Unsupervised Acoustic Classification of Bird Species Using Hierarchical Self-organizing Maps

Edgar E. Vallejo[1], Martin L. Cody[2], and Charles E. Taylor[2]

[1] ITESM-CEM, Computer Science Dept.
Atizapan de Zaragoza, Edo. de México, 52926, México
vallejo@itesm.mx
[2] UCLA, Dept. of Ecology and Evolutionary Biology
Los Angeles, CA, 90095, USA
mlcody@ucla.edu
taylor@biology.ucla.edu

Abstract. In this paper, we propose the application of hierarchical self-organizing maps to the unsupervised acoustic classification of bird species. We describe a series of experiments on the automated categorization of tropical antbirds from their songs. Experimental results showed that accurate classification can be achieved using the proposed model. In addition, we discuss how categorization capabilities could be deployed in sensor arrays.

1 Introduction

We are engaged in a research program that aims to explore the capabilities of sensor arrays for the acoustic monitoring of bird behavior and diversity. Our long term goal is to create sensor arrays that behave as a single ensemble (Taylor, 2002). In this idealization, sensor nodes can recognize concepts and discourse intelligently about them (Lee et al, 2003). Constructing autonomous sensor arrays possessing problem solving capabilities in a variety of environments remains a challenge for artificial life research.

We believe that creating adaptive sensor arrays would be a major step towards realizing the full potential of this emerging technology (Estrin et al, 2001). Similarly, we think sensor arrays are excellent platforms for studying fundamental aspects of living systems such as emergence, self-organization and the evolution of communication systems (Collier and Taylor, 2004).

Pervasive in living entities is the remarkable ability to distinguish among different elements of the environment. This involves the identification of meaningful categories describing different aspects of the environment and is often critical for the viability of an organism (Pfeifer and Bongard, 2007). Moreover, we believe that the emergence of learnable languages in sensor arrays would be contingent to the ability of associating symbolic descriptions to cognitive salient categories (Stabler et al, 2003).

Several computational models have proven to be highly effective for the accurate classification of acoustic signals, such as hidden Markov models, among

M. Randall, H.A. Abbass, and J. Wiles (Eds.): ACAL 2007, LNAI 4828, pp. 212–221, 2007.

others (Rabiner, 1993). However, the later methods possess the limitation that they often require the explicit intervention of a teacher (i. e. supervised learning). Much of the categorization is developed in living systems without the explicit intervention of a teacher (i.e. unsupervised learning). If we are to develop adaptive sensor arrays, our computational methods should adhere to unsupervised learning, whenever possible.

In this work, we explore the capabilities of self-organizing maps for categorizing different species of birds from their songs. Moreover, we would like to pose here that a hierarchy of self-organizing maps provides an effective method for the unsupervised acoustic classification of bird species.

We conducted a series of computational experiments in which bird songs are transformed into strings of symbols and then classified from this representation using hierarchical self-organizing maps. Experimental results show that the proposed method is capable of categorizing four species of antbirds accurately.

2 Methods

2.1 Hierarchical Competitive Learning

The simplest form of a self-organizing map is the competitive learning network (Kohonen, 1997). This network consists of a single layer of output units c_i, each fully connected to a set of inputs o_j via excitatory connections w_{ij} (Hertz et al, 1991). Figure 1 shows an example of such a network.

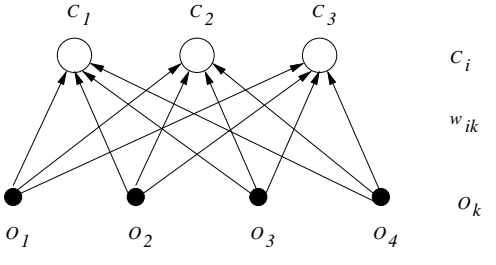

Fig. 1. Simple competitive learning network

Given an input vector $\mathbf{o}$, the winner is the unit c_{i^*} with the weight vector $\mathbf{w}_{i^*}$ that satifies the condition:

$$|\mathbf{w}_{i^*} - \mathbf{o}| \leq |\mathbf{w}_i - \mathbf{o}| \text{ (for all } i)$$

