Automated Acoustic Classification of Bird Species from Real -Field Recordings

Iosif Mporas, Todor Ganchev, Otilia Kocsis, Nikos Fakotakis
Artificial Intelligence Group
Dept. of Electrical and Computer Engineering
University of Patras
26500 Patras, Greece
{imporas, tganchev, okocsis, fakotaki}@upatras.gr

Olaf Jahn, Klaus Riede, Karl-L. Schuchmann
Zoologisches Forschungsmuseum Alexander Koenig
53113 Bonn, Germany
{o.jahn.zfmk, k.riede.zfmk, kl.schuchmann.zfmk}@uni-bonn.de

***Abstract*—We report on a recent progress with the development of an automated bioacoustic bird recognizer, which is part of a long-term project , aiming at the establishment of an automated biodiversity monitoring system at the Hymettus Mountain near Athens. In particular, employing a classical audio processing strategy, which has been proved quite successful in various audio recognition applications, we evaluate the appropriateness of six classifiers on the bird species recognition task. In the experimental evaluation of the acoustic bird recognizer, we made use of real-field audio recordings for seven bird species, which are common for the Hymettus Mountain. Encouraging recognition accuracy was obtained on the real-field data, and further experiments with additive noise demonstrated significant noise robustness in low SNR conditions.**

Keywords-bioacoustics; biodiversity informatics; acoustic bird species recognition; automatic recognition

I. INTRODUCTION

The biodiversity conservation is one of the most crucial issues that governments and international organizations have to deal with. The protection of endangered species is a must and relies primarily on the accurate monitoring of the biodiversity and secondarily on the application of conservation actions, based on the monitored biodiversity status. The observation and monitoring of birds are of major importance for the biodiversity conservation [1].

Large amount of information, concerning bird activity, has been collected by expert ornithologists. This effort includes recognition of bird species from their vocalizations, study of the interaction among them, and locating of their habitats. Such surveys are time consuming and tedious, since they require the physical presence of expert ornithologists in the field. Furthermore, these manual observations will heavily rely on the visual and acoustic abilities of the expert as well as on the degree of knowledge of the surveyor on the under investigation group of bird species. Finally, due to the difficulty of the task most of the surveys take place in infrequent time intervals, thus not allowing the long-term biodiversity monitoring of inhospitable habitats.

The disadvantages of manual observation of bird activity have led to the development and study of several approaches for automatic recognition of bird species from their vocalizations over the last decade. Automatic acoustic bird species recognition is a pattern recognition task, involving preprocessing and feature extraction of the audio signal and classification.

The first attempts in automatic bird species recognition from their vocalizations were based on template-matching techniques (dynamic time warping) [2, 3] and hidden Markov models [4], which have extensively been used in the similar task of speech recognition. Hidden Markov models have been used in more recent studies [5-7], due to their well known structure. Neural networks have also been used for the recognition of birds vocalizations using spectral and temporal parameters of the audio signal [8, 9]. Other approaches use Gaussian mixture model based schemes [6, 10, 11], support vector machines [12] and decision trees [13] for the recognition of bird songs. Other proposed classification schemes are based on sinusoidal modeling of bird syllables [14] and bird syllable pair histograms [15]. Different parametric representations have been used for audio signals with bird vocalizations, among which Mel frequency cepstral coefficients [5, 6, 16, 17] are the most widely used. Other audio features used in the literature are the linear predictive coding [16], linear predictive cepstral coefficients [16], spectral and temporal audio descriptors [12], and tonal-based features [17].

Most of the previous studies on the task of bird species recognition from their vocalizations focus on in-lab conditions of recordings, without the presence of real environmental noise [2, 3, 4, 6, 12, 13]. Exceptionally to most of the published work, in [17] waterfall noise was added to bird recordings and it was shown that the recognition of bird sounds in noisy conditions reduces significantly the recognition performance. In this paper, we evaluate a number of different machine learning algorithms on the task of bird species classification in real-field conditions, under the concept of AMIBIO project (LIFE08-NAT-GR-000539: Automatic Acoustic Monitoring and Inventorying of Biodiversity, Project web-site: http://www.amibio-project.eu/).

