A Self-Supervised Classifier Ensemble for Source Recognition in Acoustic Sensor Arrays

Edgar E. Vallejo[1] and Charles E. Taylor[2]

[1]ITESM Campus Estado de México, Atizapán de Zaragoza, 52926, México
[2]University of California, Los Angeles, Los Angeles, CA, 90095, USA
taylor@biology.ucla.edu

Abstract

In this paper, we propose a collective self-supervised learning method to be deployed in acoustic sensor arrays. We describe a series of experiments on the automated classification of tropical bird species and bird individuals from their songs by a classifier ensemble. Simulation results showed that accurate classification can be achieved using the proposed model.

Introduction

Adaptive sensor arrays provide excellent platforms for testing hypotheses about critical properties of living systems, including collective and social behavior, communication and language, emergent structures and behaviors, among others. Further, understanding the capabilities and limitations of sensor arrays are useful for understanding self-organization in its own right, and may also prove helpful in guiding the construction of artificial agents that possess problem-solving abilities.

Over the past few years, we have been concerned with developing acoustic sensor arrays for use in observing and analyzing bird diversity and behavior (Vallejo and Taylor, 2009). We would like each sensor to see and "understand" part of the situation – depending on its own location – then to fuse their experiences with other such sensors to form a single, coherent understanding by the ensemble (Taylor, 2002). The ideal is that the array will act something like a living membrane, sensitive to what is going on within it, around it and passing through it.

So far, we have developed and tested sensor arrays that can identify their own location and sense bird vocalizations in real-world settings. We have developed filters to identify species (in some instances individual birds) and software tools to localize those individuals in natural environments. In the same vein, we have determined, to some extent, the conditions under which different classification approaches, both supervised and unsupervised, would be particularly effective (Vilches, et al 2006; Escobar, et al 2007; Vallejo et al 2007; Trifa, et al, 2008; Kirschel, et al 2009).

A problem with unsupervised learning methods has been that a particular bird species might be attached to one category in one part of the array, but to another category in other parts of the array. Therefore, achieving coherence and consistency in classification at the ensemble level have remained elusive. The main goal of the learning process should not only be to allow individual nodes to classify environmental sources accurately, but also to achieve coherent and consistent classification capabilities along the entire sensor array.

Toward that goal we have devised a self-supervised classifier ensemble model in which individual nodes of the array collectively act as both learners and teachers during the learning process. At each training step, each node of the array uses the classification outcomes of its neighbor nodes as output targets and learns accordingly. Therefore, the provision of labeled data from an external teacher is not necessary as the ensemble uses self-supervision for achieving collective classification capabilities.

Here we report simulation results on birds species recognition from their songs using the proposed model. Preliminary results indicate that consistent and coherent classification capabilities could be deployed in sensor arrays using self-supervised classification. Moreover, the time required for achieving convergence in learning have been improved for unsupervised classification.

Related work

In this section, we summarize the work of our laboratories aimed at developing filters to identify species, and individual birds in natural environments. These employ a variety of supervised and unsupervised approaches, as described below.

The simplest is to calculate the power spectrum, whereby the amount of energy at each wavelength is calculated and used to form a vector, typical to that individual or species.

We obtain better results by generating a sonogram of the vocalization, then look at particular features of those sonograms that might be particular to the species or individuals. We have found it most helpful is to adapt methods from human voice recognition to create a Markov Transition Matrix appropriate to the vocalizations of each individual or species. We are also looking at other methods that appear promising, especially data mining and Self-Organizing

Maps.

A collection of software tools have proven helpful for feature extraction, by providing efficient representations of bird songs while at the same time preserving the essential information contained in the songs. The emphasis has been on feature selection and on the conversion of analog waveforms into efficient digital representations. These tools, some of which are described in Kirschel et al, (2009), are mostly built on the signal processing toolbox of MatLab. Such transformations of signals are intended to minimize the communication capacity required for transmission of bird songs over a sensor network, to minimize the storage capacity required for saving such information in databases, and to provide the simplest possible accurate descriptions of a signal so as to minimize the subsequent complexity of identification and localization of individual birds.

Following feature extraction, we explored the use of different data mining techniques for the classification of bird species. The main goal has been to understand the importance of particular features of the acoustic signal that are distinctive for the accurate discrimination of bird species. A secondary goal has been to reduce the dimensionality of the acoustic signal in order to minimize the computational resources required for its manipulation and analysis.

