animal vocalizations using averaged MFCC and linear discriminant analysis. Pattern Recognition Letters. 27 (2), 93-101.

Lee, C.H., Hyun, D.H., Choi, E.S., Go, J.W., Lee, C.Y., 2003. Optimizing feature extraction for speech recognition. IEEE Transactions on Speech and Audio Processing. 11, 80-87.

Lee, S.M., Fang, S.H., Hung, J.W., Lee, L.S., 2001. Improved MFCC feature extraction by PCA-optimized filter-bank for speech recognition. In: IEEE Workshop, Automatic Speech Recognition and Understanding. pp. 49-52.

McIlraith, A.L., Card, H.C., 1997a. Bird song identification using artificial neural networks and statistical analysis. In: Canadian Conference on Electrical and Computer Engineering, vol. 1. pp. 63-66.

McIlraith, A.L., Card, H.C., 1997b. Birdsong recognition using backpropagation and multivariate statistics. IEEE Transactions on Signal Processing. 45 (11), 2740-2748.

Minh, V.D., Lee, S.Y., 2004. PCA-based human auditory filter bank for speech recognition. In: International Conference on Signal Processing and Communications, pp. 393-397.

Rabiner, L.R., Sambur, M.R., 1975. An algorithm for determining the endpoints of isolated utterances. Bell System Technical Journal. 54 (2), 297-315.

Ricotti, L.P., 2005. Multitapering and a wavelet variant of MFCC in speech recognition. IEE Proceedings - Vision, Image and Signal Processing. Feb, 29-35.

Selouani, S.A., Kardouchi, M., Hervet, E., Roy, D., 2005. Automatic birdsong recognition based on autoregressive time-delay neural networks. In: ICSC Congress on Computational Intelligence Methods and Applications. pp. 1-6.

Skowronski, M.D., Harris, J.G., 2002. Increased MFCC filter bandwidth for noise-robust phoneme recognition. In: IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 1. pp. 801-804.

Skowronski, M.D., Harris, J.G., 2003. Improving the filter bank of a classic speech feature extraction algorithm. Circuits and Systems. 4, 281-284.

Somervuo, P., Härmä, A., 2004. Bird song recognition based on syllable pair histograms. In: IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 5. pp. V825-V828.

Somervuo, P., Harma, A., Fagerlund, S., 2006. Parametric Representations of Bird Sounds for Automatic Species Recognition. IEEE Transactions on Audio, Speech and Language Processing. 14, 2252-2263.

Takiguchi, T., Ariki, Y., 2006. Robust Feature Extraction using Kernel PCA. In: IEEE International Conference on Acoustics, Speech and Signal Processing, Vol. 1, pp. I509-I512.

Tan, J.H., 2000. On cluster validity for fuzzy clustering. Master Thesis, Applied Mathematics Department, Chung Yuan Christian University, Taiwan, R.O.C..

Wu, B.F., Wang, K.C., 2005. Robust Endpoint Detection Algorithm Based on the Adaptive Band-Partitioning Spectral Entropy in Adverse Environments. IEEE Transactions on Speech and Audio Processing. 13 (5), 762-775.

Zhang, W.J., Xie, J.Y., 2003. Endpoint detection based on MDL using subband speech satisfied auditory model. In: IEEE International Conference on Neural Networks and Signal Processing, vol. 2. pp. 892-895.

