51. IWK

Internationales Wissenschaftliches Kolloquium
International Scientific Colloquium

PROCEEDINGS

11-15 September 2006

FACULTY OF ELECTRICAL ENGINEERING AND INFORMATION SCIENCE

INFORMATION TECHNOLOGY AND ELECTRICAL ENGINEERING - DEVICES AND SYSTEMS, MATERIALS AND TECHNOLOGIES FOR THE FUTURE

Startseite / Index:
http://www.db-thueringen.de/servlets/DocumentServlet?id=12391

TECHNISCHE UNIVERSITÄT
ILMENAU

Impressum

Herausgeber: Der Rektor der Technischen Universität Ilmenau
Univ.-Prof. Dr. rer. nat. habil. Peter Scharff

Redaktion: Referat Marketing und Studentische Angelegenheiten
Andrea Schneider

Fakultät für Elektrotechnik und Informationstechnik
Susanne Jakob
Dipl.-Ing. Helge Drumm

Redaktionsschluss: 07. Juli 2006

Technische Realisierung (CD-Rom-Ausgabe):
Institut für Medientechnik an der TU Ilmenau
Dipl.-Ing. Christian Weigel
Dipl.-Ing. Marco Albrecht
Dipl.-Ing. Helge Drumm

Technische Realisierung (Online-Ausgabe):
Universitätsbibliothek Ilmenau
ilmedia
Postfach 10 05 65
98684 Ilmenau

Verlag: isle
Verlag ISLE, Betriebsstätte des ISLE e.V.
Werner-von-Siemens-Str. 16
98693 Ilrnenau

ISBN (Druckausgabe): 3-938843-15-2
ISBN (CD-Rom-Ausgabe): 3-938843-16-0

Startseite / Index:
http://www.db-thueringen.de/servlets/DocumentServlet?id=12391

51st Internationales Wissenschaftliches Kolloquium
Technische Universität Ilmenau
September 11 – 15, 2006

Stefan Dünkel, Kristina Kelber, Carlos Hernández Franco

Identification of Singing Birds Based on the Analysis of their Sounds

ABSTRACT

A system for recognition of bird species is presented based on syllables of bird songs, which are versatile among bird species and therefore suitable for recognition. The recognition system operates in several phases: pre-processing, segmentation of syllables, feature extraction, clustering, classification (design and evaluation). Time and frequency domain parameters are used as features. The classification is based on four different classifiers: minimum-distance, k-nearest-neighbor, naive Bayes and matched filter. The results of them are merged by a voting scheme. In this study the sounds of four common European songbirds are used. Experiments show that features related to the frequency band of the syllables produce best results.

1. INTRODUCTION

An automatic recognition of different bird species by their sounds provides a simplification of ornithological long-term observations and environmental monitoring applications. Furthermore, it allows a judgment of the state of the natural area since birds are indicators for the environment. Increasing possibilities of communication and information technologies can be used to control natural parks as well, e. g. guiding the visitors to current attractions in the park or keep them away from breeding places.

Therefore, a system for automatic recognition of bird songs is developed within the framework of a project dealing with the control of “El Racó de l'Olla”, a natural park of “L'Albufera” in Valencia, Spain. Recognition of bird songs is a typical problem of acoustic pattern recognition just like automatic speech recognition. However, the relative simplicity of bird songs compared to human speech can facilitate recognition of bird vocalizations.

Bird songs consist of a hierarchical assembly of discrete subunits: songs of a combination of different phrases, phrases of various similar syllables and syllables of one or more notes [1]. As an example this structure is shown in Fig. 1 by means of a part of a Greenfinch song.