The rest of this paper is organized as follows. In Section 2, the bird species recognition task in real-filed is presented. Section 3 offers description of the audio data used and the

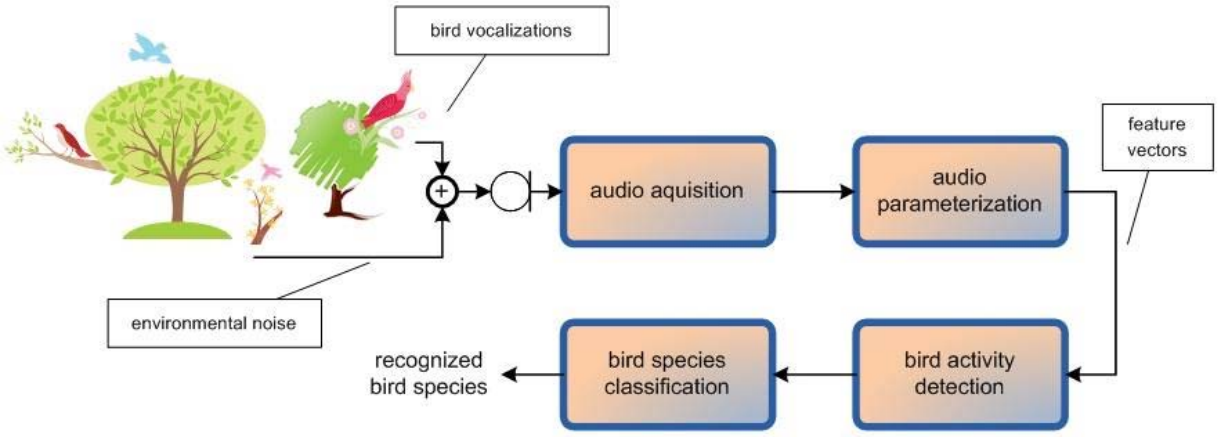

Figure 1. Block diagram of the bird species recognition scheme in real-field conditions.

experimental protocol followed in the present evaluation. In Section 4 the experimental results are presented. Section 5 concludes this work.

II. Bird Species Recognition in Real-Field

Automatic bird species recognition from audio data offers 24/7 monitoring of specific habitats and provides the information needed for biodiversity monitoring, animal species population estimation, species behavior understanding etc. The bird species recognition task falls in the category of audio pattern recognition task, and can briefly be structured in the audio acquisition stage, the audio parameterization stage and the classification stage. The recognition of the species in the birds habitats includes interferences that are additive to the bird species vocalizations. Such interferences are the wind, the sum of the leaves, vocalizations from other animal species of the habitat, sounds produces by human activities, etc. Typically, for the recognition of species in such a noisy environment detection of audio intervals with bird vocalizations precedes the species classification stage. The concept is illustrated in Figure 1.

In brief, the audio signal is captured by a microphone, next amplified and then sampled at 32 kHz, so that the wide frequency range of bird vocalizations from various species is covered. Precision of 16-bits per sample is used to guarantee sufficient resolution of details for the subsequent processing of the signal. After the audio acquisition stage the signal is decomposed to overlapping feature vectors of constant length, using spectral and temporal audio parameterization algorithms. The sequence of feature vectors is used as input to the bird activity detection block where the audio signal is binary segmented to intervals with or without bird vocalizations. Finally, the bird vocalization intervals are processed by the bird species classification block in order species-specific recognition to be performed.

The presence of non-stationary noises originating from the environment makes the species classification task more challenging. The degree of interference of the environmental noises and the actual sound-to-noise ratio are crucial for the recognition of bird species. In this work we focus on the effect of the distance of the bird from the monitoring station (field microphone) as expressed by different signal-to-noise ratios, to the species classification performance.

III. Experimental Setup

This section provides a description about the audio data used in the present evaluation, the audio parameterization algorithms used, the machine learning classification algorithms that were tested and the experimental protocol that was followed.

The dataset used in the present research consists of recordings of seven bird species which are common for the habitats of Hymettus Mountain, in Attica, Greece, namely the *bluerobin*, *buchfinck*, *chaffinch*, *chappuis*, *fringilla coelebs*, *kingfisher* and *western palaearctis*. The recordings of the vocalizations of these bird species were collected and manually labeled by expert ornithologists of the Zoologisches Forschungsmuseum Alexander Koenig (ZFMK). In order to test the bird species classification performance at different signal-to-noise ratios randomly selected recordings from the Hymettus area (from 16 different locations) were interfered to the bird vocalizations as additional noise. The amount of audio data used was approximately 14 minutes of recordings for all bird species.

