COMPARISON OF METHODS FOR AUTOMATED RECOGNITION OF AVIAN NOCTURNAL FLIGHT CALLS

Mathieu Marcarini, Geoffrey A. Williamson, and Luis de Sisternes Garcia

Department of Electrical and Computer Engineering
Illinois Institute of Technology
Chicago, IL 60616 USA
{mmarcari,williamson,ldesist}@iit.edu

ABSTRACT

Many species of birds in the Americas vocalize during nocturnal migration flights. Acoustic detection and classification of the calls shows potential for study of the natural history of these migrant birds. In particular, information about the species composition and number of birds involved in migration movements may be obtainable through acoustic techniques. Other methods such as radar monitoring may have capability only to assess the number, but not composition. Here the feasibility of classifying nocturnal flight calls of birds in the family *Parulidae* using spectragram correlation and using Gaussian mixture models of Mel frequency cepstral coefficient distributions are evaluated and compared. Tests performed on a set of recorded calls show that the techniques are complementary and may, with improvement, enable automated detection.

***Index Terms*—** Acoustic signal analysis, animals, pattern classification.

1. INTRODUCTION

Automated recognition of animal sounds is useful in a variety of situations, including population monitoring, the study of animal behavior, and prevention of harmful human/animal interactions. The cross-correlation of spectrograms has been used to recognize marine mammal vocalizations [1], [2] for the purpose of studying the animals' range and distribution. Statistical classifiers applied to various characteristics of marine mammal sounds have also been proposed [3]. Machine identification of bird sounds has been suggested to help prevent bird/aircraft collisions at airfields [4], with the classification based on speech analysis techniques.

Another application of automated acoustic monitoring is in regard to nocturnal flight calls of migrant birds [5], [6]. The determination of the species of birds to which various noctural sounds belong has only recently been made [7], thus now opening the possibility of automated classification. Automated acoustic detection of nocturnal bird calls in conjunction with other methods such as radar monitoring shows potential to provide information about the number of birds involved in migration movements, and acoustic monitoring can provide information about relative species composition in such movements while other techniques may not [8]. However, we are aware of only one attempt to automatically classify nocturnal flight calls to species: a work based on spectrogram features published only as a conference abstract [9].

Here we perform an initial study comparing two methods to classify nocturnal flight calls of nine species of birds in the family *Parulidae*, the North American warblers. The first method uses a time frequency representation of the bird call as the basic description of the audio signal. Though other time frequency representations may be more appropriate for analyzing bird vocalizations, the spectrogram is a tool commonly used by ornithologists to represent avian vocalizations. We use the correlation of spectrograms with a template spectrogram to detect bird calls. Our second method is essentially the approach used in [4], used there to classify more complex vocalizations of birds. In this approach, a Gaussian mixture model (GMM) for Mel frequency cepstral coefficients (MFCCs) utilized as the basis for the classifier.

In the following we describe the classification methods that we apply and the results that they provide on a set of test data. We conclude with a comparision of the approaches.

2. CLASSIFICATION METHODS

2.1. Spectrogram Correlation

The length of the audio files that contain bird calls were standardized to 3300 samples at 22.050 kHz (149.6 msec). The spectrogram correlation method applied to each such call proceeds in the following steps.

- Prefiltering: the signal is prefiltered using a bandpass filter with bandedges set to the upper and lower frequecy limits associated with the call that is being detected. The filter used is a 200-order FIR filter designed via Chebyshev approximation methods.

- Spectrogram calculation: we compute spectrograms using 1024-point discrete Fourier transforms (DFTs) on windows whose width is 40 samples (1.814 msec at the 22.05 kHz sampling rate). Consecutive windows are overlapped by 50%. Therefore, frequency components are calculated at time points spaced 20 samples (0.907 ms) apart.

- Scaling: the spectrograms are scaled so that the maximum value achieved equals 1000. The spectrogram values then lie in the interval $[0, 1000]$.

- Threshold saturation: the raw spectrogram is saturated to binary values ($+1$: signal present, -1: signal absent) using an iterative thresholding technique. The iterative thresholding proceeds as follows.

