AUTOMATIC IDENTIFICATION OF BIRD CALLS USING SPECTRAL ENSEMBLE AVERAGE VOICE PRINTS

Hemant Tyagi, Rajesh M. Hegde, Hema A. Murthy and Anil Prabhakar

Indian Institute of Technology Madras
Chennai, 600036, India
email: rajesh,hema@lantana.tenet.res.in
web: www.lantana.tenet.res.in

ABSTRACT

Automatic identification of bird calls without manual intervention has been a challenging task for meaningful research on the taxonomy and monitoring of bird migrations in ornithology. In this paper we apply several techniques used in speech recognition to the automatic identification of bird calls. A new technique which computes the ensemble average on the FFT spectrum is proposed for identification of bird calls. This ensemble average is computed on the FFT spectrum of each bird and is called the Spectral Ensemble Average Voice Print (SEAV) of that particular bird. The SEAV of various birds are computed and are found to be different when compared to each other. A database of bird calls is created from the available recordings of fifteen bird species. The SEAV is then used for the identification of bird calls from this database. The results of identification using SEAV are then compared against the results derived from common classifiers used in speech recognition like dynamic time warping (DTW), Gaussian mixture modeling (GMM). A one level and two level classifier combination is also tried by combining SEAV classifier with the DTW classifier. The SEAV is computationally less expensive when compared to DTW or the GMM based classifiers while performing better than the DTW technique. Several new possibilities in automatic bird call identification using SEAV are also listed.

1. INTRODUCTION

Automatic identification of bird calls from continuous recordings gathered from the field assume significance in biological studies and ornithology [1, 2]. Often these recordings are noisy or clipped which calls for the use of reliable techniques which are automatic rather than the conventional manual techniques. Manual inspection of spectrograph's is often error prone and involves multiple human experts which makes the identification unreliable. Hence there is a need for automated analysis techniques which generate reliable constituent label for each bird [3]. It is true that human and bird vocalizations are quite different. The conditions underwhich the recordings of humans and birds are carried out are also different. But the relative simplicity of bird vocalizations when compared to human vocalizations facilitates the use of simpler human speech recognition techniques to identify bird calls. The complexity of the human vocalization when compared to a typical bird vocalization is shown in Figure 1 and Figure 2. The bird call collected from the bird *koel* is shown in Figure 1 (a), while the corresponding narrow band and wide band spectrograms are shown in Figure 1 (b) and Figure 1 (c) respectively. In Figure 2 (a) is shown a typical speech signal. The corresponding narrow band and wide band spectrograms are shown in Figure 2 (b) and Figure 2 (c) respectively. It is worthwhile to note from the spectrograms in Figure 1 and Figure 2, that the frequency distributions corresponding to human vocalizations are far more complex than the bird vocalizations. It is significant to note that bird calls have been studied using mordern signal processing techniques both in the time and the frequency domains, assuming a bio-acoustic model of avian sound production [2]. Majority of birds produce sounds as a result of sound waves originating from channels of air flow within the the syrinx which is an organ located in the intersection of the main bronchi of the lungs and the trachea. The vibrations of the membranes within a birds syrinx which produce bird calls is very similar to the action of the human vocal chords which produce vowel sounds. Various signal processing techniques used in human speech recognition like the FFT spectrum, spectrograms [4], dynamic time warping (DTW) [1, 5, 6], Wigner-Ville distribution (WVD) [7], and hidden Markov models (HMM) [3], have been used for identification of bird calls. In this paper we