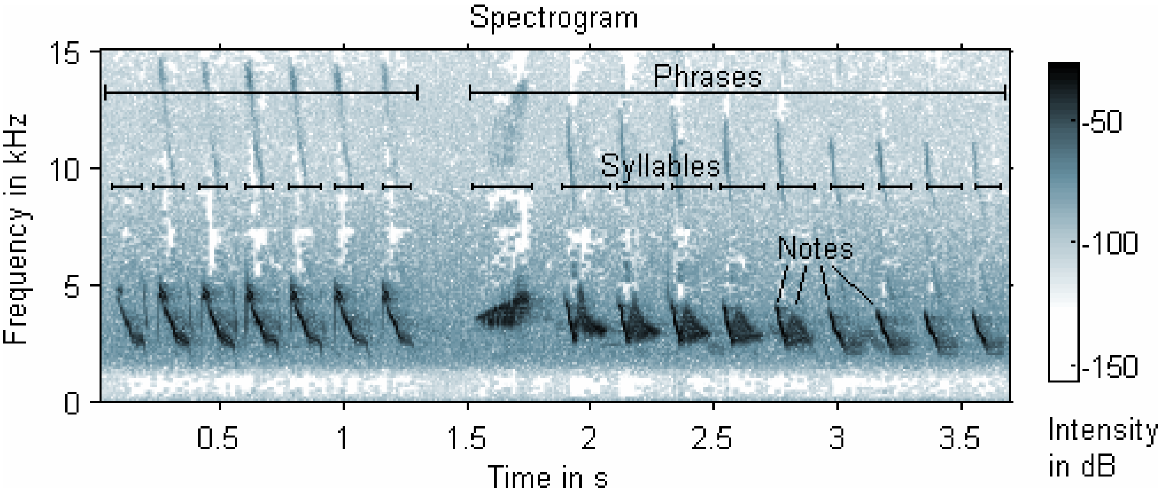

Fig. 1 Spectrogram of a part of a Greenfinch song composed of 17 syllables in 2 phrases. Subunits are marked.

For several reasons, syllables are used in this project as basic elements for recognition:

- easy to detect due to short pauses between them
- not sensitive against dialects
- no further knowledge about the huge variety of combinations to phrases and songs is required

The temporal variety of bird songs is typically orders of magnitude faster than in human sound production [2]. Therefore, for the analysis of bird songs a high temporal resolution in the range of a few milliseconds is necessary.

The problem of automatic identification of bird species is a deep and complex subject and is relatively unexplored. However, there exist some encouraging studies proving the feasibility of this task [3,4].

For this project an approach is required which is as simple as possible as the algorithm should be suitable for real-time identification by a mobile robot or a small embedded system placed in the park. Only the identification results should be transmitted by e. g. WLAN technologies. To solve this problem, a fairly straight forward method is used in this study. Based on the analysis of bird songs a recognition system is designed and evaluated. The basic principle and the architecture of the recognition system is illustrated in Fig. 2. The system operates in several phases: pre-processing, segmentation of syllables, feature extraction, clustering, classification (design and evaluation). In this paper an overview is presented. More detailed results can be found in [5].

