Ecological Informatics 4 (2009) 206–214

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

Automated classification of bird and amphibian calls using machine learning: A comparison of methods

Miguel A. Acevedo [a,*], Carlos J. Corrada-Bravo [c], Héctor Corrada-Bravo [b], Luis J. Villanueva-Rivera [d], T. Mitchell Aide [a]

[a] University of Puerto Rico, Department of Biology, Puerto Rico
[b] University of Wisconsin-Madison, Department of Computer Sciences, United States
[c] University of Puerto Rico, Department of Computer Science, Puerto Rico
[d] Purdue University, Department of Forestry and Natural Resources, United States

ARTICLE INFO

Article history:
Received 2 December 2008
Received in revised form 17 March 2009
Accepted 20 June 2009

Keywords:
Amphibian calls
Bird calls
Decision tree
Linear discriminant analysis
Machine learning
Support vector machine

ABSTRACT

We compared the ability of three machine learning algorithms (linear discriminant analysis, decision tree, and support vector machines) to automate the classification of calls of nine frogs and three bird species. In addition, we tested two ways of characterizing each call to train/test the system. Calls were characterized with four standard call variables (minimum and maximum frequencies, call duration and maximum power) or eleven variables that included three standard call variables (minimum and maximum frequencies, call duration) and a coarse representation of call structure (frequency of maximum power in eight segments of the call). A total of 10,061 isolated calls were used to train/test the system. The average true positive rates for the three methods were: 94.95% for support vector machine (0.94% average false positive rate), 89.20% for decision tree (1.25% average false positive rate) and 71.45% for linear discriminant analysis (1.98% average false positive rate). There was no statistical difference in classification accuracy based on 4 or 11 call variables, but this efficient data reduction technique in conjunction with the high classification accuracy of the SVM is a promising combination for automated species identification by sound. By combining automated digital recording systems with our automated classification technique, we can greatly increase the temporal and spatial coverage of biodiversity data collection.

1. Background

Our understanding of ecological systems is inadequate because our knowledge is based on very limited spatial and temporal scales (Levin, 1992; Condit, 1995; Porter et al., 2005). New advances in satellite imaging and sensor networks have provided invaluable tools for collecting land-cover and abiotic data over larger scales; however, collecting biodiversity data, especially for fauna is still limited due to the need for species identification by humans. Recent developments in computer science for pattern recognition and classification are providing new tools to meet this challenge.

Most classification methods used for the automated identification of species classify samples based on morphological characteristics (Gauld et al., 2000) or bioacoustic signals (Brandes et al., 2006; Nickerson et al., 2006; Fagerlund, 2007). Classification of samples based on morphological characteristics requires the collection of the organism or portions of the organism (such as wings, pollen or genitalia) which requires intensive field sampling. However, sound can be collected easily in the field through automated digital recorders, which can collect data continuously and have the ability to detect more species than traditional scientific surveys (Acevedo and Villanueva-Rivera, 2006). Furthermore, these recordings can be analyzed with classification methods to automate species identification.

The two most important tasks in the process of automated species identification with sound are signal detection and signal characterization (Rickwood and Taylor, 2008). Signal detection refers to the extraction of vocalizations of interest from the noisy environment of a continuous recording. Signal characterization refers to the classification of these extracted vocalizations into species. Signal detection has been well studied for human speech recognition; however, even though multiple classification methods have been applied for signal characterization, few studies have compared the precision and accuracy of these methods (Skowronski and Harris, 2006).

Many studies argue that supervised machine learning algorithms such as linear discriminant analysis (Simmonds et al., 1996; Parsons and Jones, 2000), decision trees (Herr et al., 1997), Artificial Neural Networks (ANN) (Balfoort et al., 1992; Boddy et al., 1994; Do et al., 1999; Chesmore et al., 2001), Hidden Markov Chains (Kogan and Margoliash, 1998) and support vector machines (SVMs) (Fagerlund, 2007) are the best choice for automated species identification because

* Corresponding author.
E-mail address: miguel_a_acevedo@yahoo.com (M.A. Acevedo).

1574-9541/$ – see front matter © 2009 Published by Elsevier B.V.
doi:10.1016/j.ecoinf.2009.06.005

of their high accuracy (>90% accuracy) when compared to human classification. The major disadvantage of these algorithms is that they require large numbers of samples (hundreds to thousands) to train the system to obtain this high accuracy. Moreover, the training stage of most of these supervised ML algorithms is computationally demanding due to the large amount of data used as input. Thus, there is a need for new ways to reduce the amount of data used for training without compromising precision or accuracy.

The objectives of this study are (1) to compare three machine learning algorithms (linear discriminant analysis, decision tree and support vector machine) in the automated classification of bird and amphibian calls and (2) to compare two methods of training data reduction, one that characterizes each call with four standard variables (minimum and maximum frequencies, call duration and maximum power) and another that included 11 call variables (minimum frequency, maximum frequency, call duration and the frequency of maximum power in eight segments of the call). To make these comparisons we used 2132 recordings from Puerto Rico that include nine species of *Eleutherodactylus* frogs and three birds.

2. Study sites

Recordings were collected from 14 montane sites in Puerto Rico (Fig. 1). One site was located in the Guajataca State Forest and two sites in the Maricao State Forest. Three other sites were located in the Toro Negro State Forest, two sites were located in the Carite State Forest and six sites in El Yunque National Forest.

3. Methods

3.1. Sound recordings

Recordings were made using an automated digital recording system (ADRS) (Acevedo and Villanueva-Rivera, 2006). This recording system was composed of a Nomad Jukebox 3 digital mp3 player and recorder (DAP-HD0003, Creative Labs, California) which recorded 16-bit wav files at a sampling rate of 48 kHz (Villanueva-Rivera, 2007). We used a Sony ECM-MS907 electret condenser microphone with a directed angle of 120°. To improve sound quality the microphone was connected to a preamplifier (SP-PREAMP, The Sound Professionals, Inc., New Jersey). The microphone was placed ~1 m above the ground. At each site the automated recorder collected 1 min of sound every 30 min for five consecutive days.

The recording dataset included samples of 12 species (9 frogs and 3 birds). These include most of the extant *Eleutherodactylus* frog community in Puerto Rico's mountains (*E. coqui*, *E. portoricensis*, *E. antillensis*, *E. hedricki, E. wightmanae, E. unicolor, E. richmondi, E. locustus* and *E. gryllus*) and three common mountain bird species (*Patagioenas squamosa*, *Loxigilla portoricensis* and *Coereba flaveola*). These 12 species vary greatly in frequency bandwidth, call duration and call structure which makes it an appropriate dataset to test classification accuracy of machine learning methods in complex acoustical communities (Fig. 2). Moreover, *E. coqui* and *E. portoricensis* are species with very similar call characteristics which provide a true classification challenge (Fig. 3).

3.2. Data collection and verification

A total of 2,132 one minute recordings were made using ADRS (Fig. 1). These recordings were analyzed twice, first by one of us (L. J. Villanueva-Rivera) to identify all the frogs present. Then, a group of trained students listened to the recordings, identified the species of frogs and birds present, and isolated three sample calls of each species, using the box tool of Raven Pro 1.2.1. FFT transformations were constructed using a Hann window with 512 samples. In each isolated call we calculated minimum frequency, maximum frequency, maximum power and call duration. These measures were chosen because they show little overlap between species (Fig. 2). In addition, we divided each call into eight segments in which we calculated the frequency of the maximum power in each segment (Fig. 4).

We compared the species list made by both observers and in case of a mismatch, a third person (M. Acevedo) listened to the recording and determined the correct species list. In addition, we created correlation plots between variables (e.g. minimum frequency vs. maximum frequency) to study outliers. These outliers were potential errors that may have been missed by the species list comparisons, thus we verified that they were correctly classified and/or that the box was correctly drawn. Once we were certain that the data set was clean, we used the variables calculated for each call to train/test the three machine learning algorithms (Fig. 1).

