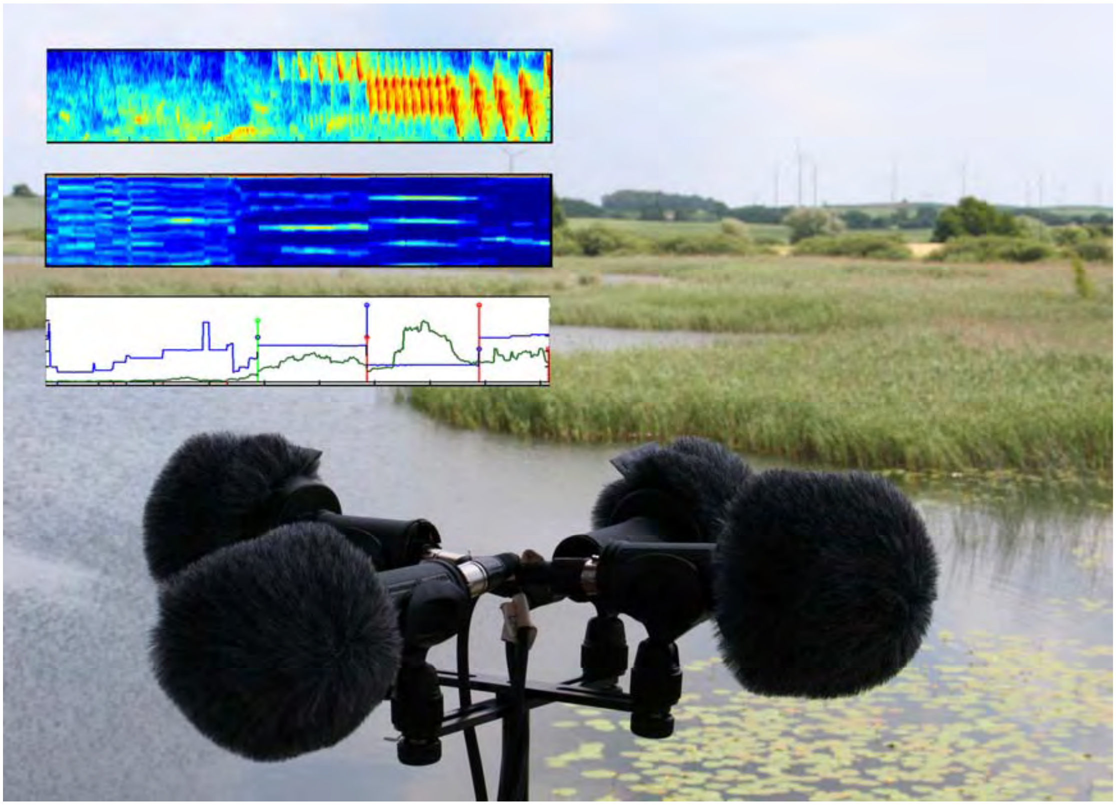
Karl-Heinz Frommolt, Rolf Bardeli and Michael Clausen (Eds.)

Computational bioacoustics for assessing biodiversity



Federal Agency for Nature Conservation

BfN - Skripten 234

2008

Computational bioacoustics for assessing biodiversity

Proceedings of the International Expert meeting on IT-based detection of bioacoustical patterns, December 7th until December 10th, 2007 at the International Academy for Nature Conservation (INA) Isle of Vilm, Germany

Editors:

Karl-Heinz Frommolt
Rolf Bardeli
Michael Clausen

BfN
Federal Agency for Nature Conservation

Cover picture: Northern part of the lake Parstein – a study site for bioacoustic monitoring (K.-H. Frommolt). The graphics in the left depict a set of features used in bioacoustic pattern recognition algorithms: spectrogram, periodicity features and a resulting automatic segmentation are plotted for the example of a Chaffinch's song (D. Wolff).

Editors' addresses:

Dr. Karl-Heinz Frommolt
Museum für Naturkunde
der Humboldt-Universität zu Berlin
Invalidenstraße 43
10115 Berlin, Germany
E-Mail: karl-heinz.frommolt@rz.hu-berlin.de

Rolf Bardeli and Prof. Michael Clausen
Institut für Informatik III
Rheinische Friedrich-Wilhelms-Universität Bonn
Römerstraße 164
53117 Bonn, Germany
E-Mail: clausen@iai.uni-bonn.de

Project coordinator BfN:

Dr. Horst Freiberg, Federal Agency for Nature Conservation (BfN)
Department Z 2.1 "Nature Conservation Information and Geographical Information"
National Focal Point CHM
E-Mail: Horst.Freiberg@bfn.de

BfN-Skripten are not available in book trade but can be downloaded in a pdf version from the internet at: http://www.bfn.de/0502_skripten.html

This publication is included in the literature database "***DNL-online***" (www.dnl-online.de)

Publisher: Bundesamt für Naturschutz (BfN)
Federal Agency for Nature Conservation
Konstantinstrasse 110
53179 Bonn, Germany
URL: http://www.bfn.de

Printed by the printing office of the Federal Ministry of Environment, Nature Conservation and Nuclear Safety.

Printed on 100% recycled paper.

Bonn, Germany 2008

Contents

Introduction

From December 7 through 10, 2007, an *international expert meeting on IT-based detection of bioacoustic patterns* was held at the facilities of the International Academy for Nature Conservation on the Isle of Vilm in the Baltic Sea. The meeting was held under the patronage of the German Federal Agency for Nature Conservation (Bundesamt für Naturschutz, BfN) in the context of a research project on bioacoustic pattern recognition.

The progress in information technology during the last years, especially in the field of pattern recognition, opens up important new perspectives for the automated bioacoustic monitoring of a multitude of animal species. The idea of the meeting was to bring together specialists from all over the world, to discuss the potential of IT-based detection of bioacoustic patterns. The topics of the meeting covered both the current knowledge on acoustic pattern recognition in bioacoustic signals and the application of bioacoustic methods for purposes of the monitoring of wild animals.

This publication contains expanded versions of the talks given during the meeting, giving an impression of the problems and methods in the field of algorithmic bioacoustics from scientists from Australia, the USA, Italy, Spain, Finland, Switzerland, the UK, the Netherlands and Germany. The articles cover a wide variety of animal species from insects and frogs to birds, from whales to bats. Monitoring locations reach from the densely populated centre of Europe to the out-backs of Australia, from the Mediterranean Sea to Antarctica.

In the first presentation, *Mapping bioacoustic phenomena over ecologically meaningful spatio-temporal scales,* Christopher Clark gave a good overview of the perspective of bioacoustic monitoring for a variety of species from elephants to whales.

The following presentations are all focused on different aspects of this general theme. They are characterised by either a certain species or group of species, a special study area, or are dedicated mainly to software systems or algorithms.

- Gianni Pavan et al. report on *Short term and long term bioacoustic monitoring of the marine environment*. Recording from an underwater test site for neutrino detection, investigations into the presence and migration of marine animals in the Mediterranean revealed a more frequent and consistent presence of sperm whales than previously believed.
- Lars Kindermann et al. describe a bioacoustics project in a truely adverse environment. *A perennial acoustic observatory in the Antarctic Ocean* consisting of hydrophones deployed under the ice shelf allows to study the acoustic repertoire of whales and seals.
- Martin Obrist et al. give a *Probabilistic evaluation of synergetic ultrasound pattern recognition for large scale bat surveys.* Residing in the realm of ultrasound, echolocation calls show strong interspecies overlap in their signal characteristics but nevertheless species recognition is often possible on the basis of these signals.
- Rafael Márquez et al. investigate the automated recording of *Anurans, the group of terrestrial vertebrates most vulnerable to climate change*. In a case study in the Iberian Peninsula, they compare results of an automated sound recording system with those of human listeners as tools in the study of the effect of temperature to Anuran vocalisations.
- Andrew Taylor et al. look back at *A decade of monitoring frog communities in northern Australia.* Their focus is on both, the pattern recognition problem and the design of autonomous recording stations. The latter are designed to operate unattended in remote areas and have to be highly robust to retain operability in adverse situations such as fires and flooding.
- David Chesmore discusses *Automated bioacoustic identification of insects for phytosanitary and ecological applications.* In particular, he uses time domain signal processing and artificial neural networks for the robust identification of taxa, concentrating on insects. He introduces the concept of time domain signal coding.

Among his results, he describes phytosanitary applications for quarantine insect larvae in timber.

- In bioacoustics research, and especially in bioacoustical pattern recognition, everybody needs a good software infrastructure. All investigation of bioacoustic signals starts from signal processing tools such as filtering, spectrogram computation and visualisation. Harold Figueroa introduces *XBAT, an open-source extensible platform for bioacoustic research and monitoring.* Here, in addition to the basic tools needed by every practitioner, new tools for feature extraction, classification and visualisation can easily be implemented based on existing algorithms. Sebastian Huebner investigates *Bioacoustic classifier system design as a knowledge engineering problem*. His system has a special focus on methods for classifier design for non-specialists in pattern recognition algorithms.
- Vlad Trifa et al. investigate *Automated methods for localization and recognition in real habitats using wireless sensor networks.* Building on recent advances in wireless networked sensing systems they use embedded systems to perform tasks such as species recognition and sound source localisation directly in the field.
- As remarked above, the meeting was organised in the context of a project on bioacoustic pattern recognition. Two articles survey results from this project. First, Karl-Heinz Frommolt et al. discuss *Advantages and disadvantages of acoustic monitoring of birds.* Their main focus is on the comparison of bioacoustic monitoring with other techniques. Based on studies in several acoustic situations, they show where acoustic monitoring can complement existing methods favourably. Second, Rolf Bardeli et al. present algorithms for *Bird song recognition in complex audio scenes.* They propose algorithms based on concepts for dealing with the noise and complexity encountered in real world monitoring scenarios.
- One of the most important tools in bioacoustics is the spectrogram. Researchers usually interpret the spectrogram as an image. Therefore, it is a natural idea to use algorithms from image analysis in this context. Scott Brandes uses *Techniques for bioacoustic signal detection using image processing* for filtering and feature extraction from the spectrogram.
- Juha Tanttu et al. use *Computational methods in analysis of bird song complexity.* They present methods for automatically assessing the song and syllable repertoire of individuals of the male Pied Flycatcher in order to study the influence of song complexity on breeding success. Environmental influences on signal-to-noise ratios make this an extremely challenging task.
- The last two presentations are dedicated to the monitoring of flight calls of migrating birds and bats. Thijs Schrama et al. investigate the *Automated monitoring of avian flight calls during nocturnal migration*. As a part of a project investigating bird migration in areas of offshore wind parks, they introduce an automated monitoring system and pattern recognition algorithms based on dynamic time warping and Euclidean distance measures. In a similar setting, Reinhold Hill et al. report on *Birds and bats: automatic recording of flight calls and their value for the study of migration.* Their automated registration revealed seasonal and daily time patterns of migration and relations of call intensity to weather parameters.

In addition to the talks, a lively discussion on the perspective of the application of technical means and pattern recognition algorithms for animal monitoring, especially in the context of conservation questions, has taken place. As a first step, the participants have agreed to form an international expert group. **A mailing list is available** for those interested in the topic and it allows to continue the exchange of ideas and algorithms. Subscription is possible at: https://mailbox.informatik.uni-bonn.de/mailman/listinfo.cgi/bioacoustic-monitoring.

March 12, 2008

K.-H. Frommolt
R. Bardeli
M. Clausen

Short Term and Long Term Bioacoustic Monitoring of the Marine Environment. Results from NEMO ONDE Experiment and Way Ahead

PAVAN G.[1], LA MANNA G.[1], ZARDIN F.[1], INTERNULLO E.[1], KLOETI S.[1], COSENTINO G.[2], SPEZIALE F.[2], RICCOBENE G.[2] & THE NEMO COLLABORATION[2]

[1] CIBRA – Centro Interdisciplinare di Bioacustica e Ricerche Ambientali, University of Pavia, Via Taramelli 24, 27100 Pavia, Italy, E-mail: gpavan@cibra.unipv.it

[2] INFN LNS - Laboratori Nazionali del Sud, Via S.Sofia 62, 95123 Catania, Italy, E-mail: riccobene@lns.inf.it

Abstract. The INFN NEMO-OνDE (Ocean Noise Detection Experiment) station, deployed on the seafloor at 2000 m depth 25 km offshore Catania (Sicily, Italy) in year 2005, was designed to continuously transmit broad-band acoustic data through optical cables to the INFN lab located in the port of Catania. It was operational until November 2006, when it was replaced by other experimental equipment. During the operational period, 5 minutes of recording (4 hydrophones, 45 kHz bandwidth, 96 kHz sampling rate at 24 bits resolution) were taken every hour. The experiment provided long-term data on the underwater noise and an unique opportunity to study the acoustic emissions of marine mammals living in, or transiting through the area east of Sicily. The recordings revealed a more frequent and consistent presence of sperm whales than previously believed.

Acoustic monitoring of the underwater environment is a key component in the study of marine mammals and in the management of the anthropogenic noise issue. Technologies now available allow to extend monitoring capabilities into the deep ocean to monitor the presence and behaviour of marine mammals as well as to long-term monitor both the local and the ambient noise background due to human activities.

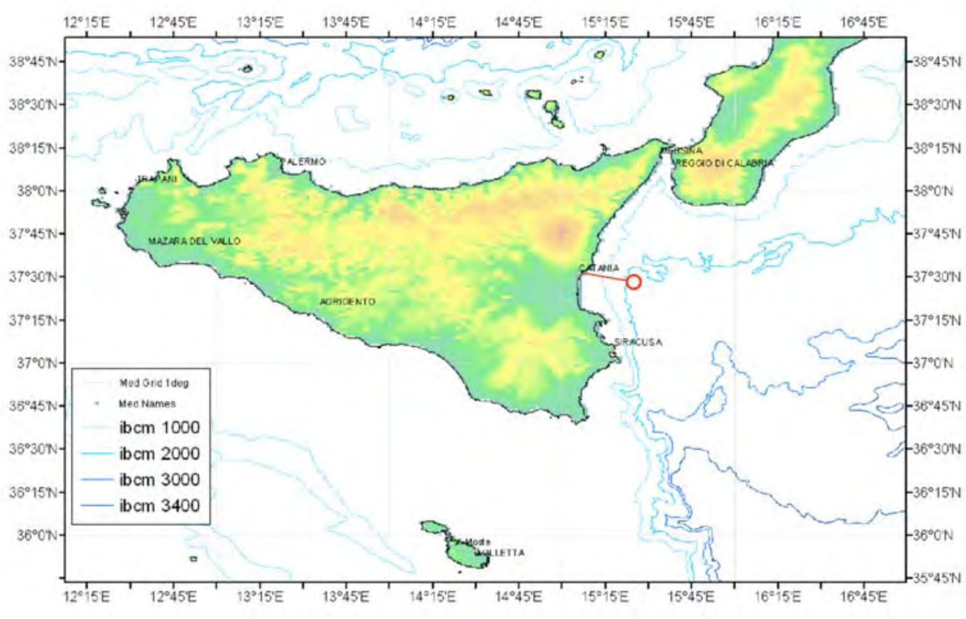


Figure 1: Location of the OνDE platform.

The goal of NEMO is the implementation of an innovative underwater telescope to search for astrophysical neutrinos (MIGNECO et al., 2006). The telescope will be deployed 100 km offshore Capo Passero (Sicily, Italy) at a depth of 3500 m. In this framework a deep sea test site at 2000 m depth has been deployed 21 km offshore Catania, connected to the shore labs through electro-optical cables to provide real-time data transfer.

The test site hosts an experimental deep station, named OνDE (Ocean Noise Detection Experiment) dedicated to the study of the underwater acoustic noise (Fig. 1). OνDE was deployed in January 2005 and was sending data since the end of March 2005 until

November 2006 when the acoustic module was replaced by other instrumentation. The acoustic experiment was concerned with the study of the underwater acoustic environment to develop strategies required for the detection of acoustic pulses that are generated by high energy neutrino interacting with water. The experiment was highly interdisciplinary and in addition to generating long term data on the underwater noise, it also provided a unique opportunity to study the acoustic emissions of marine mammals living in the area or transiting there during their movements within the Mediterranean basin. Bioacoustic research in the Mediterranean Sea dates back to 1958 (PAVAN ET AL., 1997); whereby research was mainly based on sporadic acoustic and visual surveys to monitor the presence of marine mammals in small areas. Bioacoustic research efforts have increased in the last 10 years and this is the first long term monitoring project in the Mediterranean Sea for both noise measures and marine mammals' sounds.

Experimental Setup and Methods

The hydrophones on the OνDE acoustic module, a special series produced by Reson, are hooked on the upper part of the platform frame, forming a tetrahedral array of about 1 m side (Fig. 2). Hydrophones H1, H2, H4 lie in the same plane at about 2.5 m from the seabed, H3 is placed on the top vertex at about 3.2 m from the seabed. The broad band hydrophones (30Hz-45kHz) have –175 dB peak ref 1V/μPa sensitivity (including a 20 dB gain preamplifier) and are connected to a four channels 24 bit AD board with full scale level of ± 2V; the system allows recording sound pressure levels up to 181 dB peak. The hydrophones were sampled at 96 kHz and send continuously to the shore lab; as the continuous archiving was not possible due to storage space constraints (uncompressed recording would require 124GB/day), continuous recordings were made only for the testing period and then scheduled for periods of 5 minutes every hour. For additional technical details see RICCOBENE ET AL. 2007. Data recording, totalling 2.5TB of scheduled recordings, and 1TB of continuous recordings, was made with SeaRecorder, a 4 channels software recorder developed at CIBRA that reads and synchronizes the two stereo digital streams coming from the underwater station (Fig. 3). Digital data arrives with 24 bit resolution and can be saved as standard Microsoft .wav files either in integer (16 or 32 bit/sample) or 32 bit float format. Data for noise analysis at INFN was saved in 32 bit float format; copies for bioacoustic analysis at CIBRA (PRIANO ET AL. 1997) were reduced to 16 bit to save disk space. Bioacoustic analysis was performed with SeaPro, the sound analysis software developed at CIBRA. The first phase was the classification of recorded sounds into known categories. The results have been arranged in excel tables to show, hour by hour, the occurrence of biological sounds and other relevant acoustic events such as sonar, echo sounders, sparkers or other sound.

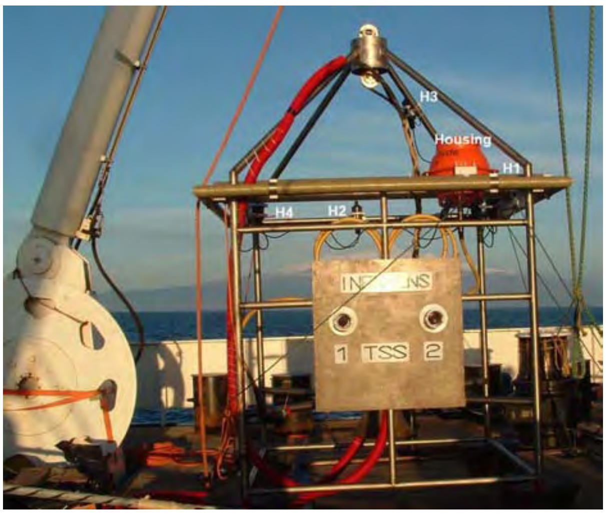


Figure 2: The titanium frame of the OνDE platform. The frontal plate hosts the connectors for the optical cables that connect the platform to the inland lab.

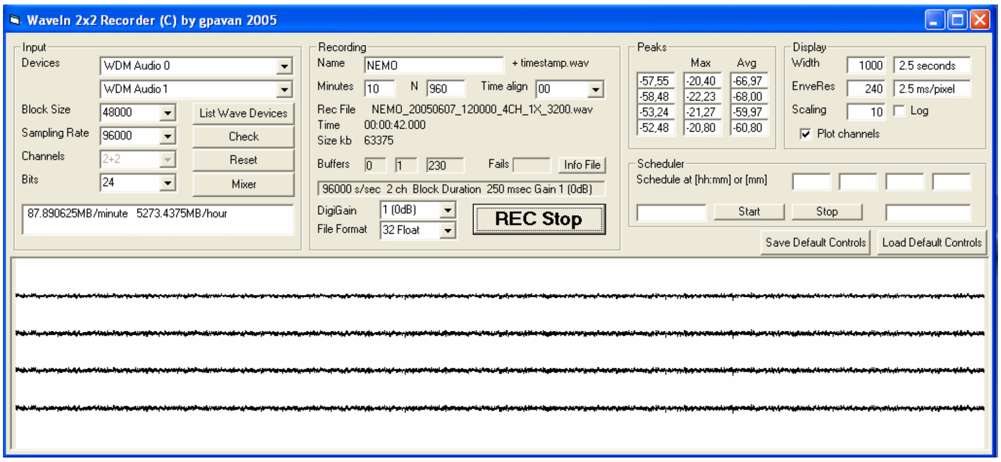


Figure 3: The SeaRecorder main panel.

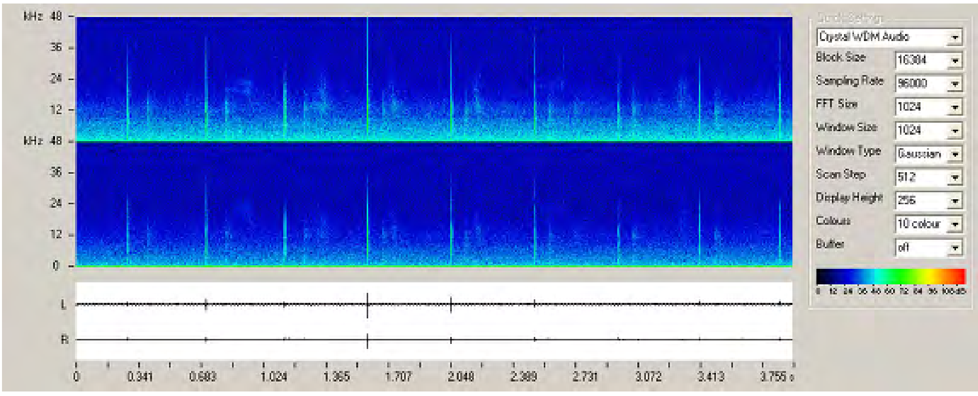


Figure 4: Two channel spectrographic display of a series of sperm whales' clicks.