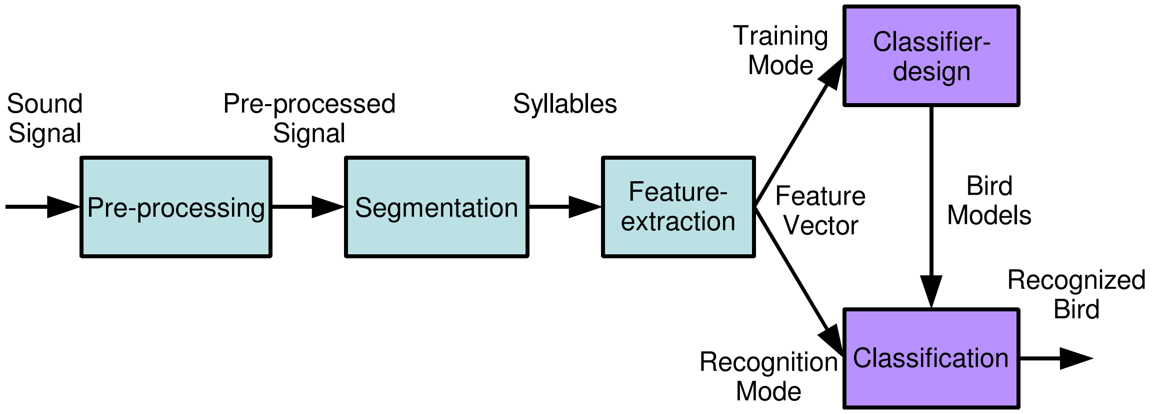

Fig. 2 Architecture of the implemented recognition system

2. DATABASE

For this project a database composed of recordings found on the internet (e. g. [6]) and of field recordings from the natural park "El Racó de l'Olla" was built. Each recording contains only one song of one individual bird. Recordings with overlapping songs from different species are tested separately. In order to get reliable results the database has to be as large as possible and it also has to contain songs of various birds of the same species (to be independent from a single individual). Due to the available data the consideration is restricted to four common European songbird species (Tab. 1).

Common name	**Latin name**	**Recordings**	**Syllables**
Blackbird	*Turdus merula*	24	421
Great Tit	*Parus major*	20	417
Greenfinch	*Carduelis chloris*	15	219
Cuckoo	*Cuculus canorus*	11	91

Tab. 1 Set of investigated birds in the study

The typical frequency range of a bird song is from 300 Hz to 10 kHz. So, a sampling frequency of 20 kHz would have been sufficient. Nevertheless, for all recordings a sampling frequency of 44.1 kHz and mono channel with 16 bit quantization are used.

3. PRE-PROCESSING AND SEGMENTATION

Signals from the database are pre-processed by filtering and normalization. According to the typical frequency range of the bird songs (300 Hz to 10 kHz), noise is attenuated by a band pass filter. Different magnitudes within a song (caused by different directions and distances of the birds to the microphone) are normalized to the maximum magnitude of the entire song.

Then, the entire recording is segmented into syllables based on the short-time energy of the signal. Syllables are defined as parts with a high energy content and pauses with a low. Decision is taken by a threshold in combination with a hysteresis. Subsequently, syllables shorter than 10 ms or longer than 700 ms are eliminated. Finally, a data set of syllables according to Tab. 1 is obtained.

4. FEATURE-EXTRACTION

For classification a feature vector for each syllable is required. It means, the highly redundant sound signal has to be described by a few distinctive features which contain all the important information for distinction between the different classes (bird species). Since bird vocalizations are dynamic and therefore non-stationary signals, it is not reasonable to analyze the complete signal at once. Instead, features are calculated for short sequences (frames) only and the information contained in their temporal change is later on used for classification too. For the analysis, an overlapping window (e. g. Rectangle and Hanning) is shifted over the complete syllable.

In this work, 17 features from time and frequency domain compose the feature space [5]. A few of them are established and approved features for audio classification [7]. In order to decide which of these features are most suitable for identification of the considered bird species, a Linear Discriminant Analysis (LDS, [8]) is employed. Using it, an objective quantitative value for the discriminating power of each feature is calculated (based on the available database).

It turns out that the most suitable ones are:

- the *duration* and the *zero crossing rate* of a syllable (derived from the time domain) as well as
- the *Spectral Centroid*

$$SC(i) = \frac{\sum_{n=0}^{K-1} n \cdot |X_i(n)|^2}{\sum_{n=0}^{K-1} |X_i(n)|^2} \qquad (1)$$

- and the *Spectral Rolloff* (both derived from the frequency domain)

$$\sum_{l=0}^{SR(i)} |X_i(l)|^2 < TH \sum_{n=0}^{K-1} |X_i(n)|^2 \qquad (2)$$

where $X_i(n)$ is the discrete Fourier transform (DFT) of the i-th frame of a syllable, the frequency bin n and the order of the DFT defined by K. The threshold TH is a value between 0 and 1 and the optimal value for this task is found by experiment at 0.8. For all features on frame basis the mean and the variance of all frames are used as features. As an example, Fig. 3 shows two dimensions of the feature space for the investigated birds.