3.3. Signal Representation

We represented each call as a pair $\langle x,c\rangle$, where $x\in\mathbb{R}^n$ is a real-valued vector of length 4 or 11 depending on the training data reduction method used to describe each call. We refer to this vector as a feature vector. Element $c\in C$ is an indicator of the species to which the signal belongs (*Eleutherodactylus coqui*, *E. portoricensis*, *E. antillensis*, *E. hedricki*, *E. wightmanae*, *E. unicolor*, *E. richmondi*, *E. gryllus*, *E. locustus*, *Patagioenas squamosa*, *Loxigilla portoricensis* or *Coereba flaveola*).

A total of 10,061 sample calls were used to train/test the system. Given that *Eleutherodactylus coqui* is the most abundant frog in the mountains of Puerto Rico, it was also the most common species in the recordings with the highest number of samples ($N=4641$) followed by *E. unicolor* ($N=1052$), *E. portoricensis* ($N=857$), *E. wightmanae* ($N=769$), *E. richmondi* ($N=663$), *E. hedricki* ($N=320$), *E. gryllus* ($N=254$), *E. antillensis* ($N=196$) and *E. locustus* ($N=128$). *Coereba flaveola* was the most common bird ($N=730$) followed by *Patagioenas squamosa* ($N=260$) and *Loxigilla portoricensis* ($N=191$).

3.4. Signal classification

The combination of variables that characterize each of the 10,061 analyzed calls composed a set of independent and identically distributed training instances $\{\langle x_1,c_1\rangle,\ldots,\langle x_l,c_l\rangle\}$ used to estimate a set of decision functions f_c, such that, given a new feature vector x_{new} classifies the call to the species defined by $\hat{c}=\arg\max_{c\in C} f_c(x_{new})$. This is the usual mathematical formulation of the supervised classification setting. The challenge in this setting is avoiding over-fitting the estimated decision functions f_c to the training instances. When the decision functions are over-fitted they classify the training instances almost perfectly but fail to classify new calls correctly. The main strategy to avoid over-fitting is to limit the complexity of the estimated decision functions. For example, by assuming that decision functions belong only to very restricted types, say linear in feature vectors x, or by allowing richer types of functions, but optimizing some trade-off of complexity and accuracy on the training instances. See Hastie et al. (2001) for more details on the over-fitting problem in the methods compared in this study.

In this paper we report the results for three classification algorithms: 1) linear discriminant analysis (LDA) which restricts functions f_c to be linear in feature vector x and assumes a probabilistic model for each class (Fig. 5a); 2) decision trees which recursively partition $\mathbb{R}^n$ into axis-parallel hyper-rectangles and assigns a class to each partition (Fig. 5b). In this case, $\arg\max_{c\in C} f_c(x)$ is given implicitly by the class assigned to the hyper-rectangle containing x. The complexity of the decision tree is given by the number of partitions and the recursion depth required to define them. This is controlled by "pruning" the tree, usually a post-processing step done after an initial tree is created from the training instances. 3) Support vector machines (SVMs) functions are allowed to be nonlinear in the feature vectors x and a trade-off between complexity and accuracy is directly optimized (Fig. 5c).

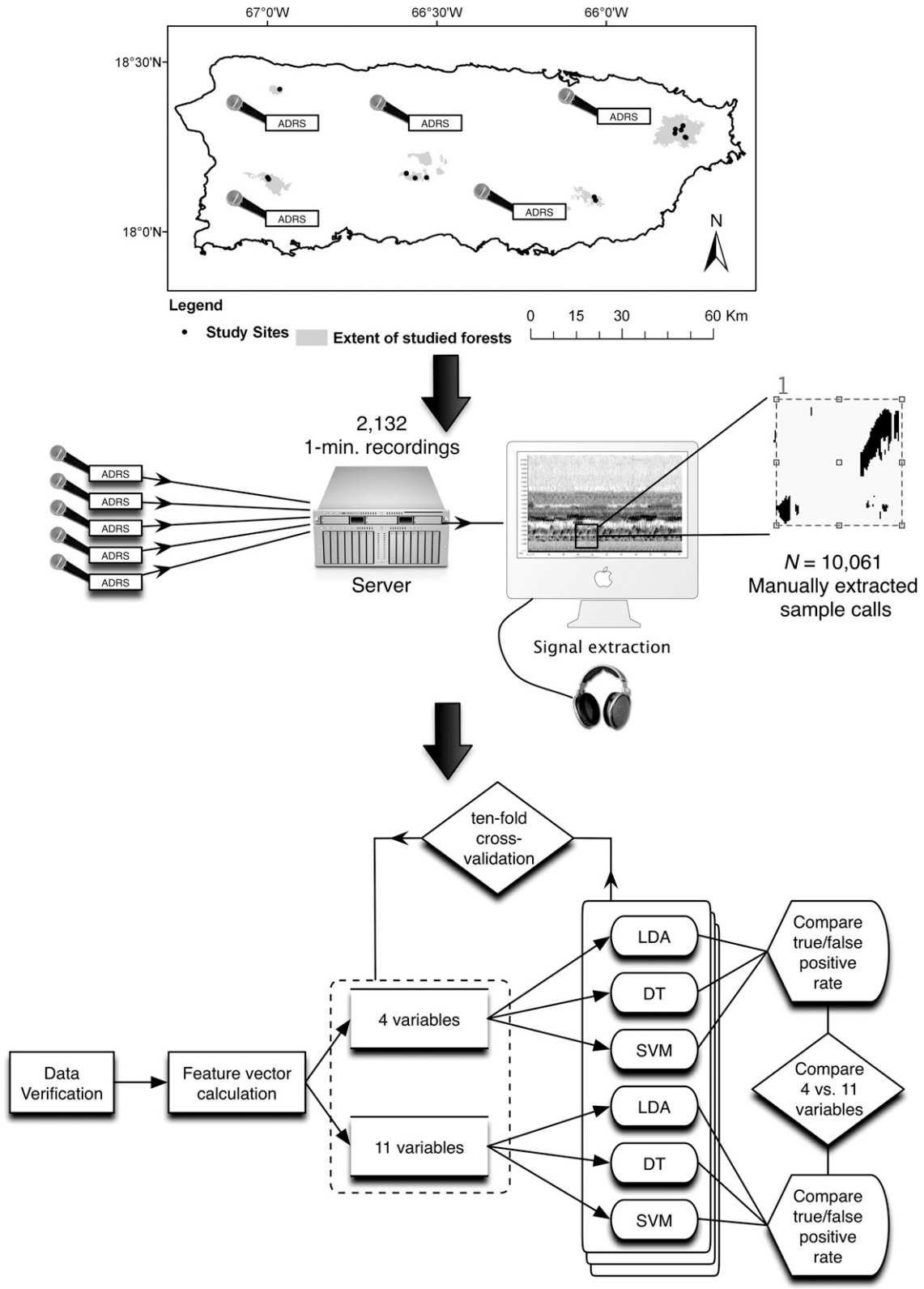

Fig. 1. Flow of information from automated field recordings to automated call classification. Automated digital recording systems (ADRS) were place in 14 field sites in the island of Puerto Rico. These recordings were digitally stored and later manually classified. This manual classification was later verified for human errors. Once the data set was cleaned we performed feature vector extraction describing each call as a pair $\langle x,c\rangle$, where $x\in\mathbb{R}^4$ or $x\in\mathbb{R}^{11}$. We did ten-fold cross-validation (using 90% of the data to train and 10% to test) to compare three machine learning methods (linear discriminant analysis; LDA, decision tree; DT and support vector machine; SVM). We compared the accuracy of each method and each data reduction technique.