Sperm Whale Detections

The most common sounds recorded are those produced by sperm whales (TELONI 2005): impulsive sounds extending in frequency to more than 30 kHz, named click (Fig. 4), arranged in regular sequences (interclick interval in the range 0.5s to 2s), or in special patterned sequences (chirrups, codas, creaks – Fig. 5 & 6). According to studies in the Mediterranean Sea, sperm whales may dive to more than 1000 meters depth, but normally travel and forage at 800-1000 meters depth. Their source level may be greater than 230 dB at 1m and on axis; with OνDE the loudest clicks were received with sound pressure levels up to 170 dB. Clicks were often recorded with high SNR; by using a high-pass filter, the SNR can be further increased to improve the detection, count and analysis of the weakest click series.

Detections indicate a presence of sperm whales more consistent and frequent than previously believed. Although the transiting of sperm whales is known since the end of XIX century (BOLOGNARI 1949), little literature is available for the area. IFAW (2004) reports a low sperm whale density in the Ionian basin with an encounter rate of 5.8 whale groups for 1000 km of transect. Lewis (2006), based on the IFAW surveys, reports a total of 16 whales detected within the truncation distance of 20 km, along 3846 km of survey transects in the Ionian Sea. PANIGADA ET AL. (2007) published a report about marine mammal presence in the area of the Strait of Messina and extending to south to 50km north of the OνDE Station. In the period from June 2005 to May 2006 they conducted 125 days of survey and sighted 80 marine mammals, of which 13.8% sperm whales and 1.2% Cuvier's beaked whales.

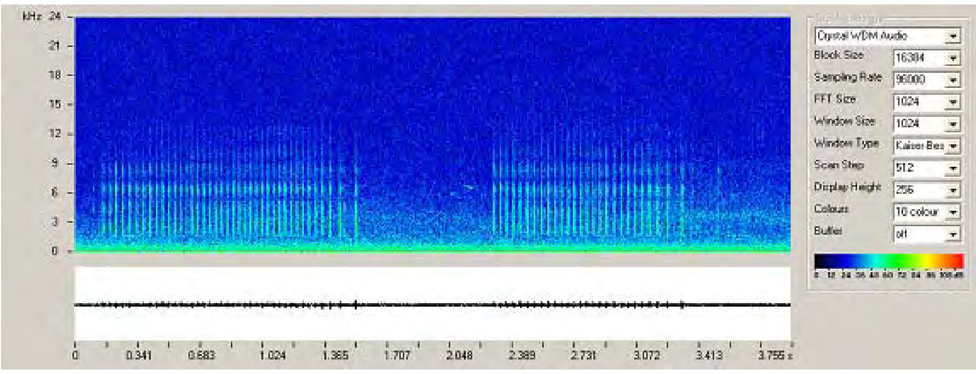


Figure 5: Spectrographic display of chirrup.

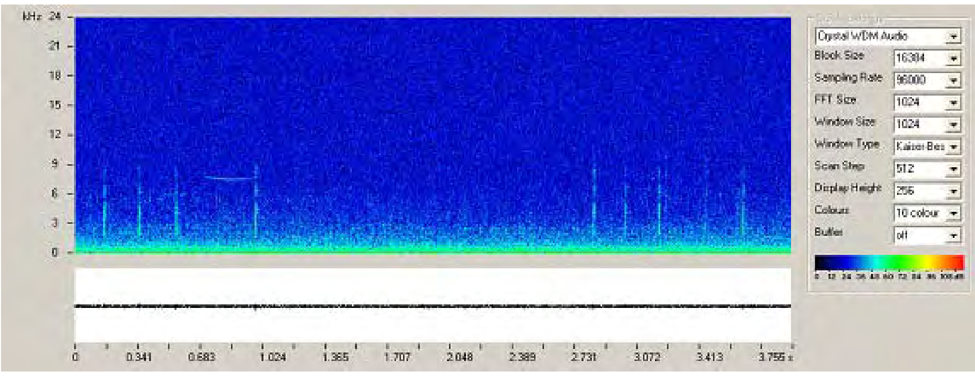


Figure 6: Spectrographic display of two codas with the typical Mediterranean pattern 3+1. A weak dolphin whistle is also shown on the first coda.

With the OνDE station, in year 2005 (Fig. 7), sperm whales have been detected in 117 of the 231 (50.6%) recorded days (1186 out of 5147 recorded hours, 11.5%). In 2006 sperm whales have been recorded in 31 of 83 recorded days (37%). Although several periods of consecutive days were characterized by the presence of sperm whales, solitary whales or groups of several individuals were often detected for few hours only and this may indicate that they were in transit. In few cases whales were present for more than one day and in some occasions, other sounds than clicks and creaks were recorded, in particular chirrups and codas indicating socializing behaviours. The coda pattern that was more frequently recorded followed the 3+1 type (PAVAN ET AL. 2000), however, the number of codas of the 2+1 type was greater than previously reported for this area.

Work in Progress

New analysis algorithms are being developed to maximize the SNR ratio and to track the movements of impulsive acoustic sources to reveal the movement of sperm whales whilst in the detection range. Estimating the TDOAs (Time Difference Of Arrivals) between the four hydrophones allows the separation of the sound arrival directions and the tracking of the sources' movements. In some cases, when surface reflections are clearly associated to individual clicks it is possible to exactly locate the animal, i.e. to know range and depth, instead of having only azimuth and elevation information. Such direct ranging will be important to assess the real detection range of the OνDE station and to improve the tracking of the animals. At the same time click details are examined to measure the Inter Pulse Interval, i.e. the intervals among the pulses that constitute the sperm whale click (PAVAN ET AL. 1997; ZIMMER ET AL. 2005). According to the latest models on sound production (ZIMMER ET AL., 2005) reliable IPI measures can be taken if on axis, either frontal or caudal, when a clean (P0)-P1-P2 structure is available (Fig. 8). By measuring the IPI it is possible to assess the whale size and, if the size is greater than 13m, the sex (females' length is up to 13m, males may grow up to 18m). By combining TDOA and IPIs it should be possible to assess the number of transiting animals and the groups' composition allowing a better estimate of population size and structure. The tracking of their movements will possibly reveal if their directions are seasonal and directed to/from the Strait of Messina, as already suggested by BOLOGNARI (1949), or if they follow other rules, if any.

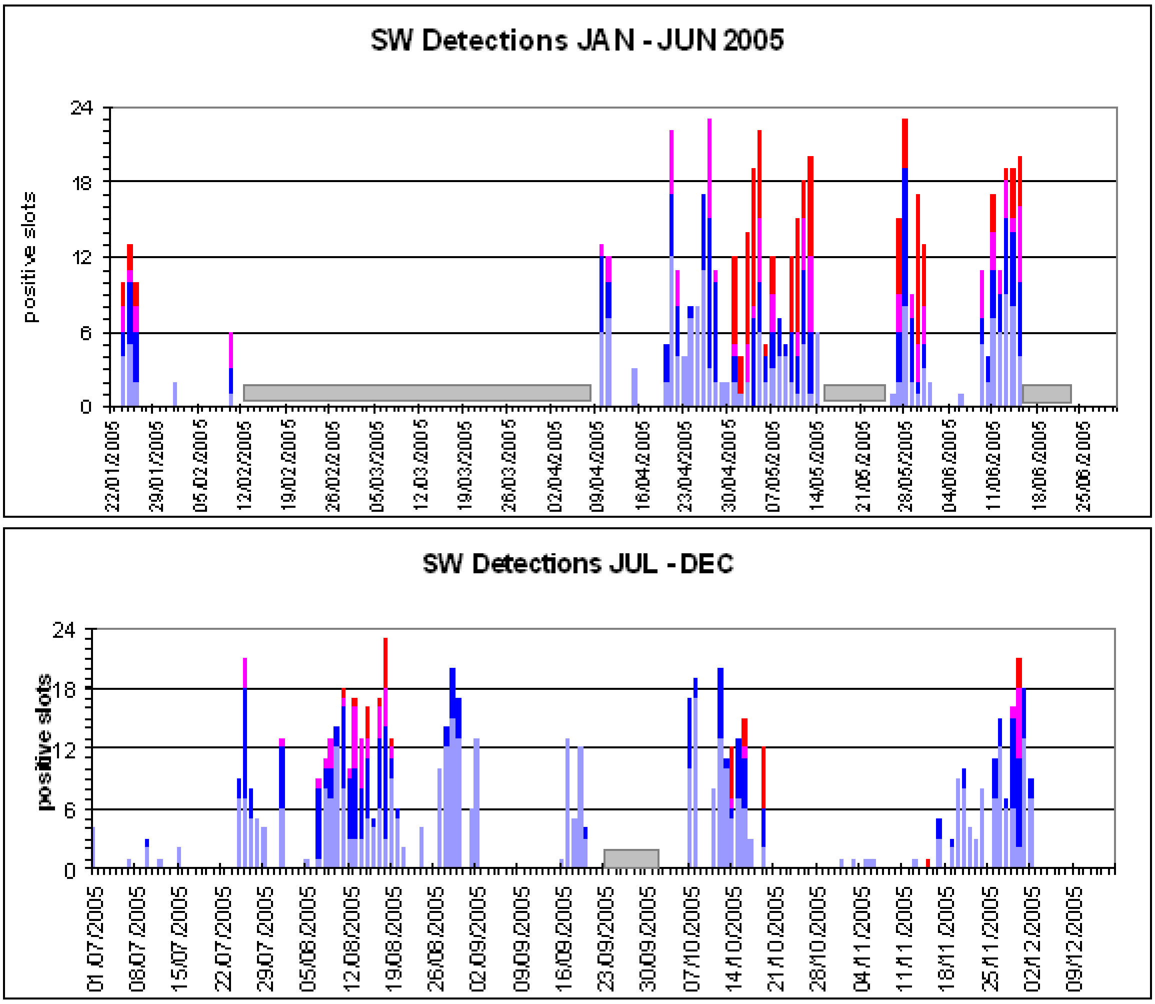


Figure 7: Sperm whales' detections in year 2005. Vertical bars indicate how many hours sperm whales have been detected each day. The stacked bars indicate how many animals were detected each hour: 1 animal detected; 2 animals detected; 3 animals detected; 3 or more animals detected. Horizontal grey bars indicate periods of recording inactivity. The longest one was due to a cable break.

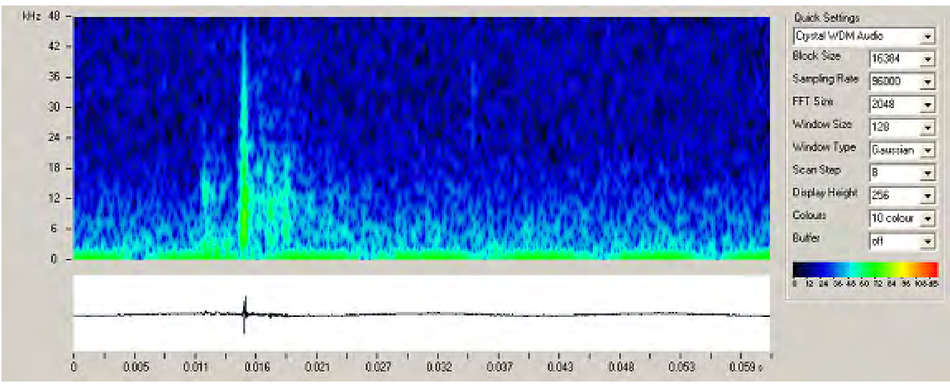


Figure 8: Sperm whale click with visible multi-pulse structure: P0-P1-P2. Based on the IPI, the size has been estimated 9.94-10.27 meters, matching a young male or a female. X-axis shows 64ms.

Cuvier's Beaked Whales

Sperm whale clicks are often loud as the animals dive at great depth, close to the receiving hydrophones; on the contrary, whistles and clicks from dolphins, which stay within few hundred meters from the surface, are recorded with much lower amplitude. Clicks similar to those emitted by Cuvier's beaked whales (JOHNSON ET AL., 2004; ZIMMER ET AL., 2005) have been detected with amplitude much greater than that of dolphins indicating a deep source that match well with the known dive depths of beaked whales. Cuvier's beaked whales have been recorded to dive to more than 1800m (JOHNSON ET AL., 2004). The habitat of this species appears to be associated to sharp continental slopes, but it was never reported for the OνDE area; few strandings have been recorded on the Calabria coasts and, recently, 10 km South of Messina (PODESTÀ ET AL. 2006, PAVAN ET AL. 2008). PANIGADA ET AL. (2007) reported few encounters in the Messina Strait area in 2005.

Received clicks match well descriptions given by the WHOI team with the D-TAG (ZIMMER ET AL. 2005) but show larger ICI (Inter Click Interval), 460-480 msec rather than 400 msec (Fig. 9 & 10). It may be worth to note that the amplitude of the received clicks is not constant but oscillating, as it can be expected by a directional source swimming and scanning the environment with left-right movements of the head. The hydrophones get the maximum amplitude when it is in the beam axis.

These detections indicate that deep acoustic sensors can be used to reveal and to monitor the presence of this species that seems sensitive to anthropogenic sound and but is difficult to detect.

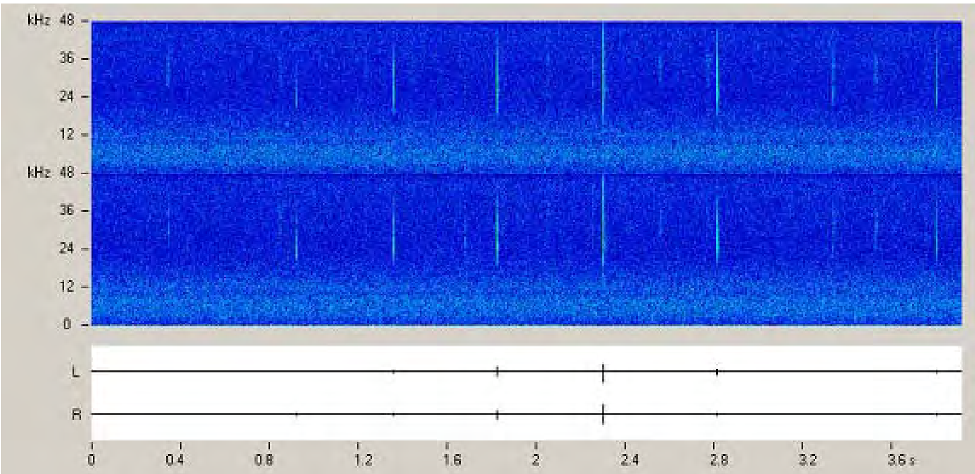


Figure 9: Series of clicks spaced 460-480 ms, much louder than clicks from shallow dolphins, with oscillating amplitude. X-axis shows 4.0 seconds.

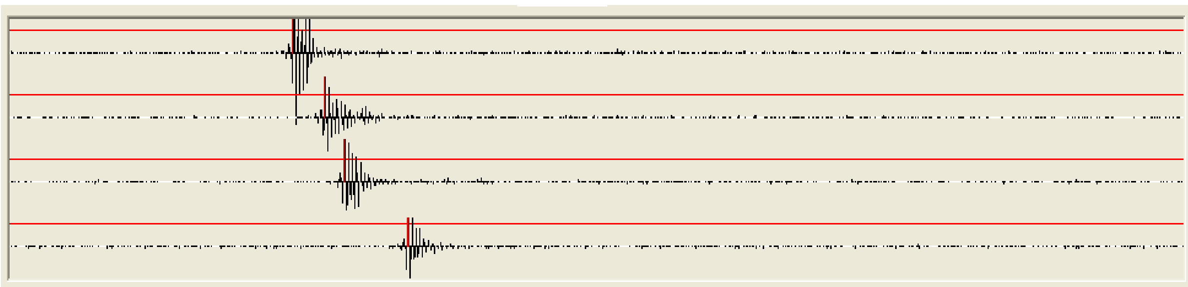


Figure 10: A click received on the 4 hydrophones. Waveform display (x-axis 9.6msec) shows a time length of 300microsec matching well with Cuvier's beaked whales' click sructure.

Other Detections

Other than recognizable biological sounds, many man-made sounds have been recorded and identified, including ship noise, sonar, echo sounders, airguns (or maybe sparkers), and explosions. Acoustic events of unknown origin have been also recorded.

Way Ahead

Based on the success of OνDE a new EEC funded project named LIDO (Listening Into the Deep Ocean) was set up with the collaboration of INGV (National Institute of Geophysics and Vulcanology) and other international partners to renew the OνDE platform and to create a Mediterranean wide acoustic monitoring network by upgrading existing underwater seismic detectors with broadband acoustic sensors.

The implementation of a number of “whale gates” to monitor the presence and movements of marine mammals in critical areas, will allow to better understand their populations dynamics and the long term changes that are possibly driven by direct or indirect human impact, including climate changes.

Following the increasing interest in autonomous underwater acoustic monitoring, a new low cost bottom recorder was developed in cooperation with Nauta-rcs, a company specializing in underwater equipment. Based on a modified commercial M-Audio Microtrack 24/96 digital audio Compact Flash recorder, custom electronics that add scheduled recording features, additional power supply with NiMH rechargeable batteries, and a customized Sensor Technology hydrophone, this system is installed into a 50cm x 9cm aluminium canister designed to operate down to -500m. In the present configuration, it allows an operating life of up to a week (depending on the recording schedule and available storage capacity); the unit can be pre-programmed from a PC to follow an extremely flexible recording scheme. Within the LIDO project, it will be used to monitor and select locations suited for a permanent monitoring platform.

Long-term acoustic monitoring programs that are made possible by present technologies generate huge amounts of recorded data, the analysis of which is a serious issue that presses for the development of reliable automatic sound recognition software. LIDO includes a project to develop and test semi-automatic and automatic classification software able to make browsing of huge amount of acoustic data easier and also to perform in real-time providing immediate feedback.

Technologies developed for the underwater environment could be easily adapted to monitor both the acoustic biodiversity and the anthropogenic noise contamination of terrestrial habitats.

Acknowledgements

A special thank to the whole NEMO team for having supported this interdisciplinary approach.

References

BOLOGNARI A., (1949): A proposito della recente cattura di alcuni esemplari di capodoglio (*Physeter macrocephalus* L.) nel Mediterraneo. Bulletin de l’Institut Océanographique, n. 949. Monaco: 1-43.

IFAW, (2004): Results of 2003 Pilot Survey and Remarks on Proposed Basin-Wide Sperm Whale Survey. - Report to ACCOBAMS 2nd Meeting of Parties, Palma de Mallorca; 9-11th November 2004.

JOHNSON M., PETER T. MADSEN P.T., ZIMMER W.M.X., AGUILAR DE SOTO N., TYACK P., (2004): Beaked whales echolocate on prey. - Proc. R. Soc. Lond. B (Suppl.) **271**: S383–S386.

LEWIS T., GILLESPIE D., LACEY C., MATTHEWS J., DANBOLT M., LEAPER R., MCLANAGHAN R. AND MOSCROP A., (2006): Sperm whale abundance estimates from acoustic surveys of the Ionian Sea and Straits of Sicily in 2003. - J. Mar. Biol. Ass. U.K. (2006), **86**, 5489/1–5.

MIGNECO E. ET AL., (2006): Stratus of NEMO. Nuclear Instruments and Methods in Physics Research A **567**: 444–451.

PANIGADA S., NOTARBARTOLO DI SCIARA G., ZANARDELLI M., AGAZZI S., BENDINONI F., COSTA M. AND PESANTE G., (2007): Habitat use and preferences of cetaceans in the Strait of Messina (Italy) through spatial modelling: implications for management. - Abstract ECS 2007.

PAVAN G., BORSANI J.F., (1997): Bioacoustic research on cetaceans in the mediterranean sea. - Mar. Fresh. Behav. Physiol. **30**: 99-123.

PAVAN G., PODESTÀ M., (2008): Banca dati nazionale sugli spiaggiamenti di mammiferi marini. Database online. http://www.unipv.it/cibra/spiaggiamenti.html.

PAVAN G., PRIANO M., MANGHI M., FOSSATI C., (1997): Software tools for real-time IPI measurements on sperm whale sounds. Proc. Underwater Bio-Sonar and Bioacoustics Symposium. - Proc. I.O.A., **19** (part 9), Loughborough, UK: 157-164.

PAVAN G., T. HAYWARD, J. F. BORSANI, M. PRIANO, M. MANGHI, C. FOSSATI, J. GORDON, (2000): Time pattern of Sperm Whale Codas recorded in the Mediterranean Sea 1985–1996. - J. Acoust. Soc. Am. **107**: 3487-3495.

PODESTÀ M., D'AMICO A., PAVAN G., DROUGAS A., KOMNENOU A. AND PORTUNATO N., (2006): A Review of Ziphius cavirostris (G. Cuvier, 1823) Strandings in the Mediterranean Sea. - J. Cetacean Res. Manage. **7**: 251–261.

PRIANO M., PAVAN G., MANGHI M., FOSSATI C., (1997): The Cetacean Sound Library of the Interdisciplinary Centre for Bioacoustics and Environmental Research of the University of Pavia. Proc. Underwater Bio-Sonar and Bioacoustics Symposium. - Proc. I.O.A. **19** (part 9), Loughborough, UK: 245-249.

RICCOBENE G., COSENTINO G., MUSUMECI M., PAVAN G., SPEZIALE F., THE NEMO COLLABORATION, (2007): Acoustic background measurements with NEMO-ONDE. - LNS Reports, 2007.

TELONI V., (2005): Patterns of sound production in diving Sperm whales in the northwestern Mediterranean. - Marine Mammal Science **21**: 446-457.