  - Step 1: select an arbitrary threshold value.
  - Step 2: separate the spectrogram in two groups – a group with energy lower than the threshold and group with higher energy than the threshold.
  - Step 3: Calculate the means of the energies in the first group, then in the second group. Average those means to produce the new threshold. Return to Step 2 until the threshold stabilizes.

- Correlation detection: the saturated spectrogram is correlated with a template signal. Correlation values above a threshold value connote a positive detection.

Figure 1 shows the spectrogram of an american redstart call together with its saturated version.

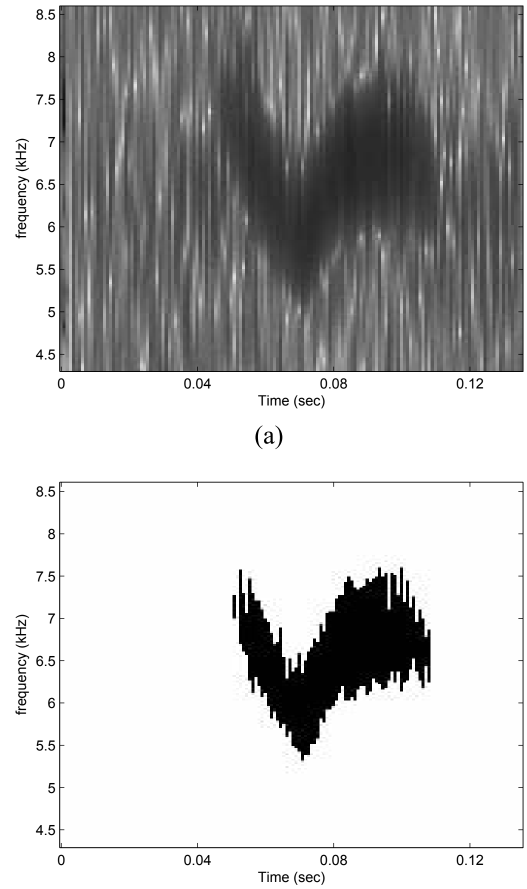

Fig. 1. (a) Spectrogram of an american redstart call. (b) Saturated version of the spectrogram.

2.2. Classification Using Gaussian Mixture Models for Mel Frequency Cepstral Coefficients

In this approach to classification of the bird calls, we following the basic method described in [4]. For each bird species, a Gaussian mixture model (GMM) is trained using a subset of the call data for that species. The GMM represents the probability density of Mel frequency cepstral coefficients (MFCCs) computed from short segments of the bird call. Each GMM is a mixture of M Gaussians

$$p(\boldsymbol{x}) = \sum_{i=1}^{M} \alpha_i \mathcal{N}(\boldsymbol{x}; \mu_i, \Sigma_i) \tag{1}$$

where $\mathcal{N}(\boldsymbol{x}; \mu, \Sigma)$ denotes a Gaussian density function with mean μ and covariance Σ. The $\{\alpha_i\}$ are weighting coefficients such that $\sum_{i=1}^{M} \alpha_i = 1$. The MFCC vector $\boldsymbol{x}$ is computed for a frame consisting of $K = 110$ samples (5 msec duration at 22.05 kHz) as

$$\boldsymbol{x} = \boldsymbol{C}\boldsymbol{y}, \tag{2}$$

where $\boldsymbol{C}$ is a discrete cosine transform (DCT) matrix and $\boldsymbol{y}$ has components

$$y_i = \log\left(\sum_{k=1}^{K} |X(k)|^2 b_i(k)\right), \tag{3}$$

where $X(k)$ is the discrete Fourier transform (DFT) of the frame and $b_i(k)$ are the magnitude response coefficients of the ith triangular bandpass filter defined on the Mel frequency scale. We use the first 13 Mel frequency coefficients ($i = 1, \ldots, 13$). Prior to the calculation of the MFCCs, the audio signals are filtered with bandpass filters with species specific band edges, in the same fashion as was done in the spectrogram correlation method.

