International Journal of Application or Innovation in Engineering and Management

ijaiem.com/March 2021/ Volume 10/Issue 1/Article No-1/40-50
ISSN: 2319-4847

Acoustic Bird Species Recognition using Data Mining and Hidden Markov Models

P.Anupama ,Orissa India, P.Anupama **@gmail.com**

Article Info

Received: 07-12-2020 ***Revised:11 -01-2021*** ***Accepted: 16-02-2021*** ***Published:13/03/2021***

Abstract

This research presents a solution to the challenge of auditory bird species identification via the merging of Data Mining and Hidden Markov Models. Initially, we demonstrate their separate applications, compare and contrast their outcomes, and then develop a model to integrate them for specific categorization tasks. Distributed sensor networks have increased computational demands due to the huge collections of spectral properties required to capture the structure of bird songs, as shown in previous work. In order to classify and minimize the dimensionality of spectral properties, data mining is used. The conventional hidden Markov models need extensive preprocessing of the music. Data mining, as we have shown, may efficiently target HMMs' input parameters and provide low-requirement outputs.

Introduction

The monitoring of animal behavior and diversity over a variety of spatial and temporal scales poses many chal- lenges to human observers. A significant amount of knowl- edge on bird diversity and behavior is the result of field observations made by expert ornithologists. Bird species identification and the study of their interactions have been developed by means of the visual and acoustic abilities of these experts. On the other hand, the identification of indi- vidual birds often requires the usage of visual aids, such as color banding.

For this work, the goal of sensor networks [13, 17, 18] is to introduce in a natural environment a certain number of small sensors or motes in order to acquire data from their surroundings. The technology of distributed senor networks also allows us to detect unusual events that oc- cur in a certain environment, such as the presence of rare or endangered bird species, their communications and social interactions without human intervention.To be able to explore the full potential of distributed sensor networks we must deal with complex processing challenges. Bird songs, when converted to the frequency domain by means of Fourier transforms, produce large amounts of data that requires processing and discrimination. Implicit relationships must be found; data must be sorted into logical groups and cleaned. Here is where data mining techniques will alleviate the computational needs for these analyses. This will allow us to incorporate these processing technologies into existing energy and processor constrained platforms, such as nodes in a sensor network.

Previous works have used canonical discriminant analy- sis [14] to demonstrate that invariant features don’t actually provide the most important recognition cues, contradicting some common assumptions in the published literature. Hid- den Markov Models have also been applied to bird song recognition due to their success in speech recognition, even though they are not appropriate for constrained platforms such as those in a distributed sensor network.

We propose to use data mining techniques to reduce the computational complexity required to classify different bird species by means of attribute reduction and to show that these techniques work better when combined with classical approaches such as Hidden Markov Models.

international journal of application or innovation in engineering and management

ijaiem.com/March 2021/ Volume 10/Issue 1/Article No-1/40-50
ISSN: 2319-4847

1. Data Acquisition and Procesing

1.1. Bird Songs

Bird songs and calls for this study were obtained from two different sources. The first was through the Cornell Lab of Ornithology, Macaulay Library [1]. We gathered samples from their collection and did initial testing. The second set of bird songs were collected from our February 2006 trip to the Reserva de la Biosfera de Montes Azules in Chiapas Mexico. We used the later songs to validate some of our results. We focused on songs from three species of antbirds: great antshrike, *Taraba major*(49 song files); dusky antbird, *Cercomacra tyrannina* (79 song files); and barred antshrike, *Thamnophilus doliatus* (76 song files). Each song file has from a few seconds to several minutes of bird calls, with either one, two or more birds singing on it.

Bird songs are normally more musical and complex than calls. Males usually produce them and they are associated with breeding. "Calls tend to be shorter, simpler and pro- duced by both sexes throughout the year. Unlike songs, calls are less spontaneous and usually occur in particular contexts" [4]. Birds use calls to communicate things to each other and between members of a flock or family.For this work, we are going to use both songs and calls to extract the most relevant features of the bird's song. We decided to include both sources, since some of their song building blocks have common elements in both songs and calls. For the remainder of this document, we will mention calls and songs as being of equal importance and will be equally used as the basis of bird song data.

The reason to choose these species is that they are abun- dant in Montes Azules, Chiapas, the ecological reserve where the sensor network will be deployed in the near fu- ture. Additionally these tropical bird species do not learn songs, which makes the job of acoustic recognition easier. Finally, the three species share common building blocks in their song organization. This makes the challenge of classi- fication very interesting, since if we compare their spectro- grams in a very close up level, we notice that some of their song components are very similar.