ZIMMER W.M.X., JOHNSON M.P., MADSEN P.T., AND TYACK P.L., (2005): Echolocation clicks of free-ranging Cuvier's beaked whales (Ziphius cavirostris). - J. Acoust. Soc. Am. **117**: 3919-3927.

ZIMMER, W.M.X., MADSEN P.T., TELONI V., JOHNSON M.P. AND TYACK P.L., (2005): Off-axis effects on the multipulse structure of sperm whale usual clicks with implications for sound production. - J. Acoust. Soc. Am. **118**: 3337-3345.

A Perennial Acoustic Observatory in the Antarctic Ocean

LARS KINDERMANN, OLAF BOEBEL, HORST BORNEMANN, ELKE BURKHARDT,
HOLGER KLINCK, ILSE VAN OPZEELAND, JOACHIM PLÖTZ & ANNA-MARIA SEIBERT

Alfred Wegener Institute for Polar and Marine Research, Am Handelshafen 12, D-27570 Bremerhaven, Germany
E-mail of the corresponding author: lars.kindermann@awi.de ; Web: http://www.awi.de/acoustics

Abstract. In December 2005 the Perennial Acoustic Observatory in the Antarctic Ocean (PALAOA, Hawaiian word for "whale") was set up on the Ekström ice shelf, Antarctica near the German Neumayer Station (Boebel et al., 2006). Since then, it almost continuously records the underwater soundscape in the vicinity of the ice shelf edge and is intended to do so over the duration of several years. These long-term recordings allow studying the acoustic repertoire of whales and seals in an environment almost undisturbed by humans. The data is analyzed to detect species specific vocalizations, infer the approximate number of animals inside the measuring range, calculate their movements relative to the observatory, and examine possible effects of the sporadic shipping traffic on the acoustic and locomotive behaviour of marine mammals. Underwater sound is recorded by means of four hydrophones located through boreholes below the 100 m thick floating ice sheet. They are attached to an autonomous, wind and solar powered station, which can record at 192 kHz / 24 Bit. A compressed data stream is transmitted in real time via wireless LAN from PALAOA to the German Neumayer Base at 15 km distance. From there, a permanent satellite link transmits an even more compressed stream to the AWI in Germany. It can be accessed live from our webpage at http://www.awi.de/acoustics. So far, Weddell seals, crabeater seals, Ross seals, leopard seals, killer whales, blue whales, fin whales and minke whales have been identified in the recordings along with several vocalizations which could not be assigned with certainty to species level yet. Additionally, many non-biological sounds were recorded, mostly generated by ice and some anthropogenic events like ships passing by and human activities on the ice. Difficulties have arisen from the sheer size of this constantly growing dataset, which consists of more than 10.000 hours so far. We develop interfaces and setups to process this stream in real time and analyze it both interactively and by means of batch processes running in parallel on a workstation cluster, for example applying detectors specific to single species, based on hidden Markov models. These recordings, which are largely free of anthropogenic noise, provide also a base to set up passive acoustic mitigation systems used on research vessels by developing automatic pattern recognition procedures to be used in the presence of interfering sounds, e.g. propeller noise.

Antarctica is one of the last areas on earth which is to a large extent untouched by human activities and it is granted a special protection status. The continent is by international agreements dedicated to nature and science. The Madrid Protocol of the Antarctic Treaty System requires a permit for all operations in the Antarctic territory beyond 60° South, to be issued by an authority of one of the respective contracting states. Applications for research permits do require a proof of the environmental soundness of the planned activities. Potential risks of the field work to marine mammals in particular must be evaluated in terms of the possible impact on species or population level. Therefore it is necessary to take into account the presence of the respective species at the proposed time and location of the operation. However, little is known about the abundances and migration patterns of many species due to the inaccessibility of this remote area which can only be assessed by ice-strengthened vessels during the austral summer months. Knowledge about the total number of animals and their spatio-temporal distribution is usually gained on ship surveys where observers count the sightings of animals and extrapolate abundance estimates from these numbers. However, marine mammals spend most of their lifetime submerged, and the probability of spotting them in the open ocean is influenced by factors such as light conditions, weather, sea state and the observer himself. Hence the results strongly depend on assumptions on detectability and behaviour. Reliable data exist only for areas which are visited by ships frequently enough to provide sufficient statistics. Recently, satellite transmitters attached to individual animals allow following their migration paths for substantial periods of time, extending the knowledge to locations and seasons where direct observation is not possible. But in general, estimations of the number of animals in a given area at a given time still suffer from a high degree of uncertainty.

As most marine mammals are vocal under water, the methods for passive acoustic monitoring which have been developed in the recent years provide a convenient and efficient way to detect marine mammals. Autonomous systems allow to record for extended periods of time without the necessity of human presence and can be deployed at sites of interest. However, the Antarctic environment imposes extraordinary demands on material and operations, so special equipment and procedures are needed here. Germany operates the year-round manned Neumayer research base on the Ekström Ice Shelf, at 70°39'S, 008°15'E, close to the shelf ice edge at the north-eastern entrance of Weddell Sea. Many species of seals and whales have been observed to inhabit this area. As this is a focal point of Germany’s Antarctic activities, it is of great interest to study the influence of human operations to the marine environment at this location. Lastly, the base also provides the necessary logistics to set up and maintain a hydroacoustic observatory, that otherwise would go far beyond the operational and financial scope of such a project. In 2005 we deployed a hydrophone array near Neumayer Base, which since then provides continuous online access to the underwater soundscape of Antarctic waters.

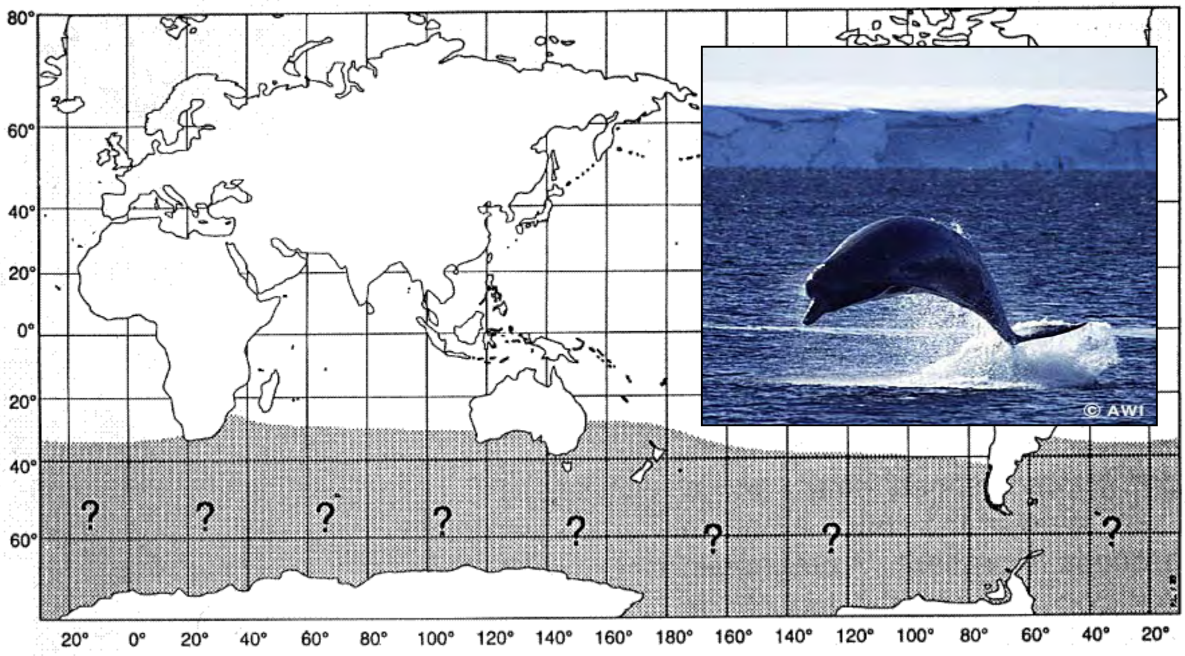


Figure 1: Proposed distribution of the Arnoux beaked whale, as given in the United Nations FAO species identification guide.

Ice Shelf Edge Habitat

Antarctica is almost completely covered by ice which can reach heights of up to 4000 metres in the inland of the continent. From this ice mass, glaciers and ice streams form which slowly move towards the coast. Reaching the ocean, the glaciers and ice streams float on top of the water and can extend hundreds of kilometres to the north, forming large ice shelves. At the edge, ice breaks off, either by calving or releasing giant table icebergs which can reach a size of hundreds of square kilometres. In austral winter the ocean is covered by sea ice, either drifting *pack ice* or *fast ice* attached to the ice shelf. However, the predominant katabatic winds drive this ice mass northwards and frequently create a stripe of open water, the coastal polynia. In summer, the sea ice disappears mostly and leaves an open ocean. This area is habitat to several species of penguins, whales and seals, year-round or seasonal (Fig 1, 2). Only the Weddell seals stay in the coastal shelf waters all the year, while Ross, leopard and crabeater seals are summer guests only. It is unclear, which whale species are present here in wintertime, taking advantage of the open waters provided by the polynia.

Figure 2: Ice shelf edge with crabeater seals on sea ice floes and penguins on an iceberg (top left), Weddell seals on fast ice (top right), a group of minke whales (bottom).

Station Location

Traditionally, hydrophones are deployed in the ocean either on anchored or drifting buoys or sea bottom moorings. In the Antarctic Ocean, buoys will be destroyed by drifting ice floes. In shallow, coastal waters moorings will be destroyed by big icebergs ploughing the seafloor. There is no way to route a cable safely from an underwater station towards the shore, like in hydrophone stations deployed at many places elsewhere in the world. The only safe place is provided by the ice itself. The massive floating ice shelf, with a thickness of up to hundreds of meters provides a perfect shelter. 15 km north of Neumayer Base, west of Atka Bay the Eckström Ice Shelf spreads to a typical, finger like structure. Under the most prominent of these ice fingers, called the *North Pier*, surrounded by the ocean on the north-west, north and east side, we were expecting to have excellent acoustic reception from the surrounding ocean (Fig 3, 4). In general it is extremely dangerous to move on the ice due to crevasses pervading the glacier, but this area is safe to access on a secured way, marked with flags. It leads to the base because ships use this site as a pier to unload the Neumayer supplies.

Figure 3: The PALAOA site ●, photographed from a helicopter (left) and from the spot marked ● (right).

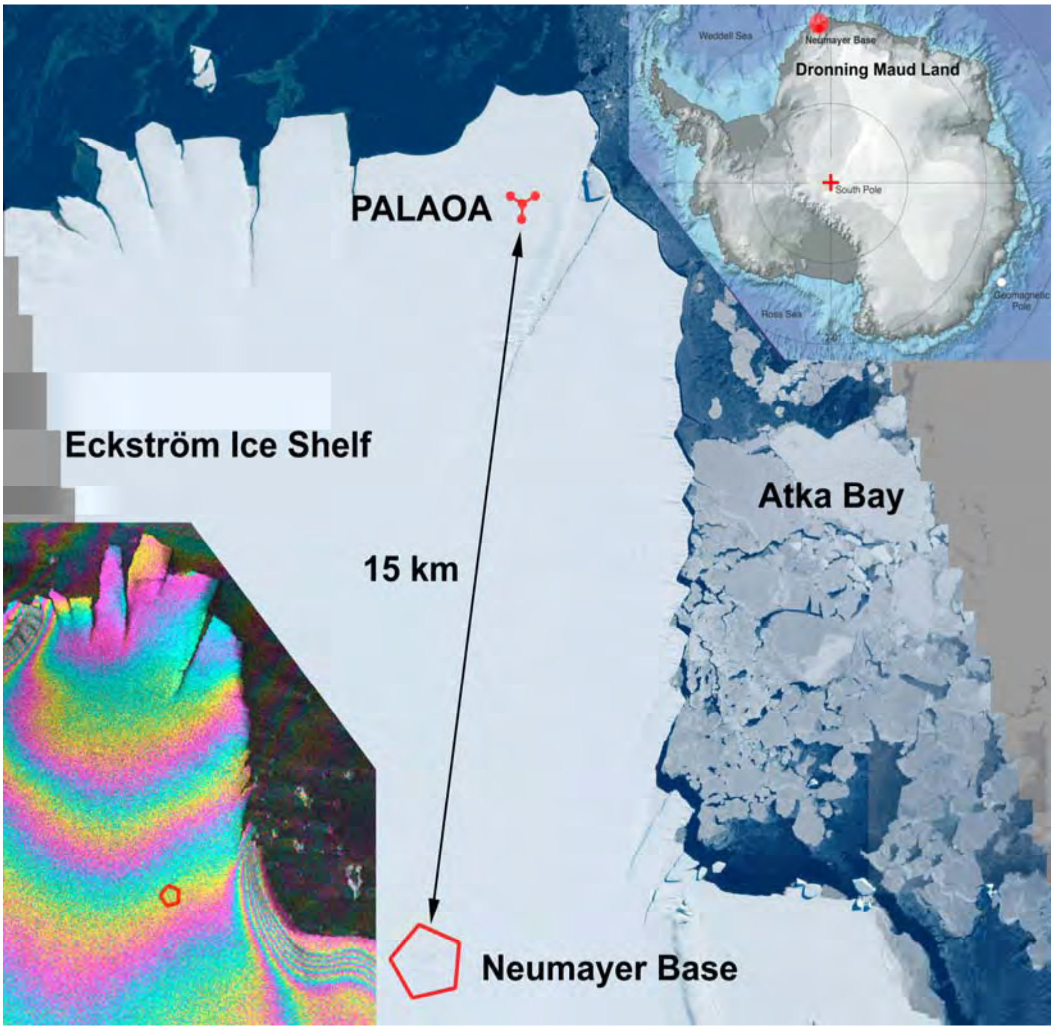


Figure 4: Location of PALAOA with an IKONOS satellite image underlay and a radar interferometer picture, showing the physiograpy of the area.

Hydrophone Deployment

Holes were drilled through the 100 m thick ice shelf to get access to the water body underneath (Fig. 5). This was achieved by a hot water drilling system developed by AWI, which is capable to melt through the ice in about 12 hours per hole. Energy consumption is enormous, requiring about 750kW to melt the ice and heat the water to 95°C.

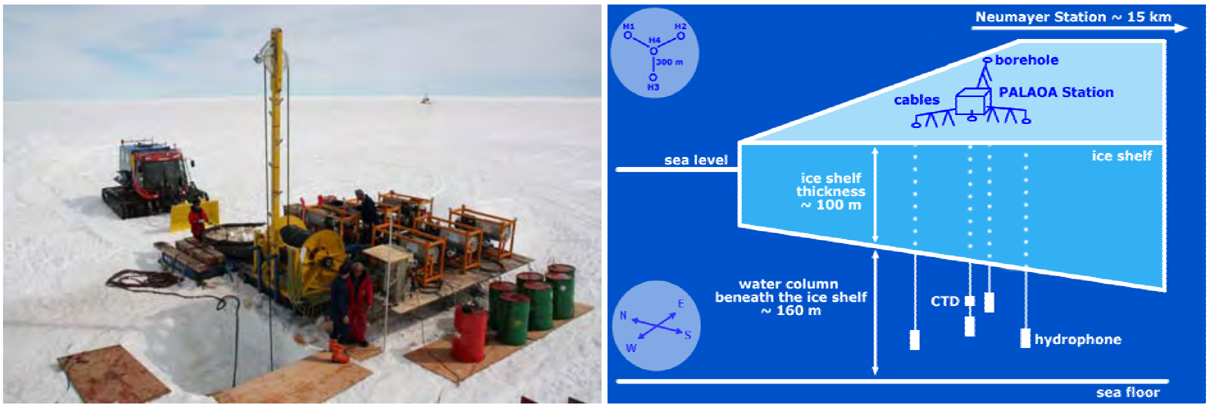


Figure 5: Hot Water drilling equipment and hydrophone array layout.

Energy and Networking

Energy is crucial for any autonomous system. PALAOA is equipped with solar cells, a wind generator, a large battery bank and a methanol fuel cell (Fig. 6). Energy consumption can be tuned according to the available supply by switching devices on or off on demand. This energy management is implemented as a BASIC program on a BARIX Barionet PLC, equipped with relays and I/O modules. Almost all devices in the observatory are switchable via a relay.

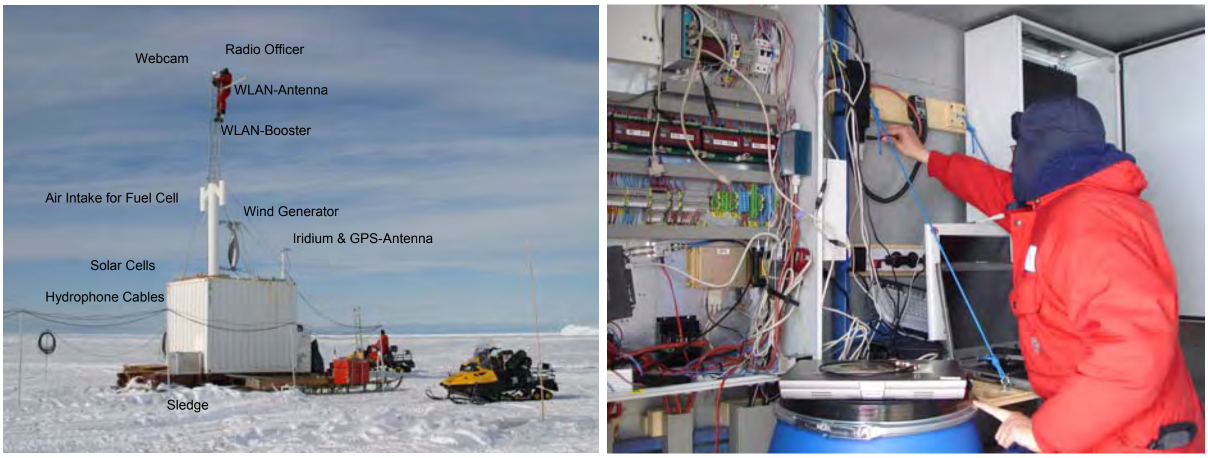


Figure 6: The PALAOA container and the electronics inside.

Hydrophone amplifiers have a very high input impedance, about 100 MOhm. This implies that they pick up interference easily which induces noise and artefacts in the audio recordings. Charge controllers, on the other hand, and DC-DC converters emit a lot of electronic noise. Attempting to power the hydrophones directly from a circuit connected to the charge controllers, results in horrible sound quality. We set up a power bus system with a main energy circuit and two audio circuits for the hydrophones. Each of 6 battery packs can be attached either to the main, charging bus or one of the audio busses. In this way it is possible to separate the audio system galvanically which reduces interference drastically. In addition, all hydrophones were also completely separated from each other. After deployment of the first hydrophone through its borehole, we received a very good signal quality. However, after attaching the second hydrophone, the signal of both channels was disturbed by spikes with higher amplitude than the audio signal. The spikes appeared immediately when any connection between the hydrophones was made. Connecting the ground of another hydrophone cable was enough. After a lot of investigations and tests we are confident now, that we accidently built an extremely large and powerful antenna. The hydrophone cables form 3 legs of a 600 x 180 m rectangle and the seawater closes the loop. Such a giant closed loop is the ideal geometry for a VLF ring antenna which picks up electromagnetic frequencies in the audible range. Lightning strikes are known to produce such signals, known as *sferics* or *whistlers*, which can travel around earth. We were listening to all the thunderstorms in the world at once. Especially during noon in Amazonas and Congo areas the spikes increased. The only solution was to completely separate every hydrophone channel galvanically from each other. However, during the built-up period we did not have the necessary equipment with us at that time and we decided to connect only one hydrophone and bring the necessary equipment to connect the other hydrophones one year later. It is not a trivial task to set up a multi-channel audio system without any wires between the components. Also each hydrophone had to be powered separately which increased the complexity of the energy system enormously. We developed an audio system with separate AD converters for each hydrophone channel which were connected via optical fibre to a digital sound card, in the PC. However when we returned one year later we discovered that two of the four hydrophones were defect. Only the central and the north eastern hydrophones were still operational. There is no chance to find out what caused this defect, as the boreholes had frozen again. With only two channels left, is was much easier to achieve galvanic separation. Standard studio DI boxes did the job for the audio signals and the power wiring could be modified to have two separate battery packs and two power busses for the hydrophone amplifiers. In minimal mode, only sending a stream it consumes 20 watts, mainly for the wireless LAN bridge, which connects the container to the local network at Neumayer Base. Data rate on this 15 km wireless link is about two MBits/s. This set up has worked flawlessly in any weather condition so far, allowing permanent access to the acoustic data and station control via the internet.

Satellite Stations

In January 2007 when the sea ice was still accessible, we set up another mobile recording station, PALAOA-S. A RESON 4032 hydrophone was deployed through a hole in the sea ice which was about two meters thick at this location (Fig. 7). This hole was excavated with a chainsaw until the seawater flooded in and finished with a core drill. Because the sea ice was already receding and it was unpredictable when this ice floe would drift away, we decided to install the electronic box on top of the ice shelf, about 200 meters away and tied it securely. It contained a portable Microtrack 2496 (24 Bit, 96 kHz) recorder, powered by external batteries. It recorded the hydrophone along with GPS signals on the second channel. The 1pps signal, a better than microsecond precise pulse once every second was mixed by a small self made circuit with the serial NMEA data, containing ASCII timestamps and positions. As the main PALAOA station has the same capability, it is possible to synchronize the recordings later at sample precision for long baseline sound source localisation. In this way we could compensate temporarily for the lost hydrophones of PALAOA.

Figure 7: Single hydrophone deployment through sea ice.