3.4.1. Linear discriminant analysis

Linear discriminant analysis (LDA) assumes that feature vectors $x\in\mathbb{R}^n$ belonging to each class $c\in C$ follow a multivariate Gaussian distribution $N(\mu_c,\sum_c)$. Furthermore, it assumes that the covariance matrices of all classes are equal, i.e. $\sum_c=\sum$ for all $c\in C$. Decision function $f_c(x)$ is the log-likelihood of x, i.e., $\log Pr(x;\mu_c,\sum)\propto x^T\sum^{-1}\mu_c-\frac{1}{2}\mu_c^T\sum^{-1}\mu_c+$

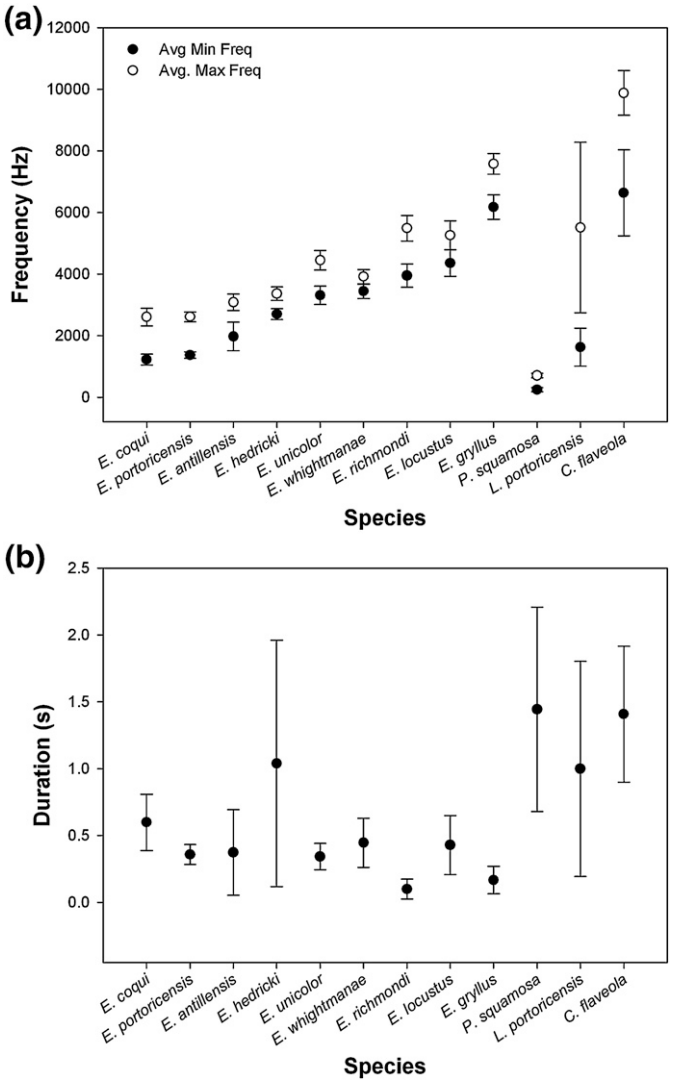

Fig. 2. Average and standard deviations for (a) minimum and maximum frequencies and (b) call duration of all amphibian and bird species identified in this study.

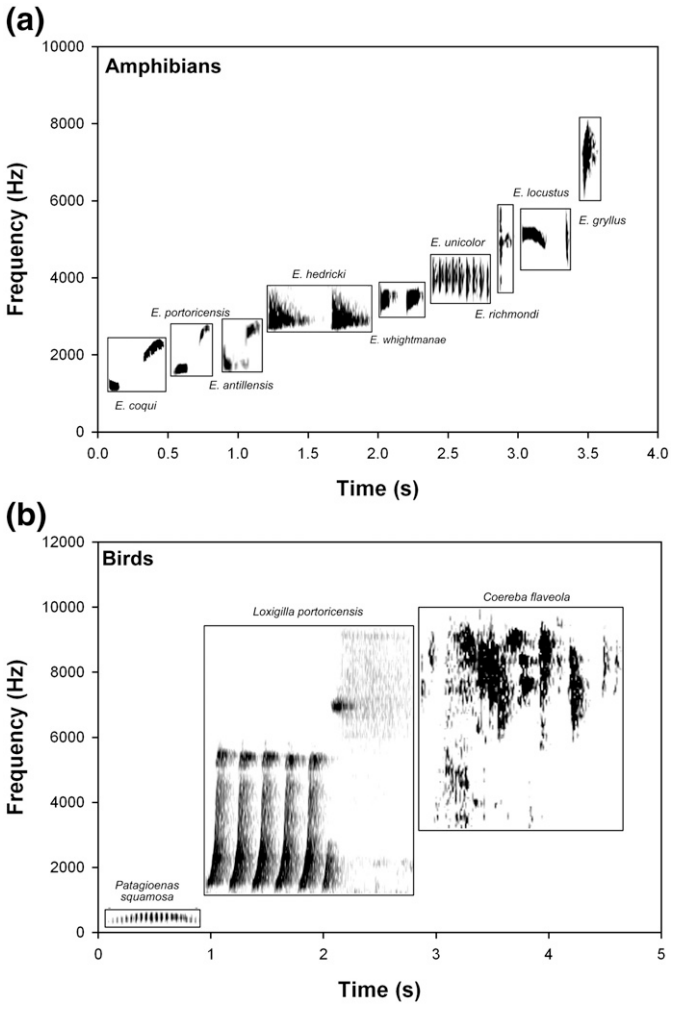

Fig. 3. Spectrograms of all (a) amphibian and (b) bird species included in this study. Boxes around calls were used to calculate call variables.

$\log \pi_c$, where π_c is a probability distribution over classes $c \in C$. Therefore, $\arg\max_{c \in C} f_c(x)$ is the class that maximizes the probability of feature vector x.

Given a set of training instances, the parameters of the functions f_c are estimated as $\pi_c = \frac{n_c}{l}$, where n_c is the number of training instances of class c, $\mu_c = \frac{1}{n_c}\sum_{c_i=c} x_i$ and the common variance $\sum = \sum_{c \in C}\sum_{c_i=c}(x_i-\mu_c)\,(x_i-\mu_c)^T/(l-|C|)$.

3.4.2. Decision trees

Decision trees (DT) recursively partition $\mathbb{R}^n$ into axis-parallel hyper-rectangles such that each of the final partitions represents a single class $c \in C$. Given a set of training instances, it selects a feature $j \in \{1,\ldots,10\}$ and cutoff value τ which splits the data into two sets: those instances for which $X_1 = \{x|\ x(j) \le \tau\}$ and those for which $X_2 = \{x|x(j) > \tau\}$. The choice of feature j and value τ is made by minimizing some measure of impurity of sets X_1 and X_2 with respect to the classes of the points in each set. A commonly used measure is the *Gini index* defined as $G(X) = \sum_{c \in C} p_c(1-p_c)$ where p_c is the proportions of instances in X of class c. Feature j and cutoff c are chosen to maximize $G(X) - (G(X_1) + G(X_2))$, where X is the current set of training instances. Once the set of instances are split into sets X_1 and X_2, decision trees are recursively built on X_1 and X_2, until some stopping criterion is met, usually a threshold of impurity, or a threshold on the minimum number of training instances in a sub-tree. In a post-processing step, partitions are heuristically merged using a complexity reduction criterion (this is usually called pruning) to avoid over-fitting.