1.2. Sound Cleaning

Table 1. Low-pass and High-pass filters per species

Filter	Taraba major	Cercomacra tyrannina	Thamnophilus dolitus
Low-pass	3597 Hz	4200 Hz	3597 Hz
High-pass	517 Hz	920 Hz	686 Hz

Field recordings can be extremely noisy, especially when performed in tropical rainforests. In these types of forests, the vegetation is densely packed causing sound reverbera- tions; there are many different bird species interacting and a huge amount of other animals producing harsh noises. The climate is also a crucial factor, rain can cause signifi- cant interference and wind causes leaves to fall and interfere through most of the acoustic frequencies. All these factors limit the quality of the sound recordings making automated

international journal of application or innovation in engineering and management

ijaiem.com/March 2021/ Volume 10/Issue 1/Article No-1/40-50
ISSN: 2319-4847

bird species recognition a more complicated process and re- quiring the introduction of different filtering techniques in order to obtain suitable results.

The songs were only preprocessed through low and high pass software filters to facilitate an accurate call and pulse recognition. These filters are species dependant as we can see in Table 1.

2. Methods

2.1. Feature Extraction

Once we finished the cleaning process, we decided to use .wav format files, since they had better resolution and were easily manipulated by computers. For this purpose, used the computer software, Sound Ruler [9]. With this software, we are able to see the oscillogram and spectrogram of the signal and within the oscillogram we are able to locate each call from the recording and each pulse within a call.

The pulse-by-pulse analytical results that Sound Ruler threw were saved as comma delimited files. These files contain the 71 attributes of each pulse from the processed samples, representing the bird's song data. The resulting datasets' size is as follows: *Taraba Major* 21,360 pulse samples, *Cercomacra Tyrannina* 5373 pulse samples, and *Thamnophilus Doliatus* 911 pulse samples.

2.2. Hidden Markov Models Theory

The basic HMM theory was published in a series of clas- sic papers by Baum and his colleagues [5]. The HMM has become one of the most powerful statistical methods for modeling speech signals. Its principles have been suc- cessfully used in automatic speech recognition, formant and pitch tracking, speech enhancement, speech synthe- sis, statistical language modeling, part-of-speech tagging, spoken language understanding, and machine translation [3, 5, 6, 7, 8, 10, 11, 12, 16].

2.3. Data Mining

When working with bird songs, we unfortunately need to deal with information that is represented as raw data. This information may contain valuable records that may be hid- den from the naked eye. We have to apply different compu- tational tools in order to extract the information we require from the raw data. The approach we took was to apply dif- ferent data mining techniques in order to obtain the most relevant information from the raw data. Once the important data is extracted, we can use only the significant informa- tion to feed our classifying algorithms in the sensor nodes in order to recognize different bird species based on their song and call production.

International Journal of Application or Innovation in Engineering and Management

ijaiem.com/March 2021/ Volume 10/Issue 1/Article No-1/40-50
ISSN: 2319-4847

2.3.1 Vector Quantization

This algorithm was implemented because of the ID3's and association rules lack of numeric support. Quantization [15] is a process in which numeric to nominal data conversion is possible. The algorithm takes an original numeric vector and returns a quantized equivalent numeric vector, which can be easily represented by nominal values.

Mainly, this process consists of the making of ranges of numbers, and assigning each range a numeric value that rep- resents the whole range: Lets say: 1 to 1000 = 1, 1000 to 2000 = 2, etc., and then interpreting the assignments as labels, instead of numbers. For the specific details, check [15].

In figure 1 we present a plot comparison from a full orig- inal signal with values from 0 to 5000 approximately versus a quantized signal with values from 0 to 6. We can clearly appreciate how the relationship among the attribute values is preserved in the quantized set, even though we can appre- ciate some information loss.

