

Feature Set Comparison for Automatic Bird Species Identification

Marcelo Teider Lopes
Federal University of Technology of Paraná
Av. 7 de Setembro 3165
Curitiba, Paraná 80230-901, Brazil
teiderlopes@gmail.com

Carlos Nascimento Silla Junior
University of Kent – School of Computing
Canterbury, Kent CT2 7NF, UK
cns2@kent.ac.uk

Alessandro Lameiras Koerich
Federal University of Paraná
Centro Politécnico CP 19.011
Curitiba, Paraná 81531-970, Brazil
alessandro.koerich@ufpr.br

Celso Antonio Alves Kaestner
Federal University of Technology of Paraná
Av. 7 de Setembro 3165
Curitiba, Paraná 80230-901, Brazil
celsokaestner@utfpr.edu.br

Abstract—This paper deals with the automated bird species identification problem, in which it is necessary to identify the species of a bird from its audio recorded song. This is a clever way to monitor biodiversity in ecosystems, since it is an indirect non-invasive way of evaluation. Different features sets which summarize in different aspects the audio properties of the audio signal are evaluated in this paper together with machine learning algorithms, such as probabilistic, instance-based, decision trees, neural networks and support vector machines. Experiments are conducted in a dataset of recorded songs of three bird species. The experimental results compare the performance of the features sets and different classifiers showing that it is possible to obtain very promising results in the automated bird species identification problem.

Index Terms—machine learning; pattern recognition; signal processing; bird species identification

I. INTRODUCTION

Ecological concerns are nowadays in the center of the contemporaneous society thoughts. In order to understand and evaluate the changes in our living environment, it is necessary to continuously obtain reliable information about the population of wild animals. In this context birds play an important role, since they are one of the most numerous classes that have direct contact with humans. Birds can be monitored indirectly by recording their songs, which greatly facilitates the surveillance process. Therefore, the use of automated methods for bird species identification – including both hardware and software – is fundamental for an effective monitoring and evaluation of the quantity and diversity of birds that are present in a specific ecosystem [2], [3].

An important ethical and moral benefit that the use of automated indirect technology brings about is the animal welfare of the birds being monitored. The awareness for animal welfare is a raising concern and an area of study in the field of Veterinary Medicine and Zoo Technology. The Canadian Council on Animal Care (CCAC)¹ has a scale to measure

the degree of invasiveness (and subsequently pain) in any procedure that involves animals. The scales goes from *A* (the least invasive) to *E* (the most invasive). According to this scale, any procedure that is indirect is classified as having an invasiveness degree *A*.

There are many challenges to develop a system capable of monitoring birds indirectly by their songs. Among these challenges the most fundamental task is to be able to perform reasonably well the task of automated bird species identification (ABSI). The use of automated methods is usually based on the machine learning / data mining fundamentals [18], [25]. In the ABSI problem, the recorded singing of the birds is the available data that must be adequately handled to serve as input of a classification system. Each of the recorded sounds can be analyzed – normally by signal processing tools – to produce a set of discriminative features that represent the original bird song regarding its species. Therefore, if a bird song database of previously classified birds is available, it is possible to construct directly a standard database to be used in the design of classification methods. Several algorithms based on different paradigms can be employed, including the probabilistic and instance-based classifiers, decision trees, neural networks and support vector machines [18]. The general framework is outlined in Figure 1.

As in most of the data mining problems, the feature set definition and the choice of a classification algorithm are crucial for the overall classification performance [25]. In the problem we are tackling it is necessary to obtain a (finite) set of features from a recorded audio signal, a problem which has infinite solutions. Additionally, when dealing with audio records a lot of signal preprocessing tools can be applied, such as filters and noise removals, among others. This is important issue because bird sound recording usually occurs in noise environments, such as gloves and forests.