Data Handling

PALAOA can generate data at an enormous rate. Two audio channels and the synchronization signal at 24 Bit and 192 kHz will produce 140 GB per day. We are not able to stream this amount continuously through the wireless link to Neumayer, instead there is a portable 500 GB USB disk attached locally, which has to be replaced by the radio officer periodically. However, single files of interest can be downloaded by FTP at any time. We record only sporadically at this quality, particularly when PALAOA-S is active too and the GPS channel is required at high precision. This mode is quite energy consuming as PC, MOTU sound interface, external hard disk and GPS consume together about 25 W. If such data needs to be transmitted online to Neumayer at a reasonable rate, a WLAN booster is required (Fig. 8). This increases the effective data rate of the wireless link from 500 kBit/s to 2 MBit/s, but requires additional 25 W. This setup can only be activated during austral summer when the solar cells are 24 hours operational. For the continuous streaming of audio we use a BARIX Instreamer device. It only requires 7 W and generates a stereo 16 Bit, 32 kHz MP3 stream at 192 kBit/s, which can be transferred via the WLAN without booster. Time stamping of the data happens at Neumayer Base, which allows only for about one second accuracy - which is not a problem as long as synchronizing with other audio sources is not required. An improved version of this device is announced and will allow to mark up the audio stream via a serial input with GPS data. It will have 24 Bit / 96 kHz capabilities and can buffer locally on an USB device in case of network congestion. We hope to deploy the improved device in January 2009. On a workstation within the Neumayer Base the audio stream is stored locally, cut down into one minute MP3 files. Thus there are 1440 files generated per day, each has a size of about 1.25 MB. Every day 1.7 gigabytes accumulate, per year up to 620 GB in half a million files. The autonomous webcam on top of the mast takes one 1280 x 960 Jpeg picture every minute, size ~300 kB, and transfers it via FTP to the Neumayer PC where it is stored along with the audio files, adding another 150 GB. In addition, we continuously collect oceanographic data from the CTD probe, temperatures and technical operating data of the meteorological readings from the Neumayer observatory, and network statistics from the satellite link. This is only 1 MB per day which is sent and logged via the standardized syslog protocol and online available at PALAOA, Neumayer, and in the Alfred Wegener Institute in Bremerhaven. Data is regularly backed-up on 200 GB LTO2 tapes which are shipped twice a year to Bremerhaven, in November when the first transport after the 8 month overwintering period is available and in February or March when the last ship leaves Neumayer just before access to the base is cut off from the rest of the world again. During the rest of the year, Neumayer is accessible only via a 128 kBit/s IntelSat link to the AWI in Bremerhaven, of which 24 kBit/s are assigned to the PALAOA project. We transmit a continuous OGG-Vorbis recompressed audio stream using the open source IceCast system and a small webcam picture every ten minutes. The data is presented and analyzed online in our lab and kept until the high quality from the tapes arrives. In total, we collected about 4 TB during the first two years of operation which are kept in our data silo in Bremerhaven with one petabyte capacity. There the data is held on LTO2 tapes in two redundant copies which are automatically loaded on demand by tape robots. To the network the system transparently appears as a simple network drive. It conforms to the requirements for reliable scientific data storage. The online audio stream and webcam pictures are publicly accessible via the internet (icecast.awi.de). Datasets are published with an open access license through the PANGAEA database (www.pangaea.de), and the World Data Centre for Marine Environmental Data, hosted by AWI. This satisfies the conditions of the Berlin Declaration on "Open Access to Knowledge in the Sciences and Humanities" as required by the AWI data policies.

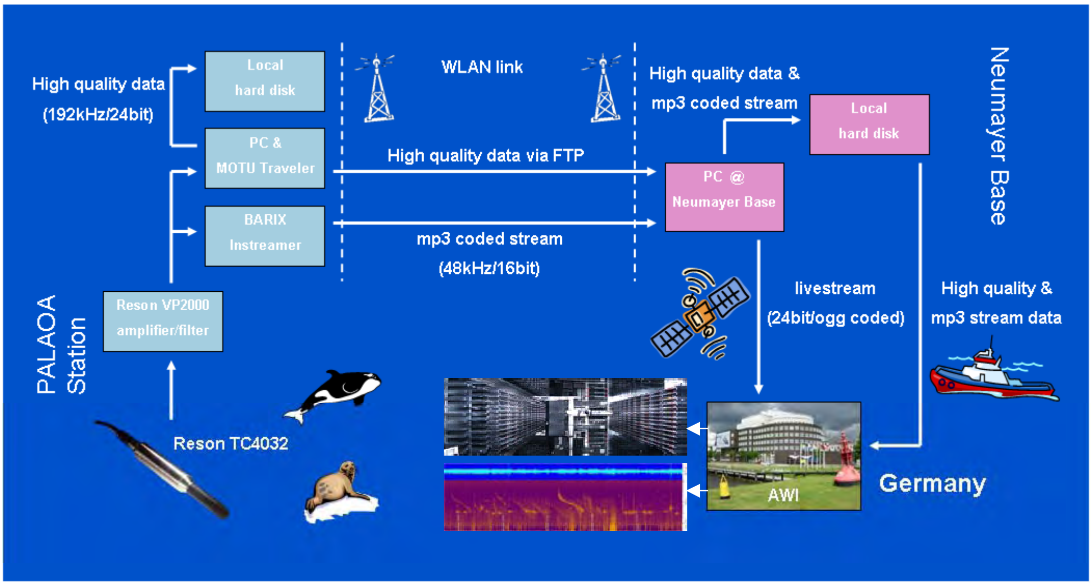


Figure 8: Audio data streams, from the water to long term archival and web broadcasting

Data Processing

All the PALAOA recordings consist of standard multimedia files (Wav, Flac, Mp3, Ogg, and Jpeg) and the additional metadata is stored in text files. As they are kept transparent on a network drive, it is possible to use standard software to easily access the dataset. However, the sheer amount of files makes it hard to analyze longer periods; no software can load a million sound files at once. We assembled an application in MATLAB, “PALAOAdb” to allow easy access to the dataset from a timeline or event oriented view. It periodically updates its database, by analyzing the recent online recordings to provide an up to date display. The initial view is a plot of several selectable parameters for the whole recording period, which is currently 2 years. Available are sound specific measures like RMS or peak sound level, external observations like air and water temperature or tidal current. Also the results of analyses like pattern recognition algorithms can be selected. One can zoom in and click to open single files, either with a built in player and spectrogram view or any external program or media player like XBAT, Ishmael, Audacity or Adobe Audition. PALAOAdb provides displays to visualize parameters and results from the whole data set (Fig. 9).

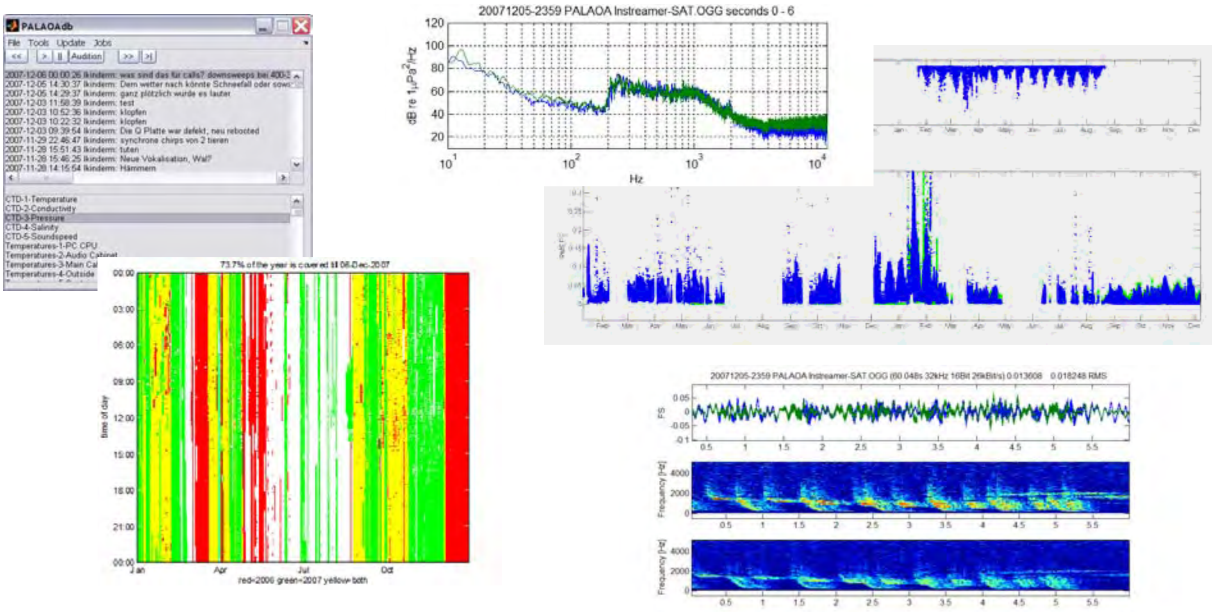


Figure 9: PALAOAdb software tool to handle long audio time series.

It takes about one second to load and process a single mp3 file with the basic procedures, additional modules will add to this. So an offline analysis can be done in up to 60 times real time speed. New algorithms have to be implemented as MATLAB functions that take a one-minute waveform array and a file information structure as arguments and deliver a structure containing the result. PALAOAdb will create timeline views from the results. To speed up processing of long-term periods like a whole year of recordings, we implemented a distributed computing system that allows sharing this task between multiple computers. PALAOAdb will compile a “worker” executable (no MATLAB licenses needed for the workers!) split the whole task into smaller jobs (defined by m-files) and just place everything into a network directory. All communication occurs via the file system: any computer in our institute network only needs to start the worker program and the computer will start to operate on a selection of the sound files. Finally, PALAOAdb collects all the results and generates a single structure array that can be accessed easily in the same way as performing all analyses on a single machine. As there is no interprocess communication between the workers, the speed scales linearly with the number of nodes as long as the network and the file server are not saturated by fetching the audio files. With a size of 1.35 MB per one-minute mp3 file, a 100 MBit network can handle up to 10 files per second, thus 500 times real time should be possible employing 10 PCs, scanning a year in less than a day. If the algorithm needs much longer than a second to analyze 60 seconds of data it will take proportionally more time.

First Results

Currently we host two years of recordings in the database, from December 2005 to December 2007 high quality files (192 kBit/s MP3 and/or 192 kHz, 24 Bit WAV), and since then compressed 24 kBit/s Ogg-Vorbis files. While we are presently developing automated pattern recognition modules for the PALAOAdb system, a first analysis of the data was done by hand, which is a necessary preparatory work also for evaluating the pattern recognition algorithms. We concentrated on seals in this phase as their vocalizations are within the human hearing range and can be analysed without further transformation of the recordings.

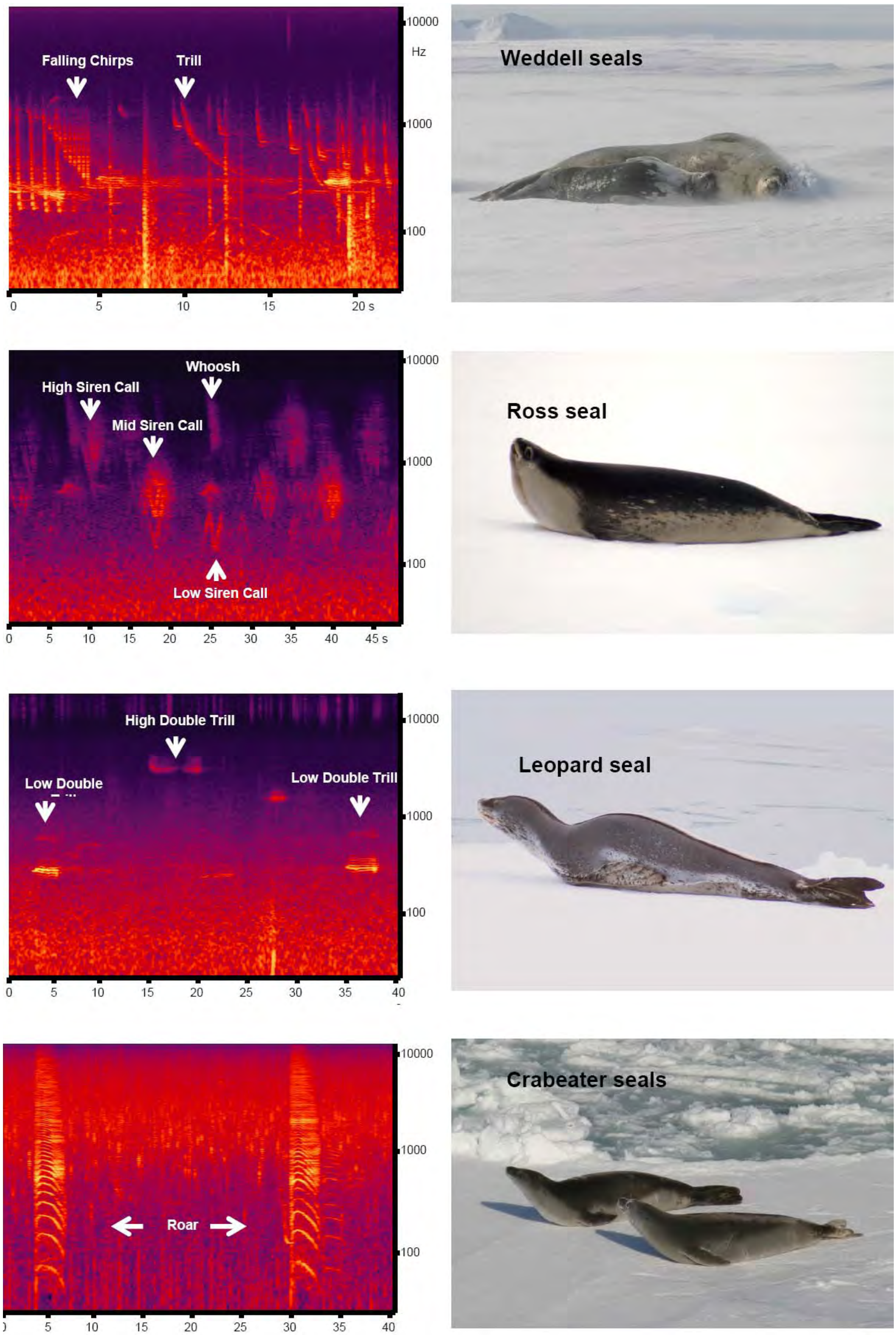


Figure 10: Spectrograms of different seal vocalizations

Seal Presence

Four species of seals are present in the vicinity of PALAOA, Weddell, Ross, leopard and crabeater seals. They can be distinguished by their underwater calls (Fig 10). Each of the species uses different sounds. A first manual scan through the recordings of 2006 shows the cycle of the presence of Weddell, Ross, crabeater and leopard seals (Fig. 11).

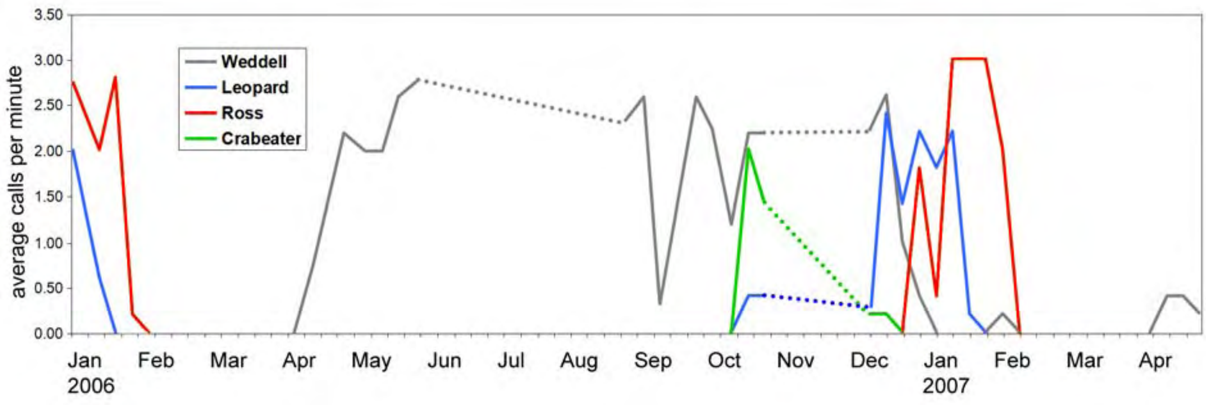
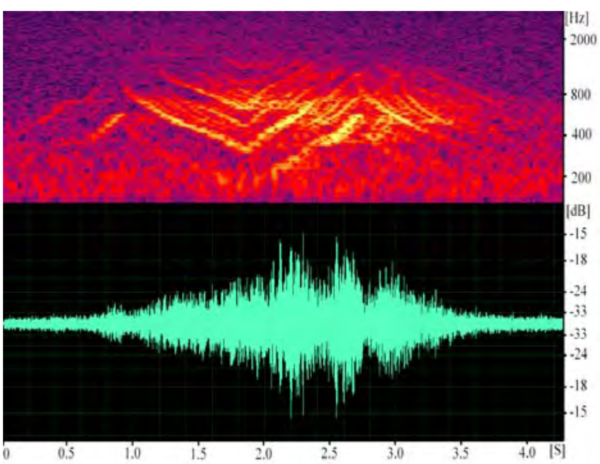


Figure 11: Seal calling activity during the year. Dotted lines bridge data gaps.

All four species show very different patterns in their acoustic presence. While it is not a priori clear that calling activity corresponds directly to the number of animals, at least recent data on the migratory behaviour of Ross seals, satellite tagged at a coastal site south-west of Atka Bay show, that their seasonal presence in this area coincides exactly with the detection of their calls.

Bioacoustics of Ross seals

The vocalizations of Ross seals, the rarest of all Antarctic pinnipeds, were analyzed for the first time in detail using the PALAOA recordings, revealing that their repertoire consists of three distinct siren like calls and a whoosh like sound. The four call types are clearly distinct from each other (Fig 12). However, the function of the different sounds remains unclear.



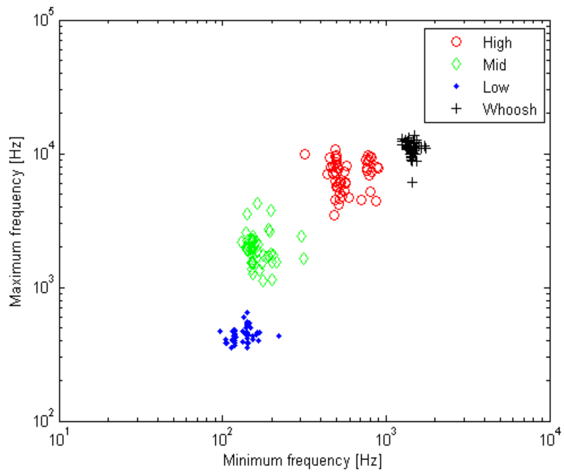


Figure 12: Spectrogram and waveform of the Ross seal high siren call (left), scatterplot of the frequency limits of all four calltypes (right).

Localization techniques allow to separate single animals from the recordings, a two-hydrophone system can reveal the bearing of an incoming sound while more hydrophones allow to detect the distance and position too (Fig. 13). By this method we are able to count the animals in acoustic range of PALAOA. In addition, individual movements of calling animals can be tracked, which enables investigation of spatial distribution of callers in relation to their vocal characteristics.

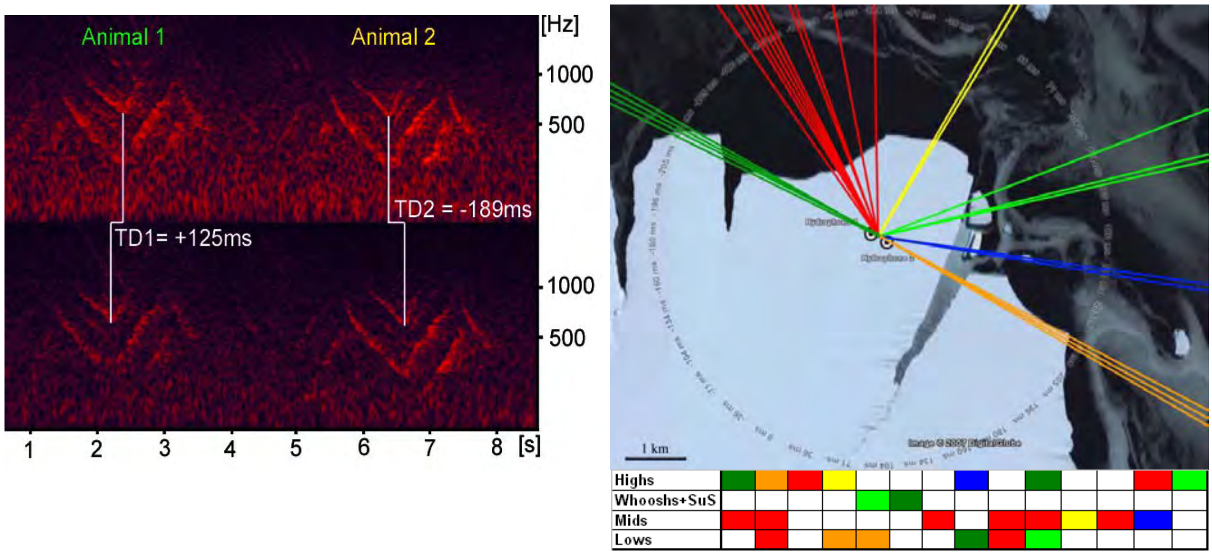


Figure 13: Localizing animals. Sounds arrive at different times at the two hydrophones (left). At least 6 or 7 individuals are vocalizing from different directions (right). The temporal structure of their "conversation" (bottom right).

Natural Sound Background

The loudest event we recorded with PALAOA so far occurred on March, 19th 2006 at 8:15. For more than 10 minutes the hydrophone input was completely overdriven, and we had to estimate the peak amplitudes from the steepness of the signal at zero crossing. From ENVISAT-ASAR satellite pictures we determined that two icebergs had collided about 20 km offshore. One of them stuck on the ground for some days while the other was approaching with the speed of the ocean current. The estimated source level of this event were well above 200 dB re 1 µPa @ 1 m and, with 10 minutes continuous duration, a sound exposure level of 228 dB SEL, this marks an iceberg collision as one of the loudest events in the oceans (Fig. 14).