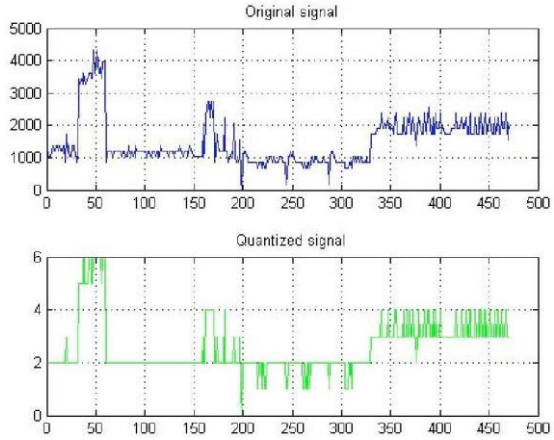

Figure 1. Original vs. quantized signal

2.3.2 ID3

Once we converted the entire species data sets into quan- tized data, we proceeded to process the information with a decision tree algorithm. The ID3 algorithm was used to generate the decision tree with Weka [21] software. ID3 is a decision tree algorithm that takes all unused attributes and counts their entropy concerning the test samples to be used. When the tree is completed, the resulting nodes will be the most significant attributes used to classify the different instances of bird species (the leaves of the tree). Once we obtained the corresponding decision tree, we only preserved in our data set the attributes that were used in the nodes of

international journal of application or innovation in engineering and management

ijaiem.com/March 2021/ Volume 10/Issue 1/Article No-1/40-50
ISSN: 2319-4847

the tree (an attribute can be repeated in many nodes). This reduced data set will be used to attempt a reliable classifi- cation with the Nave-Bayes algorithm.

2.3.3 J4.8

This algorithm is an extension of the ID3 algorithm, which solves some deficiencies that the original ID3 algorithm had. Some of the improvements are that J4.8 avoids over- fitting, uses a reduced-error pruning focus that is based on the consideration that each node of the tree is a prune can- didate, reducing this way the error. Also it performs a rule post-pruning to find high precision hypothesis and numeric attribute handling. The two main advantages that made us select this algorithm are the computational cost savings and the numeric attribute handling. Weka was used to test this algorithm with our original data sets. The surviving at- tributes in the reduced data sets were also used to attempt a reliable classification with the Nave-Bayes algorithm.

2.3.4 Nave-Bayes

We decided to introduce the Nave-Bayes algorithm usage as a final classifier because of the main disadvantages that decision tree algorithms have. One of them is that they are unstable. In effect, slight variations in the training data can result in different attribute selections at each choice point within the tree. The effect can be significant since attribute choices affect all descendent sub-trees. Another important disadvantage with decision trees is that trees created from numeric data sets can be quite complex since attribute splits for numeric data are binary. Nave-Bayes was executed in Weka, for the original, post-ID3 and post-J4.8 datasets.

In this work, we attempt to eliminate redundancy or de- pendency in data by means of decision trees (ID3 and J4.8). We use only its surviving attributes to construct the data set that will be fed into Nave-Bayes. This will assert that we are working only with independent attributes and thus as- suring that the learning process is being skewed as less as possible by redundancy and that the maximum efficiency is being obtained.

2.3.5 Association Rules

For our experiments we took into consideration the associ- ation rules that had a 100% confidence and high instance support. An important factor to consider is that different association rules express different regularities that underlie the dataset, and they generally predict different aspects of the data [21]. This is why we need to analyze and interpret rules separately.

It is important to point out that even though high confi- dence values might suggest strong association rules, there

international journal of application or innovation in engineering and management

ijaiem.com/March 2021/ Volume 10/Issue 1/Article No-1/40-50
ISSN: 2319-4847

might be some deception on these results, since the an- tecedent or consequent might have high support values giv- ing for this matter high confidence levels even if they are independent. We recall that our goal is to find relationships between the attributes that compose a bird's song.

To obtain information about the Association Rules' con- cepts used in this work (confidence, support, lift, conviction and leverage), please consult [21].

3. Results

3.1. Hidden Markov Models

Hidden Markov Models were explored in this analy- sis to contrast their accuracy on bird species' recognition against the data mining recognition approach. For this ob- jective HTK [22] was used as a HMM testing implemen- tation. HMMs [20] have been selected because they rep- resent the traditional approach used nowadays for human speech recognition. For this purpose, whole songs (dura- tion approx. 2-3 sec.) have been manually cut from audio recordings made in the natural reserve of Montes Azules in Chiapas, Mexico. Trifa, [19] a colleague and member of our laboratory, used 25 samples for each of the species of interest and different amounts of background noise. Songs have been selected with high variability's. This choice was motivated by the fact that the goal was to minimize the pos- sibility that the HMMs would model properties of the noise instead of just the bird song itself. From the 25 samples of each species, 15 samples have been used to train the HMMs and 10 samples to test the recognition performance. These samples were selected randomly from the ones available.