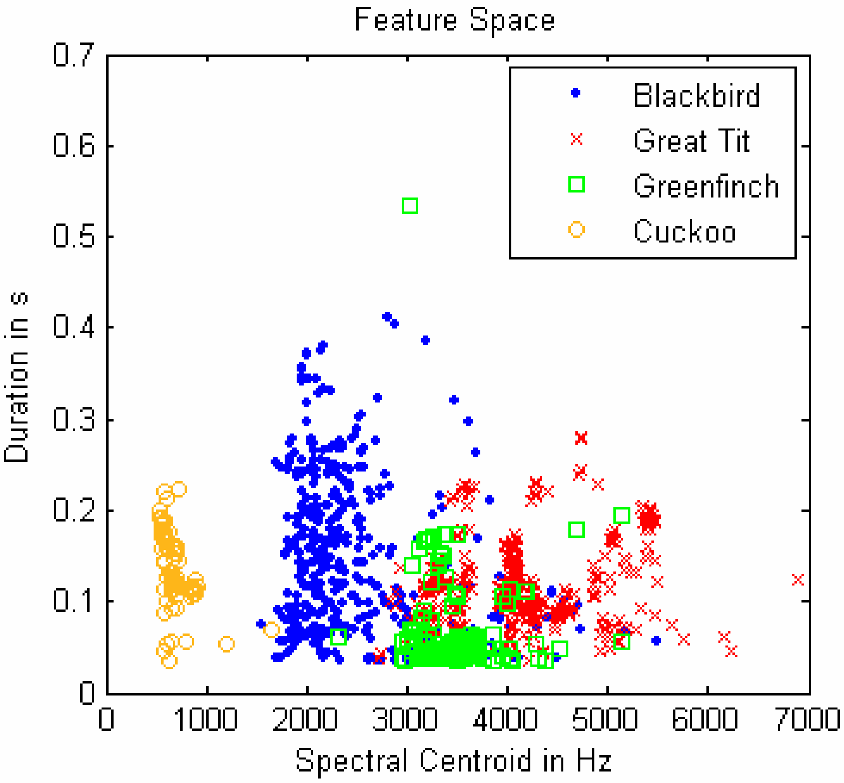

Fig. 3 2D feature space for the investigated birds

5. CLUSTERING

As depicted in Fig. 3, there exist different clusters in the feature space for each bird. Usually, each cluster belonging to a bird represents another type of syllable a bird is able to sing. In order to improve the recognition results each of these clusters is considered separately. It means the feature space is clustered into several regions (clusters) for each bird thus allowing a better classification (see Fig. 4).