For the parameterization of the audio signals we used a diverse set of audio parameters. In particular, the audio signal was blocked to frames of 20 milliseconds length and 10 milliseconds step. For audio descriptors we used two temporal audio descriptors and sixteen spectral audio descriptors. The temporal audio descriptors which were used are: the frame intensity (*Int*) and the zero crossing rate (*ZCR*). The sixteen spectral audio descriptors which were used are: the 12 first Mel frequency cepstral coefficients (*MFCCs*) as in the HTK setup [18], the root mean square energy of the frame (*E*), the voicing probability (*Vp*), the harmonics-to-noise ratio (*HNR*) by autocorrelation function and the dominant frequency (*Fd*) normalized to 500 Hz. All audio parameters were computed using the openSMILE acoustic parameterization tool [19]. After the computation of the audio parameters a post-processing with dynamic range normalization was applied to all audio features in order to equalize the range of their numerical values.

For the evaluation of the bird species classification we examined different machine learning algorithms:

- the k-nearest neighbors classifier with linear search of the nearest neighbor and without weighting of the distance – here referred as instance based classifier (*IBk*) [20].
- a 3-layer Multilayer perceptron (*MLP*) neural network with architecture 18–10–1 neurons (all sigmoid) trained with 50 000 iterations [21].
- the support vector machines utilizing the sequential minimal optimization algorithm (*SMO*) with a radial basis function kernel [22].
- the pruned C4.5 decision tree *(J48*), with 3 folds for pruning and 7 for growing the tree [23].
- the Bayes network learning (*BayesNet*) using a simple data-based estimator for finding the conditional probability table of the network and hill climbing for searching network structures [24].
- the Adaboost M1 method (*Adaboost(J48)*) using the pruned C4.5 decision tree as base classifier [25].
- the bagging algorithm (*Bagging(J48)*) for reduce of the variance of the pruned C4.5 decision tree base classifier [26].

The Weka [24] implementations of these algorithms were used. For all of the evaluated algorithms the parameters the values of which are not been defined have been set equal to the default ones.

IV. Experimental Results

In all experiments we followed a common experimental protocol as described in Section 3. Ten-fold cross validation experiments were performed on the audio data described in the previous section, thus resulting to non-overlapping training and test data subsets. The performance of the classification algorithms in frame level for various signal-to-noise ratios is shown in Table 1. The best performing algorithm for each of the evaluated signal-to-noise ratios is indicated in bold.

TABLE I. Bird Species Classification Accuracy (in percentages) at Various Signal-To-Noise Ratios

Method\ SNR	**-6dB**	**0dB**	**6dB**	**12dB**	**16dB**	**20dB**
AdaBoost(J48)	76.17	79.14	83.24	**88.14**	**90.95**	**93.02**
Bagging(J48)	**76.58**	**79.81**	**83.32**	87.64	90.18	92.13
BayesNet	69.88	70.67	74.09	77.54	79.37	80.61
IBk	70.98	75.17	80.20	87.06	89.53	90.02
J48	71.33	74.19	78.23	83.44	86.08	88.67
SMO	70.44	66.58	71.39	75.91	77.68	79.49

As can be seen in Table 1, the best performance was achieved by the two meta-classifiers used, which in agreement with [24], where it was reported that they often dramatically improve the classification performance. In particular, for high signal-to-noise ratios the boosting algorithm outperformed all the rest classification algorithms, while for low signal-to-noise ratios the bagging meta-classifier offered slightly better performance than the boosting algorithm. The meta-classifiers are followed in average performance by the rest evaluated classifiers, i.e. the k-nearest neighbor algorithm (*IBk*), the C4.5 pruned decision tree (*J48*), the Bayesian network (*BayesNet*) and the support vector machines (*SMO*).

For all the classifiers that were evaluated here there is a drop in the classification accuracy with the decrease of the signal-to-noise ratio, which is in agreement with the experimental results found in [17] for waterfall noise. It is worth mentioning that the k-nearest neighbor algorithm (*IBk*) and the decision tree algorithm (*J48*) achieved relatively high performance at noise-free conditions, i.e. for signal-to-noise ratio equal to 20 dB, approximately 3% less than the best performing meta-classifier. However, in noisy environments, such as SNR equal to 0 dB and -6 dB, they followed the best performing meta-classifier by approximately 5%. This is an indication of the advantage that the bagging and boosting algorithms can offer in real-field environments, where the presence of non-stationary interfering noises is frequent. Moreover, low signal-to-noise ratio issues are met in audio acquisition of bird vocalizations that are not close to the microphones installed in the field.