3.4.3. Support vector machines

Support vector machines (SVMs) directly optimize a trade-off of accuracy on the training instances and function complexity and are usually defined for binary classification tasks, $C = \{-1, +1\}$. It is easiest to understand SVMs in the linear setting, where the decision functions f_c are linear in the feature vectors x, and $\arg\max_c f_c(x) = \text{sign}(w^T x - \gamma)$ for some vector $w \in \mathbb{R}^n$ and scalar γ to be estimated. Notice that w and γ define a hyperplane, such that $w^T x - \gamma > 0$ corresponds to class $+1$ and $w^T x - \gamma < 0$ corresponds to class -1. The notion of complexity in this case is given by the margin of vector w defined to be the sum of the distances between the hyperplane defined by w and γ and the nearest training instances of each of the two classes. It can be shown that this margin is given by $\frac{1}{||w||_2^2}$. The larger the margin, the less over-fitting will occur.

The optimization problem to solve is then

$$\min_{w \in \mathbb{R}^n,\, \gamma \in \mathbb{R}} ||w||_2^2 + v\sum_{i=1}^{l} (1 - c_i(w^T x_i - \gamma))_+, \tag{1}$$

where $(a)_+ = \max\{a, 0\}$. For training instance i, $c_i(w^T x_i - \gamma) < 0$ if sign $(w^T x_i - \gamma) \neq c_i$, i.e. if instance i is misclassified. In that case, w and γ

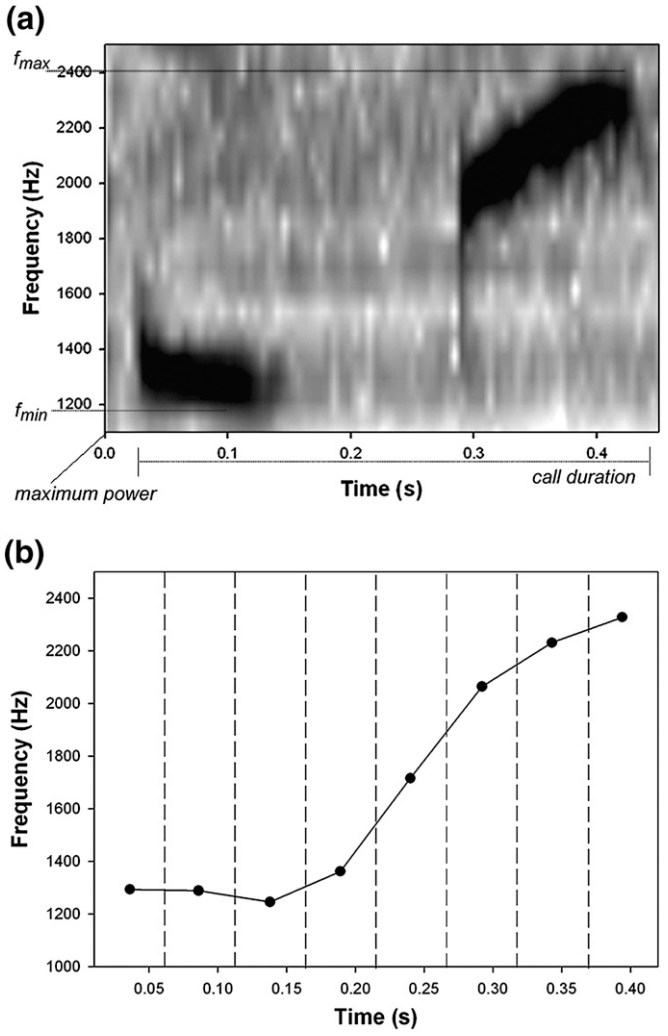

Fig. 4. We calculated 12 variables for each call: minimum and maximum frequencies, maximum power, call duration (a) and the frequency of maximum power in eight segments of the call (b). We tested two ways of representing each call. In the first method, we used four standard call characteristics (minimum and maximum frequencies, maximum power and call duration). In the second method we used minimum and maximum frequencies, call duration and the frequency of maximum power in eight segments of the call.

are penalized by $1+c_i(w^T x_i-\gamma)$. On the other hand, if $c_i(w^T x_i-\gamma)>0$ then w and γ are penalized by $c_i(w^T x_i-\gamma)$ if $c_i(w^T x_i-\gamma)<1$. v is a user-supplied parameter that trades-off complexity $||w||_2^2$ and error on the training instances. This problem can be readily solved by existing quadratic programming solvers.

Non-linearity in SVMs is achieved by assuming that decision functions have the form $f(x)=h(x)-\gamma$ where h is a function in a Reproducing Kernel Hilbert Space $\mathcal{H}$ (a generalization of Euclidean space), with corresponding kernel function k (a generalization of the dot product in Euclidean Space). The problem to optimize becomes

$$\min_{h\in\mathcal{H},\,\gamma\in\mathbb{R}} ||h||_{\mathcal{H}}^2 + v\sum_{i=1}^{l}(1-c_i(h(x_i)-\gamma))_+. \quad (2)$$

By the Kimeldorf-Wahba theorem (Kimeldorf and Wahba, 1971), the minimizer of Eq. (2) has a representation as a finite linear expansion in terms of kernel function k evaluated at the training instances, such that $h(x)=\sum_{i=1}^{l}\alpha_i k(x_i,x)$ with coefficients α_i to be estimated. The resulting optimization problem is again quadratic and the same solvers as in the linear case may be used. In practice, the user selects kernel function k, which in turns defines the corresponding space $\mathcal{H}$, in advance. In this paper we use the *Gaussian* kernel $k(x_i,x_j)=\exp\{-\sigma||x_i-x_j||_2^2\}$, where σ is some user-selected parameter. Parameters σ and v are usually selected by some form of cross-validation. We describe our choice for this paper below.

To use SVMs in multicategory classification, where $|C|>2$ as is our setting, we used the one-vs-one majority vote heuristic (Kreßel, 1999). In this case, an SVM is trained for each pair of labels, giving $\binom{|C|}{2}$ SVMs, which are then combined by majority vote. That is, given a new feature vector x_{new}, each SVM votes by classifying the new vector as belonging to one of two classes. The new vector is then classified as belonging to the class that receives the most votes. Although this is not guaranteed to converge to the optimal classifier (Lee et al., 2004), it performs well in practice.

4. Experimental setup

All experiments were carried out in R 2.7.0 (R Development Core Team, 2007). Performance statistics were estimated with ten-fold cross-validation, where folds were created at random while preserving class proportions in each fold (Fig. 1).

4.1. LDA

We used the LDA classifier implemented in the MASS 7.2-41 package (Venables and Ripley, 2002) with parameter estimates defined as in Section 3.4.

4.2. Decision trees

For decision trees we used the rpart 3.1-41 package (Therenau et al., 2007) with defaults: the gini criterion defined in Section 3.4 was used to create the tree although weighted for each class by the inverse proportion of instances in the training set; maximum recursion depth is 30; smallest node size is 20. Other defaults can be verified in the package documentation. A complexity parameter is used to avoid over-fitting by pruning subtrees that do not improve the gini index by more than a factor of 0.01.

4.3. SVM

We used the kernlab 0.9-5 (Karatzoglou et al., 2004) package's interface to libsvm (Chih-Chung and Chih-Jen, 2004) to fit SVMs, which uses the one-vs-one strategy for multicategory classification described in Section 3.4. The Gaussian kernel defined in Section 3.4 was used with parameter σ selected using the heuristic described in (Caputo et al., 2002). Trade-off parameter v was selected for each fold independently by grid-search by minimizing error on a held-out tuning set. The loss term in Eq. (2) was weighted for each class by the inverse proportion of instances in the training set.