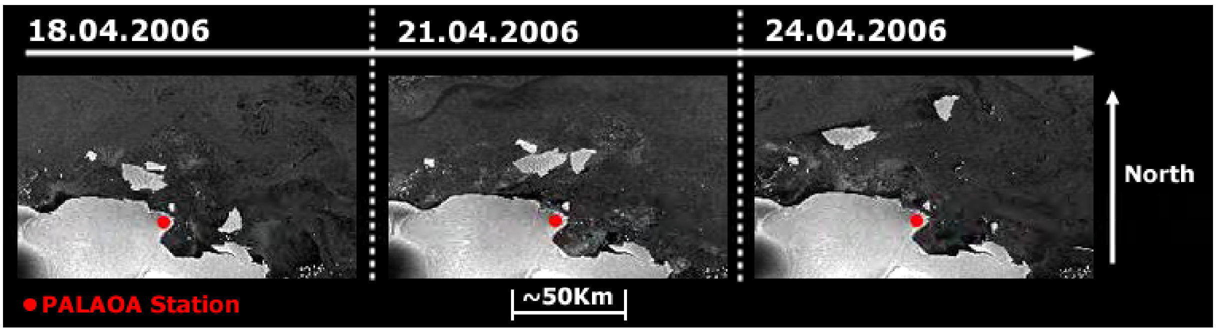


Figure 14: ENVISAT-ASAR images of an iceberg collision, provided by ESA. The large iceberg has a length of about 20 kilometers.

Perspectives

PALAOA is intended to work for several years to provide a long time series of acoustic data to look out for long term trends eventually induced by climate change. Also Atka Bay is an interesting place for studying the biology of marine endotherms. There is a local Weddell seal population of around 300 animals and an emperor penguin colony with 8000 breeding pairs, attracting predators like leopard seals and killer whales. Biologists from AWI plan to combine acoustic recordings from PALAOA with on site observations and satellite tagging programs to study the behaviour and ecology of these animals. To infer from passive acoustic recordings to real abundances of a species is a difficult task that also requires additional field work for gauging. We plan to deploy seals with acoustic tags to derive the vocal behaviour of diving individuals. Information on individual at-sea and haul-out activities, depending on time and weather and ice conditions is necessary to estimate the percentage of animals which are and which are not in the water at a given time. Then, from the calls recorded by PALAOA in a period, together with the calling rate of an individual and the percentage of animals that are in the water it will be possible to get estimates of the total number of animals in the PALAOA vicinity.

A problem in generalising the data is the question whether the location is typical for the ice shelf edge, or even for other parts of the Southern Ocean. In order to investigate this we are preparing to set up additional, smaller autonomous recorders at other locations for shorter periods to obtain acoustic recordings from other sites.

The future of PALAOA is thrilling. We intend to run the observatory for as long as possible to obtain the long time series which is necessary to detect changes and trends, both in the physical noise background and in the marine mammal presence which might be induced by changes in the Antarctic environment. However, while PALAOA is not standing on thin ice, its fate might well lie in the depth of the ocean. The break-off of the ice shelf is not predictable, and a journey on an iceberg around Antarctica might start any day. We took precautions to recover the PALAOA container in this case by equipping it with a GPS/Iridium phone, however, the chances of such a rescue operation are in the air.

References

ACKLEY S.F., BENGSTON J.L., BOVENG P., CASTELLINI M., DALY K.L., JACOBS S., KOOYMAN G.L., LAAKE J., QUETIN L., ROSS R., SINIFF D.B., STEWART B.S., STIRLING I., TORRES J., YOCHEM P.K. (2003): A top-down, multidisciplinary study of the structure and function of the pack-ice ecosystem in the eastern Ross Sea, Antarctica. Polar Record **39**: 219-230.

BESTER M.N., FERGUSON J.W.H. & JONKER F.C. (2002): Population densities of pack ice seals in the Lazarev Sea. Antarctica. Antarctic Science **14**: 123-127.

BLIX A.S. & NORDØY, E.S. (2007): Ross seal (Ommatophoca rossii) annual distribution, diving behaviour, breeding and moulting, off Queen Maud Land, Antarctica., Polar Biology **30**: 1449-1458.

BOEBEL O., KINDERMANN L., KLINCK H., BORNEMANN H., PLÖTZ J., STEINHAGE D., RIEDEL S. & BURKHARDT E. (2006): Acoustic Observatory Provides Real-Time Underwater Sounds from the Antarctic Ocean. EOS **87**: 361-372.

BOEBEL, O., KLINCK, H., KINDERMANN, L. & EL NAGGAR, S. (2008): PALAOA: Broadband recordings of the Antarctic coastal soundscape, Bioacoustics **17**: 19-21.

GREEN K. & BURTON H.R. (1988): Annual and diurnal variations in the underwater vocalizations of Weddell seals. Polar Biology **8**: 161-164.

HALL-MARTIN A.J. (1974): Observations on population density and species composition of seals in the King Haakon VII Sea, Antarctica. South African Journal of Antarctic Research **4**: 34-39.

KLINCK, H., KINDERMANN, L. & BOEBEL, O. (2007): Detection of Leopard seal (Hydrurga leptonyx) vocalizations using the Envelope-Spectrogram Technique (tEST) in combination with a Hidden Markov Model, Canadian Acoustics, in Print.

MIRHAJ M., PLÖTZ J., BORNEMANN H., KINDERMANN L., KLINCK H. & BOEBEL O. (2004): Underwater calls of Weddell seals in the Weddell sea, Antarctica. SCAR Open Science Conference “Antarctica and the Southern Ocean in the Global System”, 25-31 July 2004, Bremen, Germany. Terra Nostra, Schriften der Alfred-Wegener-Stiftung 2004/4 p 138.

ROUGET P.A., TERHUNE J.M. & BURTON H.R. (2007): Weddell Seal underwater calling rates during the winter and spring near Mawson Station, Antarctica. Marine Mammal Science **23**: 508-523.

SEIBERT, A.-M., KLINCK, H., KINDERMANN, L., BORNEMANN, H., VAN OPZEELAND, I., PLÖTZ, J. & BOEBEL, O. (2007): Characteristics of underwater calls of the Ross seal, 17th Biennial Conference on the Biology of Marine Mammals, 29 November - 3 December 2007, Cape Town, South Africa.

Probabilistic Evaluation of Synergetic Ultrasound Pattern Recognition for Large Scale Bat Surveys

MARTIN K. OBRIST[1], RUEDI BOESCH[1] & PETER FLÜCKIGER[2]

[1]Swiss Federal Research Institute WSL, Zürcherstrasse 111, CH-8903 Birmensdorf, Switzerland
[2]Bat Protection Kt SO. c/o Museum of Natural History Olten, Kirchgasse 10, CH-4600 Olten, Switzerland
E-mail of the corresponding author: martin.obrist@wsl.ch

Abstract. Vocalizations of bats have coevolved with their hearing and flight abilities to serve an orientational, autocommunicative purpose. Thus, echolocation is governed by physical rather than socio-behavioural constraints. Signal characteristics therefore show marginal communicative value and they can heavily overlap between species exploiting similar environmental niches for foraging. Nevertheless, recognition of species by their calls is possible to a large extent, even in species-rich environments as in Switzerland. Aware of the inter- and intra-specific variability of echolocation, we employ a multi-feature approach to acoustic species identification, combining repeated synergetic pattern recognition with parametric comparisons, to achieve a robust proposal of species identity to a supervising expert. We present applications on a landscape scale, results of the method in a case study, and an implementation on a national scale for the revision of the Swiss Red List of bats.

Introduction

Function and Limitations of Echolocation

Descending from nocturnal gliding insectivores (FENTON ET AL. 1995), bats have evolved an active mode of orientation, that allows them to orient, navigate and forage in the darkness of the night. During their 50 Mio years of evolution, types of vocalizations, hearing abilities, and flight characteristics have co-evolved to form in each species a unique adaptation to nocturnal feeding behaviour on the wing in a variety of habitats, and on different food types (NEUWEILER 1984).

Echolocation enables bats to accomplish tasks with comparable perfection as visually oriented animals, ranging from long-range target detection (HOLDERIED & VON HELVERSEN 2003) over the detection of water-waves indicative of fishes below the surface (SCHNITZLER ET AL. 1994) to the selective detection of an insects flutter (EMDE & SCHNITZLER 1990). Passive prey detection by their rustling or wing-beating adds to the variety of acoustic tasks performed by bats (RUSSO ET AL. 2007).

However, active acoustic orientation at ultrasound frequencies is hampered by a variety of physical constraints: quadratic spreading loss rapidly reduces available signal energy for echo generation. Adding to this, atmospheric attenuation affects signal energy progressively with higher signal frequencies (LAWRENCE & SIMMONS 1982), making echolocation in air basically a short-distance operation. As an active orientation, echolocating animals are vulnerable to eavesdropping, be it by their targeted prey or by potential predators (FENTON & RATCLIFFE 2004). The active mode can be prone to jamming (FULLARD ET AL. 1979; ULANOVSKY ET AL. 2004), and furthermore, it conveys information to conspecifics and foraging competitors (BALCOMBE & FENTON 1988).

Consequences for Recognition

Echolocation is not a fixed signal type of operation. Depending on orientational situation, bats can drastically adapt their calls: while foraging with long constant-frequency (CF) signal in open air on long ranges, the orientation along a cluttered forest edge necessitates the adoption of more broadband, shallowly frequency-modulated (FM-CF) signals, while during insect interception very short FM signals are used (GRIFFIN 1958). The active mode of orientation is a mayor clue to bat detection, and its communicative characteristics potentiates species recognition (AHLÉN 1980; FENTON & BELL 1981), but comparable niche occupation in different species, and behavioural variation (OBRIST 1995) in echolocation can seriously complicate this task.

Traditional acoustic field-identification of bats by their echolocation calls makes use of either heterodyning ultrasound detectors or countdown detectors. More recently, for offline-analysis, digital time-expansion detectors replace bulky and very expensive high-speed tape recorders (Parsons & Obrist 2004). Storage of recorded signals is a prerequisite for accurate species determination and the thorough evaluation of ambiguous signals. The step to digital signal management greatly helps in efficiently dealing with this kind of data. It speeds up data screening and extraction of temporal and spectral signal parameters, which later can be processed in statistical frameworks, artificial neural networks or similar decision finding algorithms, that identify echolocation calls to species (HERR ET AL. 1997; JONES ET AL. 2000; PARSONS & JONES 2000; PREATONI ET AL. 2005).

Aim

Echolocation studies that target temporal and spatial habitat use of bats quickly deal with a large amount of data: 1 s sampled at 500 kHz with 16 Bit translates to 1 MB/s of data; a single DVD is almost filled in one hour. Analysis of such amounts of data asks for automatisms, be it at the recording side, the analysing side, or both. We aimed at developing an efficient combination of hardware and software to sample echolocation signals for extended periods of time, preferably all night, and at multiple sites concurrently. Furthermore, ideally the data should be analysed to the species level in an automatic and unsupervised process and be applicable at the Swiss bat fauna, which comprises 30 species, covering a wide range of signal types (OBRIST ET AL. 2004a).

Material and Methods

Detectors and Recording

For automated acoustic bat monitoring, sophisticated bat detectors are not required for recording purposes. However, high sensitivity and linear frequency response curves of acoustic transducers are mandatory. We use condenser microphone capsules from Ultrasound Advice (London, UK), connected to an amplifier stage custom-built by the same company. The amplifier is able to drive the signals through more than 200 m of coaxial cable and holds two lead-acid gel cells for prolonged continuous operation.

Recording was initially performed with a PCMCIA data acquisition card (PCCARD-DAS16/330, Measurement Computing Corporation, Middleboro, MA, USA) in an Apple Macintosh PowerBook G3 laptop computer. Present developments of a ultrasound data logger allow for autonomous intelligent (self-triggering) recording stations, which are equipped with electret microphones (Knowles, Itasca, Ill, USA) and digitally store the data on SD memory cards.

Preprocessing

The recordings consist of continuous echolocation sequences of 5 to 20 sec duration in binary format, sampled at 312.5 kHz with 12 bit. Prior to analysis, the sequences are high-pass filtered at 7.5 kHz. Thus conditioned, a peak detection and centering routine extracts single echolocation call cut-outs, each of 8192 data points in length (26.21 ms duration). Of these, spectrograms are calculated, consisting of 159 spectra (0.17 ms resolution), which contain 128 spectral points (1.22 kHz resolution) and 81% overlap in data points (OBRIST ET AL. 2004b). All subsequent analyses are performed on these 20352-point spectrograms (Fig. 1).

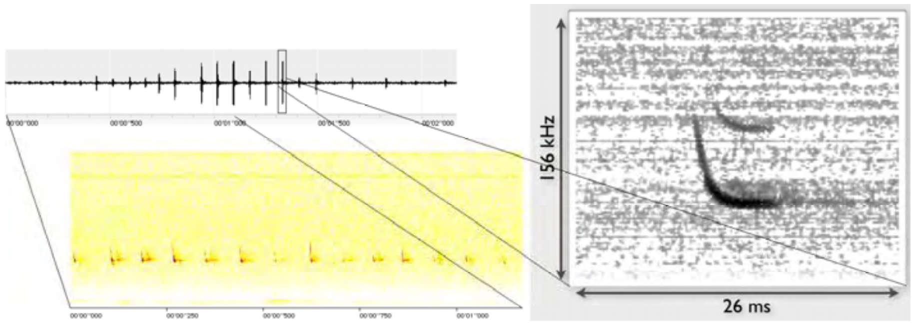


Figure 1: Cutting-out of single echolocation calls and subsequent spectrogram formation.

Synergetic Pattern Recognition

Synergetics is an interdisciplinary field which deals with self-organizational phenomena in nature (HAKEN 1978; KOHONEN 1984). These phenomena have in common that many microscopic parts in an unsorted order (chaos) transform themselves in a sorted order. The importance of each part is minor; only the properties of the whole system are relevant and can be described through synergetic differential equations.

The new set of algorithms emanating from this interdisciplinary field has only recently been used for classification tasks (HAKEN 1988, 1996; WAGNER ET AL. 1993; WAGNER ET AL. 1995). For the classification of bat calls we used an algorithm termed SC-MELT (DIECKMANN 1997). This algorithm combines several training patterns per class into one feature vector without losing any information about the training patterns. The training patterns are melted into one prototype. The prototype has the same dimension as the training vectors and is normalized to length 1. Because of this ability the algorithm can handle big dimensions in contrast to artificial neural networks (ANN). ANN also can handle big dimension, but the computational power needed to train an ANN with input vector of 16384 features is prohibitive. The employed algorithm is termed ‚synergetic computer using adjoint prototypes' (SCAP, HOGG & TALHAMI 1996; WAGNER ET AL. 1993). A most interesting property of the algorithm is its ability to emphasize pattern content that is unique among all others, at the same time neglecting pattern content common to all others. The learning time of the algorithm is very fast, taking only seconds on a 2.6 GHz Intel-Mac. The classification is even faster because it is simply a scalar or dot product.

Training and classifying can be described as follows: assume we train the algorithm with learn 4 calls of each of 3 classes (species), resulting in 3 prototype feature vectors. We now test 3 echolocation calls of species unknown, but contained in the training base. The algorithm computes the scalar product of each test call with each class prototype, resulting in 3 values per call, varying between 1 (identical call as in training base) and 0 (no resemblance to any of the training calls). The training class with the highest scalar product identifies the species where the calls most likely came from (circled in Fig. 2).

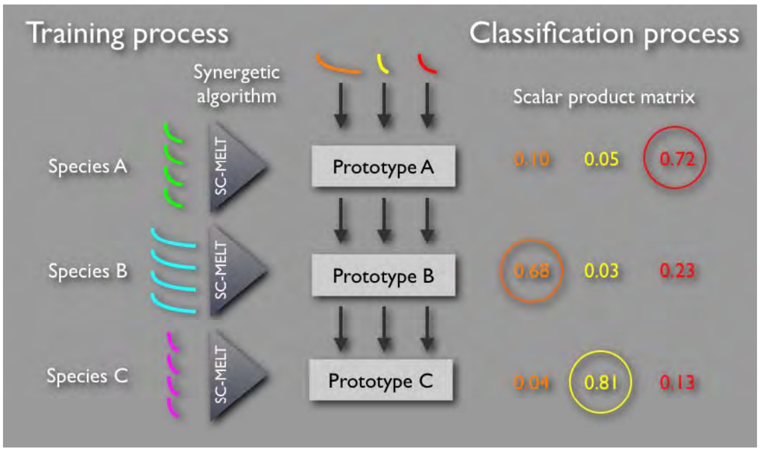


Figure 2: Learning and classifying echolocation calls with the synergetic algorithm.

Parameter Extraction

To also achieve parametric values for each extracted echolocation call, we calculate the summations of the spectrogram over the temporal and spectral dimensions, thus obtaining a power spectrum and an amplitude representation respectively. In these arrays, we search for peaks (frequency of peak energy Fpeak, time of peak amplitude Tpeak) and spectral and temporal values at -6, -12 and -18 dB below the peak, thus achieving corresponding values for maximum frequency (Fmax), minimum frequency (Fmin), start time (Tstart) and end time (Tend). From these we finally derive bandwidth (Fmax - Fmin) and duration (Tend - Tstart) of the signal at corresponding intensities below the peak (Fig. 3).

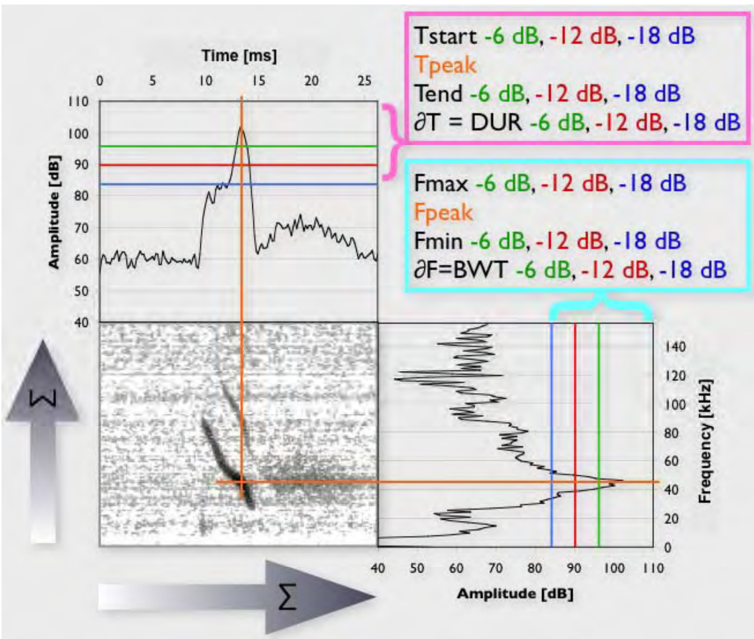


Figure 3: Calculation of spectral and temporal parameters from the spectrogram (see Fig. 1).

Field Recording

Field recording and analysis initially was restricted by processor speed (PPC G3) and hard disk capacities. Thus, analysis was shifted to offline processing. During monitoring 10 s blocks were recorded, saved, filtered and preprocessed in terms of peak detection. If four or more peaks where detected in the 10 s sequence, the file was kept otherwise erased and in either case correspondingly noted in a log-file. Subsequent recordings happened sequentially and circularly from additional microphones connected to the computers data acquisition card. Thus storage space could be saved for retaining the raw-data, at the cost of duty-cycle, as the processing used about 20 s per file, resulting in a stations duty-cycle of 30% divided by the number of connected channels.

In later inventories, data loggers were employed. To save memory space on the 2 GB SD-cards, small sound sequences of 5 s duration are only recorded real-time, after a trigger event occurred. In a first stage, a real-time amplitude peak detector triggers a more time consuming (40 ms) second stage, an FFT and spectral peak detection. Only these then triggered the recording, thus increasing the duty-cycle to >99% but incorporating a slight delay in recording start, which equals roughly to one missed echolocation call.

Optimization of Classification Methodology

Signal preprocessing and synergetic pattern recognition as outlined in the previous sections was carried out offline in the lab.

In a pilot study, classification of bats from a training set of 12 known species resulted in a high correct classification rate of the synergetic algorithms alone. For training purposes varying sets of training calls per species were used. In average we reached 45% correct classifications when including 5 training calls per species (Figure 4). Increasing the number of training calls to 9 improved this value considerably. As a classification always consists of a frequency distribution of a calls matches (scalar product S) to all prototypes (classes) included in the training base, we could further select only those classifications, that reached in those histograms a critical value of Smin > 0.5 and exhibited a good separation from other species ∂S > 0.2. Thus we finally rejected 33% of signals to be classified as unqualified, at the same time increasing average classification rates to 80% (see Table 64.2 in OBRIST ET AL. 2004b).

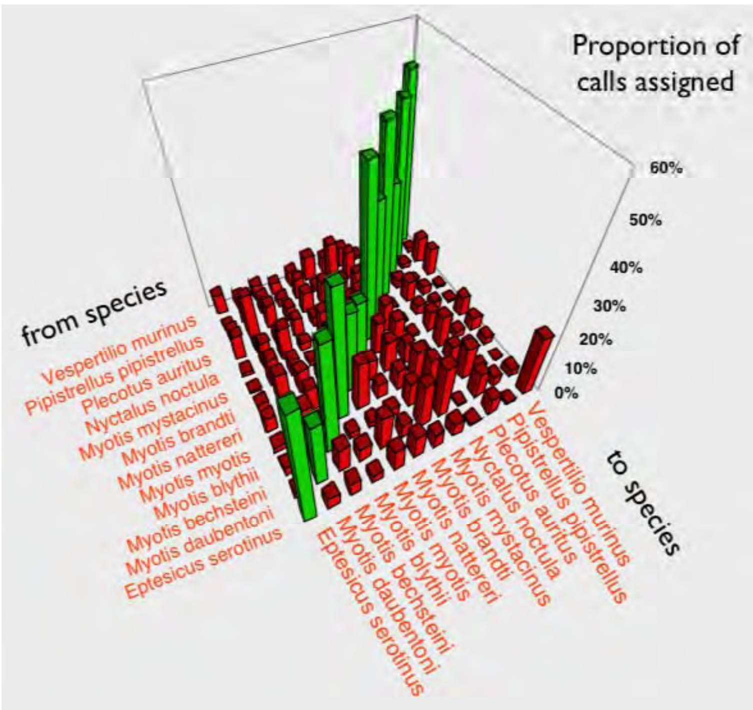


Figure 4: Classification rates with 12 species in training set, 5 training calls of each. Misclassifications become obvious e.g. in the genus *Myotis* or between *Eptesicus serotinus* and *Vespertilio murinus*.

