BIRD SONG IDENTIFICATION USING ARTIFICIAL NEURAL NETWORKS AND STATISTICAL ANALYSIS

Alex L. McIlraith, Howard C. Card

Department of Electrical and Computer Engineering
University of Manitoba
Winnipeg, Manitoba
R3T 5V6
Email: mcilrth@ee.umanitoba.ca, hcard@ee.umanitoba.ca

ABSTRACT

A system for automatically identifying six bird species by their songs was implemented. Pre-processing of sampled songs extracted temporal measurements of periods of sound and silence within songs. Power spectral densities were used to extract spectral information. Statistical methods were used to reduce data dimensionality and for identification tasks. An artificial neural network was also used for identification. Quadratic discriminant analysis achieved a 93%, and a backpropagation neural network 82% overall accuracy.

1. INTRODUCTION

Birds have been of interest to people since history began to be recorded. Both the social and ecological importance of birds is reflected in the laws we have instituted to protect them. In Canada and the United States, migratory bird legislation makes it illegal to kill or collect migratory species without a permit [1]. In addition to ongoing bird research (ornithology), a growing number of people are taking up bird-watching as a hobby. Biologists are often called upon to predict or assess the impact of human development on plants and animals. In the process of executing such an assessment, the biologist may have to identify and count birds in a site. As many of the birds in an area may be heard and not seen, it is often quicker and easier to rely on sounds. Thus biologists must study the sounds of birds for that area and be able to identify them by sound alone. Since this can be a difficult task, it can be helpful to record unknown sounds so that they can be identified later.

Recently, interest has arisen in the possibility of building a device that could not only record but automatically identify bird species by their sounds. In order to fully appreciate the problem of automatic species identification using sound, background in the areas of biology, computer engineering and statistics is beneficial. Biology provides information about why and how animals make sounds. Knowing this, in turn suggests some features of sounds that might be important in distinguishing species.

The systems animals use to generate and receive signals, of which sound is one type, are the product of evolution. The successful use of a particular signal also depends on the nature of the habitat and channel conditions (e.g. the presence of noise or attenuation). It is possible make some broad generalizations [2] that are relevant to discrimination of species with sound. For example, it makes sense that different species sharing the same habitat might in some cases use signals differing in frequency and pattern in order to reduce interference. Frequency modulation is likely to predominate since it is less affected by fading and echoing. Sound repetition rates should also vary among species and could vary with channel conditions. We also know that bird species do not all use the same cues to recognize members of their own species [3]. Published data that exists on all six species used in this study (e.g. [4-6]) suggests that the degree of variation observed within a bird species probably exceeds that observed in the problem of human word recognition. Songs do vary between geographic regions, local groups or even individuals and over time.

Any method of automatic identification must be more sensitive to among-species variation than to within-species variation. Since it is reasonable to assume that some features of even the most variable songs are stereotyped, the task of automatic identification only becomes possible if the system is capable of extracting such stereotyped features. Knowing this, in turn suggests some features of sounds that might be important in distinguishing species.

Computer engineering provides the theory required to extract these features. Like human speech, which has been thoroughly studied, bird songs have characteristic temporal and spectral qualities. Power spectral density (PSD) analysis has been used successfully in the past for human speech recognition, and is also useful for analysis of bird sounds. Spectral variables combined with temporal ones can form input vectors for pattern recognition methods like

0-7803-3716-6 /97/$5.00

artificial neural networks (ANN) and discriminant analysis. These can be used to perform the ultimate task of classification and identification.

2. METHODS

In this study, 133 recorded bird songs of six species (Song Sparrow (SSP), Fox Sparrow (FOX), Marsh Wren (MWR), Sedge Wren (SWR), Yellow Warbler (YLW) and Red-winged Blackbird (RWB)) were sampled (11.025kHz, 8bits/sample) from audio tapes and compact disks [7-12]. Levels were adjusted to give a maximum amplitude without clipping.

Temporal processing of sampled data was performed using a 'leaky-integrator' algorithm that parsed songs in a such a manner that silences too short in duration to be perceptible were ignored. The number of elements (periods of sound) in a song was determined and the mean and variance of both element and silence lengths calculated. Means and standard deviations for measured quantities, and the element count constituted the five temporal variables extracted from each song. The amplitude of each element was normalized prior to spectral analysis. PSDs were calculated for each element using the Welch method [13], a triangular window and a 16-point FFT. Magnitudes for nine spectral bands were averaged across elements. Band averages were normalized with respect to the band containing the largest average to reduce the effect of amplitude variations. Means and standard deviations for the nine spectral bands constituted eighteen spectral variables for each song. Twenty-three variables were thus generated for each song. SAS software was used for all subsequent statistical analyses [14, 15] of this data set.

Preliminary examination of the data correlation structure indicated complex inter-correlation of variables. Since this is known to adversely affect analysis techniques, the number of variables had to be reduced. Stepwise discriminant analysis was performed to obtain a smaller set of variables that still contained enough information for discrimination; the significance criterion used to enter a variable and to retain it during the elimination phase was α=0.15. This method (see results) suggested that good discrimination performance could be obtained with only eight variables. Further statistical analyses (principal components analysis, canonical discriminant analysis) of the data supported the contention that the eight variable data set contained considerable structure.