In this paper we present a study that compares different feature sets which were previously employed for audio signal

¹http://www.ccac.ca/en/About_CCAC/About_CCAC_Main.htm

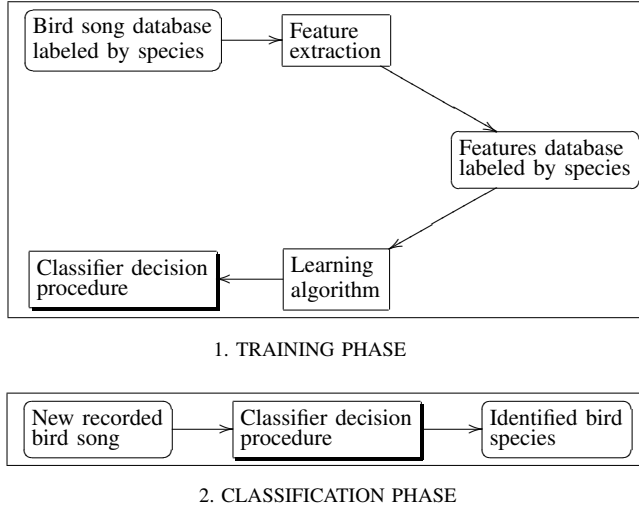


Fig. 1. The general classification framework

processing in the bird species identification problem, as well as different approaches to handle the input signal. We employ three feature sets: (a) the one produced by the MARSYAS framework [23]; (b) the IOIHC feature set, that taps into rhythmic properties of sound signals, using a particular rhythm periodicity function, the Inter-Onset Interval Histogram [12], [13]; and (c) the feature set obtained from the audio processing tool Sound Ruler [22], an open code and free tool for measuring and graphing sounds. The first two feature sets were mainly employed in automated music genre classification problems [12], [13], [19], [20], [23], while the last one was employed in previous experiments in bird species identification [24]. Two scenarios are considered: one employs the bird songs as recorded in the field; another considers the audio signals split according to pulses. We consider as pulses relatively short sound intervals, with high amplitudes, which better characterize the bird vocalization. We argue that the use of pulses outperforms the use of the complete sound recording in the ABSI problem. The employed feature sets are evaluated in a specific database with contains bird songs from three species. Therefore, the main contributions of this paper are: evaluating different types of features and classifiers for the ABSI task; verifying for this task if it is better to use the full recorded song signals or segmenting them into pulses.

This paper is organized as follows: Section II presents a formal view of the problem and Section III presents the employed feature sets, preprocessing steps and classification algorithms; Section IV describes the dataset we use and the experimental results obtained in the classification experiments; Section V summarizes the results of similar research works in automatic bird species identification; Finally, Section VI presents the conclusions of our work and proposes future research directions.

II. BIRD SPECIES IDENTIFICATION

The identification of bird species is a standard problem for ornithologists. In recent years, the use of autonomous devices and computers greatly facilitate this task. Acoustic communication in birds is rich and it is one of the most direct ways for humans to detect them, even though when it is difficult to see the bird itself. Therefore, the use of sounds can be considered as one the most efficient means for monitoring birds. In addition, acoustic surveillance of birds also allows quick evaluations of the biodiversity of specific regions [3].

A. Problem definition

In this paper we define the automated bird species identification problem (ABSI) as the task of finding the species of a specific bird from its recorded singing sounds. For most species the bird songs are also beautiful and interesting, which makes this task still more pleasant. In general, bird sounds can be classified as songs and calls. Songs are related to the coupling, and calls are usually employed as a danger signal or for other communication activities. Calls usually correspond to very short and transient sounds; bird songs are longer and melodious, and are considered by experts to be more adequate for bird species classification [6], [24]. So, we restrict our problem to bird songs.

With the advent of digital recording an audio signal is no longer analogous to the original sound wave. The analog signal is sampled, several times per second, and transformed by an analog-to-digital converter into a sequence of numeric values in a convenient scale. This sequence represents the digital audio signal, and can be employed in reproduction and analytical processes [14].