5. Results and discussion

5.1. Results

The choice of classification method showed a much greater effect on the accuracy of the system compared to the effect of call representation. Overall, there was a >20% difference between the overall true positive rates of LDA and SVM ($F=8.91$, $P=0.001$), while the difference between 4 and 11 input variables had an overall

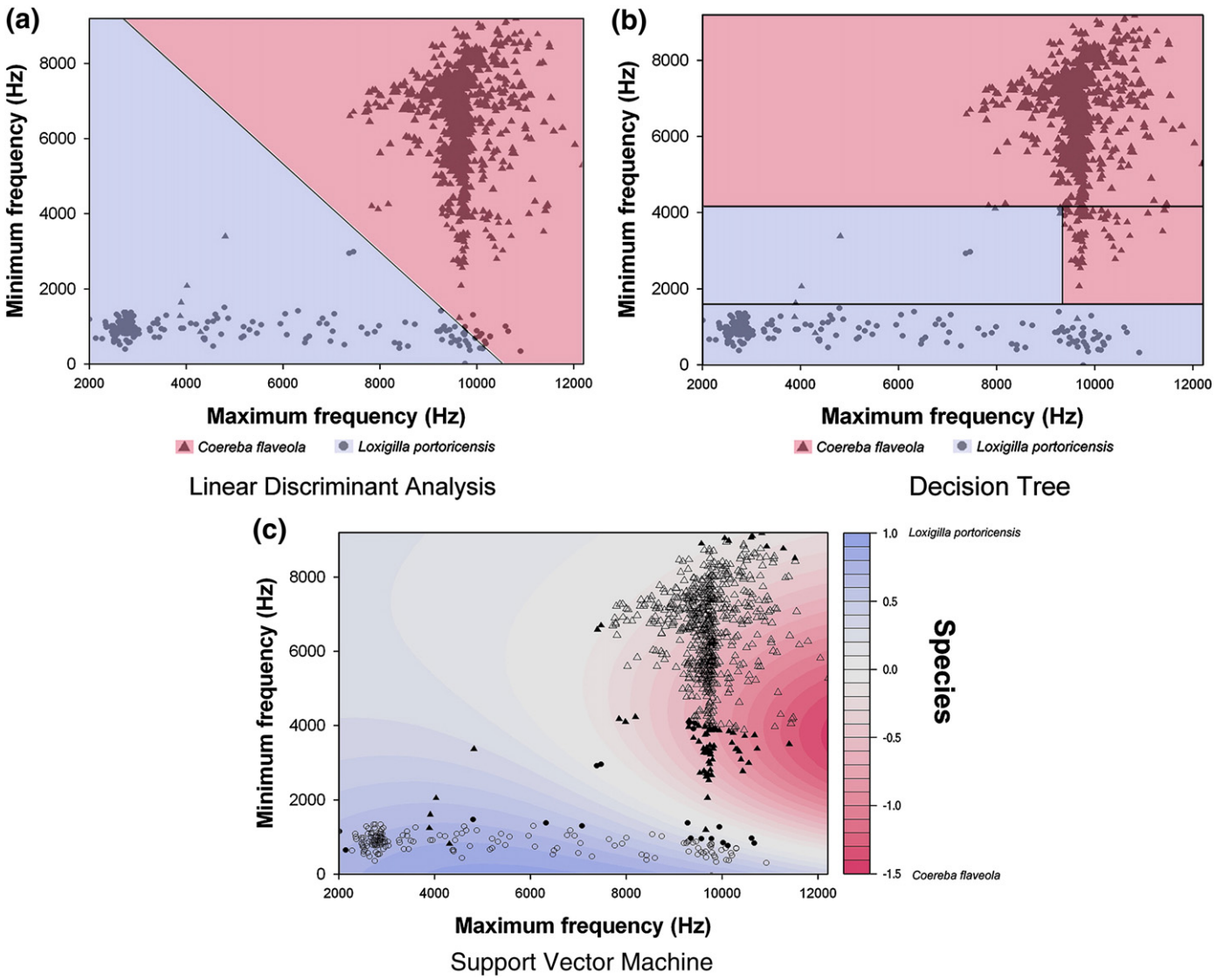

Fig. 5. Graphical examples of the three machine learning algorithms compared in this study. They show the decision functions f_c used to distinguish *Loxigilla portoricensis* and *Coereba flaveola*. Note that for simplification the figures only show two species and two variables (minimum and maximum frequencies). The actual system was trained/tested with 12 species and used 4 or 11 variables.

effect of <1% ($F=0.074$, $P=0.786$; Tables 1–3). Despite the small overall effect, the call representation of 11 variables was more accurate for most species and classification methods, thus for simplification purposes, the comparisons of classification methods made below will refer to results from training with calls characterized by 11 variables. The SVM (Table 3) had the highest average true positive rate (94.95%) and the lowest average false positive rate (0.94%) followed by DT (average TP = 88.20% and FP = 1.25%; Table 2) and LDA (average TP = 71.45% and FP = 1.98%; Table 1). SVM true positives varied from 86.99 to 100% depending on species and false positives from 0.00 to 1.62%. SVM had the highest true positive rate for all species. Moreover, this algorithm had positive rate >90% for all species except *E. antillensis*. Thus it was the most accurate classification algorithm overall (Table 3). An examination of a confusion matrix of the actual and predicted classification by SVM shows that the two species most often confused are *E. coqui* and *E. portoricensis* (Table 4). Most of the remaining errors respond to confusion between species with similar minimum and maximum frequencies (Table 4, Fig. 2).

The LDA had the highest variability of true (0.00 to 99.99%) and false positive (0.00 to 20.83%) rate. This algorithm performed poorly identifying *E. portoricensis*, *E. antillensis* and *Loxigilla portoricensis* with true positive rate <50%. However, it had true positive rate >90% for *E. wightmanae*, *E. richmondi*, *E. gryllus*, *Patagioenas squamosa* and *Coereba flaveola*. Moreover, *Patagioenas squamosa* was identified with true positives of 100.00% and false positives of 0.00%. In addition the LDA had a true positive rate of 99.84% identifying *E. coqui*, but its false positive rate of 18.41% was the highest for all species and classification methods (Table 1).

The DT true positive rates varied from 76.94 to 99.44% and false positive rates from 0.02 to 3.98%. This algorithm had true positive rate >90% for *E. hedricki*, *E. locustus*, *E. gryllus*, *Patagioenas squamosa*, and *Coereba flaveola*.

Call representation using 11 variables (minimum and maximum frequencies, call duration and the frequency in the highest energy point in eight segments of the call) had slightly higher true positive rate in all classification algorithms in comparison with the 4 variables method (minimum and maximum frequencies, call duration and maximum power), but this difference was not statistically significant ($F=0.074$, $P=0.786$). Overall species average true positive rate in the LDA increased from 69.89% to 71.45%, in the DT from 89.00% to 89.20% and in the SVM from 93.80% to 94.95%. True positive rate for many species increased when using 11 input variables, however, true positive rate also decreased for some other species in all classification

Table 1
Average and standard deviation of percent accuracy of linear discriminant analysis (LDA) in identifying nine *Eleutherodactylus* frogs and three bird species.