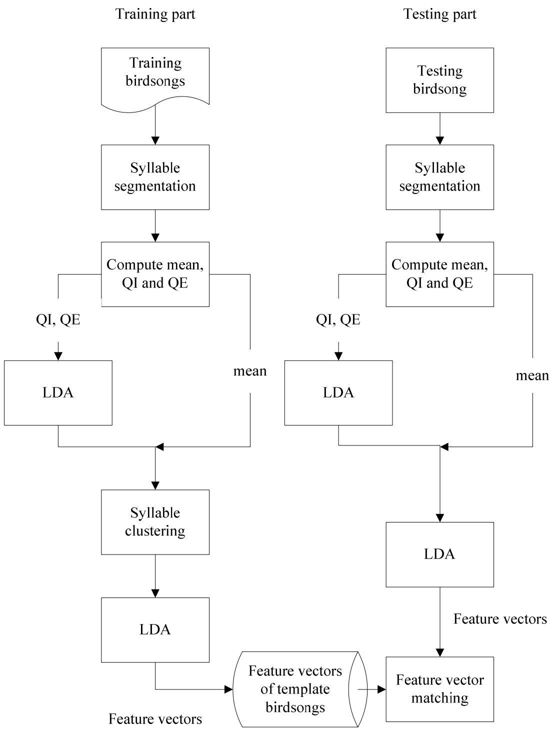

Figure 1.1 Block diagram of the proposed system

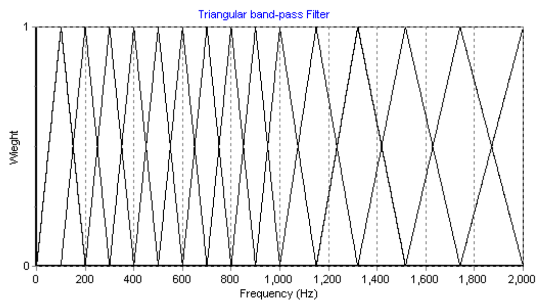

Figure 1.2 Applied triangular filters for computing the MFCCs

Table 2.1 *RRs* of using the proposed structure under various values of θ

Dim: dimension of feature vectors

features / θ	mean, QI		Mean, QE		mean, QI, QE	
	RRs	Dim	*RRs*	Dim	*RRs*	Dim
0.95	81.85	38	82.18	27	82.09	66
0.90	82.03	32	82.41	22	82.7	53
0.85	82.1	28	82.4	19	83.04	43
0.8	82.07	24	82.09	17	83.17	36
0.75	81.99	22	81.59	16	83.30	31
0.7	81.87	20	81.11	15	83.28	27
0.65	81.58	18	80.70	14	83.13	24
0.6	81.22	16	80.67	14	82.94	20

Table 2.2 Statistics of *RRs* using the proposed structure under $\theta = 0.75$

RR(%)	Max	Min	Avg	S
mean, QI	82.93	79.51	81.99	0.92
mean, QE	82.68	78.98	81.59	1.08
mean, QI, QE	84.34	81.02	83.30	0.81

Table 3.5 Comparison of LDA and PCA in the first stage dimension reduction.

Threshold of LDA or PCA in the first stage	RR(%) by using LDA	RR(%) by using PCA
0.95	82.8142	68.1191
0.9	83.3573	67.7633
0.85	83.5445	68.3376
0.8	83.8067	68.3001
0.75	83.9878	68.2627
0.7	84.0252	67.9006
0.65	83.7193	68.1129

計畫成果自評：

1. 研究內容與原計畫相符程度

計畫書中，本年度欲完成之目標：

1) 針對單一音節為單位，進行音節之MFCCs 特徵改良、開發新的音節特徵，並結合形成音節特徵向量。
2) 針對一段聲音為單位，進行MFCCs 特徵改良，並開發新的特徵向量。
3) 結合上述之特徵擷取法，與前兩年計畫所研究的音節切割法及分類器，開發完整的鳥鳴聲辨識系統。

自評：本計劃成果確實針對 MFCC 進行改良，並以 MFCC 為基礎進行特徵擷取的研究。

2. 達成預期目標情況

自評：本計劃成果提出兩個新的特徵向量，實際應用於鳥鳴聲辨識，並提升辨識效果。

3. 研究成果之學術或應用價值

自評：本方法為一個新的觀念，計算過程簡單，對於即時的辨識應用，有相當的可行性。

4. 是否適合在學術期刊發表或申請專利

自評：目前正攥寫成投稿型式，準備投稿中。

5. 主要發現或其他有關價值

自評：MFCC 為語音辨識中，最重要的聲音特徵之一。基於 MFCC 特徵的語音辨識，基本上可以得到某種程度的效果。本計劃針對每一階(order)MFCC，觀察其隨時間的變化的時間序列。由時間序列的變化特性中，萃取出特徵向量。此向量結合原本的 MFCC 係數，可以明顯的提升單純由 MFCC 所做的辨識率。