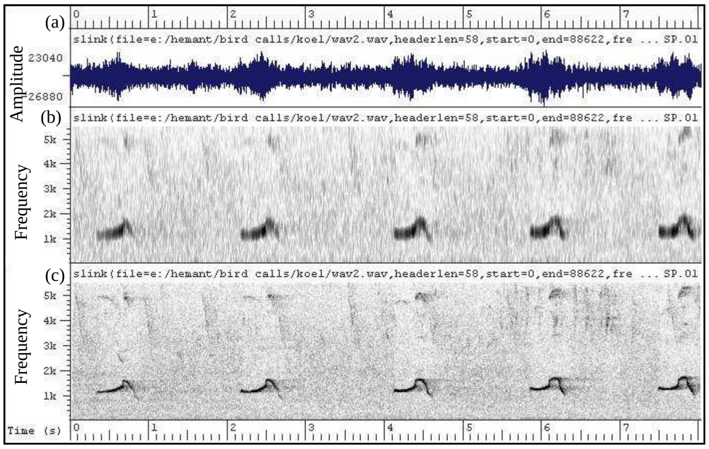

Figure 1: Wide-band and narrow-band spectrograms of a typical bird call. a) a typical bird call signal, b) the wide-band spectrogram of the signal shown in (a), and c) the narrow-band spectrogram of the signal shown in (a).

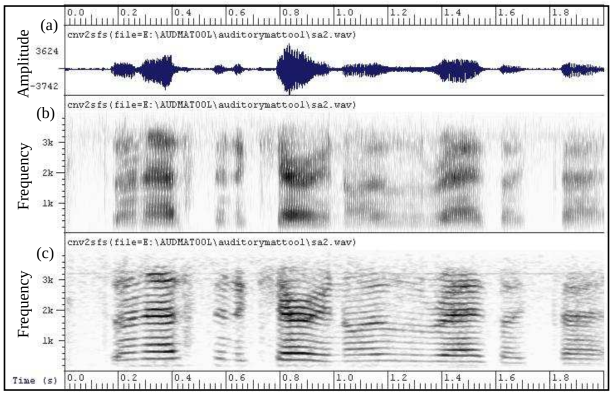

Figure 2: Wide-band and narrow-band spectrograms of a typical human speech signal. a) a typical speech signal, b) the wide-band spectrogram of the signal shown in (a), and c) the narrow-band spectrogram of the signal shown in (a).

propose a new technique that uses the ensemble average computed on the FFT spectrum to build a system that automatically identifies bird calls.

We begin with a discussion on the preparation of the database of individual bird calls from a set of recordings collected from fifteen species of birds. The recordings are available at different sampling rates. They are first re sampled such that all the calls are converted to the same sampling rate. Regions of activity corresponding to each call within a single bird are then detected automatically using both frame energy and the zero crossing rate. Each such region of activity is considered as a template belonging to that bird species. Many such templates derived from each bird are used as patterns for training and testing. The procedure of computing an ensemble average on the FFT spectrum for each bird call is then described. The spectral ensemble average thus derived for each bird is called the spectral ensemble average voice print (SEAV) of that particular bird. The SEAV is then used to identify bird calls using a simple distance measure. The performance results of identifying birds using SEAV are also compared to performance results obtained using a modified version of the DTW algorithm. The results are also compared to that derived using the Gaussian mixture modeling (GMM) technique using various features like the MFCC, PLPCC and also RASTA-PLPCC. The SEAV classification results are also combined with the DTW classification results at both measurement and rank level. A two level classifier combination of (SEAV + DTW), first at rank level and then at measurement level is also tried, to further improve the recognition performance. The paper concludes with a discussion on the significance of the SEAV in automatic identifiaction of bird calls and future scope of the work presented.

2. SPECTRAL ENSEMBLE AVERAGE VOICE PRINTS (SEAV)

Conventional techniques used in automatic bird identification use the technique of Dynamic time warping and Hidden Markov Models. Both these techniques are computationally expensive. We propose a new technique called the Spectral Ensemble Average Voice Print (SEAV) for the automatic identification of bird calls. The bird call is first windowed into a number of frames with a frame rate of 20 ms and an overlap of 10 ms. A Hamming window is used. For each frame a N point FFT is computed. The value of N is computed as equal to or greater than the product of the sampling rate and the window length. An ensemble average is then computed on the FFT spectrum, by taking the average of the corresponding FFT co-efficients of each frame of the bird call. Hence an ensemble average vector of length $N/2$ is computed for an N point FFT spectrum. The ensemble average vector thus derived is called the Spectral Ensemble Average Voice Prints (SEAV) of that particular bird. The algorithm for computing the SEAV is as follows