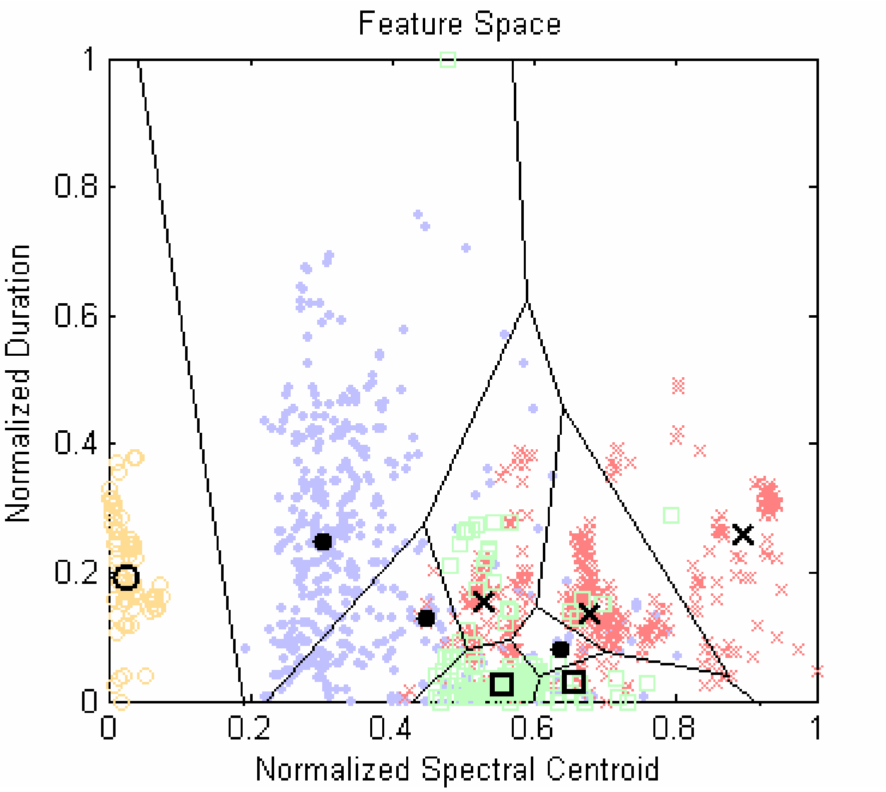

Fig. 4 2D feature space with cluster center points and boundaries

Clustering is one of the unsupervised learning methods. In this work an iterative bottom-up clustering algorithm [9] is applied to discover sub-classes (clusters) formed by agglomerations of feature vectors in the feature space.

Additionally a rejection class is introduced. It is necessary to avoid a false recognition of sounds not belonging to the trained classes (e. g. other birds or cars sounds) by allocation of the doubtful samples to this rejection class.

Of course there exists the possibility to wrongly reject a known bird which is called *false rejection*. The other extreme is the *false acceptance* when an unknown sound is classified as one of the trained birds. Both probabilities should be as low as possible, but it is not feasible to minimize them independently from each other. The optimal value for the rejection threshold (an intern classifier parameter) is determined by the equal error rate of both error probabilities.

For that reason, a set of songs from other birds is tested. This collection consists of various birds from different families and orders. For example the European Serin (*Serinus serinus*), the Starling (*Sturnus vulgaris*) and the House Sparrow (*Passer domesticus*) are classified. All together nine species with a total number of 205 syllables are tested. Some of them belong to the same family as the trained birds and are similar in vocalizations too.

6. CLASSIFICATION

The goal of the classification step is to associate a given unknown input pattern (syllable of a bird song) with a class (bird species). In this work four fairly simple classifiers are used*: minimum-distance, k-nearest-neighbor*, *naive Bayes* and *matched filter* [10]. For instance, application of the minimum-distance classifier to the data in Fig. 3 results in nine clusters which are depicted in Fig. 4 including corresponding cluster boundaries and centers. Recognition rates for all classifiers are summarized in Tab. 2. It shows that three of the classifiers are rather similar in their performance.

Minimum distance	**kNN**	**Naive Bayes**	**Matched-filter**
71.2 %	78.9 %	77.5 %	52.5 %

Tab. 2 Recognition rates for the different classifiers

For all investigations the leave-one-out cross validation is applied, which allows a reliable judgment of the recognition system in case of a sparse database.

As an example, the classification rate for the kNN classifier is summarized in Tab. 3. The poor recognition rate for Greenfinch can be explained with the few training data for this species. Fortunately, most of the false classifications are assigned to the rejection class (‘Unknown’). Blackbird has a rich vocabulary with similar syllables to many other species. Cuckoo and Great Tit use relative simple structured and unique syllables and are therefore easy to identify.

	Unknown	**Blackbird**	**Great Tit**	**Greenfinch**	**Cuckoo**
Unknown	**74.6**	13.8	2.4	23.7	2.2
Blackbird	14.2	**77.7**	9.8	6.9	6.6
Great Tit	8.3	6.2	**85.8**	7.3	0
Greenfinch	2.9	1.9	1.9	**62.1**	0
Cuckoo	0	0.5	0	0	**91.2**

Tab. 3 Confusion table for kNN classifier (k=3) with recognition rate in %. Columns give the percentage of syllables from a species indicated in the top row being identified as a syllable of a species indicated in the leftmost column. Species 'Unknown' represents the rejection class.

7. POST-PROCESSING

A significant improvement of the recognition rate is accomplished by merging the prognosis for a single syllable to an overall prognosis for the recording. Usually, a recording contains a phrase or even a song of a special bird species and thus at least a few syllables from the same bird.

Based on the syllables detected in the recording, an independent prognosis is calculated for each bird. It is composed of the syllable frequencies in the recording and their individual reliabilities estimated by the classifier. Using this approach several species can be detected independently in one recording.