The learning process consists of updating weights w_{i^*j} for the winning unit c_{i^*} only, using the standard competitive learning rule:

$$\Delta w_{i^*j} = \eta(o_j - w_{i^*j})$$

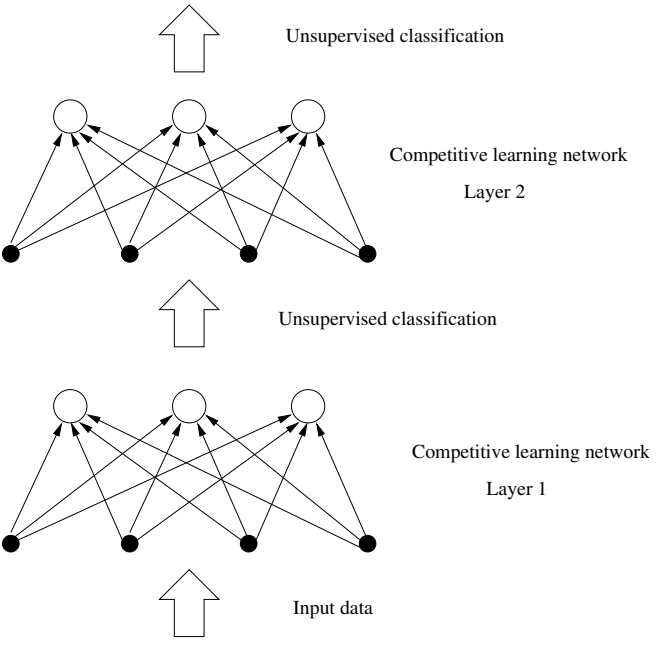

Fig. 2. Hierarchical competitive learning network

where $\eta \in [0, 1]$ is the learning constant. This learning rule moves w_{i^*j} directly towards o_j.

A hierarchical competitive learning network has two or more layers, each consisting of a simple competitive learning network (Kohonen, 1997). In such a network, each layer produces a new representation of the input data. Each layer in the network is expected to elucidate features that are implicit in the original representation. A two-layer hierarchical competitive learning network is depicted in Figure 2.

2.2 The Model

We propose a model for the unsupervised acoustic classification of bird species. The overall approach consists of transforming the acoustic signal of bird songs into strings of symbols. This transformation is achieved by the unsupervised classification of syllables of the original acoustic signal using a competitive learning network. Unsupervised species classification is achieved using a second competitive learning network that classifies strings of symbols from their syllable structure features.

A block diagram describing the proposed model is shown in Figure 3. Each module composing the diagram will be described in the experiments section of this paper.

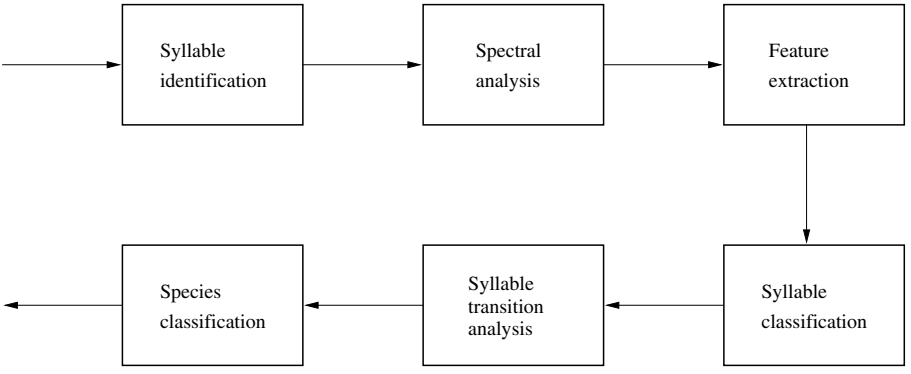

Fig. 3. Block diagram

3 Experiments and Results

3.1 Dataset

The samples used in the experiments reported here came from two sources: the Macaulay Library of Natural Sounds of the Cornell Laboratory of Ornithology and from recordings obtained in the field by one of the co-authors of this paper. The dataset consists of songs from four different antbird species that are abundant at the Montes Azules Biosphere Reserve in Chiapas, Mexico. They are listed in Table 1.