Species	True positive (%)	False positive (%)
Eleutherodactylus coqui	99.99 ± 0.22*	20.83 ± 2.24
(*N* = 4641)	99.74 ± 0.33*	18.41 ± 2.17
E. portoricensis	0.00 ± 0.00	0.00 ± 0.00
(*N* = 857)	0.33 ± 1.05	0.04 ± 0.11
E. antillensis	0.00 ± 0.00***	0.03 ± 0.07
(*N* = 196)	31.24 ± 13.40***	0.05 ± 0.08
E. hedricki	64.81 ± 10.91	0.17 ± 0.21
(*N* = 320)	71.09 ± 12.34	0.42 ± 0.18
E. wightmanae	93.29 ± 4.09	2.39 ± 0.41
(*N* = 769)	95.39 ± 3.84	2.94 ± 0.85
E. unicolor	78.71 ± 4.41*	0.70 ± 0.38
(*N* = 1052)	72.21 ± 6.15*	0.88 ± 0.41
E. richmondi	98.85 ± 1.94**	0.79 ± 0.47
(*N* = 663)	94.08 ± 5.02**	0.80 ± 0.36
E. locustus	51.01 ± 22.53	0.03 ± 0.07
(*N* = 128)	53.03 ± 18.34	0.00 ± 0.00
E. gryllus	100.00 ± 0.00	0.22 ± 0.11
(*N* = 254)	100.00 ± 0.00	0.22 ± 0.13
Patagioenas squamosa	97.85 ± 2.89*	0.02 ± 0.05
(*N* = 260)	100.00 ± 0.00*	0.00 ± 0.00
Loxigilla portoricensis	45.89 ± 9.42	0.03 ± 0.04
(*N* = 191)	46.51 ± 12.64	0.05 ± 0.08
Coereba flaveola	96.26 ± 2.07*	0.05 ± 0.09
(*N* = 730)	93.76 ± 2.53*	0.05 ± 0.12
Total average 4 var	69.89 ± 37.50	2.11 ± 5.94
Total average 11 var	71.45 ± 32.54	1.98 ± 5.24

The first line of the values for each species represents the result using the 4 input variables (minimum and maximum frequencies, call duration and maximum power) the second line represents the result using the 11 input variables (minimum and maximum frequencies, call duration and the frequency in the highest energy point in 8 segments of the call) Cells marked with an asterisk (*) have *t*-test statistical difference between 4 and 11 input variables of $P<0.05$, those with ** $P<0.01$ and those with *** $P<0.005$.

Table 2
Average and standard deviation of percent accuracy of decision tree (DT) in identifying nine *Eleutherodactylus* frogs and three bird species.

Species	True positive (%)	False positive (%)
Eleutherodactylus coqui	86.27 ± 2.33	1.64 ± 0.79
(*N* = 4641)	85.58 ± 1.58	1.64 ± 0.48
E. portoricensis	87.07 ± 7.03	4.58 ± 1.04*
(*N* = 857)	82.07 ± 5.94	3.68 ± 0.67*
E. antillensis	85.09 ± 9.91	1.42 ± 1.07
(*N* = 196)	83.22 ± 10.84	0.80 ± 0.47
E. hedricki	93.96 ± 6.27	0.72 ± 0.26
(*N* = 320)	93.45 ± 5.97	0.57 ± 0.32
E. wightmanae	89.92 ± 5.52*	1.72 ± 0.46
(*N* = 769)	84.96 ± 3.14*	1.63 ± 0.53
E. unicolor	68.25 ± 5.16**	0.95 ± 0.54
(*N* = 1052)	76.94 ± 7.37**	1.23 ± 0.32
E. richmondi	96.81 ± 2.15***	0.02 ± 0.06
(*N* = 663)	88.77 ± 3.83***	0.16 ± 0.38
E. locustus	98.75 ± 3.95	0.82 ± 0.39
(*N* = 128)	94.60 ± 13.62	1.02 ± 0.39
E. gryllus	96.64 ± 9.00	0.17 ± 0.16
(*N* = 254)	99.17 ± 2.64	0.23 ± 0.16
Patagioenas squamosa	99.50 ± 1.58	0.00 ± 0.00**
(*N* = 260)	99.44 ± 1.76	0.02 ± 0.02**
Loxigilla portoricensis	67.59 ± 19.54***	1.96 ± 1.02**
(*N* = 191)	89.14 ± 7.82***	3.98 ± 1.26**
Coereba flaveola	96.87 ± 1.83*	0.25 ± 0.29*
(*N* = 730)	93.08 ± 4.24*	0.04 ± 0.08*
Total average 4 var	89.00 ± 10.99	1.19 ± 1.27
Total average 11 var	89.20 ± 6.97	1.25 ± 1.33

The first line of the values for each species represents the result using the 4 input variables (minimum and maximum frequencies, call duration and maximum power) the second line represents the result using the 11 input variables (minimum and maximum frequencies, call duration and the frequency in the highest energy point in 8 segments of the call) Cells marked with an asterisk (*) have *t*-test statistical difference between 4 and 11 input variables of $P<0.05$, those with ** $P<0.01$ and those with *** $P<0.005$.

Table 3
Average and standard deviation of percent accuracy of support vector machine (SVM) in identifying nine *Eleutherodactylus* frogs and three birds species.

Species	True positive (%)	False positive (%)
Eleutherodactylus coqui	97.34 ± 0.76*	3.34 ± 1.35***
(*N* = 4641)	96.38 ± 1.07*	1.62 ± 0.77***
E. portoricensis	80.63 ± 6.39***	1.12 ± 0.26*
(*N* = 857)	90.13 ± 4.43***	1.70 ± 0.70*
E. antillensis	84.78 ± 8.62	0.20 ± 0.17
(*N* = 196)	86.99 ± 9.04	0.25 ± 0.20
E. hedricki	95.41 ± 5.25	0.27 ± 0.09
(*N* = 320)	96.04 ± 5.10	0.25 ± 0.16
E. wightmanae	95.02 ± 3.15	0.51 ± 0.41
(*N* = 769)	96.18 ± 1.41	0.80 ± 0.48
E. unicolor	94.70 ± 4.47**	0.35 ± 0.74
(*N* = 1052)	90.63 ± 1.06**	0.39 ± 0.22
E. richmondi	98.52 ± 1.72	0.06 ± 0.08
(*N* = 663)	99.35 ± 1.06	0.04 ± 0.08
E. locustus	91.19 ± 10.93	0.07 ± 0.09
(*N* = 128)	94.24 ± 7.66	0.03 ± 0.07
E. gryllus	100.00 ± 0.00	0.00 ± 0.00
(*N* = 254)	100.00 ± 0.00	0.00 ± 0.00
Patagioenas squamosa	99.38 ± 1.98	0.02 ± 0.05
(*N* = 260)	100.00 ± 0.00	0.00 ± 0.00
Loxigilla portoricensis	89.12 ± 13.62	0.22 ± 0.16
(*N* = 191)	90.33 ± 6.86	0.13 ± 0.13
Coereba flaveola	99.47 ± 0.85	0.04 ± 0.08
(*N* = 730)	99.18 ± 1.35	0.14 ± 0.16
Total average 4 var	93.80 ± 6.22	0.52 ± 0.45
Total average 11 var	94.95 ± 4.47	0.94 ± 0.61

The first line of the values for each species represents the result using the 4 input variables (minimum and maximum frequencies, call duration and maximum power) the second line represents the result using the 11 input variables (minimum and maximum frequencies, call duration and the frequency in the highest energy point in 8 segments of the call). Cells marked with an asterisk (*) have *t*-test statistical difference between 4 and 11 input variables of $P<0.05$, those with ** $P<0.01$ and those with *** $P<0.005$.

methods. For example, the true positive rate of LDA classification of *E. coqui*, *E. unicolor* and *E. richmondi* decreased from 99.99 to 99.74%, from 78.71 to 72.21% and from 98.85 to 94.08% respectively. The DT true positive rate of *E. wightmanae* decreased from 89.92 to 84.96%, *E. richmondi* decreased from 96.81 to 88.7%, and *Coereba flaveola* decreased from 96.87 to 93.08%. In the SVM, true positive rate for *E. coqui* decreased from 97.34 to 96.38%, *E. unicolor* decreased from 94.70 to 90.63%. Although the average true positive rate increase for each method was low (~1%) there were significant increases in classification accuracy for some species in all three classification

Table 4
Confusion matrix showing the accuracy of the SVM in the classification of 9 amphibian and 3 bird species calls based on 11 call variables.