Hence the digital audio signal can be represented by a sequence $S = \langle s_1, s_2 \dots s_N \rangle$ where s_i stands for the signal sampled in the instant i and N is the total number of samples of the signal. This sequence contains a lot of acoustic information, and several features can be extracted from it. Therefore, a feature vector $\vec{X} = \langle x_1, x_2 \dots x_D \rangle$ can be generated, where each feature x_j is extracted from S (or some part of it) by an appropriate extraction procedure, defined as a function $\chi_j : S \rightarrow X_j$, where X_j is the feature domain.

Now we can formally define the ABSI problem as a pattern classification problem, using segment-level features as input. From a finite set of bird species \mathcal{B} we must select one class \hat{b} which best represents the species of the bird that produces the song associated to the input signal S .

From a statistical perspective the goal is to find the most likely $\hat{b} \in \mathcal{B}$, given the feature vector $\vec{X} = \chi(S)$, that is

$$\hat{b} = \arg \max_{b \in \mathcal{B}} P(b|\vec{X}) \quad (1)$$

where $P(b|\vec{X})$ is the *a posteriori* probability that the song belong to a bird of the species b given the features expressed by \vec{X} . Using the Bayes' rule the equation can be rewritten as

$$\hat{b} = \arg \max_{b \in \mathcal{B}} \frac{P(\vec{X}|b) \cdot P(b)}{P(\vec{X})} \quad (2)$$

where $P(\bar{X}|b)$ is the probability in which the feature vector \bar{X} occurs in class b , $P(b)$ is the *a priori* probability of the bird species b , that can be estimated from frequencies in the database, and $P(\bar{X})$ is the probability of occurrence of the feature vector \bar{X} . The last probability is usually unknown, but if the classifier computes the likelihoods of the entire set of bird species, then $\sum_{b \in \mathcal{B}} P(b|\bar{X}) = 1$ and we can obtain the desired probabilities for each $b \in \mathcal{B}$ by

$$P(b|\bar{X}) = \frac{P(\bar{X}|b) \cdot P(b)}{\sum_{b \in \mathcal{B}} P(\bar{X}|b) \cdot P(b)} \quad (3)$$

III. THE ABSI SYSTEM FRAMEWORK

One of the contributions of this paper is to evaluate different feature sets and classification algorithms applied to the ABSI problem. In this section we briefly describe the three feature sets used in this work as well as the classification algorithms.

A. The MARSYAS feature set

Initially we employ the MARSYAS framework [17] for feature extraction from each audio segment. This framework originally implements the feature set proposed by Tzanetakis and Cook [23] for music genre classification.

The employed features encompass means and variances of timbral features, calculated in the intervals, for the spectral centroid, rolloff, flux, the time-domain zero crossings, including the twelve initial Mel-Frequency Cepstral coefficients (MFCC) in each case. The complete set includes 64 features. More details can be found in [17].

B. The IOIHC feature set

In the Inset-Onset Interval Histogram Coefficients, features are related to rhythmic properties of sound signals [12], [13]. The features are computed from a particular rhythm periodicity function (IOIH) that represents normalized salience with respect to period of inter-onset intervals which are present in the signal.

The IOIH is further parameterized by the following steps: (a) projection of the IOIH period axis from linear scale to the Mel scale, of lower dimensionality, by means of a filter; (b) computation of the IOIH magnitude logarithm; and (c) computation of the Inverse Fourier Transform, keeping the first 40 coefficients. These steps produce features analogous to the MFCC coefficients, but in the domain of rhythmic periods rather than in signal frequencies. The overall procedure generates a 40-dimensional feature vector that is employed for classification.

C. The Sound Ruler feature set

The Sound Ruler [22] is a free tool for measuring and graphing sounds and for teaching acoustics. It provides an interactive visual interface that allows a productive analysis, since it conjugates the control of manual analysis and the objectivity and speed of the automated analysis.

Among several signal processing options, including oscillogram and spectrogram visualizations, Sound Ruler can calculate 36 features from a pulse sound signal. The feature set includes: peak time, relative pulse peak, pulse durations and shapes, pulse intervals and periods, duty cycle, crest factor, several energy measures in the subintervals, dominant, fundamental, min and max frequencies, and relative amplitudes. More detailed explanation is available in the Sound Ruler web site [22]. This feature set was also employed by Vilches et al. [24].