The spectrograms describing the songs of each species are shown in Figure 4. It can be appreciated that the songs from different species posses a similar structure. In effect, they consist of repetitive segments of sounds that spawns over similar

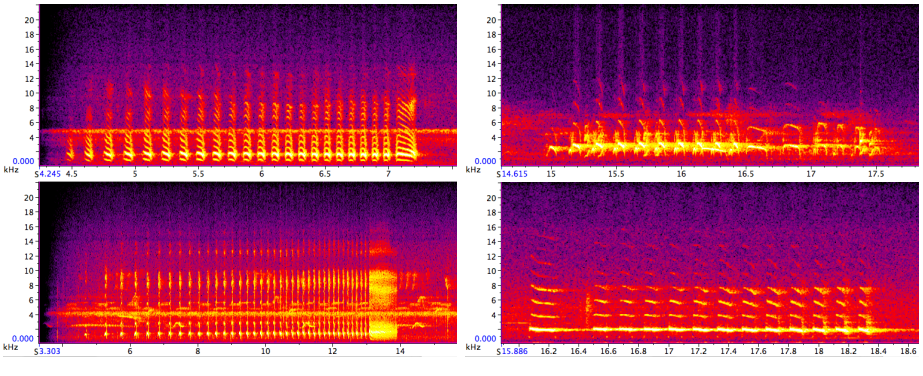

Fig. 4. Spectrograms for BAS (upper left), DAB (upper right), GAS (lower left) and MAT (lower right)

Table 1. Bird species used in the experiments

label	species	samples
BAS	Barred antshrike	12
DAB	Dusky antbird	12
GAS	Great antshrike	12
MAT	Mexican antthrush	12

frequency spectra. These similarities pose challenges for automated species recognition; especially for those methods that rely on unsupervised classification.

3.2 Syllable Identification

Bird song is thought to possess a hierarchical organization similar to that used for describing human language. In effect, bird song is typically described as consisting of phrases, syllables and elements (Catchpole and Slater, 1995). Compared to that of other singing birds, the structure of antbird songs is relatively simple. As a consequence, we believe that a two-level description consisting of songs and syllables would provide sufficient information elements for accurate automated recognition.

For experiments reported here, songs were segmented using the Raven bird song analysis program (Charif, 2004). Syllables were identified by small discontinuities in the corresponding spectrogram as shown in Figure 5.

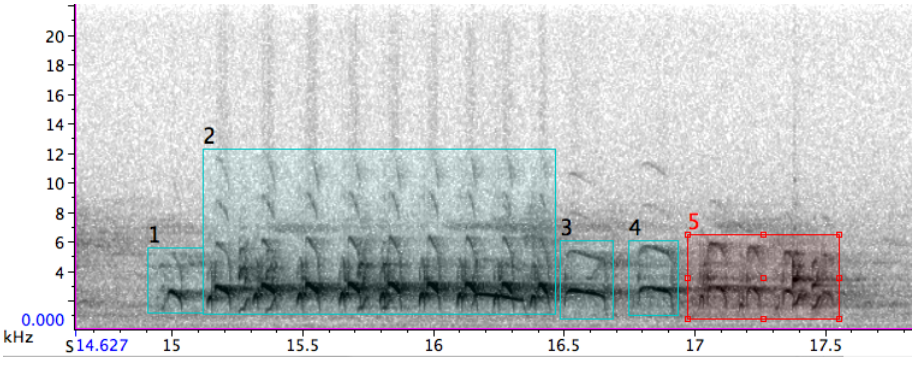

Fig. 5. Syllable identification

3.3 Spectral Analysis

Using the procedure described above, we obtained a collection a syllable samples as listed in Table 2. For each sample, we obtained a series of temporal and spectral measurements using Raven. These parameters are extracted from the sound signal using the short-time Fourier transform (STFT) (Charif, 2004).

Table 2. Number of syllable samples per species

label	samples
BAS	216
DAB	129
GAS	339
MAT	117

3.4 Feature Extraction

Previous work on species recognition using discriminant feature analysis have demonstrated the existence of minimal subsets of features for accurate discrimination of different bird species (Nelson, 1989). These subsets of features have proven to depend heavily on the species to be discriminated. However, some parameters such as high frequency and duration are commonly present in these results. From these observations, we select a collection of measurements for each syllable. These measurements are described in Table 3.

Table 3. Acoustic features

parameter	description
Low frequency	The lower frequency bound of the syllable
High frequency	The upper frequency bound of the syllable
Delta time	The duration of the syllable
Max amplitude	The upper amplitude bound of the syllable
Max power	The upper power bound of the syllable

A normalization process was applied to this data as the selected measurements spawn different orders of magnitude. Using the mean and the standard deviation of each measurement, we obtained a collection of feature vectors described as z-scores.

3.5 Syllable Classification

The collection of feature vectors describing syllables were classified using a simple competitive learning network. Once the syllables have been categorized we proceeded to represent the original songs as strings of symbols using the label from each syllable category. Table 4 shows the string representation of a subset of the songs obtained using a two-unit competitive learning network, each representing an hypothetical syllable, with $\eta = 0.1$ and epochs $= 1000$.