Both quadratic discriminant analysis (QDA) and a multilayer perceptron with back propagation learning (MLP-BP) were used for classification. Records were divided in a stratified random manner into a test set containing one record of each species, and a training set which contained the remaining 127 records. Training data were used to generate discriminant functions, which were then used to classify the six 'unknown' test songs. In order to reduce the impact of atypical records, results were presented as the average (n=25) of several independent randomizations.

Backpropagation without momentum or higher order derivative information was employed as the learning model [16, 17]. An initial learning rate of 0.2, and target values of 0.1 and 0.9 were used to accelerate learning [18]. Eight inputs and six outputs were used. To determine the best network configuration, the number of hidden nodes was varied from three to eight, and the number of training epochs from 20 to 500. The best accuracy was obtained with 6 hidden units and 200 epochs. This configuration was also used with 25 randomizations, as described above, to generate the final results. In testing, any output for a given song that had an output in excess of 0.6 was considered to be active. If any incorrect output was active or if no outputs were active above this threshold, a misidentification was recorded. To make the results comparable to those derived from classification by QDA, training and test sets were generated in the same manner for both.

3. RESULTS

Table 1. Stepwise discriminant analysis of temporal and spectral variables. SD is standard deviation, ASCC is average squared canonical correlation.

Step	Variable added	ASCC
1	Mean element length (XE)	0.14
2	SD of spectral band 4 (S4)	0.24
3	Mean silence length (XS)	0.33
4	Number of elements (NE)	0.42
5	SD of element length (SE)	0.46
6	Mean for spectral band 3 (X3)	0.49
7	Mean for spectral band 2 (X2)	0.52
8	Mean for spectral band 8 (X8)	0.56
9	SD of spectral band 7 (S7)	0.57

The average squared canonical correlation (Table 1) in stepwise discriminant analysis is an indication of the contribution a new variable makes in discrimination at a given step. This measure increased only slightly (0.01) between steps eight and nine of the procedure, suggesting that little could be gained in discrimination by adding more than eight variables.

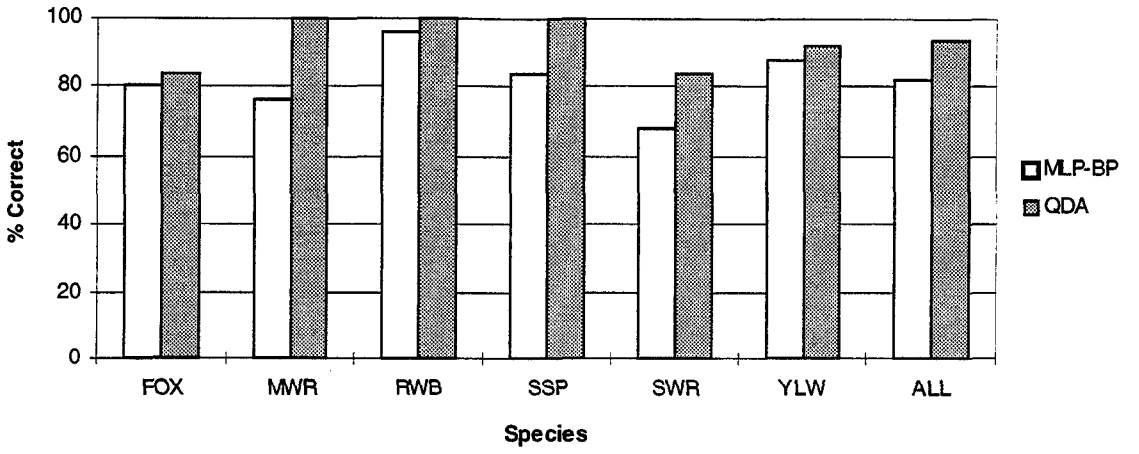

Fig. 1. Classification accuracy for quadratic discriminant analysis (QDA), and a backpropagation neural network (MLP-BP).

The results of QDA (Fig. 1) for the *test* records were excellent, with an overall accuracy of 93.3%. No errors were made in assigning *training* records to their proper categories. The MLP-BP results (Fig. 1) were are not as good as those indicated for QDA but did show similar trends across the species; Red-winged Blackbirds show the best performance, and Sedge Wrens the poorest. Overall accuracy was 82%.

4. DISCUSSION

As suggested by James and McCulloch [19] care must be taken in speculation about the theoretical significance of the variables chosen by stepwise procedures. In this study, such methods were used to choose a subset of the original 23 variables since there was much inter-correlation among the original ones. Had all variables been retained, computational loads would have been higher in subsequent analyses, and performance would have been lower due to the large number of parameters that would have been involved in the models. It is not surprising that both temporal and spectral variables were discovered to be good predictors of species identity since these cues are important for species recognition among birds [3].