To be able to recognize all Swiss bat species, we performed an extensive collection of reference signals from animals of known species affiliation, recording 26 species from Southern Germany to Northern Italy between 1999 and 2000. Thus the only recently described species *Plecotus macrobullaris* (KIEFER & VEITH 2001; SPITZENBERGER ET AL. 2003) and *Myothis alcathoe* (VON HELVERSEN ET AL. 2001) are not included. The same holds true for *Nyctalus lasiopterus*, which is extremely rare in Switzerland.

When including calls of all 26 species in the training of the synergetic algorithm, rates of correct classification dropped considerably to 55%. Increasing the number of training signals lead to mixed success: above 10 training calls per species the classification rates only marginally increase and start to drop again beyond 30 training calls. Furthermore, the previously evaluated delimiting values for ∂S and Smin showed negative effects on the classification rates. After numerous iterations we settled on a very conservative delimiting quality measure of ∂S > 0.05 and dropped Smin as a condition altogether.

Bat species show a considerable variability in their echolocation calls: depending on environment and behavioural situation, calls may change in duration, frequency content and emphasis (OBRIST 1995). To decrease variation and thus accuracy of matching in our training classes, we divided echolocation calls types of every species in sub-classes depending on duration and frequency structure, thus achieving more homogeneous training patterns (Fig. 5), but increasing their number from 26 to 85.

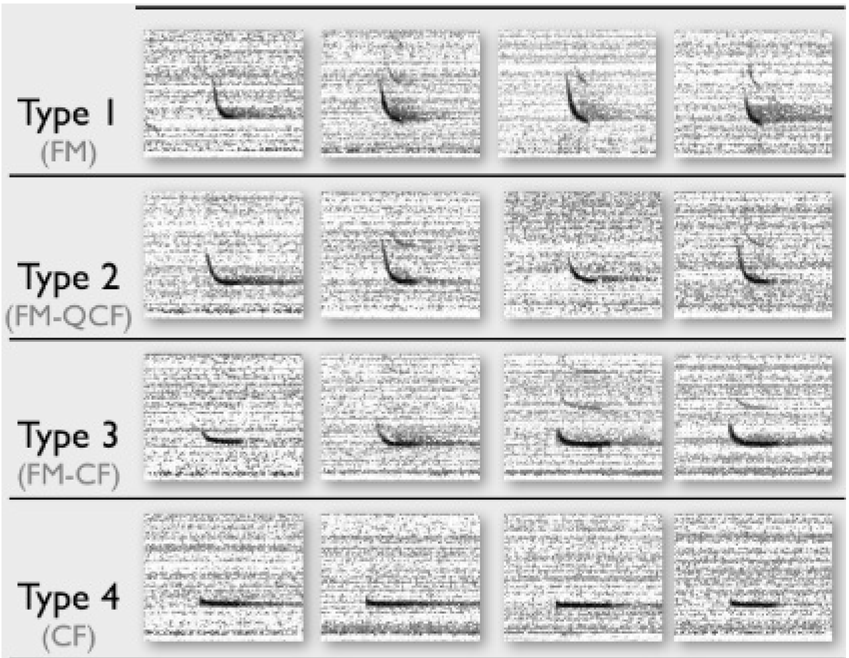


Figure 5: Dividing echolocation calls of *Pipistrellus pipistrellus* into four distinct type classes; frequency modulated (FM), frequency modulated – quasi constant frequency (FM-QCF), frequency modulated – constant frequency (FM-CF), and constant frequency (CF) type.

Probabilistic Evaluation

To train a synergetic algorithm well balanced, the number of patterns per class have to be identical. As we have between 20 and 1393 calls per type class, we randomly selected N calls per type class to calculate a training base. We then repeated each random selection 99 times for every N from 2 to 15, achieving 1386 training bases. We tested each of these against all remaining calls not considered in the respective training. The 5 training bases performing best where finally chosen as reference for further species identification.

Comparing species assignment between the 5 best training bases still revealed differences. Thus, we decided to test every signal of unknown provenance five times, and focus on the joint outcome. At least 4 of the 5 classifications had to result in a $\partial S > 0.05$, and at least 3 of the 5 classifications had to point to the same species. Subdividing species' calls in sub-types combined with the multi-classification and the prerequisites increased average classification rates but still proved dissatisfactory.

When plotting temporal and spectral parameters of correctly and erroneously classified signals, parameter ranges could be identified, which separated the two classes (Fig. 6).

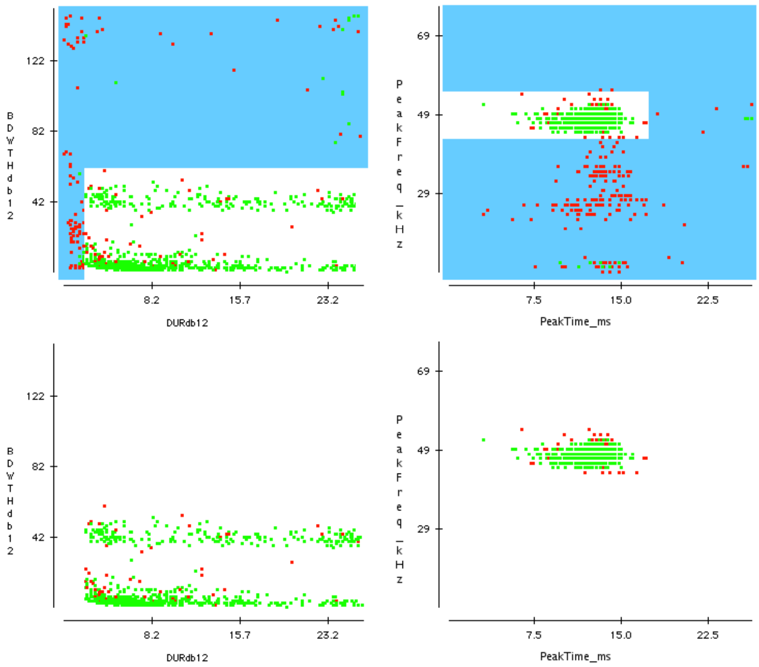


Figure 6: Call type specific restriction of temporal and spectral parameters. Correctly classified call types (green), erroneously classified call types (red) and parameter ranges for rejection (blue) are illustrated for calls of *Pipistrellus pipistrellus* and two arbitrary parameter combinations (see also Fig. 3).

Rejecting the parameter ranges underlined blue (Figure 6) drastically decreased the number of erroneously classified call types (red) as it disproportionately favors correctly classified call types (green). In a ultimate refinement step, we can confidently judge species when considering the many calls classified to species, that occur in each recorded echolocation sequence. Species of the genus *Myotis* are correctly identified in 74% of the cases, other species in 92% of the cases (Average 86%, see Table 3 in OBRIST ET AL. 2004a).

Still the recognition process is not fully automated and asks for expert interaction after the calculations, mainly to screen existing ambiguous sequences visually. Few identified calls or many different species per sequence are indicative of such cases and advise a check for plausibility.

Applications and Results

Cross-Validation

After development of the recognition method, its applicability was tested with cross-validating studies, prior to implementing it in large scale inventories. During 9 nights a total of 17 locations were sampled with mist nets, our computers and with bat detectors to compare the monitoring outcome of the different methods (Fig. 7).

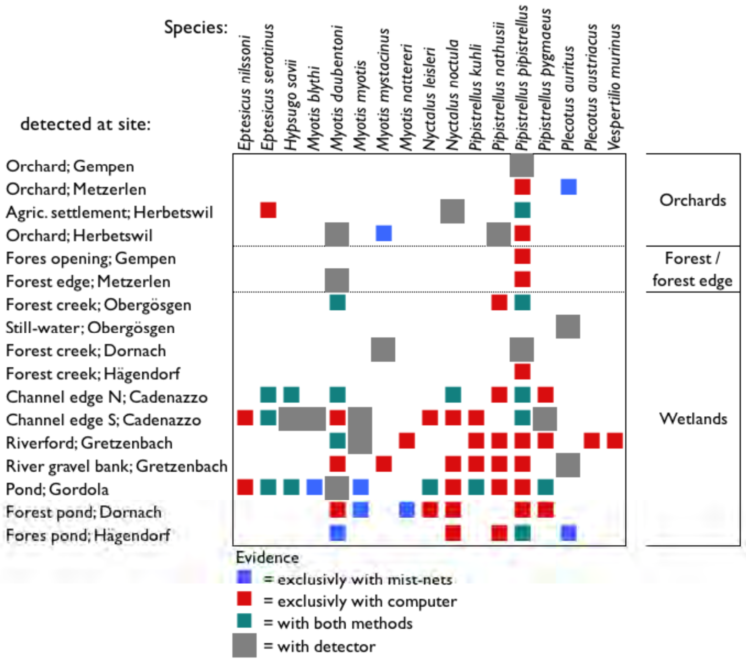


Figure 7: Species registrations with different methods and different habitat types during cross-validation of automated recognition method.

The computer recordings clearly detected most evidences of species occurrences in the diverse habitat types. *Plecotus sp*. and some *Myotis species* where the only ones, for which mist nets proved adequate too. Due to their rarity and especially low intensity echolocation calls (*Plecotus sp.*) the acoustic methods may be less superior for these species. Overall the computer detected 14 out of 17 registered species, mist nets 13 and the acoustic monitoring 7. The outcome encouraged us to focus in future monitoring projects on the automated recording and analysing with computers, all the more as running mist nets necessitates costly constant human attendance.

Habitat Requirements of Bat Species Assemblages

To assess the habitat requirements of bats, we designed a survey of bat activity in habitat types of value to nature conservation, as lakeshores, riverbanks, floodplains, forests, forest edges and old orchards but contrasted them to agricultural areas and settlements. The inventory design encompassed simultaneous sampling with to 2 computers, each connected to 5 microphones positioned in different habitat types. Thus we aimed at the ability to concurrently compare activity e.g. on a transect from a riverbank through a floodplain into a forest and its edge. In nine regions of Northern Switzerland, the recordings where repeated two to three times at the same microphone locations during different seasons of the year, totalling to 55 microphone locations with 150 microphone nights recorded.

Bats predominantly used two types of habitat, water related and settlements (Fig. 8).

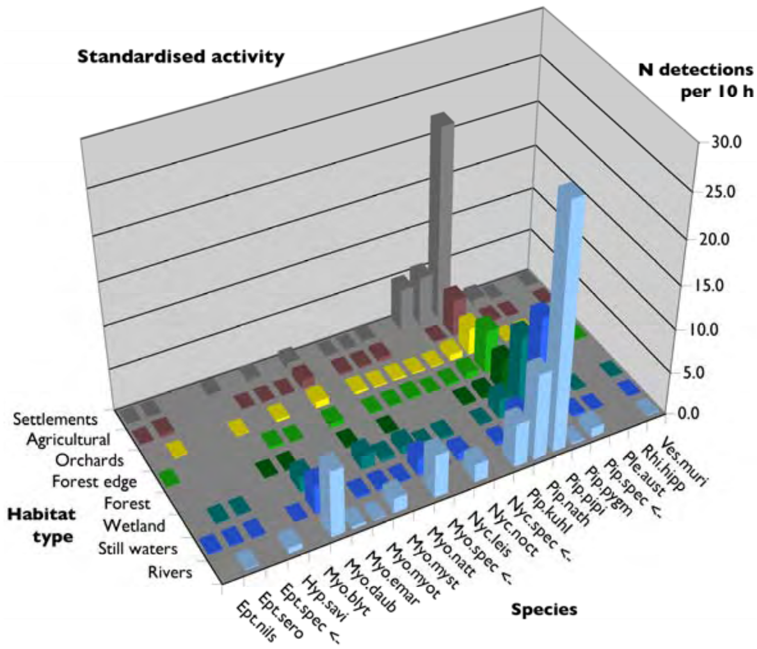


Figure 8: Species specific habitat use. Activity is standardized in number of detections per 10 hours of observation. Species names are abbreviated.

Some species clearly segregate in favored environment like *Myotis daubentonii*, which almost exclusively is found in water related sites, *Nyctalus noctula* mainly over water. While *Pipistrellus pygmaeus* is found preliminary in wetlands, *P. pipistrellus* shows a preference for water related sites and settlements, but it is found troughout and proved to be the most common species registered in almost all microphone sites.

The sampling design of this study, which included repeated sampling of the same locations, allowed to calculate occupancy rates of sites and the detection probabilities of species. Only recently new models where developed to account for imperfect detection probability in site occupancy models (MACKENZIE ET AL. 2002, 2003, 2004; DORAZIO & ROYLE 2005; MACKENZIE & ROYLE 2005). Once these values are known for a given species, different occupancy study designs can be evaluated re. their power and error rates depending on the number of repetitions and/or sites (MACKENZIE & ROYLE 2005). We calculated the necessary values with the help of the software PRESENCE (http://www.mbr-pwrc.usgs.gov/software/presence.html) and evaluated several survey designs with the software GENPRES (http://www.mbr-pwrc.usgs.gov/software/presence.html) to propose a design for the revision of the Red List of bats in Switzerland (OBRIST & BONTADINA in prep). See below.

Response of Bat Diversity to Change in Forest Management

In the recent past in Switzerland, intensification of agricultural and forestry practices coupled with human depopulation of rural areas lead to abandonment of traditional pastoral and forestry practices. In this context, we designed a study in Southern Switzerland which aimed at assessing the effect of the abandonment of chestnut orchards (*Castanea sativa*) on bat species richness and foraging activities to find practical recommendations for bat conservation (RATHEY ET AL. in prep).

We sampled bats in traditionally managed and unmanaged chestnut orchards with acoustic surveys, thereby testing the hypothesis, that management would open up access to more, and less clutter resistant bat species. The study took place in 32 pairs of managed and unmanaged chestnut orchards (64 sites in total) on the Southern slope of the Swiss Alps.

Comparable to the study mentioned before, we surveyed bats by ultrasound registrations during 32 nights with two computer sets, each with 4 microphones connected, recording simultaneously in a pair of managed and unmanaged orchards of close proximity. Vegetation structure (Fig. 9) was characterized at the places where microphones were installed, and food availability (nocturnal aerial insects) was assessed during recording nights using non-directional light traps.

We found twice the number of bat species (12 species) and a fourfold foraging activity (530 ultrasound sequences) in the managed chestnut orchards compared to the unmanaged ones (6 species, and 132 sequences). Within the managed habitat, bats visited only the most open orchards, free of undergrowth. Bat species of low flight manoeuvrability, although rarely detected at all, were 14 times more common in managed than unmanaged orchards. Bats of medium to high manoeuvrability resisted clutter better and preferred managed only 5 times over the unmanaged orchards. The stand structure in the managed orchards was significantly different from that in unmanaged ones, the latter being denser and closer grown over with young shoots (≈ 4 cm) than the former.

Figure 9: Illustration of the difference in overgrowth of the understory in managed (top) and abandoned (bottom) traditional chestnut orchards.

Management showed no affect on food availability (i.e. aerial insect number and biomass). The study demonstrates that the abandonment of chestnut orchards leads to a decline in bat species richness and foraging activities, due to restricted access into the overgrown forests. Restoring and keeping chestnut orchards open thus maintains endangered bat species in the Southern Alps.

Red List Inventory

The pending update of the Swiss red list of endangered species for the group Chiroptera (SCHWEIZERISCHE KOORDINATIONSSTELLEN FÜR FLEDERMAUSSCHUTZ OST UND WEST ET AL. 1994) required a scientific based and sound procedure to collect reliable data on species occurrence. It is obvious that a nation-wide evaluation of a large group of elusive and secret mammals as bats will need a huge effort, and because the resources are limited, an optimization of sampling design is needed. Thus, we could profit from the habitat assessment study presented above, to evaluate a sampling layout, that promised a required accuracy, and was temporally and economically realistic (OBRIST & BONTADINA in prep).

Presently our proposed sampling design is implemented in the field: 100 squares of 1 km2 each are surveyed four times between 2007 and 2009, twice in early, twice in late summer. In each square the observers identify and describe (botanically and structurally) 10 locations, which they monitor for 15 min in every survey. Observers are equipped with digital bat detectors, that allow them to protocol acoustically identified species on the spot, or store short sequences of doubtful origin in slowed down mode to a digital recorder. Additionally, each observer is equipped with a second ultrasound microphone and a digital ultrasound data-logger, which autonomously triggers on and stores ultrasound signals. A GPS-logger completes the apparatus for the field operators. Catching bats with mist-nets complements the acoustic methods.

Results from the first year are ready for analysis and prove the feasibility of the ambitious approach. We are confident to be able to calculate the proportion of area occupied (PAO according to MACKENZIE ET AL. 2002) for most species. The PAO can then be used to calculate a valid Red List criterion, the change in the area of occupancy (IUCN criteria for red lists B, IUCN 2001) according to the procedure applied in Amphibians (MACKENZIE ET AL. 2002; PELLET & SCHMIDT 2005).

Discussion

Quality and Taxonomy

Several recent publications tackle acoustic bat species recognition with a variety of methods, statistics, artificial neural networks or decision trees (HERR ET AL. 1997; OBRIST ET AL. 2004b; PARSONS & JONES 2000; RUSSO & JONES 2002), all presenting good performance but on incomparable data sets. However, some species' groups invariably prove very difficult to identify by their calls, e.g. the genus *Myotis* emits very similar signals, *Pipistrellus nathusii* and *Pipistrellus kuhlii* often overlap in their signal types and worse still, with *Eptesicus serotinus*, *Nyctalus leisleri*, and *Vespertilio murinus* three genera may be mixed. For conservation purposes, the quality of the recognition must not be measured in terms of % accuracy but with 100% accuracy and in variable terms of species or species groups assignment. Thus, in such cases high classification accuracy may only be reached with a coarser taxonomic level, e.g. a species or even a genus complex. After extensive work with bat species identification by their calls, we come to the conclusion, that acoustic identification of all species in species-rich bat assemblages likely can not be achieved with a single method whatsoever, but it will gain from a probabilistic approach, which combines candidate methods as exemplified in this paper.

Merits and Pitfalls of Automated Methods

Present digital recording techniques allow for an autonomous acoustic monitoring in large scale bat surveys. Automated recording of bats features many advantages over acoustic monitoring in the field with bat-detectors: recordings are far less personel-intensive, potential disturbance of foraging bats by observers becomes negligible, ideally a duty-cycle of close to 100% can be achieved, data are immediately stored digitally, and archiving and later offline-analysis with a variety of methods becomes available. Thus they better account for the scientific requirement of reproducibility. Synergetic pattern recognition is one of these analysis methods, which we apply successfully to large data-sets. In combination with supplementary methods it allows for the recognition of the majority of the Swiss bat species. However, expert knowledge and experience is still highly advisable to omit mistaking interpretations, when checking species identities proposed by automatic algorithms. Furthermore, at the recording end, pitfalls linger, like inadvertent recording of insects or setting of automatic trigger conditions inappropriate to the species in focus.

To conserve the value of long term collections of digital recordings of ultrasound signals from bats, it is also necessary to demand standards for the journaling of meta-information of recordings, which surpass self-evident parameters like time, location and observer. Values like type of microphones used, filter settings, noise-levels, possibly frequency response and if not already contained e.g. in AIFF-data, sampling rate and bit depths among others. Correspondingly stored acoustic data will become more universally explorable and will increasingly gain value for biodiversity conservation.

Acknowledgement

The projects mentioned are supported by the Canton of Solothurn, the Federal Office for the Environment (FOEN) and the Swiss Bat Protection Foundation (SSF). We thank the many people who helped us catch and record bats, the Forester Service of the Canton Ticino and the many private land owners who allowed us to access their property.

References

AHLÉN, I. (1980): Field identification of bats and survey methods based on sounds. - Myotis **18-19**: 128-136.

BALCOMBE, J. & M. B. FENTON. (1988): Eavesdropping by bats: the influence of echolocation call design and foraging strategies. - Ethology **79**: 158-166.

DIECKMANN, U. (1997): SESAM: A biometric person identification system using sensor fusion. - Pattern Recognition Letters 18:827-833.

EMDE, G. V. D. & H.-U. SCHNITZLER (1990): Classification of insects by echolocating greater horseshoe bats. Journal of Comparative Physiology A **167**: 423-430.

FENTON, B., & J. RATCLIFFE (2004): Animal behaviour - Eavesdropping on bats. - Nature **429**: 612-613.

FENTON, M. B., D. AUDET, M. K. OBRIST & J. RYDELL (1995): Signal strength, timing, and self-deafening: The evolution of echolocation in bats. Paleobiology **21**: 229-242.

FENTON, M. B. & G. P. BELL (1981): Recognition of species of insectivorous bats by their echolocation calls. Journal of Mammalogy **62**: 233-243.

FULLARD, J. H., M. B. FENTON &J. A. SIMMONS (1979): Jamming bat echolocation: the click of arctiid moths. - Canadian Journal of Zoology **57**: 647-649.

GRIFFIN, D. R. (1958): Listening in the Dark. The Acoustic Orientation of Bats and Men. Yale University Press, New Haven. (1986 reprint by Cornell University Press, Ithaca, New York).

HAKEN, H. (1978): Synergetics. An introduction. Nonequilibrum phase transitions and self-organization in physics, chemistry and biology. Springer, Berlin, Heidelberg, New York.

HAKEN, H. (1988): Learning in synergetic systems for pattern recognition and associative action. Zeitschrift für Physik B **71**: 521-526.