For the training subset of the bird calls, we extract frames from each call with an overlap of 90% between successive frame samples. MFCCs are computed for each frame, and we then train the GMMs using the `gmmem.m` routine of Netlab [10]. The training produces for each species of warbler

Species	Code	# of calls
American Redstart	ar	102
Black-throated Blue Warbler	bb	69
Blackpoll Warbler	bp	97
Black-and-white Warbler	bw	31
Cape May Warbler	cm	67
Common Yellowthroat	cy	87
Northern Waterthrush	nw	73
Ovenbird	ob	90
Northern Parula	p	62

Table 1. Data set – warbler type and number of calls.

a set of values $\{\alpha_i\}$, μ_i, and Σ_i (for $i = 1, \ldots, M$) for the GMM in (1).

Once the GMMs have been developed, classification of a bird call is straightforward, assuming that one knows that the bird call belongs to one of the nine species that have been modeled. The technique that we apply proceeds as follows. We presume that the start and end points of the bird call within the audio signal have been identified, and then we segment the call into a sequence of frames as described above (5 msec in duration with 90% overlap between successive frames). From the sequence of frames we compute a sequence $\{\boldsymbol{x}_k\}$ of MFCC vectors. The density functions of the nine GMMs are evaluated for each $\boldsymbol{x}_k$ and are then ranked in decreasing order. For each frame index k, we assign points for the top three ranked species: seven points for the highest ranked, three points for the second highest ranked, and one point for the third. The points are totalled over the sequence of frames, and call is classified as being produced by the species with the highest total of points.

3. RESULTS

The data set on which this work was performed consists of audio recordings of nocturnal flight calls of nine different species of warblers. These calls were recorded, digitized, and identified by William R. Evans, who provided the audio files to the authors. The audio signals are sampled at 22.05 kHz with 16-bit resolution. Each has a duration of about 0.25 sec. We show the species type and the number of calls for that species in Table 1. We determined the upper and lower frequencies of the range of the call signature by visual inspection, and we report these in Table 2.

We tested the spectrogram correlation method by determining a detection rate for a given false alarm rate, and varying the false alarm rate from 1% to 10% in steps of 1%. The false alarm rate was set by choosing the detection threshold to be the value that would produce the given false alarm rate. The detection threshold was set separately for each species. Because the detection and false alarm rates depend on what is used as the template spectrogram, we computed the performance for each species using each individual call as the template in turn. The call which produced the best detection rate for a given false alarm rate, when used as the template, was selected for the results presented below. The achieved detection rate values as functions of false alarm rate and species type are shown in Tables 3 and 4.

	Frequency range	
Species	f_ℓ (kHz)	f_u (kHz)
ar	4.31	8.61
bb	5.17	9.47
bp	5.60	8.83
bw	5.17	9.04
cm	5.38	9.04
cy	4.31	8.61
nw	5.38	7.97
ob	5.60	9.04
p	5.17	8.61

Table 2. Band edges of the frequency range occupied by the spectral energy of the warbler calls.

	Species				
false alarm rate	**ar**	**bb**	**bp**	**bw**	**cm**
10%	66.6	89.8	23.7	38.7	64.1
9%	66.6	86.9	18.5	35.4	62.6
8%	65.6	86.9	14.4	35.4	58.2
7%	63.7	84	13.4	25.8	56.7
6%	62.7	82.6	12.3	22.5	53.7
5%	61.7	81.1	11.3	22.5	52.2
4%	61.7	79.7	7.2	19.3	44.7
3%	58.8	79.7	5.1	12.9	40.2
2%	58.8	78.2	1	9.6	32.8
1%	55.8	78.2	1	3.2	23.8

Table 3. Results for the spectrogram method, first five species.

To test the GMM method, we first needed to choose the number of Gaussians in the mixture. The `gmmem.m` routine used to train the GMM also allows to constrain Σ_i to be diagonal or to have full flexibility (subject to a positive semidefinite constraint). We varied the number M of Gaussian components in the model (see (1)) and used both diagonal and fully flexible covariance matrices. For the amount of data available, using $M = 15$ and constraining each Σ_i to be diagonal produced the best results.