Our approach has been to obtain large collections of temporal and spectral attributes using signal processing software tools to characterize bird songs and to use data mining to extract implicit and potentially useful information from these data. In this way, we have obtained a collection of association rules that describe correlations among features that appear to be inherent to a group of individuals and their conspecifics (Vilches et al, 2007).

Particularly, we used decision tree-based ID3 and J.48 algorithms for the identification of the most informative attributes and then use the selected attributes for species discrimination using a Naive Bayes classifier. Experimental results showed considerable dimensionality reduction can be achieved without significant loss in species classification accuracy with respect to alternative methods (Vilches et al, 2006).

In addition, we have explored the use of Self-Organizing Maps (SOMs) for the acoustic classification of bird species and individuals. The overall goal has been to examine the scope in which unsupervised learning is capable of conferring meaningful categorization abilities and increasing autonomy to sensor arrays.

Despite its preliminary character, our experiments with SOMs indicate that accurate unsupervised categorization of bird species can be achieved using two-dimensional SOMs (Escobar et al, 2007). However, unsupervised classification of bird individuals have proven to be extremely difficult for SOMs so we are beginning to explore complementary approaches such as semi-supervised and supervised classification.

Bird song is thought to possess a hierarchical organization similar to that used for describing human language. As a result bird song is typically described as consisting of phrases, syllables and elements (Catchpole and Slater, 1995). We have drawn inspirations from the structure of bird song to formulate a hierarchical approach for species and individual unsupervised classification.

The overall approach has been to transform 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 (i. e. syntactical) features (Vallejo et al, 2007).

Our experiments suggested that using different abstraction levels for the description of bird song provides a convenient approach for analyzing different aspects of the 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.

Despite of their obvious advantages, unsupervised learning methods have shown important limitations in practice. For example, even though individual nodes have been competent at discriminating bird species, and in some cases individual birds, achieving consistency and coherence in classification along the entire sensor array has been less satisfactory. In this paper, we further elaborate on this particular aspect of source recognition.

Methods and tools

Biological context

The principal field site for our work has been the rainforest environment at the Estacion Chajul in the Reserva de la Biosfera Montes Azules, in Chiapas Mexico (approximately $16°6'44''$ N and $90°56'27''$ W). The species of birds in our analysis have been antbirds from the suboscine families *Thamnophilidae* and *Formicariidae*. The songs of suboscines are less complicated than those of some others, and are thought to be largely determined genetically, rather than learned, making them more stable and appropriate for testing methods of classification. The species toward which we have directed most of our attention are Barred Antshrikes (BAS) (*Thamnophilus doliatus*), Dusky Antbirds (DAB) (*Cercomacra tyrannina*), Great Antshrikes (GAS) (*Taraba major*), and the Mexican Antthrushs (MAT) (*Formicarius analis*). The spectrograms describing the songs of each species are shown in Figure 1. It is apparent that the songs from different species posses a similar structure. In effect, they consist of repetitive segments of sounds that span similar frequency spectra. These similarities pose challenges for

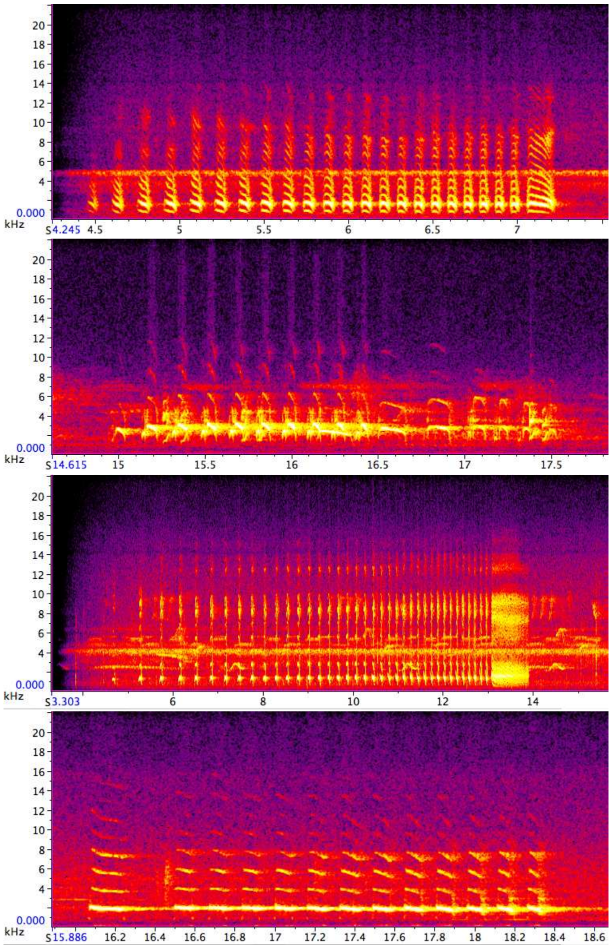

Figure 1: Spectrograms for antbirds in this study. From top, BAS, DAB, GAS, and MAT. The spectrograms were obtained from the Raven sound analysis software tool (Charif et al., 2004).

automated species recognition; especially for those methods that rely on unsupervised classification.