Together, the four temporal variables selected provide information on the number of elements, their mean length and their variability, and the mean length of silent periods. The fact that this almost exhausts the list of measured temporal variables suggests that these species can be distinguished largely by their patterns of sound versus silence. Spectral measurements were required to sort out finer distinctions.

QDA provided excellent results. When the nature of the classification errors made with discriminant analysis was examined, certain records (five of them) were found to be responsible for the errors. Examination of these suggested that they did indeed sound different from the others, but that it was difficult to discern what the critical differences were. These pathological samples were not incorrectly labeled, but could have represented atypical or incomplete songs. They also may have represented the presence of uncommon dialects among the sample songs used.

The neural network classifier performed well, but somewhat poorer than QDA. The pattern among species was consistent with that obtained with QDA. Examination of trends in overall accuracy with respect to the number of training runs indicated that performance declined between 15 and 25 runs. It is possible that pathological cases were more prevalent among samples. Further tuning and a larger number of runs could, in theory, result in network performance that is similar to that for discriminant analysis. Even so, the network worked with a low number of hidden units and short training periods. These facts suggest that the data contained clear structure and that pre-processing methods were successful in retaining this structure while at the same time reducing dimensionality.

5. CONCLUSIONS

The approach of extracting temporal and spectral features from bird songs achieved excellent results in the identification of six species. Use of an inter-disciplinary approach drawing information from biology, computer engineering and statistics provided both the tools and the clues that made it possible to design an automatic recognition system. In fact, classical statistics provided the means to critically evaluate the utility of the variables and provided an objective benchmark for gauging ANN performance. There is considerable potential for future

research on this topic, which could lead to the development of products for biological research and education.

6. ACKNOWLEGEMENTS

We wish to thank members of the Ganglion research group for their helpful input. This research was supported by NSERC and by a University of Manitoba Fellowship to A.L. McIlraith.

7. REFERENCES

[1] Canadian Wildlife Service, *Migratory Birds Convention Act*, R.S., 1970, c. M-12 and the migratory birds regulations established by C.R.C., c. 1035 and amendments, Canadian Wildlife Service, Department of the Environment, Ottawa, Ontario: Published by Minister of Supply and Services, Canada, 1980.

[2] J.A. Endler, *Signals, signal conditions and the direction of evolution*, American Naturalist, Vol. 139 (Supplement), pp. S125 - S153, 1992.

[3] P.H. Becker, *The coding of species-specific characteristics in bird sounds*, in Acoustic Communications in Birds, Vol. I, D.E. Kroodsma, E.H. Miller and K. Ouellet, eds., N.Y.: Academic Press, 1982, pp. 213-252.

[4] D.M. Weary, R.G. Weisman, R.E. Lemon, T. Chin and J. Mongrain, *Use of relative frequency of notes by Veeries in song recognition and production*, Auk, Vol. 108, pp. 977-981, 1991.

[5] M.C. Baker, *Sharing of vocal signals among Song Sparrows*, Condor, Vol. 85, pp. 482-490, 1983.

[6] D.E. Kroodsma and J. Verner, *Complex singing behaviors among Cisthothorus wrens*, Auk, Vol. 95, pp. 703-716, 1978.

[7] M. Brigham, Bird Sounds of Canada, Mount Albert, Ontario: Holborne Dist. Co. Ltd, no date.

[8] D. J. Borror, Songs of Eastern Birds, New York: Dover, 1970.

[9] D. J. Borror, Common Bird Songs, New York: Dover, 1967.

[10] L. Elliot and T. Mack, *Wild Sounds of the Northwoods*, Ithaca, N.Y.: NatureSound Studio, 1990.

[11] R. K. Walton and R. W. Lawson, Birding by Ear (Eastern / Central) - A Guide to Bird-song Identification, Boston: Houghton - Mifflin, 1989.

[12] P. P. Kellogg, R. T. Peterson and W. W. H. Gunn, A Field Guide to Western Bird Songs, Boston: Houghton - Mifflin, 1975.

[13] W.H. Press, W.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, Numerical Recipes in C, 2nd ed., N.Y: Cambridge University Press, 1992.

[14] SAS Institute Inc., SAS User's Guide: Basics, Version 5 edition, Cary, NC: SAS, 1985.

[15] SAS Institute Inc., SAS User's Guide: Statistics, Version 5 edition, Cary, NC: SAS, 1985.

[16] D. E. Rumelhart, F. E. Hinton and R. J. Williams, *Learning representations by back-propagation of errors*, Nature, Vol. 323, pp. 533-536, 1986.

[17] J.L. McLelland and D.E. Rumelhart, Explorations In Parallel Distributed Processing, Cambridge, Mass.: MIT Press, 1989.

[18] S. Haykin, Neural Networks: a Comprehensive Foundation, N.Y.: Macmillan College Publishing Co, 1994.

[19] F.C. James and C.E. McCulloch, *Multivariate analysis in ecology and systematics: a panacea or Pandora's box?*, Annual Review of Ecology and Systematics, Vol. 21, pp. 129-166, 1990.