The performance metric used to measure the efficiency of the HMMs is the one proposed in the HTK Book [22]. For each species or individual i tested, performance p is measured simply as the ratio of correctly recognized sam- ples C_i over the total amount of samples N_i used for testing, and is defined as:

$$p_i = \frac{Ci}{N_i} \times 100$$

For each experiment, 50 runs were performed using ran- dom partitioning of the files. They were also randomly split into training and testing sets. After several runs, a win- dow size of 25ms was selected and an overlapping range for these windows of 15ms. These values were chosen because they obtained better performance results.

As it was expected, the performance of HMMs is slightly superior when the frequency range of the bird's song is bounded according to each species spectral parameters. This happens because the noise located outside this fre- quency is not taken into consideration by the HMMs.

international journal of application or innovation in engineering and management

ijaiem.com/March 2021/ Volume 10/Issue 1/Article No-1/40-50
ISSN: 2319-4847

With these settings, it was found that the average HMM recognition performance for these species is about 93%. Several aspects cause this 7% lack of accuracy, among them it was identified that sometimes, birds do not sing their songs completely, they sing just a part of them. Also it was noted that ambient noise is one of the main error sources.

To find the optimal number of states required to model the species was a complex task. It was difficult, if not im- possible, to relate this parameter with any physical property of bird songs. A value of 5 states was selected, since it was noted to work better on average for all tests.

Performance did not change in a significant way when using between 5 and 15 states, but adding more than 15 states actually degraded the performance. We realized that using the standard feature extraction proposed by HTK is a very generic procedure and might not reflect the features relevant for bird species' classification [19].

3.2. Data Mining

In figure 2 we can observe that the most accurate algo- rithm is J4.8 (without Nave-Bayes) obtaining a 98.39% of accuracy. The original attribute number was 71 which this algorithm reduced to 47. We can also appreciate that regard- ing Nave-Bayes, the reduced data sets produce a slightly better performance, up to 4.5% improvement.

Besides the reliable accuracy preservation, the required processing power is also directly affected from the attribute reduction, since the number of calculations needed to clas- sify in a smaller data set is lower and so is the power con- sumption.

In the J4.8 tree, the main attribute was pulse dominant frequency, the root of the tree. In the next level we find the width of the dominant frequency peak at half of its height divided by the frequency of the peak. One more level down, we find the maximum of dominant frequency in the pulse, the total number of pulses in the call and the dominant fre- quency at final 50% peak amplitude. These five attributes which J4.8 identified as the main ones, contrast with the song duration, number of phrases and number of notes iden- tified by Nelson [14] and also the speed, duration, frequency range, and center frequency identified by Bard [2]. The rea- sons of these disagreements probably are the use of songs from different bird species and different algorithms for at- tribute selection, such as canonical discriminant analysis.

3.2.1 Association Rules

Our tests were performed considering only results valid for instances with a 100% of confidence and a high degree for support. We knew that there was a high degree of associ- ation between some attributes of our instances. From our results, we can confirm that there is a high degree of associ- ation among attributes.

International Journal of Application or Innovation in Engineering and Management

ijaiem.com/March 2021/ Volume 10/Issue 1/Article No-1/40-50
ISSN: 2319-4847

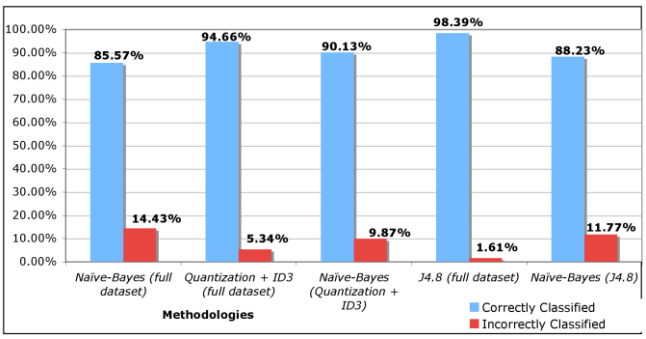

Figure 2. Data Mining accuracy percentages plot

Rule A: If the duration of a pulse on the initial axis (0%) of a call is high, the time in which half of the frequency modulation is low, and the second harmonic's relative am- plitude is low, then the duration of a pulse during the first 10th% of the call is high.