Sensor arrays

Th sensor arrays we are using consist of Acoustic ENSBox subarrays (Girod et al, 2006), pictured in Figure 2. These are ARM-based embedded platform designed for rapid development and deployment of distributed acoustic sensing applications. Each subarray node is self contained, with an embedded processor and a four channel microphone array that can process data locally as well as archive it and forward to other nodes wirelessly.

Typically, 5 - 8 nodes are deployed concurrently to form a distributed system of sensor sub-arrays. They are typically placed 10 - 30m apart encompassing the area to be monitored. They are automatically calibrated, to determine their node locations and orientation, then activated to perform streaming event recognition and acquire data when triggered by animal vocalizations.

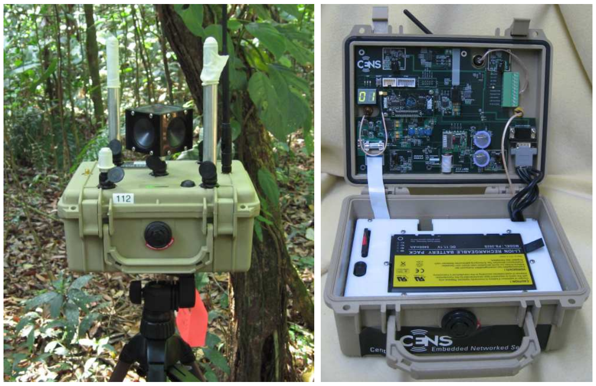

Figure 2: The Acoustic ENSBox Version 2, shown deployed near Chajul Station at left. A detailed description of both the hardware and software of this platform may be found in Collier (2010).

This approach provides greater sensor coverage, and creates a multi-hop wireless network for forwarding data and results back to a base station where data can be archived and displayed. Since each sub-array is small and has a fixed geometry, data from a single sub-array can be processed using algorithms that rely on coherence. Data from several sub-arrays can be fused to perform source localization (Ali, et al 2008). Mre detailed descriptions of the hardware and software of this platform may be found in Collier (2010) and Collier et al (2010a).

Self-supervised classifier ensemble

For this study, we devised a self-supervised classifier ensemble model (El Gayar, 2004). Different versions of self-supervised learning have been increasingly used for modeling different aspects of life-like behavior such as pattern classification, sensory motor coordination and motion planning, among others (Cohen, 2007; Lieb, 2005).

The proposed classifier ensemble consists of a collection of competitive neural networks in which classification is achieved by self-supervised learning as described below. Each competitive learning network, in turn, 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}. Figure 3 shows an example of such a network.

The presence of an external source initiates the operation of those nodes of the ensemble that perceived the external stimulus. Particularly, if a node of the ensemble detects an input stimulus, it proceeds to determine the output unit that most resembles the input signal. Formally, given an input vector $\mathbf{o}$, the winner is the unit C_{i^*} with the weight vector $\mathbf{w}_{i^*}$ as follows:

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

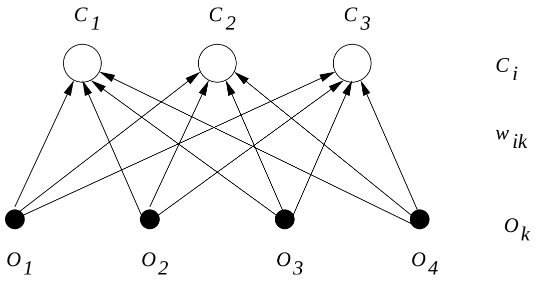

Figure 3: Simple competitive learning network. Each unit C_i can be seen as possessing a prototype that is used to represent a collection of inputs belonging to the same category.