		Actual												
		EC	EP	EA	EH	EW	EU	ER	EL	EG	PS	LP	CF	Total
Predicted	EC	286	5	1	1	0	0	0	0	0	0	0	1	294
	EP	3	40	0	0	0	0	0	0	0	0	0	0	43
	EA	1	0	17	1	0	0	0	0	0	0	0	0	19
	EH	0	0	0	18	0	1	0	0	0	0	0	0	19
	EW	0	0	0	0	55	3	0	0	0	0	0	0	58
	EU	0	0	0	0	8	33	0	0	0	0	0	0	41
	ER	0	0	0	0	0	0	40	0	0	0	0	0	40
	EL	0	0	0	0	0	0	0	5	0	0	0	0	5
	EG	0	0	0	0	0	0	0	0	14	0	0	0	14
	PS	0	0	0	0	0	0	0	0	0	18	0	0	18
	LP	1	0	0	0	0	0	0	0	0	0	12	1	14
	CF	0	0	0	0	0	0	0	0	0	0	0	49	49
	Total	291	45	18	20	63	37	40	5	14	18	12	51	614

This is the result of one of ten iterations that were made with 10% of the data (i.e. 1351 calls). Rows represent model classification while columns represent the actual identity of the call. Correct identifications are indicated on the diagonal. EC: *E. coqui*, EP: *E. portoricensis*, EA: *E. antillensis*, EH: *E. hedricki*, EW: *E. whightmanae*, EU: *E. unicolor*, ER: *E. richmondi*, EL: *E. locustus*, EG: *E. gryllus*, PS: *Patagioenas squamosa*, LP: *Loxigilla portoricensis*, CF: *Coereba flaveola*.

algorithms. For example, in the LDA, true positives for *E. antillensis* and *Patagioenas squamosa* increased from 0.00 to 31.24% and from 97.85 to 100.00% respectivley. In the DT, *E. unicolor* increased from 68.25 to 76.94% and *Loxigilla portoricensis* increased from 67.59 to 89.14%. In the SVM true positive rate for *E. portoricensis* increased from 80.63 to 90.13% (Table 3).

Similarly to true positives, false positives for many species decreased with an increase in input variables in all classification algorithms, but there were also significant increases in false positive detections. For instance, in the LDA false positive detections increased in *E. hedricki* and *E. wightmanae* from 0.17 to 0.42% and from 2.39 to 2.94% respectively. In the DT, the false positive rate for *Patagioenas squamosa* increased from 0.00 to 0.02 and for *Loxigilla portoricensis* increased from 1.96 to 3.98%. There were no significant increases in false detections in the SVM. There were no significant decreases in false positive detections in the LDA, however in the DT *E. portoricensis* there was a decreased from 4.58 to 3.68% and *Coereba flaveola* decreased from 0.25 to 0.04%. In the SVM significant decreases in false positive detections were found for *E. coqui* which decreased from 3.34 to 1.62%.

5.2. Discussion

The three most important characteristics of this study include (1) the large amount of call samples (10,061) available to train/test the system, (2) the small number of variables used to represent each call and (3) the higher accuracy (>90% for most species) of the support vector machine in comparison to the other classification algorithms.

Even though the discrimination of species with similar calls is an important problem when automating species classification by sound using machine learning, decreasing the amount of data used in each sample to train the system is an important issue. This is especially important when the training samples are high quality sound recordings that may include thousands of data points (e.g. 44100/s) (Skowronski and Harris, 2006; Oswald et al., 2007; Trifa et al., 2008). In this study, four standard call variables (minimum and maximum frequencies, call duration and maximum power) were enough for the SVM to accurately identify most species, however true positive rate for *E. portoricensis* increased from 80.63 to 90.13% when the call was represented by 11 variables. *E. coqui* and *E. portoricensis* have similar minimum and maximum frequencies, thus the inclusion of a coarse representation of call structure (i.e. the frequency of maximum power in eight segments of the call) became an important factor that helped the SVM discriminate between these species (Figs. 2 and 3).

Frog calls are usually simple in terms of call structure (Drewry and Rand, 1983) and their calling frequencies overlap less than birds (Fig. 3). These acoustical characteristics of the frog community aided our automated classification process. For instance, most of the frog species included in this study were accurately classified with only 4 variables because of little overlap between the combined characteristics (i.e. minimum and maximum frequencies, maximum power and call duration) of calls. Even though *Coereba flaveola*, and *Patagioenas squamosa* had unique values of maximum frequency and minimum frequency respectively, bird calls are usually more complex and few will call in an exclusive bandwidth of frequencies. Thus, we expect that the 11 variable classification method will be more useful in sites with high diversity of birds which have more complex calls.

The addition of more variables to characterize each call (4 vs 11 variables) did not always improve the classification. In these cases, the additional variables may increase the similarities between species decreasing the classification accuracy. However, in the case of SVM, these decreases were minimal (<1% true positive decrease and <0.6% false positive increase) if compared to the increases in classification accuracy for *E. portoricensis* (9.5% true positive increase). Nevertheless, characterizing each call with only 11 variables is a significant reduction in the amount of training data, and in this case, were enough for the SVM to accurately discriminate between a highly variable dataset of acoustical signals.

We tested three machine learning methods (LDA, DT and SVM) for the automated classification of amphibian and bird calls and SVM was the most accurate and precise method (high true positives >90% and low false positives <1.5%). A major difference between LDA and SVM is that model decision functions are nonlinear functions. This non-linearity is especially important when discerning between two very similar classes in terms of call features. For instance, *E. coqui* and *E. portoricensis* have very similar call structure (Fig. 4), but *E. coqui* in the mountains has a lower minimum frequency and longer call duration than *E. portoricensis* (Drewry and Rand, 1983). The LDA performed poorly in discerning between these species with an average true positive for *E. portoricensis* of 0.33% and average false positive for *E. coqui* of 18.41%. The LDA restriction on functions to be linear is the main reason for this poor classification.

Support vector machines have proven to be successful in a number of varied settings showing high discrimination accuracy. In addition to being the most accurate method for the classification of bird and amphibian calls in comparison to LDA and DT, it has been demonstrated to be the most accurate method in image classification in comparison to neural network, decision trees, naive Bayes and k-nearest neighbor (Tan et al., 2008). It has also been tested against artificial neural networks for drug classification with similar results (Byvatov et al., 2003). In addition, SVM has been more accurate than maximum likelihood classifier in the classification of Landsat ETM + images (Sanchez-Hernandez et al., 2007). The non-linearity of the function in vector x is one of the principal characteristics of the SVM that gives it advantage over other methods. However, this increased representation power is balanced by a complex control method which has the effect of making the decision functions depend only on the support vectors. Thus, the effort required to classify new calls can be ameliorated since only a subset of the training set needs to be stored.

Even though SVM outperformed DT and LDA for most species, these other linear methods were accurate classifiers for some species. For example, *Patagioenas squamosa* was classified by LDA and DT with an average accuracy of 100.00% and 99.44% respectively. This species had the lowest minimum frequency of all thus it was easy for these classifiers to use this variable to accurately discriminate it from the others (Fig. 4). Similarly the high accuracy of LDA classifying *Coereba flaveola* and *E. gryllus* is explained by their unique ranges of minimum frequency (6636–8032 Hz for *C. flaveola* and 6170 - 6563 Hz for *E. gryllus*). *E. richmondi* was also accurately discriminated by this linear method, but in this case this species had unique ranges of call durations (0.08–0.10 s). Therefore in some settings LDA and DT can be used in accurate systems that are extremely efficient to deploy.