Similarly, Table 5 shows the string representation of a subset of the songs obtained using a four-unit competitive learning network, each representing an hypothetical syllable, with $\eta = 0.1$ and epochs $= 1000$.

Table 4. String representation of songs with 2 syllables

label	string
BAS_1	BBBBBBBBBBAAABABBBBBBA
BAS_2	BBBBBBAAAABBBAAABBBBAAAAA
BAS_3	BBBBBBBBBAABAABBAABBBBBBA
DAB_1	BBBBBBBBBBBBBB
DAB_2	BBBBBBBBBBBBBB
DAB_3	BBBBBBBBBBBBBB
GAS_1	BBAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAB
GAS_2	BBAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAB
GAS_3	BBBAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAB
MAT_1	BBBBBBBBBBBBB
MAT_2	BBBBBBBBBBBBB
MAT_3	BBBBBBBBBBBBB

Table 5. String representation of songs with 4 syllables

label	string
BAS_1	DDDAAAAAABAAAAAAAAABAC
BAS_2	DDDAAAAAAAABBBABBBBBBBAAB
BAS_3	DDDAAAAAAAAABAAAAAAAAAAC
DAB_1	DBBBBBBBBBDDDDD
DAB_2	DBBBBBBDDDDDDD
DAB_3	DDBBBBBBBBDDDD
GAS_1	DDDDCCCDCCCCCCCCCCCCBCCACBAACCAAAABACCD
GAS_2	DDACCCCCCBCCCCBCCCCCACAACABBABBBBAAD
GAS_3	DDDCDCCCCCCBCCCCBBBCCBCCAABAABABBBBBCD
MAT_1	DDDDDDDDDDDDD
MAT_2	DBBBBBBDDDBBB
MAT_3	BBBBBBBBBBBBB

3.6 Syllable Transition Analysis

Once represented as strings of syllables, bird songs are more amenable to their syntax analysis. This abstract representation of songs hides much detail of the acoustic signal and emphasizes others. For example, the calculation of syllable composition of songs is straightforward from this representation.

In this work, we propose to describe the structure of each song using the length (l), the number of different syllables (Σ) and the frequency of each pair of syllable combinations in the song. In this way, we obtained an additional collection of feature vectors as shown in Table 6 for the two-syllable experiment.

Similarly, Table 7 shows the feature vectors obtained for the four-syllable experiment. For this experiment, there are 16 (4×4) two-syllable combinations.

These feature vectors were again normalized as z-scores.

Table 6. Syllable transition analysis with two syllables

label	l	Σ	AA	AB	BA	BB
BAS_1	22	2	2	2	3	14
BAS_2	25	2	9	2	3	10
BAS_3	25	2	3	3	4	14
DAB_1	15	1	0	0	0	14
DAB_2	14	1	0	0	0	13
DAB_3	14	1	0	0	0	13
GAS_1	39	2	35	1	1	1
GAS_2	36	2	32	1	1	1
GAS_3	38	2	33	1	1	2
MAT_1	13	1	0	0	0	12
MAT_2	13	1	0	0	0	12
MAT_3	13	1	0	0	0	12

Table 7. Syllable transition analysis with four syllables

label	l	Σ	AA	AB	AC	AD	BA	BB	BC	BD	CA	CB	CC	CD	DA	DB	DC	DD
BAS_1	22	4	13	2	1	0	2	0	0	0	0	0	0	0	1	0	0	2
BAS_2	25	3	8	3	0	0	2	8	0	0	0	0	0	0	1	0	0	2
BAS_3	25	4	18	1	1	0	1	0	0	0	0	0	0	0	1	0	0	2
DAB_1	15	2	0	0	0	0	0	8	0	1	0	0	0	0	0	1	0	4
DAB_2	14	2	0	0	0	0	0	5	0	1	0	0	0	0	0	1	0	5
DAB_3	14	2	0	0	0	0	0	7	0	1	0	0	0	0	0	1	0	4
GAS_1	39	4	4	1	3	1	2	0	1	0	2	2	16	2	0	0	2	3
GAS_2	36	4	2	2	3	0	2	3	2	0	3	2	12	0	1	0	0	1
GAS_3	38	4	2	3	0	0	2	5	4	0	1	3	10	2	0	0	2	2
MAT_1	13	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12
MAT_2	13	2	0	0	0	0	0	5	0	1	0	0	0	0	0	2	0	2
MAT_3	13	1	0	0	0	0	0	12	0	0	0	0	0	0	0	0	0	0