- Enframe the bird call signal with a frame size of 20 ms and frame rate of 10 ms
- Compute the N point FFT of each frame of the windowed bird call signal x(m) as

  $$X(k) = \sum_{m=0}^{N-1} x(m) e^{j2\pi m \frac{k}{N}} \qquad (1)$$

  where k = 0,1,..., (N-1).
- Compute the ensemble average of the FFT spectrum across all the frames . If there are J frames in the bird call signal then the spectral ensemble average (SEAV) is computed as
  $X_{seav}(0) = X_1(0) + X_2(0) + X_3(0) + ... + X_J(0)$
  $X_{seav}(1) = X_1(1) + X_2(1) + X_3(1) + ... + X_J(1)$
  .... = ... + ... + ... + ... + ...
  $X_{seav}(N/2-1) = X_1(N/2-1) + X_2(N/2-1) + ... + X_J(N/2-1)$
- The vector $\{X_{seav}(0), X_{seav}(1), ..., X_{seav}(N/2-1)\}$ of length N/2 is the SEAV corresponding to each bird.

The SEAV of various birds are illustrated in Figure 3. The value of N is equal to 1024 in all the illustrations in Figure 3. It is significant to note from Figure 3, that the SEAV for various birds is different when compared to each other. This makes the SEAV a potentially suitable candidate for automatic bird call identification.

3. PRE-PROCESSING THE RAW BIRD CALL DATA AND DATABASE PREPARATION

The collection of bird calls used in this work are recorded from fifteen birds. Each bird has recordings ranging from three to seven in number. The recordings are done at different sampling rates ranging from 8 KHz to 44.1 KHz. The recordings are first resampled to a uniform sampling rate of 8 KHz. It was also clear from the spectrograms of each bird call recording that each of these calls could be split into several template calls by separating the regions of actual bird calls from the silence and noisy regions (background wind, chirping of other birds etc.) of the recordings. Both the energy and the

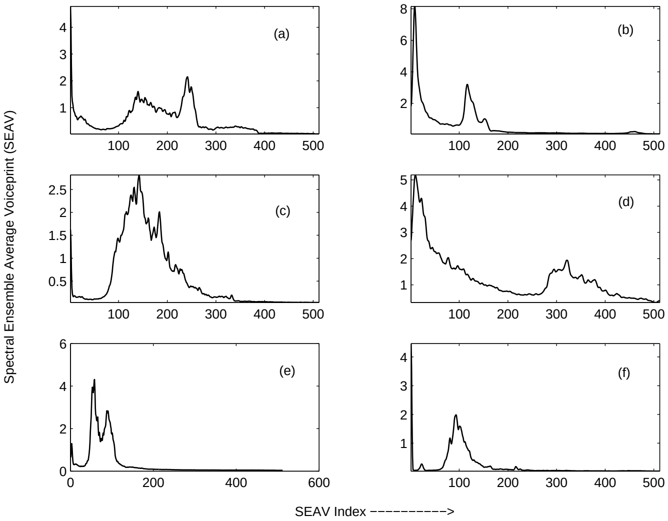

Figure 3: The Spectral Ensemble Average Voice Prints (SEAV) of Various birds. a) Hairy Woodpecker, b) Koel, c) White Eyed Vireo, d) Ashy Drongo, e) Bluebird, and f) Blackbird.

zero crossing rate function were used to separate the regions of noise and silence from the actual call. Figre 4, illustrates the separation of the bird call template regions from the complete bird call recording, using the energy function.

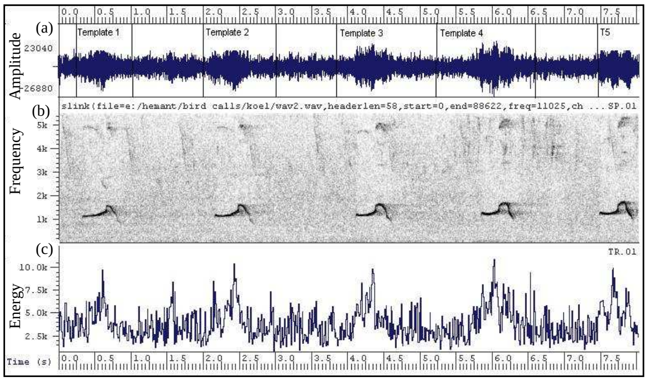

Figure 4: Bird call template preparation using the energy function. a) The complete bird call, b) Spectrogram of the bird call in (a), and c) The energy function

After such a separation from each bird call recording, several template bird calls were generated. Table 1, shows the details of the database of template bird calls, created after applying the aforementioned preprocessing technique.