行政院國家科學委員會補助國內專家學者出席國際學術會議報告

99 年 6 月 9 日

附件三

報告人姓名	周智勳	服務機構及職稱	中華大學資訊工程系
時間 會議 地點	2010/5/16~2010/5/19 俄羅斯-海參崴	本會核定補助文號	計劃編號： NSC98-2221-E-216-033
會議名稱	（中文） （英文）The 2nd Russia-Taiwan Symposium on Methods and Tools of Parallel Programming Multicomputers (MTPP 2010)		
發表論文題目	（中文） （英文）A Compound Scheduling Strategy for Irregular Array Redistribution in Cluster Based Parallel System		

報告內容應包括下列各項：

一、參加會議經過

這是個跟演算法相關的會議，會議時間在 5/16~5/19。由於首爾至海參崴班機時間配合的問題，必須在首爾過夜後，再搭乘早晨的班機，於當地時間兩點多抵達海參崴機場，隨即至飯店 check in，並兌換當地貨幣。

會議第一天白天為參觀行程，包括西伯利亞鐵路終點站、武器博物館、自然科學博物館及潛艇等。晚上則為迎賓酒會。

圖一 迎賓酒會之照片

會議開幕及議程於第二天開始共三天，地點在遠東科技大學圖書館大樓。包括該國科學院的官員及該校的副校長皆前來致詞，感覺出對此研討會之重視。由於是新大樓，議場的設備不錯，也提供無線網路上網。論文發表包括上台報告及海報兩種，我的論文排在議程的第一天，與會者大部分來自俄羅斯及台灣、韓國及德國的學者，主要針對演算法及其應用做學術交流。議程中我也曾問了兩個問題，互動還不錯。

圖二 議場入口海報

圖三 會議室一隅

研討會第三天晚上為會議晚宴，餐廳位於十九樓，能俯視整個海參崴視野相當不錯。食物以當地特色為主，冷盤居多，過程中播放影片介紹海參崴未來的遠景，並邀請了提琴手及歌手前來表演，場面相當熱烈。研討會於第四天中午結束，除了紀念品，大會也提供交通車機場接送，相當貼心。

圖四 晚宴一隅

圖五 與主辦單位主管合影

二、與會心得

1. 會場的佈置似乎不像國內辦研討會的熱鬧，國內辦研討會，會花不少心思在會場佈置上，感覺比較有那麼個氣氛。
2. 此次研討會沒有參展的攤位，不太能了解該地區的科技產業如何。
3. 國內也常舉辦國際性研討會，可以於議程中安排半日遊，讓外國學者增加認識台灣的機會，或於晚宴時安排有代表性的表演，如此對推展觀光也許有一些幫助。
4. 近年來大陸方面參加研討會的學者漸多，國內方面出國留學的學生日漸減少，因此應該鼓勵國內的研究生，參加國際研討會。

三、考察參觀活動(無是項活動者省略)

包括海港、西伯利亞鐵路終點站、武器博物館、自然科學博物館及潛艇內部等，算是相當特別，有歷史意義的景點。

四、建議

1. 大會提供的交通及餐飲算是充足，值得學習，但是接待方面，時間拿捏的不是很好，有時候會 delay。

五、攜回資料名稱及內容

Proceeding 直接於 IEEE 網頁中。

六、其他

1. 俄羅斯人不太講英文，晚上買東西時，不太好溝通。
2. 晚宴採歐洲家庭自助式方式，蠻有意思的。
3. 海參崴 150 年前曾是中國的領土，博物館中感覺得出來，但市容上感覺不太出來。
4. 海參崴正在大興土木，幾年後會有新風貌，但可能也會減少原來的俄羅斯氣息。
5. 感謝國科會工程處的補助。