Table 8. Unsupervised classification results with two-syllables

Species	**classification**
BAS	100%
DAB	100%
GAS	100%
MAT	92%

Table 9. Unsupervised classification results with four-syllables

Species	**classification**
BAS	100%
DAB	92%
GAS	100%
MAT	83%

3.7 Species Classification

The collection of feature vectors describing the structure of songs were classified, again, using a simple competitive learning network. Table 8 shows the accuracy of classification obtained for the two-syllable experiment using a four-unit competitive learning network, each representing a different species, with $\eta = 0.1$ and epochs = 1000.

Similarly, Table 9 shows the accuracy in classification obtained for the four-syllable experiment using a four-unit competitive learning network each representing a different species, with $\eta = 0.1$ and epochs = 1000.

It should be noted that the proposed model has been tested for generalization using an additional test set with similar results.

4 Discussion

Despite its preliminary character, the results shown here seem to indicate that meaningful acoustic categorization of bird species can emerge using hierarchical self-organizing maps. They also show that the accuracy in classification depends on the number of syllables describing the bird songs. This suggests the existence of a particular number of syllables for representing bird songs that is optimum for accurate species classification.

We show that using different abstraction levels for the description of bird song provides a convenient approach for analyzing different aspect of acoustic signals. On the one hand, temporal and spectral features have proven to be useful for the categorization of song segments. On the other hand, compositional features of syllables have proven to be sufficiently informative for species classification.

We think the proposed method could be extended in several ways. For instance, different sources of information could be combined within the same layer (e. g. acoustic localization of signal and the signal itself). Other extensions, such as adding higher layers would combine information from lower-layers for describing more abstract scenarios (e. g. two birds in the same territory at the same time).

It should be noted that the proposed model has only been tested in a simple simulated setting. We will test the proposed model in real settings in the near future.

Acknowledgements

This work was supported by the National Science Foundation under Award Number 0410438 and by Consejo Nacional de Ciencia y Tecnología under Award Number REF:J110.389/2006. Any opinions, findings and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the sponsoring agencies.

References

Catchpole, C.K., Slater, P.L.B.: Bird song biological themes and variations. Cambridge University Press, Cambridge (1995)

Charif, R.A., Clark, C.W., Fistrup, K.M.: Raven 1.2 user's manual. Cornell Laboratory of Ornithology, Ithaca, NY (2004)

Collier, T.C., Taylor, C.E.: Self-Organization in Sensor Networks. Journal of Parallel and Distributed Computing 64(7), 866–873 (2004)

Estrin, D., Girod, L., Pottie, G.: Srivastava, M.: Instrumenting the world with wireless sensor networks. In: Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP (2001)

Hertz, J., Krogh, A., Palmer, R.G.: Introduction to the theory of neural computation. Addison-Wesley, Reading (1991)

Kohonen, T.: Self-organizing maps, 2nd edn. Springer, Heidelberg (1997)

Lee, Y., Riggle, J., Collier, T.C., et al.: Adaptive communication among collaborative agents: Preliminary results with symbol grounding. In: Sugisaka, M., Tanaka, H. (eds.) AROB 8th. Proceedings of the Eighth International Symposium on Artificial Life and Robotics, Beppu, Oita Japan, January 24-26, 2003, pp. 149–155 (2003)

Nelson, D.A.: The importance of invariant and distinctive features in species recognition of bird song. Condor 91(1), 120–130 (1989)

Pfeifer, R., Bongard, J.: How the body shapes the way we think. MIT Press, Cambridge (2007)

Rabiner, L., Juang, B.H.: Fundamentals of speech recognition. Prentice-Hall, Englewood Cliffs (1993)

Stabler, E.P., Collier, T.C., Kobele, G.M., et al.: The learning and emergence of mildly context sensitive languages. In: Banzhaf, W., Ziegler, J., Christaller, T., Dittrich, P., Kim, J.T. (eds.) ECAL 2003. LNCS (LNAI), vol. 2801, Springer, Heidelberg (2003)

Taylor, C.E.: From cognition in animals to cognition in superorganisms. In: Bekoff, M., Allen, C., Gurghardt, G. (eds.) The Cognitive Animal. Empirical and Theoretical Perspectives on Animal Cognition, MIT Press, Cambridge (2002)