Table 1: Database of bird calls. Second column indicates the original number of recordings and the third column indicates the number of template calls generated after pre-processing

Name of the Bird	No. of recordings	No. of templates
Hairy Woodpecker (HW)	3	8
Koel (K)	3	10
Magpie Robin (MR)	3	11
White Eyed Vireo (WEV)	8	14
Ashy Drongo (AD)	3	10
Black Capped Chickadee (BCC)	3	2
Blackbird (BB)	4	10
Blue bird (BL)	7	15
Chestnut Bulbul (CB)	4	9
Common Tailor Bird (TB)	3	6
Great Tit (GT)	4	9
Hermit Thrush (HT)	6	18
Jamaican WEV(JW)	5	7
Northern Parula Warbler (PB)	6	6
Pileated Woodpecker (PW)	5	5

4. PERFORMANCE EVALUATION

In this section the results of performance evaluation of automatic bird call identification using the SEAV and various other techniques are presented. The results of the SEAV technique on the database described in Section 3, are compared with that of the dynamic time warping (DTW) and the gaussian mixture modeling techniques. The results are also improved by using classifier combination techniques. A novel two level classifier combination is also attempted to further improve the recognition performance.

4.1 The SEAV technique

The SEAV is computed for each bird as described in Section 2, using the reference templates. During testing the SEAV of the test template corresponding to the bird being identified is computed first. A simple Euclidean distance between the SEAV of the reference and the test template is then computed. The Euclidean distance between the test SEAV and the reference SEAVs of various birds in the database is evaluated. The reference template that gives the minimum Euclidean distance is declared as the correct match.

4.2 The Dynamic time warping technique

The Dynamic Time Warping (DTW) technique is widely used to compare two speech templates by addressing the problem of time alignment by non linearly stretching one of the templates in order to synchronize similar acoustic segments in the test and reference templates using the Bellman optimality principle. In this work we first compute the spectrograms of the test and the reference bird calls. The spectrogram matrices are used as the templates in the succeeding warping process. A similarity matrix is formed from the test and reference templates such that each element of the matrix tends to take a zero value when the test and the reference templates are similar in terms of the Euclidean distance. During the testing phase, a warping distance is calculated between the test bird call and all the reference template calls. The reference template which gives the minimum warping distance is declared as the correct match.

4.3 Gaussian mixture modeling

The third technique used in this work for identification of bird calls is the Gaussian mixture modeling technique (GMM) [4]. In this method we generate a GMM for each bird in the database using the template calls in the database. Various features like the Mel frequency cepstral co-efficients (MFCC), perceptual linear prediction co-efficients (PLP), and Relative-Spectral PLP (RASTA-PLP) are tried in this work. The dimensionality of the feature is varied from 6 to 13 to evaluate the its significance in the bird call identification scenario. During the testing phase, features are extracted from the test bird call template and the probability that the feature belongs to one of the bird classes in the database evaluated. The bird class that gives the highest probability is declared as the correct match.

4.4 Experimental results

The results of bird call identification using the SEAV, DTW, and the GMM techniques is listed in Table 2. The recognition performance quoted here is for four tests on fifteen birds amounting to a total of sixty tests. It is significant to note that the SEAV performs better than the DTW. The GMM technique as expected performs the best among the three techniques. It is also significant to note that among the various features the MFCC gives the best performance when compared to that of the PLP and the RASTA-PLP.