In a second step, we applied a post-processing smoothing window filter to the recognized labels of each frame. This post-processing aims at eliminating sporadic erroneous labeling of the current audio frame, e.g. due to momentary burst of interference, and thus contributes for improving the overall classification accuracy. A simple and computationally effective rule for post-processing is smoothing each decision (or score) with respect to its closest neighbors. In particular, when the N preceding and the N successive audio frames are classified to one bird species vocalization then the current frame is also (re)labeled as of this bird species. The length w of the smoothing window is subject to investigation and in the general case is equal to $w=2N+1$, where $N\geq 0$. The case $N=0$ corresponds to eliminating the post-processing of the classified labels, while the cases $N=1, 2, 3$ correspond to window size $w=3, 5, 7$. The effect of the smoothing window in the bird species classification performance for the best performing at low signal-to-noise ratios Bagging(J48) algorithm and at various SNRs is shown in Table 2. The best performing post-processing smoothing window length for each evaluated signal-to-noise ratios (SNRs) is indicated in bold.

TABLE II. Bird Species Classification Accuracy (in percentages) of the Bagging (J48) Algorithm at Various Signal-To-Noise Ratios for Different Length of the Smoothing Window

w\ SNR	**-6dB**	**0dB**	**6dB**	**12dB**	**16dB**	**20dB**
w=1 (N=0)	76.58	79.81	83.32	87.64	90.18	92.13
w=3 (N=1)	**79.83**	**82.23**	**84.81**	**88.72**	91.76	92.70
w=5 (N=2)	78.82	80.92	84.10	88.20	**91.94**	**92.89**
w=7 (N=3)	76.96	79.82	83.15	87.76	90.16	92.17

As can be seen in Table 2, the effect of the post-processing smoothing window is significant for all signal-to-noise ratios and especially in the case of noisy environment, i.e. for low signal-to-noise ratios. In detail, the window length equal to three offers the best or close to the best performance across all the evaluated signal-to-noise ratios. The application of this window length ($w=3$) achieved approximately 3.5% absolute improvement of the bird species classification accuracy at -6 dB of signal-to-noise ratio. The improvement of the classification accuracy for SNR equal to 20 dB is approximately 1% in terms of absolute recognition accuracy.

The evaluated performance indicates the importance of the smoothing window post-processing in the noisy real-field environment.

V. Conclusion

Evaluation of six classification techniques, employed in a widely-used scheme for audio processing, was performed on the bird classification task. In our setup this evaluation involves identification of seven species of birds, which are commonly found at the Hymettus Mountain. The highest recognition accuracy was achieved by a bagging and a boosting meta-classification algorithm, which used the pruned C4.5 decision tree as base classifier.

Experiments with additive noise, for several sound-to-noise values, demonstrated the robustness of the bird species recognizer in noisy conditions, such as the ones found in the real field. Furthermore, the use of post-processing on decision level per frame proved to offer significant improvement for low sound-to-noise ratios, which are the case for the noisy real-field conditions.

Acknowledgment

The research reported in the present paper was supported by the AmiBio project (Automatic Acoustic Monitoring and Inventorying of Biodiversity -- LIFE08 NAT/GR/000539), which is implemented with the contribution of the LIFE+ financial instrument of the European Union. Project web-site: www.amibio-project.eu.

The authors wish to acknowledge the contribution of Dr. Stavros Ntalampiras, and Dr. Theodoros Kostoulas from the University of Patras and also the entire team of the Association for Protection and Development of Hymettus (SPAY), who supported the implementation of the data collection campaign at the Hymettus Mountain.

References

[1] D.K. Dawson, M.G. Efford, “Bird population density estimated from acoustic signals,” Journal of Applied Ecology, vol. 46, 2009, pp. 1201–1209.

[2] S.E. Anderson, A.S. Dave, and D. Margoliash, “Template-based automatic recognition of birdsong syllables from continuous recordings,” J. Acoust. Soc. America, 100(2), pp. 1209-1219, Au-gust 1996.

[3] K. Ito, K. Mori, and S. Iwasaki, “Application of dynamic programming matching to classification of budgerigar contact calls,” J. Acoust. Soc. America, 100(6), pp. 3947-3956, December 1996.

[4] J.A. Kogan, and D. Margoliash, “Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden Markov models: A comparative study,” Journal of Acoust. Soc. America, 103(4), pp. 2185-2196, April 1998.

[5] V.M.. Trifa, A.N.G. Kirschel, and C.E. Taylor, “Automated species recognition of antbirds in a Mexican rainforest using hidden Markov models,” J. Acoust. Soc. America, 123(4), pp. 2424-2431, April 2008.