6. Conclusion

We compared three machine learning algorithms (LDA, DT and SVM) for the automated classification of bird and amphibian calls. The SVM had the highest accuracy (highest true positive and lowest false positive rates) for most of the species. SVM, as well as other machine learning algorithms, requires large amounts of data to be trained to attain high levels of accuracy. Training SVM with digital sound files (which may contain thousands of data points) results in a computationally demanding method, however, we provide an alternate method of training data reduction that characterizes each species with 11 call variables (minimum frequency and maximum frequency, maximum power, call duration and frequency in the highest energy point in 8 segments of the call). This efficient data reduction technique in conjunction with the high classification accuracy of the SVM is a promising combination in automated species identification by sound, which if coupled with ADRS can increase the temporal and spatial coverage of biodiversity data collection.

Acknowledgments

This project was funded by NSF grant BDI 0640143 and DoD Legacy RMP 07345. The work of Héctor Corrada Bravo was funded in part by NSF grant DMS 0604572. We thank the ARBIMON team for their help, especially Alberto Estrada, Josi Mar Figueroa, Carlos J. Milán, Beatriz Otero, Sharissa Ramírez and Neysha Sánchez for analyzing thousands of field recordings. Two anonymous reviewers gave us insightful comments that greatly improved this manuscript.

References

Acevedo, M.A., Villanueva-Rivera, L.J., 2006. Using automated digital recording systems as effective tools for the monitoring of birds and amphibians. Wildlife Society Bulletin 34, 211–214.

Balfoort, H.W., Snoek, J., Smits, J.R.M., Breedveld, L.W., Hofstraat, J.W., Ringelberg, J., 1992. Automatic identification of algae: neural network analysis of flow cytometric data. Journal of Plankton Research 14, 575–589.

Boddy, L., Morris, C.W., Wilkins, M.F., Tarran, G.A., Burkill, P.H., 1994. Neural network analysis of flow citometric data for 40 marine phytoplankton species. Cytometry 15, 283–293.

Brandes, T.S., Naskrecki, P., Figueroa, H.K., 2006. Using image processing to detect and classify narrow-band cricket and frog calls. Journal of the Acoustical Society of America 120, 2950–2957.

Byvatov, E., Fechner, U., Sadowski, J., Schneider, G., 2003. Comparison of support vector machine and artificial neural network systems for drug/nondrug classification. Journal of Chemical Information and Computer Sciences 43, 1882–1889.

Caputo, B., Sim, K., Fursejo, F., Smola, A., 2002. Appearance-based object recognition using SVMs: which kernel should i use? Proceedings of NIPS workshop on Statistical Methods for Computational Experiments in Visual Processing and Computer Vision. Whistler.

Chesmore, E.D., Femminella, O.P., Swarbrick, M.D., 2001. Automated analysis of insect sounds using time-encoded signals and expert systems - a new method for species identification. In: Bridge, P., Jeffries, P., Morse, D.R., Scott, P.R. (Eds.), Information Technology, Plant Pathology and Biodiversity. CAB International, Wallinford, pp. 273–287.

Chih-Chung, C., Chih-Jen, L., 2004. LIBSVM: a library for support vector machines. http://www.csie.ntu.edu.tw/cjlin/libsvm/.

Condit, R., 1995. Research in large, long-term tropical forest plots. Trends in Ecology and Evolution 10, 18–22.

Do, M.T., Harp, J.M., Norris, K.C., 1999. A test of a pattern recognition system for identification of spiders. Bull. Entomol. Res. 89, 217–224.

Drewry, G.E., Rand, A.S., 1983. Characteristics of an acoustic community: Puerto Rican frogs of the genus Eleutherodactylus. Copeia 1983, 941–953.

Fagerlund, S., 2007. Bird species recognition using support vector machines. EURASIP Journal of Advances in Signal Processing 1–8.

Gauld, I.D., O'Neill, M.A., Gaston, K.J., 2000. Driving Miss Daisy: the performance of an automated insect identification system. In: Austin, A.D., Dowton, M. (Eds.), Hymenoptera: Evolution, Biodiversity and Biological Control. CISRO, Collingwood, VIC, pp. 303–312.

Hastie, T., Tibshirani, R., Friedman, J., 2001. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer.

Herr, A., Klomp, N.I., Atkinson, J.S., 1997. Identification of bat echolocation calls using a decision tree classification system. Complexity International 40.

Karatzoglou, A., Smola, A., Hornik, K., Zeileis, A., 2004. Kernlab – An S4 package for kernel methods in R". Journal of Statistical Software 11, 1–20.

Kimeldorf, G.S., Wahba, G., 1971. Some results on Tchebycheffian spline functions. Journal Mathematical Analysis Applications 33, 82–95.

Kogan, J.A., Margoliash, D., 1998. Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden Markov models: a comparative study. Journal of the Acoustical Society of America 103, 2185–2196.

Kreßel, U., 1999. Pairwise classification and support vector machines. Advances in Kernel Methods. Support Vector Learning. 255–268.

Lee, Y., Lin, Y., Wahba, G., 2004. Multicategory support vector machines: theory and application to the classification of microarray data and satellite radiance data. Journal of the American Statistical Association 99, 67–82.

Levin, S.A., 1992. 'The problem of pattern and scale in ecology. Ecology 73, 1943–1967.

Nickerson, C.M., Bloomfield, L.L., Dawson, M.R.W., Sturdy, C.B., 2006. Artificial nerual network discrimination of black-capped chickadee (*Poecile atricapillus*) call notes. Journal of the Acoustical Society of America 120, 1111–1117.

Oswald, J.N., Rankin, S., Barlow, J., Lammers, M.O., 2007. A tool for real-time acoustic species identification of delphinid whistles. Journal of the Acoustical Society of America 122, 587–595.

Parsons, S., Jones, G., 2000. Acoustic identification of twelve species of echolocating bat by discriminant function analysis and artificial neural networks. Journal of Experimental Biology 203, 2641–2656.

Porter, J., Arzberger, P., Braun, H., Bryant, P., Gage, S., Hansen, T., Hansen, P., Lin, C., Lin, F., Kratz, T., Michener, W., Shapiro, S., Williams, T., 2005. Wireless sensor networks for ecology. BioScience 55, 561–572.

R Development Core Team, 2007. R: a language and environment for statistical computing. http://www.r-project.org.

Rickwood, P., Taylor, A., 2008. Methods for automatically analyzing humpback song units. Journal of the Acoustical Society of America 123, 1763–1772.

Sanchez-Hernandez, C., Boyd, D.S., Foody, G.M., 2007. Mapping spe- cific habitats from remotely sensed imagery: support vector machine and support vector data description based classification of coastal saltmarsh habitats. Ecological Informatics 2, 83–88.

Simmonds, E.J., Armstrong, F., Copland, P.J., 1996. Species identification using wideband backscatter with neural network and discriminant analysis. ICES Journal of Marine Science 53, 189–195.

Skowronski, M.D., Harris, J.G., 2006. Acoustic detection and classification of micro-chiroptera using machine learning: lessons learned from automatic speech recognition. Journal of the Acoustical Society of America 119, 1817–1833.

Tan, C.P., Lani, N.F.M., Lai, W.K., 2008. Application of support vector machine classifier for security surveillance system. Advances in Computer Science and Technology 605, 101.

Therenau, T.M., Atkinson, B., Ripley, B., 2007. rpart: recursive partitioning. R package version 3, 1–38.

Trifa, V.M., Krischel, A.N.G., Taylor, C.E., 2008. Automated species recognition of antbirds in a Mexican rainforest using hidden Markov models. Journal of the Acoustical Society of America 123, 2424–2431.

Venables, W.N., Ripley, B.D., 2002. Modern Applied Statistics with S. 4ed. Springer, New York, USA.

Villanueva-Rivera, L.J., 2007. Digital recorders increase detection of *Eleutherodactylus* frogs. Herpetological Review 38, 59–63.