HAKEN, H. (1996): Future trends in synergetics. Nonlinear physics of complex systems. Current status and future trends. Springer, Berlin.

HERR, A., N. I. KLOMP, & J. S. ATKINSON (1997): Identification of bat echolocation calls using a decision tree classification system. Complexity International. Charles Sturt University, Australia.

HOGG, T., & I. TALHAMI (1996): A competitive non-linear approach to object recognition: the generalised synergetic algorithm. Pages 47-50. Australian New Zealand Conference on Intelligent Information Systems. IEEE, New York, NY, USA, Adelaide, South Australia.

HOLDERIED, M. W., & O. VON HELVERSEN (2003): Echolocation range and wingbeat period match in aerial-hawking bats. - Proceedings of the Royal Society of London Series B-Biological Sciences **270**: 2293-2299.

IUCN (2001): IUCN Red List Categories and Criteria: Version 3.1. Page ii + 30pp. IUCN Species Survival Commission, IUCN, Gland, Switzerland and Cambridge, UK.

JONES, G., N. VAUGHAN, & S. PARSONS (2000): Acoustic identification of bats from directly sampled and time expanded recordings of vocalizations. - Acta Chiropterologica **2**: 155-170.

KIEFER, A., & M. VEITH (2001): A new species of long-eared bat from Europe (Chiroptera: Vespertilionidae). Myotis **39**: 5-16.

KOHONEN, T. (1984): Self-organization and assoziative memory. Springer-Verlag, Berlin.

LAWRENCE, B. D. & J. A. SIMMONS (1982): Measurements of atmospheric attenuation at ultrasonic frequencies and the significance for echolocating bats. The Journal of the Acoustical Society of America **71**: 585-590.

MACKENZIE, D. I., J. D. NICHOLS, G. B. LACHMAN, S. DROEGE, J. A. ROYLE & C. A. LANGTIMM (2002): Estimating site occupancy rates when detection probabilities are less than one. - Ecology **83**: 2248-2255.

MACKENZIE, D. I. & J. A. ROYLE (2005): Designing occupancy studies: general advice and allocating survey effort. Journal of Applied Ecology **42**: 1105-1114.

NEUWEILER, G. (1984): Foraging, Echolocation and Audition in Bats. - Naturwissenschaften **71**: 446-455.

OBRIST, M. K. (1995): Flexible bat echolocation: the influence of individual, habitat and conspecifics on sonar signal design. - Behavioral Ecology and Sociobiology **36**: 207-219.

OBRIST, M. K., R. BOESCH & P. F. FLÜCKIGER (2004a): Variability in echolocation call design of 26 Swiss bat species: consequences, limits and options for automated field identification with a synergetic pattern recognition approach. - Mammalia **68**: 307-322.

OBRIST, M. K., R. BOESCH, P. F. FLÜCKIGER & U. DIECKMANN (2004b): Who's calling? Acoustic bat species identification revised with synergetics. - In J. Thomas, C. Moss, and M. Vater (editors). Echolocation in Bats and Dolphins, Proceedings of the Biosonar Conference 1998. University of Chicago Press, Chicago, 484-491.

OBRIST, M. K. & F. BONTADINA (in prep.): Estimation of bat species detectability and site occupancy: towards an optimal sampling design for the Swiss red list bat survey.

PARSONS, S. & G. JONES (2000): Acoustic identification of twelve species of echolocating bat by discriminant function analysis and artificial neural networks. - Journal of Experimental Biology **203**: 2641-2656.

PARSONS, S. & M. K. OBRIST (2004): Recent methodological advances in the recording and analysis of chiropteran biosonar signals in the field. - In J. Thomas, C. Moss, and M. Vater (editors). Echolocation in Bats and Dolphins, Proceedings of the Biosonar Conference 1998. University of Chicago Press, Chicago, 468-477.

PELLET, J. & B. R. SCHMIDT (2005): Monitoring distributions using call surveys: estimating site occupancy, detection probabilities and inferring absence. - Biological Conservation **123**: 27-35.

PREATONI, D. G., M. NODARI, R. CHIRICHELLA, G. TOSI, L. A. WAUTERS & A. MARTINOLI (2005): Identifying bats from time-expanded recordings of search calls: Comparing classification methods. - Journal of Wildlife Management **69**: 1601-1614.

RATHEY, E., P. CHRISTE, M. K. OBRIST, F. BONTADINA, A. MARTINOLI, M. CONEDERA & M. MORETTI (in prep): Response of bat diversity to change in forest management.

RUSSO, D. & G. JONES (2002): Identification of twenty-two bat species (Mammalia: Chrioptera) from Italy by analysis of time-expanded recordings of echolocation calls. - Journal of Zoology, London **258**: 91-103.

RUSSO, D., G. JONES & R. ARLETTAZ (2007): Echolocation and passive listening by foraging mouse-eared bats *Myotis myotis* and *M-blythii*. - Journal of Experimental Biology **210**: 166-176.

SCHNITZLER, H. U., E. K. V. KALKO, I. KAIPF & A. D. GRINNELL (1994): Fishing and Echolocation Behavior of the Greater Bulldog Bat, Noctilio leporinus, in the Field. - Behavioral Ecology and Sociobiology **35**: 327-345.

SCHWEIZERISCHE KOORDINATIONSSTELLEN FÜR FLEDERMAUSSCHUTZ OST UND WEST, E. A. D. MAUR, R. ARLETTAZ, J. BARANDUN, A. BECK, W. D. BURKHARD, P. F. FLÜCKIGER, J. GEBHARD, S. GLOOR, M. GRAF, R. GÜTTINGER, M. HAFFNER, A. KELLER, R. KISTLER & M. LUTZ (1994): ROTE Liste der gefährdeten Fledermäuse der Schweiz. – In: W. u. L. Bundesamt für Umwelt (ed.). Rote Liste der gefährdeten Tierarten in der Schweiz. Eidgenössische Drucksachen- und Materialzentrale (EDMZ), Bern: 22-23.

SPITZENBERGER, F., P. STRELKOV & E. HARING (2003): Morphology and mitochondrial DNA sequencies show that Plecotus alpinus Kiefer & Veith, 2002 and Plecotus microdontus Spitzenberger, 2002 are synonyms of Plecotus macrobullaris Kuzjakin, 1965. - Natura Croatica **12**: 39-53.

ULANOVSKY, N., M. B. FENTON, A. TSOAR & C. KORINE (2004): Dynamics of jamming avoidance in echolocating bats. Proceedings of the Royal Society of London Series B - Biological Sciences **271**: 1467-1475.

VON HELVERSEN, O., K. G. HELLER, F. MAYER, A. NEMETH, M. VOLLETH & P. GOMBKOTO (2001): Cryptic mammalian species: a new species of whiskered bat (*Myotis alcathoe n sp*) in Europe. - Naturwissenschaften **88**: 217-223.

WAGNER, T., F. G. BOEBEL, U. HASSLER, H. HAKEN & D. SEITZER (1993): Using a synergetic computer in an industrial classification problem. - In R. F. Albrecht, C. R. Reeves, and N. C. Steele, editors. International Conference on Artificial Neural Nets and Genetic Algorithms. Springer-Verlag, Berlin, Innsbruck, Austria, 206-212.

WAGNER, T., U. SCHRAMM & F. G. BOEBEL (1995): Synergetic learning for unsupervised texture classification tasks. Physica D **80**: 140-150.

Anurans, the Group of Terrestrial Vertebrates Most Vulnerable to Climate Change: A Case Study with Acoustic Monitoring in the Iberian Peninsula

RAFAEL MÁRQUEZ[1], DIEGO LLUSIA[1], JUAN FRANCISCO BELTRÁN[2], JOSÉ PEDRO DO AMARAL[3] & RICHARD G. BOWKER[3]

[1] Fonoteca Zoológica, Dept. de Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias Naturales (CSIC), José Gutiérrez Abascal 2, 28006 Madrid, Spain, E-mail: rmarquez@mncn.csic.es
[2] Departamento de Fisiología y Zoología, Facultad de Biología, Universidad de Sevilla, Avda. Reina Mercedes 6, 41012 Sevilla, Spain.
[3] Dept. of Biology. Western Kentucky University. 1906 College Heights Blvd. Bowling Green Kentucky, KY 42101. USA.

Abstract. We report preliminary analyses from an ongoing sound monitoring project that involves five species of anurans: two species of tree frogs in the genus *Hyla* (Hylidae) and three species of midwife toads in the genus *Alytes* (Discoglossidae) in the Iberian Peninsula. Each station was monitored with an automated recording system based on solid state recorders, coupled with programmable temperature and relative humidity probes. We present comparative data of vocal activity of two populations of *Alytes cisternasii* from thermal extremes of the species range using human detection and commercial automated sound recognition software. Parameters such as duration of reproductive season, preferred temperatures for calling activity and relation with relative humidity are discussed. We compare the performance of analysing the recordings between an automated system of detecting the presence of *Alytes cisternasii* calls and listening of the recordings by non-expert personnel.

The most recent reviews of world wide anuran species conservation indicate that anurans are among the most threatened groups, even moreso than mammals or birds (STUART ET AL., 2004). Among the many factors that are related to that threat, global climatic change appears to be related to the declining populations. Temperature influences the physiology, ecology and behavior of anurans. These ectothermal vertebrates can select their body temperature from a mosaic of temperatures in their environment. Furthermore, temperature indirectly determines the availability of surface water, an element essential for the development of their larvae and for the survival of the adults. Therefore, an increase in temperature related to global climatic change is likely to impact the biology and conservation of anurans.

The geographic location of the Iberian Peninsula makes it particularly vulnerable to these changes. The different predictions of climatic change indicate a general increase in temperatures as well as long term changes in water availability, resulting from a decrease in rainfall for a number of regions (MMA, 2005).

TEMPURA is the acronym for the project “Adaptations of anurans to climate change: comparative study of population in thermal extremes”, funded by the Ministerio de Educación y Ciencia, Spain (Programa Nacional I+D+i), and involves researchers of 4 institutions: Western Kentucky University (USA), Universidade de Lisboa (Portugal), Universidad de Sevilla (Spain) and the Museo Nacional de Ciencias Naturales-CSIC, (Spain).

Within the project TEMPURA we investigate what will be the response of some anuran species to variations in the temperature and water availability of their habitats, among other parameters, we investigate the relationship between acoustical activity of anurans and weather variables in 10 populations of 5 species of anurans. Two sites were selected per species, one of them in the coldest area of their range, and another one in the hottest area of their range. The species selected for the study were three continental species of midwife toads in the genus *Alytes* (*A. obstetricans, A. cisternasii* & *A. dickhilleni*), and two species of treefrogs in the genus *Hyla* (*H. arborea* & *H. Meridionalis*; Fig. 1). The main goals of this project are 1) to determine the reproductive phenology of the species using male calling activity as the indicator; 2) compare the relationship between temperature and humidity and phenology between the populations occuring in warm (xeric) and the cold (mesic) habitat extremes (determining whether any between populations differences respond to phenotypic plasticity or to adaptation), and to complete the comparison comparing the results between species within genera, and between genera, to get an indication of the evolutionary history of the adaptations. In this paper we particularly consider the problem of the analysis of the arrays of recordings to detect acoustical activity of a focal species, and we compare the performance of an automated software-based system and listening of the recordings by non-expert personnel.

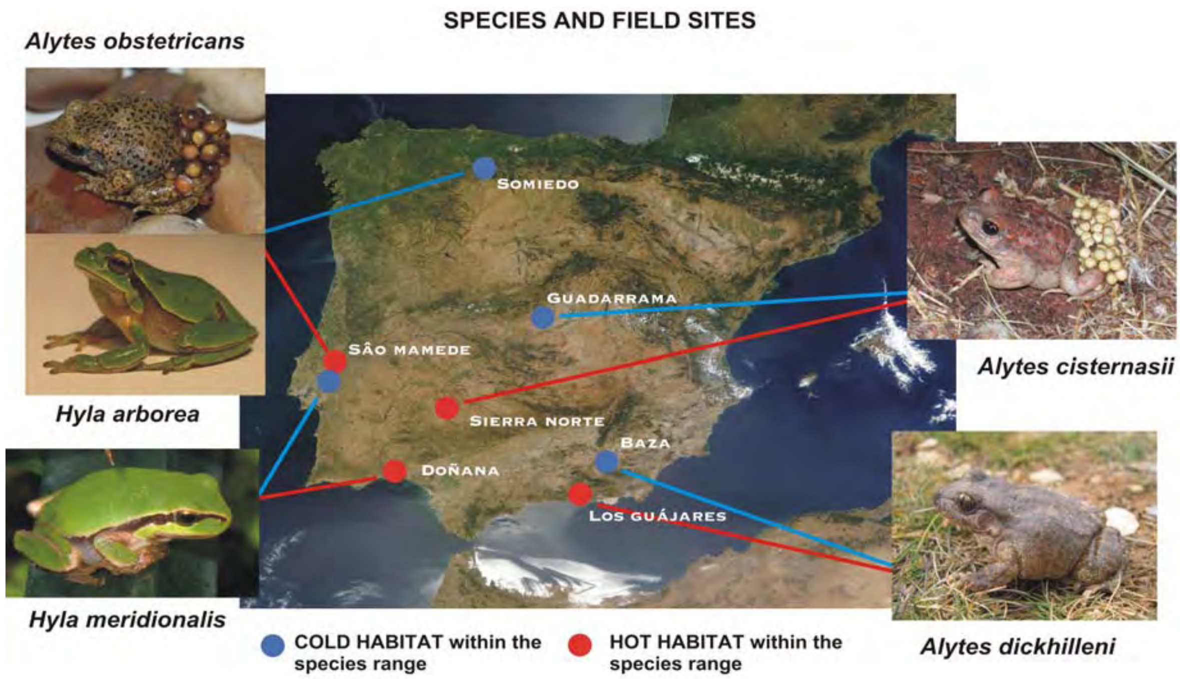


Figure 1: Location of monitoring stations.

Call Characteristics

The spectral and temporal characteristics of the species involved in this study are shown in Fig. 2, (sounds from MÁRQUEZ & MATHEU, 2004). While the two hylid calls are mainly pulsed sounds with wide frequency spectra, the three discoglossids have short and tonal sounds with no frequency modulation and extremely simple harmonic structure (pure tones, MÁRQUEZ & BOSCH, 1995). Consequently the choice of genera allows interspecific and intergeneric comparisons, it also allows exploration of environmental effects on very different sound types.

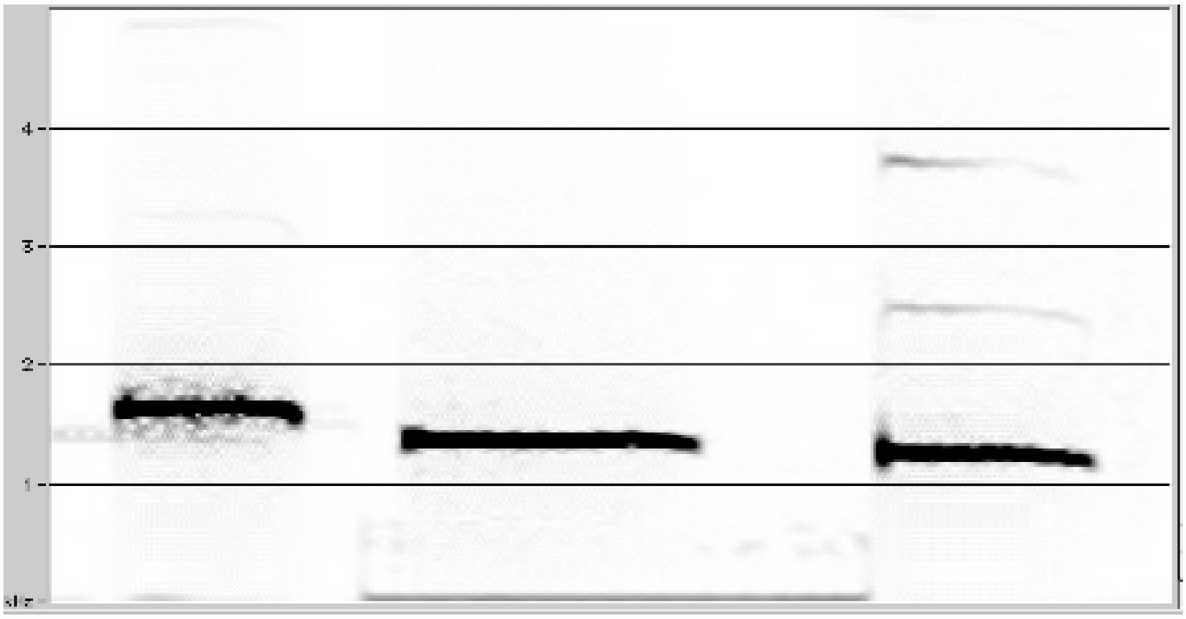


Figure 2A: Audiospectrogram of one advertisement call of *A. cisternasii*, one of *A. dickhilleni*, and one of *A. obstetricans* (FFT window size, 1024 points; total duration 900 milliseconds).

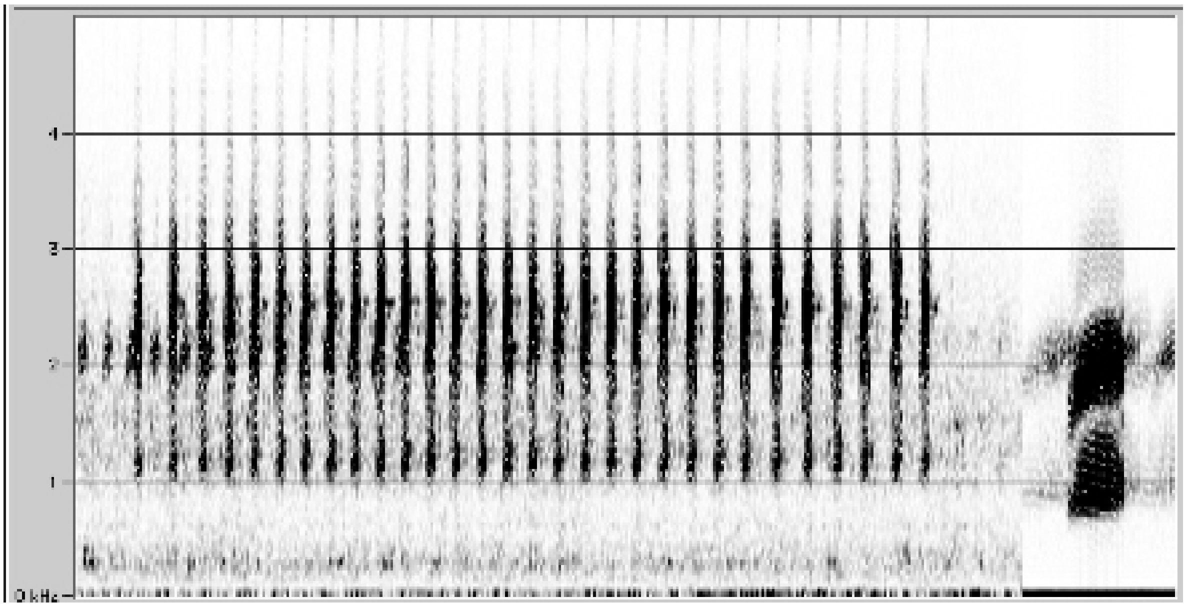


Figure 2B: Audiospectrogram of one advertisement call of *H. arborea* and one of *H. meridionalis* (FFT window size, 1024 points; total duration 9.9 secs).

Methods: Automated Monitoring System

The recording system includes four main elements (Fig. 3): 1) An omnidirectional condenser microphone Fonestar FCM-62 (powered by a AA alkaline battery); 2) a digital recorder Marantz MPD-660 controlled by; 3) an "Amphibulator", a custom made programmable timer developed by the Department of Biology and the Department of Engineering of Western Kentucky University, and 4) a 12v battery (40 Ah or more). The recording protocol was record 3 min per hour, 24 hours per day.

Maximizing the time of autonomous functioning of the recording system is crucial to diminish maintenance costs, and this is an important issue if you consider that the 10 stations are located in distant points of the Iberian Peninsula. This requirement justified the choice of MP3 compression in the sound recording format. With the largest Compact Flash cards available at the beginning of the project (2GB) MP3 compresses sound in MONO theoretically allows for more than 72 hours of recordings (more than 60 days, recording 3 minutes per hour every hour). The Marantz PMD-660, however, has a lower maximum recording limit, which is determined by the three digits that encode the recordings which limit the number of recorded tracks to 999 or 41 days.

Each recording station is also equipped with environmental dataloggers that monitor periodically the air temperature and RH (Onset Hobo H08) as well as the soil temperature at 5-10 cm depth, and water temperature (Onset Pendant). Data are stored every five minutes during the calling activity of the monitored species and every 30 minutes during the rest of the year.

Figure 3: Diagram of the automated recording system.

Protection of the recording stations. - In order to prevent the theft or destruction of the recording stations in sites with public access, the units were housed in iron boxes which were attached to solid elements of the landscape or inserted in a concrete bed. The microphones were housed in an opening on the box or were placed nearby (Fig. 4).

Figure 4: Examples of location and protection of recording systems.

Calibration. - The calibration of the recording system including the methodology for setting similar recording levels in all the units, determining the shape and extension of the area covered by the station, and ultimately determining what percentage of the total population and the corresponding number of adults have been recently described (MÁRQUEZ ET AL., 2008).

Results

The first results regarding two populations of *A. cisternasii* were recently published (TEJEDO ET AL., 2008; MARQUEZ ET AL., 2008,). During 2006 these populations, which are located 400 kms apart in the Iberian Peninsula showed a remarkable degree of synchrony in the beginning of their reproductive season. In both cases acoustical activity started in the second week of September, coinciding with the first rainfalls of fall after the summer drought (this observation is in agreement with previous reports, CRESPO, 1982; RODRÍGUEZ-JIMÉNEZ, 1984, 1988; MÁRQUEZ, 1992; GARCÍA-PARÍS ET AL., 2004). However, there are important differences in the temperature and relative humidity between the periods of acoustical activity of the two populations. In the first weeks of activity air temperature was between 13-20ºC in the warm population during acoustical activity, whereas in the northern population males called at air temperatures of 8-15ºC. Regarding relative humidity, the maximae were similar in both populations reaching 100%, contrasting with the minimae which showed substantial differences between populations reaching below 50% in the southern population while always being above 74% in the northern population (Fig. 5).