D. Classification algorithms

A set of classifiers based on different paradigms was used to evaluate diverse possibilities. We employ the classical probabilistic Naïve Bayes algorithm, the instance-based k nearest neighbors (kNN) with $k = 3$, the decision tree classifier J4.8, a multi-layer perceptron neural network trained with the back-propagation momentum algorithm, and the support vector machine classifier (SVM), using the Platt's Sequential Minimization Algorithm (SMO) implementation, with two different kernel functions: polynomial and Pearson VII function-based universal kernel.

IV. EXPERIMENTS

The experiments were carried out on bird songs obtained from the Xeno-Canto website [26]. We have limited our experiments to three species of birds: great antshrike (*Taraba major*) – 32 audio records, dusky antbird (*Cercomacra tyrannina*) – 34 audio records and barred antshrike (*Thamnophilus doliatus*) – 35 audio records. All the files are in “.wav” format and are formed only by bird songs. The audio records were obtained directly in real environments, without any filtering or similar preprocessing, and thus contain noise and sounds of other animals and from the environment. Therefore, our database is very similar to the one used by Vilches et al. [24].

In the following subsections we describe two sorts of experiments. In the first experiment our aim is to evaluate different feature sets and different classifiers using bird song pulses. In the second experiment our aim is to evaluate whether it is better to use the full recorded bird signal or to divide the bird songs into pulses, as used in the first experiment. Also, all experiments reported in this section were carried out considering a 5-fold cross-validation procedure, that is, the presented results are obtained from five randomly independent experiment repetitions.

A. Evaluation of different feature sets and classifiers

For the experiments reported in this subsection, the audio records were split into separate pulses, eliminating some parts of the audio record where the bird songs were not present, as shown in Figure 2. The Audacity audio processing tool [1], an open source software tool for digital audio, was employed to split the bird songs into pulses. A new dataset was generated, with 74, 136 and 102 audio records for *Taraba major*, *Cercomacra tyrannina* and *Thamnophilus doliatus*, respectively. A similar procedure was employed in [24].

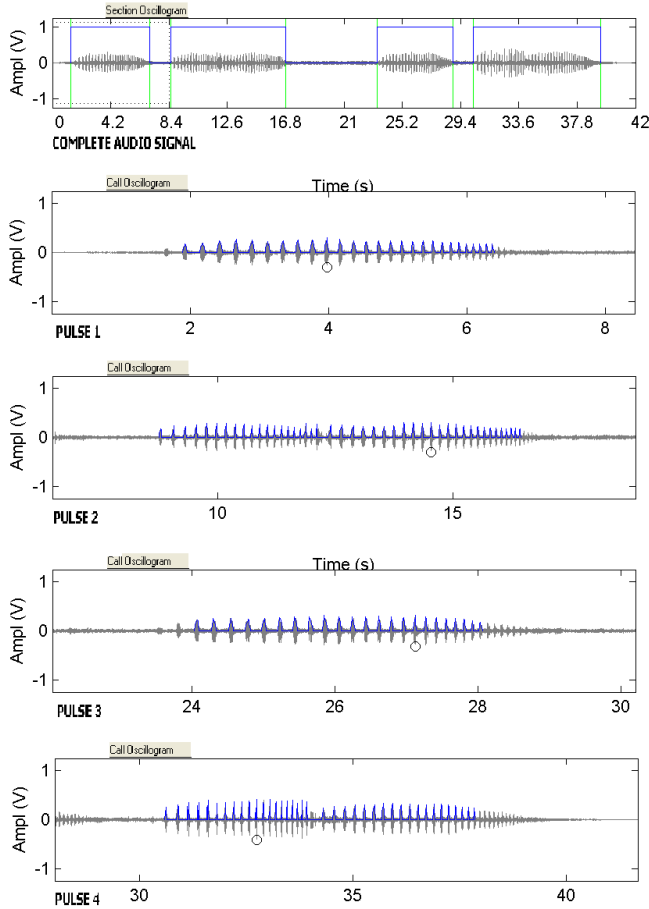


Fig. 2. Audio partition in song pulses

From this new dataset, we have extract the MARSYAS, IOIHC and Sound Ruler feature sets (described in Section III), and we applied the different classification algorithms. As previously mentioned, the feature set sizes are 64, 40 and 36 respectively. Table I summarizes the results obtained with the different feature sets and classifiers. The values represent the weighted average for the three classes.