To further enhance the system's performance, a kind of voting scheme was implemented to combine the different classifier's results to a single result with respect to their performance (Tab. 2). The final recognition rates for the four investigated birds are given in Tab. 4.

	Unknown	**Blackbird**	**Great Tit**	**Greenfinch**	**Cuckoo**
Unknown	**58**	0	0	7	0
Blackbird	17	**100**	0	0	0
Great Tit	25	0	**100**	7	0
Greenfinch	0	0	0	**87**	0
Cuckoo	0	0	0	0	**100**

Tab. 4 Confusion table for the final implementation. Recognition rate for recordings in %.

8. SUMMARY AND FUTURE WORK

This study discusses a simple method for the automatic recognition of bird species analyzing their sounds. The presented approach reveals the feasibility of an automatic recognition with these methods in principle. However, it has to be taken into consideration that the results have been obtained for a small number of species which are members of different families.

Features describing the stationary properties as well as the short-time-behavior of syllables are employed. In general, features based on measures from the frequency band yield best discriminative power for the investigated birds. However, for a larger number of bird species in the training set, recognition will be harder.

Another aspect is the result, that most features have different classification ability in context of different species. So there exists no feature which is suitable for all species and a smart combination of several features is necessary.

Problems for real applications are:

- that the rejection parameters for the classifiers are very sensitive to changes in feature selection, investigated birds and number of clusters and therefore have to be estimated for each new training set.
- that segmentation of syllables is complicated when birds are singing competitively and their songs overlap in time (may be separable in frequency domain).

Some possibilities for future developments are:

- providing more training data to get more reliable estimations of the class models.
- using song-level contextual information, to describe typical arrangements of syllables in phrases or songs (grammar).
- Using superior classification algorithms like Hidden Markov Models or neural networks.

9. ACKNOWLEDGEMENT

S. D. acknowledges the scholarship of the DAAD (German Academic Exchange Service). Many thanks also to the people working at the "El Racó de l'Olla", especially to José Luis Terrasa Nebot, for their support during the data recording.

References:

[1] C. K. Catchpole and P. J. B. Slater, *"Bird Song: Biological Themes and Variations"*, Cambridge University Press, Cambridge, UK, 1995.

[2] C. P. H. Elemans, *"How do Birds Sing?"* PhD thesis, Wageningen University, Wageningen, The Netherlands, 2004.

[3] A. Härmä, *"Automatic Identification of Bird Species Based on Sinusoidal Modeling of Syllables"*, Int. Conf. Acoust. Speech and Signal Processing, 2003.

[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"*, J. Acoust. Soc. Am., vol 103, pp. 2185-2196, 1998.

[5] S. Dünkel, *"A Contribution to the Identification of Singing Birds Based on their Sounds"*, Final thesis, University of Applied Sciences Dresden, Dresden, Germany, 2005.

[6] http://www.birding.dk/

[7] T. Andersson, *"Audio Classification and Content Description"*, Master thesis, Luleå University of Technology, Luleå, 2004.

[8] K. Fukunaga, *"Introduction to Statistical Pattern Recognition"*, 2. Ed., Academic Press, San Diego, 1990.

[9] R. Hoffmann, *"Signalanalyse und -erkennung"* [*"Signal Analysis and Recognition"*], Springer, Heidelberg, 1998.

[10] D. W. R. Paulus and J. Hornegger, *"Applied Pattern Recognition"*, 4. Ed., Vieweg, Wiesbaden, 2003.

Authors:

Dipl.-Ing. (FH) Stefan Dünkel
Prof. Dr.-Ing. Kristina Kelber
Dept. of Electrical Engineering, University of Applied Sciences Dresden,
Friedrich-List-Platz 1, D-01069 Dresden, Germany
Phone: +49 351 462 2313
Fax: +49 351 462 2193
E-mail: stefan.duenkel@et.htw-dresden.de, kelber@et.htw-dresden.de,

Dr. Carlos Hernández Franco
Dept. of Communications, Polytechnic University of Valencia,
Ctra. Natzaret-Oliva s/n, 46730, Gandia, Spain
chernan@dcom.upv.es