Table 2: Results of bird call identification using the SEAV, DTW, and the GMM techniques

Identification Technique	% Recognition Performance
SEAV	87
DTW	67
GMM with MFCC	100
GMM with PLP	54
GMM with RASTA-PLP	47

4.5 Experimental results using classifier combination techniques

In an effort to investigate the complementary nature of the DTW and the SEAV techniques, classifier combination techniques both at rank level and measurement level have been attempted in this work. A two level classifier combination is also tried in this paper.

4.5.1 Classifier combination at rank and measurement level

The DTW and the SEAV classifiers are combined at rank level by considering the first three ranks of the individual classifier decisions. In this method, the top three reference templates, ranked according to the euclidean distances between the test and the reference template, as indicated by the DTW and SEAV, are considered. From these two sets of templates a score was assigned to each bird based on its rank. If a bird occured twice (once in each set) then its total score is equal to the sum of its two scores. The bird with a minimum score is identified as the correct match.

In the combination at measurement level the procedure adopted is very similar to the combination at rank level. The first three Euclidean distances output by each classifier are considered. If a particular bird is present in both sets, as classified by the two classifiers, then its score is calculated as the mean of the weighted sum of its normalized Euclidean distance values. The bird with a minimum score is identified as the correct match.

4.5.2 Two level classifier combination

Another approach was tried in which the combination of classifiers was first done at the rank level and the scores calculated for all the birds which were common among the two sets as classified by the SEAV and DTW. In case of a tie between two or more birds a measurement level combination was done only for those birds. Finally the bird with a minimum score is identified as the correct match.

4.5.3 Results

The results of bird identification using different classifier combination techniques is listed in Table 3. It is significant to note that there is an improvement in recognition performance due to a combination at rank level, while there is degradation due to a combination at measurement level. There is also a significant improvement due to a two level classifier combination.

Table 3: Results of bird call identification using various classifier combination techniques

Combination Technique	% Improvement in Recognition
SEAV + DTW RANK LEVEL	+ 5.7
SEAV + DTW MEASUREMENT LEVEL	- 5.7
SEAV + DTW TWO LEVEL	11.4

5. CONCLUSIONS

This paper looks at automatic bird identification techniques from a signal processing and pattern recognition perspective. A new technique based on the Spectral ensemble average voice prints is proposed for automatic bird call identification. The results using such an approach are compared with other conventional approaches like the DTW and the GMM techniques. The SEAV is found to be computationally inexpensive and performs better than the DTW technique. Classifier combination techniques at rank and measurement level are tried. A two level classifier combination is also attempted. The SEAVs of various birds are considerably different from each other, which calls for further analysis to see if this technique could be used to identify bird songs based on syllable identification. There is also a need to analyze the theoretical basis of using the SEAV for various bird recognition tasks.

REFERENCES

[1] J. A. Kogan and D. Margoliash, "Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden Markov models : A comparative study," *J. Acoust. Soc. Amer.*, vol. 103(4), pp. 2185–2196, April 1998.

[2] A. Harma, "Automatic identification of bird species based on sinusoidal modeling of syllables," in *Proceedings of IEEE Int. Conf. Acoust., Speech, and Signal Processing*, Hong Kong, April 2003, vol. 5, pp. 545–548.

[3] J. A. Kogan and D. Margoliash, "Automated bird song recognition using dynamic warping and hidden Markov Models," *J. Acoust. Soc. Amer.*, vol. 102(5), pp. 3176, November 1997.

[4] Douglas O' Shaughnessy, *Speech communications*, Universities press, Univ. press(India) limited, India, 2001.

[5] S. E. Anderson and A. S. Dave, and D. Margoliash, "Template-based automatic recognition of bird song syllables from continuous recordings," *J. Acoust. Soc. Amer.*, vol. 100(2), pp. 1209–1219, August 1996.

[6] S. E. Anderson, C. A. Staicer, S. Inoue, and D. Margoliash, "Objective analysis of song learning in birds : Towards automated techniques," *J. Acoust. Soc. Amer.*, vol. 99(4), pp. 2534–2574, April 1996.

[7] C. Rogers, "High resolution analysis of bird sounds," in *Proceedings of IEEE Int. Conf. Acoust., Speech, and Signal Processing*, Detroit, May 1995, vol. 5, pp. 3011–3014.