The analysis of the results presented in Table I shows that the IOIHC feature set is the one that provides the worst results. One possible reason for this is the fact that we are dealing with bird vocalizations, where the rhythmic content of the songs originated from different species are not very different. However, both the features sets extracted from Sound Ruler and MARSYAS provide interesting results.

The Friedman test with the post-hoc Shaffer's static procedure was employed to evaluate if there are statistically significant differences between the results originated from the three feature sets. This procedure is strongly recommended by Garcia and Herrera [11] for the comparison of classifiers over multiple datasets. Table II presents the results of this test using the values of F-measure. The first column of Table II presents which feature sets are being compared. The second

TABLE I
F-MEASURE ON THE BIRD SONGS DATASET (%)

Classifier	Feature Set		
	Sound Ruler	IOIHC	MARSYAS
Naïve Bayes	99.7	43.5	86.9
kNN ($k = 3$)	96.8	57.4	98.4
J4.8	99.0	61.0	99.7
MLP	98.7	68.0	99.7
SVM (Polynomial)	97.8	53.5	99.4
SVM (Pearson)	99.4	64.3	99.4

column presents the p value of the statistical test, which needs to be lower than the corrected critical value shown on the third column to have a statistically significant difference between the two feature sets at 95% confidence level.

TABLE II
RESULTS OF STATISTICAL TESTS FOR $\alpha = 0.05$

Feature Set	p	Shafer
IOIHC vs. MARSYAS	0.0024	0.0166
Sound Ruler vs. IOIHC	0.0303	0.0500
Sound Ruler vs. MARSYAS	0.3864	0.0500

The analysis of the statistical tests presented in Table II shows that there is no statistically significant difference between the results obtained from MARSYAS and Sound Ruler features sets. However, for practical reasons we recommend that the MARSYAS framework is used as its current implementation already has a non-graphical mode which allows the software to be used to process files in batch, while the current version of Sound Ruler does not have this feature. For this reason, and since there is no significant difference between the two feature sets results, all the experiments presented in the next section were performed using MARSYAS features only.

B. Comparison of signal segmentation: pulses x full audio

In this section we perform experiments that use the complete audio bird songs, directly as recorded in the field. The main motivation of this experiment is to verify if using the full audio signal it is possible to achieve similar results to the ones obtained in the previous experiment. If this occurs a considerable preprocessing effort can be eliminated. We use the MARSYAS feature set and the same classifiers. Table III presents the comparison of using the complete audio signal and the corresponding pulses for the MARSYAS feature set.

TABLE III
FULL AUDIO SIGNAL X CORRESPONDING PULSE SIGNALS: F-MEASURE FOR THE MARSYAS FEATURE SET (%)

Classifier	Full audio	Pulses
Naïve Bayes	49.6	86.9
kNN ($k = 3$)	57.3	98.4
J4.8	62.2	92.2
MLP	79.2	99.7
SVM (Polynomial)	76.0	99.4
SVM (Pearson)	71.3	99.4

The analysis of the results in Table III shows that splitting the bird song into pulses is a better approach than using the

full audio signal. This is probably due to the fact that the bird song files employed in the experiments present several silent intervals, where the bird sing itself is not present, but only the environmental noise. Also, we argue that pulses encompasses the most significant parts of the audio signal in terms of bird song characteristics, which reflects in the features values extracted from them.

V. RELATED WORKS

In recent years several papers have dealt with the ABSI problem using pattern recognition and machine learning techniques. We summarize the results of some of these works.