[6] P. Somervuo, A. Harma, and S. Fagerlund, “Parametric repre-sentations of bird sounds for automatic species recognition,” IEEE Transactions on Audio, Speech and Language Processing, vol. 14, no. 6, pp. 2252–2263, 2006.

[7] I. Agranat, “Automatically Identifying Animal Species from their Vocalizations,” Wildlife Acoustics, Inc., Concord, Massachusetts, March 2009.

[8] S.A. Selouani, M. Kardouchi, E. Hervet, and D. Roy, “Automatic birdsong recognition based on autoregressive timedelay neural networks,” In Congress on Computational Intelligence Methods and Applications, pp. 1–6, Istanbul, Turkey, 2005.

[9] A. L. Mcllraith and H. C. Card, “Birdsong recognition using backpropagation and multivariate statistics,” IEEE Transactions on Signal Processing, vol. 45, no. 11, pp. 2740–2748, 1997.

[10] C. Kwan, K.C. Ho, G. Mei, et al., “An automated acoustic system to monitor and classify birds,” EURASIP J. on Applied Signal Processing, vol. 2006, Article ID 96706, 19 pages, 2006.

[11] H. Tyagi, R.M. Hegde, H.A. Murthy, and A. Prabhakar, “Automatic identification of bird calls using spectral ensemble average voiceprints,” In Proc. of the 13th European Signal Processing Conference, Florence, Italy, September 2006.

[12] S. Fagerlund, “Bird Species Recognition Using Support Vector Machines,” EURASIP Journal on Advances in Signal Processing, Vol. 2007, Article ID 38637, 8 pages, doi:10.1155/2007/38637.

[13] E. Vilches, I.A. Escobar, E.E. Vallejo, and C.E. Taylor, “Data mining applied to acoustic bird species recognition,” In Proceedings of the 18th International Conference on Pattern Recognition, vol. 3, pp. 400–403, Hong Kong, August 2006.

[14] A. Harma, “Automatic identification of bird species based on sinusoidalmodelling of syllables,” In Proceedings of IEEE Interna-tional Conference on Acoustics, Speech, and Signal Processing, vol. 5, pp. 545–548, Hong Kong, April 2003.

[15] P. Somervuo and A. Harma, “Bird song recognition based on syllable pair histograms,” In Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 5, pp. 825–828, Montreal, Canada, May 2004.

[16] C.H. Lee, Y.K. Lee, and R.Z. Huang, “Automatic Recognition of Bird Songs Using Cepstral Coefficients,” Journal of Information Technology and Applications, vol. 1, no. 1, pp. 17-23, May 2006.

[17] P. Jancovic and M. Kokuer, “Automatic Detection and Recognition of Tonal Bird Sounds in Noisy Environments,” EURASIP Journal on Advances in Signal Processing, Vol. 2011, Article ID 982936, 10 pages, doi: 10.1155/2011/982936.

[18] S. Young, G. Evermann, M. Gales, T. Hain, D. Kershaw, X. Liu, G. Moore, J. Odell, D. Ollason, D. Povey, V. Valtchev, and P. Woodland, The HTK book (for HTK Version 3.4), Cambridge University Engineering Department.

[19] F. Eyben, M. Wollmer, and B. Schuller, “OpenEAR - intro-ducing the Munich open-source emotion and affect recognition toolkit,” In Proc. of the 4th International HUMAINE Association Conference on Affective Computing and Intelligent Interaction (ACII 2009).

[20] D. Aha, D. Kibler, "Instance-based learning algorithms", Machine Learning, 6 (1991), pp. 37–66.

[21] T.M. Mitchell, Machine Learning, McGraw-Hill International Editions (1997).

[22] S.S. KeerthiS.S., S.K. Shevade, C. Bhattacharyya, K.R.K. Murthy, "Improvements to Platt's SMO algorithm for SVM classifier design, Neural Computation, 13 (3) (2001), pp. 637–649.

[23] R. Quinlan, C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers, San Mateo, CA (1993).

[24] H.I. Witten, and E. Frank. Data Mining: practical machine learning tools and techniques. Morgan Kaufmann Publishing.

[25] Yoav Freund, Robert E. Schapire: Experiments with a new boosting algorithm. In: Thirteenth International Conference on Machine Learning, San Francisco, 148-156, 1996.

[26] Leo Breiman (1996). Bagging predictors. Machine Learning. 24(2):123-140.