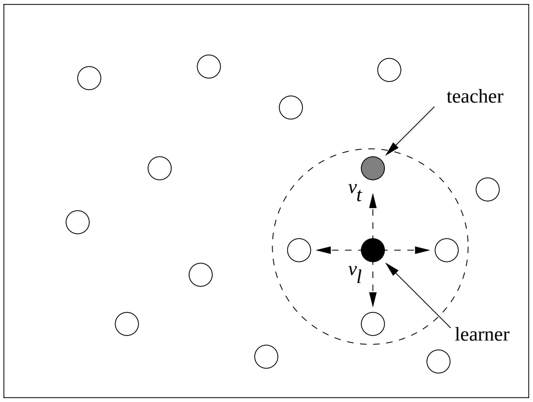

Figure 4: The learning procedure. Learner node v_l interacts with teacher node v_t and then iterates over all of the neighbor nodes.

Once the output nit for a given input has been determined, the node of the ensemble becomes a learner and its neighbor nodes become teachers, as shown Figure 4. For example, a learner node v_l of the ensemble detects an input $\mathbf{o}$ from the environment and determines the winner unit C_{i^*l}. The learner node v_l then communicates with the teacher node v_t to use the teacher's winner unit C_{i^*t} as label for the input $\mathbf{o}$.

The learner node v_l then updates the weights w_{i^*j} for the winning unit C_{i^*} only, as follows:

$$\Delta w_{i^*j} = \begin{cases} +\eta(o_j - w_{i^*j}) & \text{if } C_{i^*l} = C_{i^*t} \\ -\eta(o_j - w_{i^*j}) & \text{if } C_{i^*l} \neq C_{i^*t} \end{cases}$$

where $\eta \in [0, 1]$ is the learning constant.

A prediction derived from the formulation of the learning algorithm is that learning at the node level would be accelerated by the interaction of the learner node with a group of teacher nodes instead of using a target output provided by an external teacher. Furthermore, coherence and consistency of classification at the ensemble level would be incidental to the collective learning process.

The operation of the collective self-supervised learning procedure is described using the pseudocode in Table 1.

```
1. Create a set $N$ of neural networks with initial random weights
   (one for each node)
2. Do until number of simulation steps $k$ is met
  (a) For each node $v_l \in N$ that detects an input signal do
     i. Determine the winner unit $C_{i^*l}$ of $v_l$
    ii. Select a set $T \subseteq N$ of networks in the neighborhood of $v_l$
   iii. For each node $v_t \in T$ do
        Modify the weights of $v_l$ using the learning rule:
        $\Delta w_{i^*j} = \begin{cases} +\eta(o_j - w_{i^*j}) & \text{if } C_{i^*l} = C_{i^*t} \\ -\eta(o_j - w_{i^*j}) & \text{if } C_{i^*l} \neq C_{i^*t} \end{cases}$
        End for
      End for
   End do
```

Table 1: Training algorithm.

Parameter	Value
Nodes	16-32
Neighbors	2-8
Categories	4-8
Learning constant	0.01-0.1
Simulation steps	100-2000

Table 2: Parameters for the simulations. The values of the learning constant and simulation steps were determined empirically.

Experiments and results

Bird species recognition

We conducted simulations in order to explore the capabilities of the proposed classifier ensemble on the discrimination of bird species from their songs. We use recordings obtained by Martin L. Cody at our field site. From these recordings, we generated a collection of unlabeled training and validation sets using the procedure described in (Vallejo, et al 2007). Twelve training and twelve validation samples for each species of antbirds (BAS, DAB, GAS and MAT) were used in our experiments.

Multiple simulations were conducted using different combinations of parameter values as shown in Table 2. The following were the major results:

1. The classifier ensemble produced a meaningful classification of the unlabeled training sets. Table 3 shows the accuracy in classification in a typical simulation.

2. The classifier ensemble produced acceptable generalization performance when confronted to labeled validation sets, as shown in Figure 5.

3. Reasonable numbers of training steps (˜500) are required

procedure	accuracy	classified	misclassified
training	93.75%	45	3
testing	91.66%	44	4

Table 3: Classification results

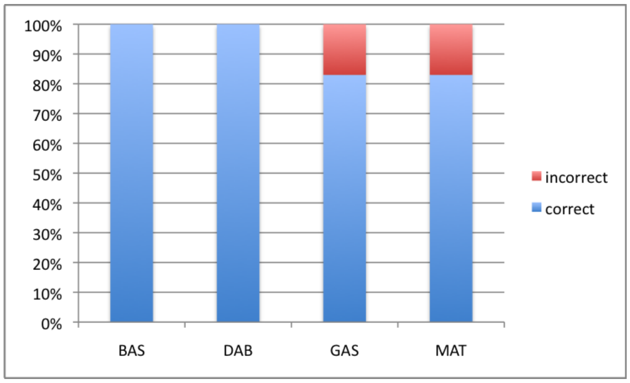

Figure 5: Classification results during validation. Misclassified samples are false negatives

for achieving coherent and consistent classification along the entire classifier ensemble.