Kwan et al. [15] have proposed a bird classification system for identify dangerous birds near airports. They have employed hidden Markov models and Gaussian mixture models in the classification, achieving poor results due to the low signal-to-noise ratio in the airport environment as well as due to the limited performance of the acoustic monitoring device.

Somervuo, Härmä and Fagerlund [21] have developed some signal processing techniques for the ABSI problem. They have used a sinusoidal modeling together with the Mel-frequency cepstral coefficients. They have evaluated their proposal in 14 common North-European bird species, and the best accuracy achieved was about 71.3 %.

Vilches et al [24] have tackled the ABSI problem using some data mining algorithms such as the ID3, J4.8, and Naïve Bayes. They have considered that the identification of distinctive features is crucial in resource constrained applications. For this reason, they have investigated dimensionality reduction in relation to classification accuracy. Using a database containing 154 song files from three bird species, their best result was 98.39 % using the J4.8 classifier with a full feature dataset, obtained with the use of the Sound Ruler audio processing tool.

Chou, Lee and Ni [7] have proposed a bird recognition system where bird songs are segmented into many syllables, from which a frequency spectrum is obtained. Syllables are obtained by clustering, using a fuzzy C-means method, and each syllable group is modeled by a hidden Markov model to characterize the songs of each bird species. In the experiments they obtained a recognition rate of 78 % in a database of 420 kinds of bird species.

Fagerlund [10] has employed a global decision tree with support vector machines classifiers in each node, each one employed to separate two species. The author has employed two feature sets: Mel-frequency cepstral coefficients and a set of low level signal parameters. His best overall result was 98 %, obtained in a database with 8 bird species.

Cai et al. [5] have used a neural network classifier with different sets of features to investigate the ABSI problem. They have employed a neural network architecture that considers the dynamic nature of the bird songs. Additionally, noise reduction algorithms were used. They test their proposal in a database which contains data from 14 bird species and the best accuracy was achieved using a neural network with 160 hidden units and a Mel-frequency cepstral coefficients feature set.

Chou, Liu and Cai [8] have proposed an enhanced syllable segmentation method based on Rabiner and Sambur endpoint detection method. This method is combined with a Mel-frequency cepstral coefficients based feature vector to deal with two problems: syllable detection and birdsong section recognition. They have used songs from a commercial CD with bird calls and songs of 420 bird species, with recordings made in the field. The best obtained recognition rate result was 73.19 %, using a neural network trained with the backpropagation algorithm.

Lee, Han and Chuang [16] have presented a method for automatic classification of bird species that splits the original signal into syllable segments, which are considered as the basic recognition unit. Then Mel-frequency cepstral coefficients are calculated, and Gaussian mixture model and vector quantization are employed to find the most appropriate number of Gaussian mixture model components and cluster number of vector quantization for each species. In the experiments they obtain the best classification accuracy of 84 % for 28 bird species.

Briggs, Raich and Fern [4] have proposed a probabilistic model for audio features and use a risk-minimizing Bayes classifier, showing it is closely approximated by a nearest-neighbor classifier that uses Kullback-Leibler divergence to compare histograms of features. The proposed classifiers have achieved an accuracy over 90 % on a dataset that contains 6 species of birds.

Chou and Liu [9] have used a wavelet transformation to transform sections of the bird songs. Then, the first five order Mel-frequency cepstral coefficients are computed, and the Mel-frequency cepstral coefficients of the same order are aligned. They use a neural network classifier in a database with 420 bird species, achieving 73.41 % for the recognition rate.

From the state of the art it is clear that several papers have attacked the ABSI problem with various feature sets and different machine learning algorithms. However, to the best of our knowledge, no direct comparison of the approaches over the same database is available, and there is no accepted theory of which features are adequate in audio classification. Also, there is no clear idea whether one should divide the bird songs into pulses or to use the full audio signal in classification. Our paper aims to fill this gap.