Rule B: If the duration of a pulse on the initial axis (0%) of a call is high then the pulse duration during the first 10th% of the call is high, the time in which half of the fre- quency modulation is acquired in the pulse is low and the relative amplitude of the second harmonic is low.

Rule C: If the duration of a pulse on the initial axis (0%) of a call is high, the energy during the 90th% of the initial call is high and the relative amplitude of the second har- monic is low, then the duration of the pulse during the first 10th% of the call is high and the time in which half of the frequency modulation is acquired in the pulse is low.

We can observe that there is a mild degree of association from attributes from the consequent and the antecedent, re- flected by a low leverage value. For example in rules A and B only 16% of the total instances (21,360) covered by the antecedent and consequent of those rules are totally inde- pendent. We can also observe lift values above one, indicat- ing that the rules generated are valid.

We know that leverage and lift measure similar charac- teristics in a rule but leverage measures more precisely the level of co-occurrence of X in Y as independent probabili- ties. In other words, leverage measures the proportion of the cases covered by X and Y over the expected cases of X and Y if they were considered independent from each other. We can also observe that there is a high association among the elements on either side of the rule, noted by the conviction values.

Interesting relationships have been found among the at- tributes observed such as that a high pulse duration at the edge of a call can imply a high pulse time until the peak of the same call, helping species classification. tributes that have a high dependency on each other, which can lead us to do further Nave Bayes testing with a dataset, in which these attributes have been removed.

4. Conclusions and Future Work

The increase of performance obtained through the com- bination of decision tree algorithms and Nave Bayes is due to the elimination of redundant information performed by these algorithms, although an 88.23% or 90.13% are still not enough for reliable classification. Further work is re- quired and a combination of association rules with these al- gorithms might prove valuable.

We can mention some advantages of data mining over HMMs, which might be crucial in order to obtain high pre- cision reliable results. The first of these advantages is that Hidden Markov Models

international journal of application or innovation in engineering and management

ijaiem.com/March 2021/ Volume 10/Issue 1/Article No-1/40-50
ISSN: 2319-4847

required us to partition the sound files and to place bird songs close together in order to get efficient results. Different audio files had to be chopped down and cut in order to compose a multi call file for the HMMs. Although this is not an extremely complex proce- dure it does limit the performance of applying these models on live sensor networks, since the general goal for a near future is to do analysis in real time. These HMMs were also extremely sensitive to ambient noise in the recordings during the training phase.

On the other hand, the data mining approach suffered very little with these inconveniences. Songs were processed in full without cutting or placing calls together. Only low and high pass software filters were applied to the original recordings when used with the data mining algorithms. In this way we removed some of the ambient noise, which left little margin for errors. This is because we were looking for certain attributes from each species, which repeated con- stantly among the recordings. We also noticed that the com- putational power to use very large data sets when working with data mining was very low in contrast with the one re- quired by the HMM approach.

As an integration of the two methods, we propose based on our current experiments, to use feature extraction to tar- get HMMs input parameters in order to optimize the recog- nition process. We propose for future work, to vary the in- put parameters given to HTK in order to train the HMMs with different data models. These models will be obtained through Data Mining. As a final approach, the recognition results of the HMMs will be compared amongst each other in order to see how efficient is the targeting selection per- formed by the Data Mining on their input parameters. The proposed model is as follows:

1. Run HTK.
2. Register the results in a small database.
3. A *C* program will perform the feature extraction from the sound signal through the FFTs of the signal instead of using the Perceptual Linear Prediction and MFCC used normally by HTK [22].
4. The *C* program will get the parameters extracted in the previous step and will process them by means of gen- erating decision trees. With the decision trees we will be able to prune down the attributes from the extracted base. The output of this step will include the most sig- nificant attributes used for classification.
5. Finally the program will convert the output from the previous step into a binary file to be used with HTK with an optional modification.

Finally we plan to take this work into the field and test our algorithms using adapted beamforming microphones with sensor networks and performing live monitoring and classification, expecting to see if our results hold, consider- ing ambient noise and tropical weather interference.