The data sets for each warbler species were separated into calls used for training (80% of the total) and calls used for testing (20% of the total). Training calls were selected randomly from the available data. After training the GMMs, the testing calls were classified by the model, and detection rates and false alarm rates were calculated. The detection rate for

false alarm rate	Species			
	cy	nw	ob	p
10%	6.8	64.3	74.4	40.9
9%	6.8	60.2	72.2	40.9
8%	5.7	56.1	70	40.9
7%	3.4	46.5	66.6	37.7
6%	3.4	46.5	63.3	37.7
5%	2.2	45.2	56.6	36
4%	1.1	41	54.4	22.9
3%	1.1	27.3	51.1	19.6
2%	0	19.1	40	13.1
1%	0	10.9	27.7	4.9

Table 4. Results for the spectrogram method, remaining four species.

Species	detection rate (%)	false alarm rate (%)
ar	7.14	1.6
bb	33.9	1.9
bp	81.2	26.2
bw	32.1	2
cm	55.3	8.9
cy	66.6	0.6
nw	33.3	1.8
ob	81.9	11.6
p	25	4.1

Table 5. Results for the GMM/MFCC method.

a given species of warbler was determined as the number of that species' calls that were identified correctly divided by the total number of test calls for that species. The false alarm rate for a given species was computed as the number of calls of other species falsely identified as this one, divided by the total number of other species' calls. These detection rates and false alarm rates were computed and averaged over four random selections of the training set. The achieved detection rates together with the corresponding false alarm rates are shown in Table 5.

4. DISCUSSION

For four of the nine species, the spectrogram method achieves correct categorization of 50% or more of the calls with no greater than 5% false alarm. The GMM/MFCC method correctly identifies greater than 50% of four species, yet for three of those there is greater than 5% misclassification. However, its performance should improve with the availability of more than the limited amount of training data that was used here. Both methods show some promise but each needs to be improved before practical application.

Note that the results show that each method may be appropriate for certain species, and thus the two are complementary. For instance, considering false alarm rates around the value of 2% as reference for comparison, species *ar* is detected with a rate of 58.8% using the spectrogram method, and with 7.14% using the GMM/MFCC method. Something similar happens with *bb* (with 78.2% and 33.9% respectively). However for false alarm rates around 1%, *cy* is detected with a rate of 66.6% using the GMM/MFCC, and with 0% using spectrograms.

5. REFERENCES

[1] D.K. Mellinger and C.W. Clark, "Recognizing transient low-frequency whale sounds by spectrogram correlation," *J. Acoust. Soc. Am.*, vol. 107, no. 6, pp. 3518–2529, June 2000.

[2] D. Chabot, "A quantitative technique to compare and classify humpback whale (*magapter novaeangliae*) sounds," *Ethology*, vol. 77, pp. 89–102, 1988.

[3] B. Pinkowski, "Robust fourier descriptors for characterizing amplitude-modulated waveform shapes," *J. Acoust. Soc. Am.*, vol. 95, pp. 3419–3423, 1994.

[4] C. Kwan, K.C. Ho, G. Mei, Y. Li, Z. Ren, R. Xu, Y. Zhang, D. Lao, M. Stevenson, V. Stanford, and C Rochet, "An automated acoustic system to monitor and classify birds," *J. Appl. Signal Process.*, vol. 2006, pp. 1–19, 2006.

[5] R.R. Graber, "Nocturnal migration in illinois: different points of view," *Wilson Bull.*, vol. 80, pp. 36–71, 1968.

[6] W.R. Evans and D.K. Mellinger, "Monitoring grassland birds in nocturnal migration," *Studies in Avian Biol.*, vol. 19, pp. 219–229, 1999.

[7] W.R. Evans and M. O'Brien, *Flight Calls of Migratory Birds: Eastern North American Landbirds* (CD-ROM), Oldbird, Inc., Ithaca, NY, 2002.

[8] A. Farnsworth, "Flight calls and their value for future ornithological studies and conservation research," *The Auk*, vol. 122, no. 3, pp. 733–746, 2005.

[9] A. Taylor, "Bird flight call discrimination using machine learning," *J. Acoust. Soc. Am.*, vol. 97, pp. 3370(A), May 1995.

[10] I. Nabney and C. Bishop, "Netlab neural network software," `http://www.ncrg.aston.ac.uk/netlab/`, 2007.