無衍生研發成果推廣資料

98 年度專題研究計畫研究成果彙整表

計畫主持人：周智勳			計畫編號：98-2221-E-216-033-			
計畫名稱：以鳥鳴聲為基礎之鳥種辨識系統架構之研究與開發(III)						
成果項目			量化			單位	備註（質化說明：如數個計畫共同成果、成果列為該期刊之封面故事...等）

成果項目			實際已達成數（被接受或已發表）	預期總達成數(含實際已達成數)	本計畫實際貢獻百分比	單位	備註（質化說明：如數個計畫共同成果、成果列為該期刊之封面故事...等）
國內	論文著作	期刊論文	0	0	100%	篇	
		研究報告/技術報告	0	0	100%		
		研討會論文	0	1	100%		
		專書	0	0	100%		
	專利	申請中件數	0	0	100%	件	
		已獲得件數	0	0	100%		
	技術移轉	件數	0	0	100%	件	
		權利金	0	0	100%	千元	
	參與計畫人力（本國籍）	碩士生	50	0	100%	人次	
		博士生	0	0	100%		
		博士後研究員	0	0	100%		
		專任助理	0	0	100%		
國外	論文著作	期刊論文	0	1	100%	篇	
		研究報告/技術報告	0	0	100%		
		研討會論文	0	1	100%		
		專書	0	0	100%	章/本	
	專利	申請中件數	0	0	100%	件	
		已獲得件數	0	0	100%		
	技術移轉	件數	0	0	100%	件	
		權利金	0	0	100%	千元	
	參與計畫人力（外國籍）	碩士生	0	0	100%	人次	
		博士生	0	0	100%		
		博士後研究員	0	0	100%		
		專任助理	0	0	100%		

其他成果 (無法以量化表達之成果如辦理學術活動、獲得獎項、重要國際合作、研究成果國際影響力及其他協助產業技術發展之具體效益事項等，請以文字敘述填列。)	無

	成果項目	量化	名稱或內容性質簡述
科教處計畫加填項目	測驗工具(含質性與量性)	0	
	課程/模組	0	
	電腦及網路系統或工具	0	
	教材	0	
	舉辦之活動/競賽	0	
	研討會/工作坊	0	
	電子報、網站	0	
	計畫成果推廣之參與（閱聽）人數	0	

國科會補助專題研究計畫成果報告自評表

請就研究內容與原計畫相符程度、達成預期目標情況、研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）、是否適合在學術期刊發表或申請專利、主要發現或其他有關價值等，作一綜合評估。

1. 請就研究內容與原計畫相符程度、達成預期目標情況作一綜合評估

■達成目標

□未達成目標（請說明，以 100 字為限）

□實驗失敗

□因故實驗中斷

□其他原因

說明：

2. 研究成果在學術期刊發表或申請專利等情形：

論文：□已發表 ■未發表之文稿 □撰寫中 □無

專利：□已獲得 □申請中 ■無

技轉：□已技轉 □洽談中 ■無

其他：（以 100 字為限）

3. 請依學術成就、技術創新、社會影響等方面，評估研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）（以 500 字為限）

MFCC 為語音辨識中，最重要的聲音特徵之一。基於 MFCC 特徵的語音辨識，基本上可以得到某種程度的效果。本計劃針對每一階(order)MFCC，觀察其隨時間的變化的時間序列。由時間序列的變化特性中，萃取出特徵向量。此向量結合原本的 MFCC 係數，可以明顯的提升單純由 MFCC 所做的辨識率。

本方法為一個新的觀念，計算過程簡單，對於即時的辨識應用，有相當的可行性。