4. Low communication bandwidth would be required for data transmission between nodes of a sensor arrays during self-supervised learning.

5. Coherence and consistency in classification along the entire classifier ensemble is achieved without compromising the accuracy of classification of individual nodes.

Bird individuals classification

It is sometimes possible to distinguish individual singers. Songs were recorded from each of 5 Mexican Antthrushs (MAT) (*Formicarius analis*) bird individual during December 2006, by Martin Cody. The identification of each singer was inferred from timing and location. The individuals were identified by labels PMPa, PMPb, PBEa, AVEa, and SNWa, Samples of 16 songs from each of the 4 territories they occupied (labeled PMP, PBE, AVE, SNW) were included. The sonogram of each song was measured for 7 traits, including length and maximum or minimum frequency at various parts of the song, so that each song was represented by a vector. From this dataset, it is apparent that some individuals are clearly distinguished while others are much less so, at least by inspection.

Multiple simulations were conducted using different combinations of parameter values as the previous experiment. The classification results obtained in a typical simulation are shown in Table 4. Specific results during validation are shown in Figure 6.

procedure	accuracy	classified	misclassified
training	77.50%	33	7
testing	72.50%	31	9

Table 4: Classification results

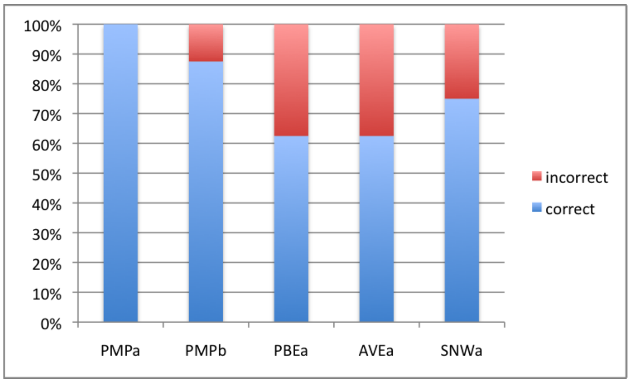

Figure 6: Classification results during validation. Misclassified samples are false negatives

Conclusions and future work

Our long term goal is to provide sensor arrays with the adaptation capabilities required to identify the meaning of bird vocalizations in the social context of the vocalizing animals. This requires event recognition, symbol grounding and adaptive communication in order for the array to arrive at a collective understanding (Lee et al, 2003). Previous studies have established plausible scenarios for the emergence of these capabilities in sensor arrays (Collier and Taylor, 2005).

Several methods for event recognition have been suggested, e.g. (Nolfi, 2005). We are currently examining methods based on information theory, among others (Kobele et al, 2004). Symbol grounding, identifying and binding semantically meaningful events to symbols, then communicating that information among parts of the arrays is of great importance.

Once events have been recognized then we can use the unsupervised classification to categorize the songs . A problem has been that new events might be attached to one symbol in one part of the array, but to another symbol in other parts of the array. Our future efforts will be directed at testing the prediction that coherence and consistency in communication could be achieved in sensor arrays using the method proposed here.

Finally, we are developing the linguistic structure that is necessary to describe these songs and events in an expressive, learnable manner, based on the ideas developed by Stabler et al (2003).

Overall, adaptive sensor arrays seem promising platforms for monitoring applications. In the near future, our efforts will be directed towards enabling sensor arrays with increasing adaptability and cognitive abilities. To accomplish this we will build largely on the results reported here.

Acknowledgements

This work was supported by the US National Science Foundation under Award Number 0410438 and by Consejo Nacional de Ciencia y Tecnología under Award Number I0110/127/2008. The authors would like to thank Lewis Girod, Travis C. Collier, Alex Kirschel, Martin L. Cody and Kung Yao for their invaluable contribution to this paper.

References

Ali, A. M. S., Asgari, T. C. Collier, M. Allen, L. Girod, R. E. Hudson, K. Yao, C. E. Taylor,Blumstein, D. T. (2008) An empirical study of collaborative acoustic source localization. *J. Sign. Process Syst.* **57**:415-436.