VI. CONCLUSIONS AND FUTURE WORK

In this paper we compare the performance of several feature sets and several classifiers in the automated bird species identification problem. Our primary evaluation metric was the F-measure. Results show that MARSYAS and Sound Ruler feature sets present good performance in the experiments, for almost every classifier. On the contrary, IOIHC feature set seems to be inadequate for the ABSI problem. We argue that this occurs due to the rhythmic characteristics of the bird songs being different than the ones for music, were this feature set is employed with success.

Best results for the MARSYAS feature set were obtained with J4.8 and multilayer perceptron neural network classifiers,

whereas for the Sound Ruler feature set the best results were obtained using the Naïve Bayes probabilistic classifier. However, the conducted statistic tests show that the differences in the resulting F-measure values are not significant. This situation can be explained because most of the features in both sets are similar in nature, consisting mainly in Mel-frequency cepstral coefficient values calculated over specific time intervals.

We also compare the application of the MARSYAS feature extraction procedures to the full audio recording signal and to a series of pulse signals identified in the original signal. Experimental results show that the second option is the most adequate to the problem, since in this case the feature values are extracted from the most representative parts of the audio signal.

In order to compare our results with the ones listed in section V, we indicate our best accuracy results: for the MARSYAS feature set we obtain 99.7 % of correct classification using the multilayer perceptron neural network classifier, whereas for the Sound Ruler feature set we obtain the very same accuracy using the Naïve Bayes classifier. These values are superior to most of the results presented in the previous section. The fairest comparison can be done with the work of Vilches et al. [24], that uses the same bird species and the Sound Ruler feature set, and achieves an accuracy of 98.39 % in its best result.

We plan to apply feature selection procedures to attack the ABSI problem, in order to discover which are the most important sound characteristics to identify a bird species. Similarly, new sets of features can be applied to the problem; we intend to do so in the future.

As shown by our experimental results, the selection of the time intervals of the audio record to be considered as input for the feature extraction procedure is very important to improve the classification accuracy. Therefore, we intend to elaborate an automatic procedure to detect pulses from bird song signals, that can be customized according to the environment noise.

This study is also part of a large project that encompasses hardware and software development to monitor the diversity of bird species in specific ecosystems, with wide applications, such as ecological impact project evaluation and bird population control. A database containing bird songs from 76 species in a specific ecosystem is under construction, where we plan to apply the presented techniques.

ACKNOWLEDGMENT

This research work is supported by CNPq and Fundação Araucária Brazilian Agencies.