References:

[1] Macaulay Library, Cornell Lab of Ornithology. URL: http://birds.cornell.edu/MacaulayLibrary/index.html.
Those authors are S. C. Bard, M. Hau, M. Wikelski, and J. C. Wingfield. Spotted antbirds are a suboscine species native to the Neotropics, and they have vocal uniqueness and a reaction to music played by other species. Volume 104, pages 387–394, 2002, edited by T. C. O. Society, published in The Condor.
[3] Campbell, J. P. A guide to speaker recognition technologies. Publication date: 1997, volume 85, pages 1437–1462, edited by the IEEE.
C. K. Catchpole and P. J. Slater were cited (4). Themes and Variations in Bird Song, a Biological Perspective. [April 1995]: Cambridge University Press.
It was written by E. Chang, J. Zhou, S. Di, C. Huang, and K.-F. Lee. Mandarin voice recognition with a large

ijaiem.com/March 2021/ Volume 10/Issue 1/Article No-1/40-50
ISSN: 2319-4847

vocabulary using several tonal modeling approaches. Publication date: October 2000 in Beijing, China, in the proceedings of the International Conference on Spoken Language Processing, 2000 (ICSLP-2000), pp 983-986.
[6] M. Eichner and M. Wolff [authors]. The German Verbmobil Project's data-driven production of pronunciation dictionaries: a review of the experiments. Volume 3, pages 1687-1690, Istanbul, Turkey, June 2000, in Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP '00), editorially edited by IEEE.
Handbook of Screen Format Design, by W. O. Galitz [7]. Information Sciences (Q.E.D.), 1981.
J. Gao, H.-F. Wang, M. Li, and K.-F. Lee were the authors of the piece [8]. Making statistical language modeling for Chinese a cohesive effort. Volume 3, pages 1703–1706, Istanbul, Turkey, June 2000, in Acoustics, Speech, and Signal Processing, 2000 (ICASSP '00), edited by IEEE.
Mr. Gridi. Visit http://soundruler.sourceforge.net/ to use the sound ruler.
Researchers X. Huang, A. Acero, and H.-W. Hon. A Handbook of Theory, Algorithms, and System Development for Spoken Language Processing. Spring 2001, Prentice Hall PTR, first edition.
Citation: [11] X. Huang et al. In the next generation of personal digital assistants, there is Mipad. In volume 3, pages 33–36, Beijing, China, October 2000, International Conference on Spoken Language Processing (ICSLP-2000).
K. Laurila and **P. Haavisto. How practical is name dialing? Volume 6, pages 3731–3734, Istanbul, Turkey, June 2000, in Acoustics, Speech, and Signal Processing, 2000 (ICASSP '00), edited by IEEE.
The authors of this work are Mainwaring, Mainwaring, Polastre, Szewczyk, Culler, and Anderson [13]. Monitoring habitats with wireless sensor networks. Publication date: September 2002 in Atlanta, Georgia, USA, in the proceedings of the International Symposium on Wireless Sensor Networks and Applications (WSNA'02), edited by ACM.
This is D. A. Nelson. The significance of both unique and constant characteristics for identifying bird species via their songs. Volume 91, pages 120–130, 1989, edited by T. C. O. Society. The Condor.
S. Russell and P. Norvig [15] written the following. Embracing AI: A Contemporary Perspective. Publishing House: Prentice Hall, Englewood Cliffs, New Jersey, 2002, 2nd anniversary edition.
Yukio Sagisaka and Lee Lee. Speech recognition of asian lan- guages. Snowbird, Utah, December 1995, pages 55–57, in Automatic Speech Recognition and Understanding, 1995 (ASRU '95), edited by IEEE.
This sentence is a citation of work by R. Szewczyk, E. Osterweil, J. Polastre, M. Hamilton, A. Mainwaring, and D. Estrin. Habitat monitoring using sen- sor networks. Publication date: June 2004, volume 47, pages 34–40, edited by ACM.
P. Hamilton, C. Taylor, and E. Stabler [18]. Arrays of sensors for the aural tracking of avian variety and behavior. from 2004 to 2009, see http://www.nsf.gov/awardsearch/showAward.do? Award- Number=0410438.
[19] Trifa, of V. A structure for detecting, recognizing, and localizing bird songs using acoustic sensor networks. In 2006, I completed my master's thesis at the E'cole Polytechnique Fedérale de Lausanne.
In their work, M. Wilde and V. Menon provide [20]. Hiding Markov Models for Bird Call Recognition... Tulane University's EECS Department produced a technical report in 2003.
By I. H. Witten and E. Frank, [21] we mean. Data Mining: Practical Machine Learning Tools and Techniques. Second edition published in June 2005 by Morgan Kaufmann in San Francisco, California.
[22] is aS. Young, together with others. A guide to using HTK with Version 3.4. The Department of Engineering at Cambridge University, December 2006.