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

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

Chen, C-E. , A. Ali, W. Asgari, H. Park, R. E. Hudson, K. Yao, Taylor, C. E. (2006) Design and testing of robust acoustic arrays for localization and enhancement of several bird sources. In *Fifth International Conference on Information Processing in Sensor Networks*

Coen, M. H. (2007) Learning to sing like a bird: Self-supervised acquisition of birdsong. In *Proceedings of the Twenty Second National Conference on Artificial Intelligence (AAAI'07)*

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

Collier, T. C. (2010) Wireless sensor network-based acoustic localization for studying animal communication in terrestrial environments. PhD Thesis, Department of Ecology and Evolutionary Biology, University of California, Los Angeles.

Collier, T. C., Kirschel, A. N. G., and C. E. Taylor (2010) Acoustic localization of antbirds in a Mexican rainforest using a wireless sensor network. *The Journal of Accoustical Society of America* In press.

El Gayar, N. (2004) An Experimental Study of a Self-Supervised Classifier Ensemble. *International Journal of Information Technology*.

Escobar, I. A., Vilches, E., Vallejo, E. E., Cody, M. L., Taylor, C. E. (2007) Self-organizing acoustic categories in sensor arrays. In F. Almeida e Costa, M. L. Rocha, E. Costa et al (eds.) *Advances in Artificial Life, 9th European Conference, ECAL 2007.* LNAI 4648, pp. 1161-1160, Springer-Verlag.

Girod, L., Lukac, M., Trifa, V., Estrin, D. (2006) The Design and Implementation of a self-calibrating distributed acoustic sensing platform. In *ACM SenSys*.

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

Kirschel, A. N. G. , D. A. Earl, Y. Yao, I. A. Escobar, E. Vilches, E. E. Vallejo, and C. E. Taylor (2009) Using songs to identify individual Mexican antthrush *Formicarius moniliger*: Comparison of four classification methods *Bioacoustics* **19**:1-20

Kobele, G. M., J. Riggle, R. Brooks, D. Friedlander, C. Taylor, E. Stabler (2004) Induction of Prototypes in a Robotic Setting Using Local Search MDL. in M. Sugisaka and H. Tanaka, (eds.), *Proceedings of the Ninth International Symposium on Artificial Life and Robotics* Beppu, Oita Japan, pp 482-485.

Lee Y., Riggle, J., Collier, T.C. et al (2003) Adaptive communication among collaborative agents: Preliminary results with symbol grounding. In M. Sugisaka, H. Tanaka (eds), *Proceedings of the Eighth International Symposium on Artificial Life and Robotics* (AROB8th), Beppu, Oita Japan, Jan 24-26, pp.149-155.

Lieb, D., Lookingbill, A. and Thrun, S. (2005) Adaptive road following using self-supervised learning and reverse optical flow. *Proceedings of robotics: science and Systems*.

Nolfi, S. (2005) Categories Formation in Self-Organizing Embodied Agents. in H. Cohen and C. Lefebvre (eds), , *Handbook of Categorization in Cognitive Science* Elsevier, Amsterdam.

Stabler, E. P., Collier, T. C., Kobele, G. M., et al (2003) The learning and emergence of mildly context sensitive languages. In W. Banzhaf, T. Christaller, P. Dittrich, et al (Eds.) *Advances in Artificial Life, 7th European Conference, ECAL 2003.* Springer-Verlag.

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

Trifa, V. M., Kirschel, A. N., Taylor, C. E., Vallejo, E. E (2008) Automated species recognition of antbirds in a Mexican rainforest using hidden Markov models. *The Journal of Accoustical Society of America* **123**:2424-2431.

Vallejo, E. E. and Taylor, C. E. (2004) A simple model for the evolution of a lexicon. In *Proceedings of the International Symposium on Artificial Life and Robotics AROB9th.*

Vallejo, E. E. and Taylor, C. E. (2009) Adaptive sensor arrays for acoustic monitoring of bird behavior and diversity: preliminary results on source identification using support vector machines. *Artificial Life and Robotics,* vol 14, num 4, pp. 485-488. Springer-Verlag.

Vallejo, E. E., Cody, M. L., Taylor, C. E. (2007) Unsupervised acoustic classification of bird species using hierarchical self-organizing maps. In M. Randall, H. A. Abbass, and J. Wiles (Eds.) *Progress in Artificial Life, Third Australian Conference, ACAL 2007.* LNAI 4828, pp. 212-221, Springer-Verlag, 2007.

Vilches, E., Escobar, I. A., Vallejo, E. E., Taylor, C. E. (2006) Data mining applied to acoustic bird species recognition. In *18th International Conference on Pattern Recognition, ICPR 2006* Volume 3, pp. 400-403., IEEE.