REFERENCES

- [1] Audacity Web Site, available in <<http://audacity.sourceforge.net/>> accessed in june 26th, 2010.
- [2] R. Bardeli, D. Wolff, F. Kurth, M. Koch, K-H. Tauchert and K-H. Frommolt, "Detecting Bird Songs in a Complex Acoustic Environment and Application to Bioacoustic Monitoring", *Pattern Recognition Letters*, Vol. 31, No. 12, pp.1524–1534, 2010.
- [3] T.S. Brandes, "Automated Sound Recording and Analysis Techniques for Bird Surveys and Conservation", *Bird Conservation International*, Vol. 18, pp. 163–173, 2008.
- [4] F. Briggs, R. Raich and X.Z. Fern, "Audio Classification of Bird Species: a Statistical Manifold Approach", *Proceedings of the 9th International Conference on Data Mining (ICDM'2009)*, Miami, USA, pp. 51–60, december 2009.
- [5] J. Cai, D. Ee, B. Pham, P. Roe and J. Zhang, "Sensor Network for the Monitoring of Ecosystem: Bird Species Recognition", *Proceedings of the 3rd IEEE International Conference on Intelligent Sensors, Sensor Networks and Information (ISSNIP'07)*, Melbourne, Australia, pp.293–298, december 2007.
- [6] C.K. Catchpole and P.J.B. Slater, *Bird Songs: Biological Themes and Variations*, Cambridge University Press, 1995.
- [7] C-H. Chou, C-H. Lee and H-W Ni, "Bird Species Recognition by Comparing the HMMs of Syllables", *Proceedings of The 2nd International Conference on Innovative Computing, Information and Control (ICICIC'07)*, Kumamoto City, Japan, pp. 143–147, september 2007.
- [8] C-H. Chou, P-H. Liu and B. Cai, "On the Studies of Syllable Segmentation and Improving MFCCs for Automatic Birdsong Recognition", *Proceedings of the Asian Pacific Services Computing Conference (APSCC'08)*, Yilan, Taiwan, pp. 745–750, december 2008.
- [9] C-H. Chou and P-H. Liu, "Bird Species Recognition by Wavelet Transformation of a Section of Birdsong", *Proceedings of the Symposia and Workshop on Ubiquitous, Autonomic and Trusted Computation (UIC-ATC'09)*, Brisbane, Australia, pp. 189–193, july 2009.
- [10] S. Fagerlund, "Bird Species Recognition Using Support Vector Machines", *EUSASIP Journal on Advances in Signal Processing*, Vol. 2007, Article ID 38637, pp. 1–8, 2007.
- [11] S. Garcia and F. Herrera, "An extension on "Statistical comparisons of classifiers over multiple data sets" for all pairwise comparisons", *Journal of Machine Learning Research*, vol. 9, pp. 2677–2694, 2008.
- [12] F. Gouyon, P. Herrera and P. Cano, "Pulse-dependent analysis of percussive music", *Proceedings of the 22th International AES Conference on Virtual, Synthetic and Entertainment Audio*, Espoo, Finland, 2002.
- [13] F. Gouyon, S. Dixon, E. Pampalk and G. Widmer, "Evaluating rhythmic descriptions for music genre classification", *Proceedings of the 25th International AES Conference on Virtual, Synthetic and Entertainment Audio*, London, UK, 2004.
- [14] S. Hacker, *MP3: The Definitive Guide*, O'Reilly Publishers, 2000.
- [15] C. Kwan, X. Zhao, Z. Ren, R. Xu, V. Stanford, C. Rochet, J. Aube and K.C. Ho, "Bird Classification Algorithms: Theory and Experimental Results", *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'04)*, Montreal, Canada, Vol. 5, pp 289–292, may 2004.
- [16] C-H. Lee, C-C Han and C-C. Chuang, "Automatic Classification of Bird Species from their Sounds using Two-Dimensional Cepstral Coefficients", *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 16, No. 8, pp. 1541–1550, 2008.
- [17] Marsyas Web Site, available in <<http://marsyas.info/>> accessed in june 24th, 2010.
- [18] T. M. Mitchell, *Machine Learning*, McGraw-Hill, 1997.
- [19] C. N. Silla Jr., C.A.A. Kaestner and A. L. Koerich, "Automatic Music Genre Classification using Ensemble of Classifiers", *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics (SMC'07)*, Montreal, Canada, pp. 1687–1692, 2007.
- [20] C. N. Silla Jr., A. L. Koerich and C. A. A. Kaestner, "A Machine Learning Approach to Automatic Music Genre Classification", *Journal of the Brazilian Computer Society*, Vol. 14(3), pp. 7–18, 2008.
- [21] P. Somervuo, A. Härmä and S. Fagerlund, "Parametric Representations of Bird Sounds for Automatic Species Recognition", *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 14, No. 6, pp. 2252–2263, november 2006.
- [22] Sound Ruler Web Site, available in <<http://soundruler.sourceforge.net/>> accessed in june 24th, 2010.
- [23] G. Tzanetakis and P. Cook, "Musical Genre Classification of Audio Signals", *IEEE Transactions on Speech and Audio Processing*, Vol. 10, pp. 293–302, 2002.
- [24] E. Vilches, I.A. Escolbar, E.E. Vallejo and C.E. Taylor, "Data Mining Applied to Acoustic Bird Species Recognition", *Proceedings of the 18th IEEE International Conference on Pattern Recognition (ICPR'06)*, Hong Kong, China, pp. 400–403, 2006.
- [25] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann, San Francisco, 2005.
- [26] Xeno-Canto Web Site, available in <<http://xeno-canto.org/>> accessed in june 26th, 2010.