Even the soil temperatures measured 5 cm deep were significantly different (Fig. 6) (t-Student=-4.79; p=0.000; n=48).

These preliminary results suggest that the toads are active at different temperatures although, in order to discriminate to what extent the differences result from a process of adaptation or simply reflect phenotypic plasticity, further studies are necessary.

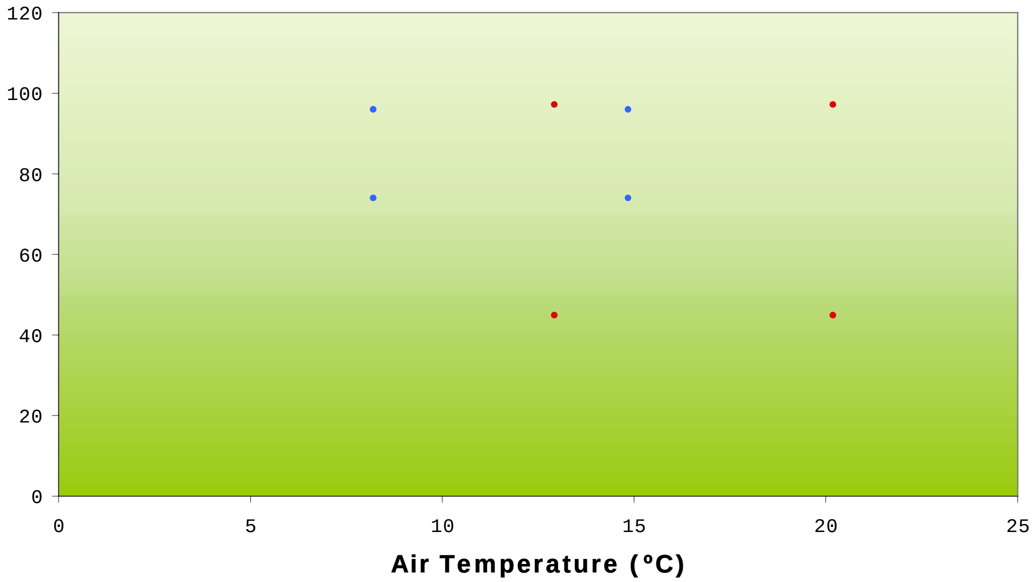


Figure 5: Air temperature and relative humidity during the hours of acoustical activity of the Iberian midwife toad (*Alytes cisternasii*) in the two study sites (fall 2006). Blue: Sierra de Guadarrama, Madrid (cold habitat). Red: Sierra Norte, Sevilla (warm habitat).

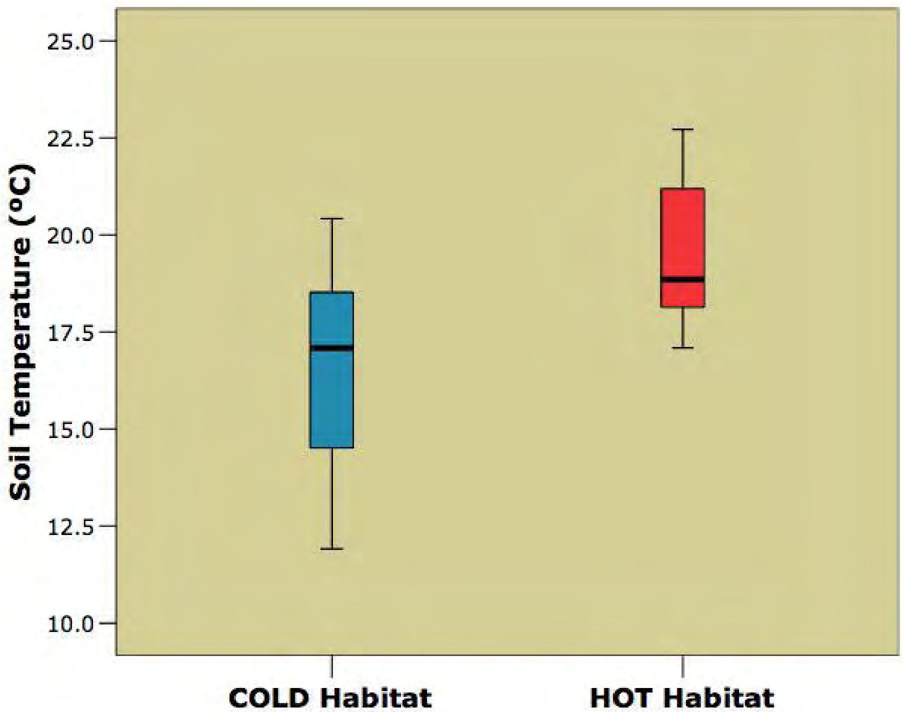


Figure 6: Mean soil temperature during the hours of acoustical activity of the Iberian midwife toad (*Alytes cisternasii*) in the two study sites (fall 2006). Blue: Sierra de Guadarrama, Madrid (cold habitat). Red: Sierra Norte, Sevilla (warm habitat).

Comparative Results of an Automated Sound Recognition System and Human Listening

We used the data from the first season of the two study populations of *Alytes cisternasii* to compare the effectiveness of an automated system when compared to the inspection of the files by students. The automated software used was Song Scope Bioacoustics Monitoring Software, a commercial program made by WildLife Acoustics. According to the developper of this software this program operates as follows “Song Scope builds Hidden Markov Models using algorithms specially designed to consider both the spectral/temporal components of individual syllables, but also how these syllables are arranged into songs for more complex avian vocalizations (frogs are generally single-syllable songs). A "recognizer" is really an HMM -- for the layman, a statistical model with "states" representing spectral fingerprints and probabilities of transitioning between states. Plus a little more...” (Ian Agranat, pers. com.). Again, the populations of *Alytes cisternasii* followed were: cold habitat (Sierra de Guadarrama, Madrid) and warm habitat (Sierra Norte, Sevilla).

The software used required first the construction of models of recognition. Three different sound files were analyzed to build a recognizer, one of them included high quality recordings from a sound guide (MÁRQUEZ & MATHEU, 2004) whereas the other two files were recordings from the automated recording system where *Alytes* sounds were confirmed to exist. More than 8000 recordings were scanned corresponding to daytime and night recordings from fall 2006 to early summer 2007. SongScope yielded 200 positive recordings (presence of at least one call) with the sensitivity level that was selected (9 in a scale of 10).

Also, a subsample of the recordings were listened by two students with no prior experience in call detection, the subsample selected was the first three hours after dusk during the months of the reproductive season (the recordings made at the times with a maximum probability of midwife toads calls). More than 1000 3-min recordings were listened by the students and 96 of the recordings scored positive (presence of at least one call).

All the files that scored positive were listened again by experienced researchers to confirm the presence of the species call (Table 1). Only 3.6 % y 7.1% of the recordings that scored positives in the human listening process were confirmed to have *Alytes* calls in them. Song Scope was more effective since 15.3 % and 19.7% of the positive files were confirmed by subsequent expert screening of the recordings. However, the large percentage of false positives (Type II error) with both methodologies is remarkable, and only partially explainable by the simplicity of the call. The presence of “acoustical mirages” or of “oversensitive acoustical search image” have been noted previously by the authors as a psycho-acoustic phenomenon which is triggered in sound environments which are far away from the midwife toads breeding grounds and results in the sensation of detecting midwife toad calls erroneously, presumably when a burst of energy occurs within the frequency range of these species.

Another important measure of the efficiency of the sound recognition system is the number of recordings that included midwife toad calls that did not score positively in the automated test (Type I error, Table 2). In order to determine the importance of that type of error we had expert ears listen to 86 tracks selected from the dates and hours of maximum probability of calling. The software used fail to detect 35% of the files that actually contained *A. cisternasii* calls (24% in the hot habitat and 50% in the cold habitat). These figures are somewhat disappointing and may be related to the limit of 10 dB sensitivity (sound/noise) of the software or to the variability of the frequency of the call within the population that may exceed the range established for the software scanning (if this is the case, the problem would probably be solved by further improving the recognizer model).

Table 1: Type II error (call detected in track not containing the call). Comparative results between non-expert human listener and automatic detection system (Song Scope) for mating calls of *Alytes cisternasii*. Pooled recordings of hot and cold populations.

Detection system	Population	Nº tracks analyzed	Total tracks with positive scores	Confirmed positive	False positive	% Type II Error	% Accurate
Non-expert human listener	Hot	563	42	3	39	92.9	7.1
Automatic	Hot	2364	124	19	105	84.7	15.3
Non-expert human listener	Cold	532	55	2	53	96.4	3.6
Automatic	Cold	6597	76	15	61	80.3	19.7

Table 2: Type I error (call not detected in a track that included a call). Results of automatic detection system (Song Scope) for mating calls of *Alytes cisternasii*. Pooled recordings of hot and cold populations.

Detection system	Population	Nº tracks analyzed	Total tracks with positive scores	Tracks with calls detected	Tracks with calls not detected	% Type I Error	% Accurate
Automatic	Cold	34	34	8	26	50.0	50.0
Automatic	Hot	52	52	16	36	23.8	76.2
Automatic	Total	86	86	24	62	35.1	64.9

Conclusions

Of all of the species studied in TEMPURA, *A. cisternasii* is the species that occurs throughout a more homogeneous habitat and has the least fragmented distribution. It was therefore, a priori, the species that was less likely to show substantial thermal differences between populations at thermal extremes of its distribution. However, the siginificant differences of activity temperatures observed in this preliminary study, together with the old evolutionary age of the lineage (the differenciation of *A. obstetricans* and *A. cisternasii* has been estimated to have occurred 16 m. y. ago with the formation of the Betic Strait, Arntzen & García-París, 1996; Martínez-Solano et al. 2004) indicate that there may have been a substantial selective pressure related to thermal conditions, and ample opportunity for adaptations to evolve. Further insight on these processes may be produced by ulterior progress in TEMPURA research.

Acknowledgements

We are grateful for the help of many friends and collaborators, among others, Alberto, Arturo, Borja, Catarina, Gema, Helder, Maribel, Jaime, Lidia and the staff of P.N. de Sierra Norte de Sevilla, forest wardens of Fresnedillas de la Oliva, P. N. Baza and P. N. de Somiedo. We are particularly grateful to Daniel García Arnero, who permited us to access to his property in Fresnedillas de la Oliva; and to Drs. Albert & Ouida Meier of Western Kentucky University, who lead a similar acoustical monitoring project in “Upper Green River Biological Preserve”, Kentucky, USA. We are particularly grateful to Dr. Mark Cambron, Dept of Engineering, Western Kentucky University, who developed the amphibulator for this research project.

References

ARNTZEN, J. W. & GARCÍA-PARÍS, M. (1995): Morphological and allozyme studies of midwife toads (Genus *Alytes*), including the description of two new taxa from Spain. *Bijdragen tot de Dierkunde* **65**: 5-34.

CRESPO, E. G. (1982): Contribuição para o conhecimento da biología das espécies Ibéricas de *Alytes*, *Alytes obstetricans boscai* (Lataste 1879) e *Alytes cisternasii* (Boscá 1879) (Amphibia Discoglossidae). Ovos, posturas (Epocas de reprodução). *Arquivos do Museo Bocage. Serie A* **1**: 453-466.

GARCÍA-PARIS, M., MONTORI, A. & HERRERO, P. (2004): *Fauna Ibérica. Volumen **24**. Amphibia.* CSIC. Madrid.

MÁRQUEZ, R. (1992): Terrestrial paternal care and short breeding seasons: Reproductive phenology of the midwife toads *Alytes obstetricans* and *A. cisternasii*. *Ecography* **15**: 279-288.

MÁRQUEZ, R. & BOSCH, J. (1995): Advertisement calls of the midwife toads *Alytes* (Amphibia, Anura, Discoglossidae) in continental Spain. *J. Zool. Syst. Evol. Research* **33**: 185-192.

MÁRQUEZ, R., LLUSIA, D., & BOWKER, R. G. (2008): Terrestrial sound monitoring systems, a methodology for calibration and data processing. *Proceedings IBAC meeting, Pavia 2007*. In Press.

MÁRQUEZ, R. & MATHEU, E. (2004): Sound Guide of the frogs and toads of Spain and Portugal (Audio CD). (Ed. by Fonoteca Zoológica). Barcelona: Alosa.

MARTÍNEZ-SOLANO, I., GONÇALVES, H. A., ARNTZEN, J. W. & GARCÍA-PARÍS, M. (2004): Phylogenetic relationships and biogeography of midwife toads (Discoglossidae: *Alytes*). *J. Biogeogr.* **31**: 603-618.

MMA (2005): *Principales conclusiones de la evaluación preliminar de los impactos en España por efecto del cambio climático*. Informe Técnico. Ministerio de Medio Ambiente. Madrid.

RODRÍGUEZ JIMÉNEZ, A. J. (1984): Fenología del sapo partero ibérico (*Alytes cisternasii* Boscá. 1879). *Alytes (España)* **2**: 9-23.

RODRÍGUEZ JIMÉNEZ, A. J. (1988): Fenología de una comunidad de anfibios asociada a cursos fluviales temporales. *Doñana, Acta Vertebrata* **15**: 29-43.

STUART, S.N., CHANSON, J.S., COX, N.A., YOUNG, B.E, RODRIGUES, A. S. L., FISCHMAN, D. L., WALLER, R.W. (2004): Status and Trends of Amphibian Declines and Extinctions Worldwide. *Science* **306**: 1783-1786.

TEJEDO, M., RAFAEL M., BELTRÁN, J.F., MARANGONI, F., LLUSIA, D., CAMBRON, M., GÓMEZ-MESTRE, I., MEIER, A., EEKHOUT, X., BOWKER, R., BENÍTEZ, M., MOREIRA, C., CRESPO, E., SOUSA DO AMARAL, J.P. Y PENNA, M. (IN PRESS): Las Investigaciones sobre Anfibios en el Parque Natural de la Sierra Norte. En: Menor, A. y Cuenca, I. (Coords.). *Investigación científica y conservación en el Parque Natural Sierra Norte de Sevilla*. Junta de Andalucía, Consejería de Medio Ambiente. Sevilla.

A Decade of Monitoring Frog Communities in Northern Australia

GORDON GRIGG[1], HAMISH MCCALLUM[2], AND ANDREW TAYLOR[3]

[1]Integrative Biology, University of Queensland, E-mail: g.grigg@uq.edu.au
[2]Zoology, University of Tasmania, E-mail: hamish.mccallum@utas.edu.au
[3]Computer Science & Engineering, University of NSW, E-mail: andrewt@cse.unsw.edu.au †
† Author to address correspondence to

Abstract. We have bioacoustically monitored frog populations at 16 sites in tropical Australia for over a decade using autonomous solar-powered computers. System software, in real-time, identifies the calls of over 20 species of frog. Call frequency statistics for each species are logged at 10 minute intervals. These systems operate year-around unattended with only annual visits for data retrieval and maintenance. We discuss the techological success and failures during the development and deployment of these systems which can reliably operate in a remote and difficult environment with high humidity, extreme temperatures, dry season fires, wet season flooding and occasional cyclones.

The Cane Toad (*Bufo marinus)* is a large toxic opportunistic predator native to a wide variety of habitats extending through the Americas from northern Argentina to southern Texas (LEVER 2001). It has been introduced to over 40 countries outside its native range largely motived by a belief its consumption of insects would benefit agriculture. Its release in 1935 near Gordonvale in Australia began a spectacular an invasion which has seen Cane Toad occupy a vast area of northern Australia. Before its release there was public concern at the impact the Cane Toad might have on Australia's native fauna (FROGGATT 1936) and this concerns remains seven decades later. In the 1990s the possibility of biological controls becoming feasible led the Australian federal government to fund research systematically examining for the first time the impact Cane Toads might have on native fauna as they invaded the Northern Territory (GRIGG 2000).

Anecdotal reports (COVACEVICH & ARCHER 1975) suggested adverse impacts on a number of taxa including goannas, quolls, snakes and frogs. None of these groups are easy to census in the wet tropics but the conspicuous advertisement calls of male frogs accompanying their wet season breeding suggested bioacoustic monitoring might be feasible, although not without major obstacles. Flooding from storms and cyclones may prevent access during precisely those parts of the wet season when frog breeding occurs. The calling activity of some species is associated with the inundation produced by the first wet season storms (TYLER ET AL. 1983) but the time of arrival of these storms varies between years and is too unpredictable to allow scheduling of field surveys. It is also necessary to capture sufficient data so that the natural variability of frog populations does not camouflage any changes due to the arrival of Cane Toads.

We overcame these obstacles by building and deploying a series of 16 solar-powered computer systems to perform the bioacoustic monitoring. For the last decade these systems have operated unattended through the wet season monitoring frog. System software analyses sound from a microphone and identifies 22 Top End frog species by their call. Calling statistics for each species are stored for retrieval after the wet season. As this is the first and to date the most extensive deployment of automated recognition systems for biological monitoring, we are describing our experience and some of the pragmatics of this type of field work.

Automatic Recognition Software

The automatic recognition software is built using the C4.5 machine learning system (QUINLAN 1993) from manually prepared positive and negative training sets. Spectral and a few time domain (waveform) attributes are used to classify every spectral peak. These unreliable classifications are aggregated using a manually-constructed per-species hierarchical voting system to produce reliable identification. The structure of the voting system for each species is chosen to be similar to its typical vocalization patterns. A fuller description of the recognition software can be found in (TAYLOR ET AL. 1996).

Study Locations

In 1997-1998 we deployed 10 monitoring systems in savannah woodland in the Northern territory's Roper valley along a 125 km transect roughly perpendicular to the arriving toad front. The experimental design placed sites in pairs 1-2 km apart with approximately 25 km between pairs. Figure 1 shows habitat typical of our Roper sites. In 1999-2000 we deployed 6 sites further north in Kakadu National Park, 2 sites each in savannah woodland, rocky stream and floodplain habitats. In 2005 we deployed 3 sites further west on the Northern Territory/Western Australian border, ahead of the current invasion front.

Figure 1: a frog monitoring system in the Roper valley.

Mark I Hardware

The wet tropics are a difficult environment for long-term deployment of electronics. Major concerns in the design of our system enclosures included vandalism, dry season fires, wet season flooding (Fig. 2), extreme heat and humidity, lightning, violent thunderstorms and even cyclones. Our enclosures utilize recycled 5-6m high hollow steel poles formerly used for street lighting. A 30 watt solar panel, balanced dynamic microphone (Shure SM58 clone), rain gauge and temperature sensors are installed on top of the poles. System electronics, a solar charge regulator and four 12V 7Ah sealed-lead-acid batteries are raised by a pulley system into the centre of the pole, lifting them several metres above ground level. This (hopefully!) provides protection against both flood, fire, vandalism and damage by animals. Data is downloaded by removing a steel plate at the base of the pole, lowering the electronics using the pulley system and attaching a serial cable to an environmentally-sealed connector (Fig. 3).

Figure 2: checking on a system during the wet season.

Computer hardware needs to be carefully chosen to operate in this type of field environment. Our first generation of systems were based on the AT304 industrial single board computer built by JED Microprocessors . It has a 25 MHz Cyrix 486SLC processor, 1 MB of RAM and 2 MB of flash memory. As this board consumes 10 watts it could not be operated continuously and was instead powered on 4 hours per day (8pm-midnight). A smaller CPU operated continuously saving data from temperature & rain gauges and powering the main board on when needed for frog monitoring. This smaller board used a 2 MHz 8-bit 68HC11 CPU, 256 bytes of ram, 2 kilobytes of EEPROM and consumed 1 watt, Other electronics included a microphone preamp, a Soundblaster16 card on a passive backplane to digitize the sound and a DC-DC convert to provide the +12 V, +5 V and –12 V supply required by the AT302 board. The electronics are enclosed in 1 metre of 150 mm diameter PVC (sewer) pipe sealed at both ends with 3 Bulgin environmental-sealed sockets for power, sensor and data download.

Figure 3: downloading data during the dry.

Mark II Hardware

The Mark I hardware was serviceable but less than ideal and after 4 years we changed to a new hardware platform based on the *Pleb* single board computer developed at UNSW (WIGGINS 2001). It has a 200 MHz StrongArm SA-1100 processor, 32 MB of RAM and 4 MB of flash memory. A daughter board with an ATMEL AT90LS8535 AVR processor interfaces to an LCD display, temperature and rainfall sensors. Also on the daughter board is a MAX146 ADC used to digitize the incoming sound at 16 kHz and 12 bits. The only ancillary electronics required is a small PCB containing a microphone preamp (SSM2017) and voltage regulator (LP2951). The power consumption of these systems is just over 1 watt, an order of magnitude less than our Mark I platform.

Mark III Hardware

In 2004 off-the-shelf hardware became available with similar to our custom-built mark II hardware and in 2005 we deployed 3 systems based on Technologic System TS-7260 single board computer. It has 200 MHz Arm-9 system-on-a chip CPU, 64 MB of RAM, 128 MB of flash memory and consumes 2 watts. It has a USB host interface which we have used to attach a 2.5" 80 GB disk allowing a large amount of sound to be captured, in addition to frog calling statistics (stored in flash). These systems use 2 Knowles waterproof microphones (MR-8406) and sound is digitized by a USB-attached Griffin iMic. The Mark III systems were deployed at short notice using a simpler mounting system, a 2.5 metre galvanized steel pole on a pegged base plate and guyed to star pickets (Fig. 4).

Figure 4: A Mark III monitoring system.

Reliability

We have now well over 100 system years deployment. The overall failure rate has been approximately 20% per year. This failure rate is better than expected at the commencement of the study but still creates difficulties when analysing the data. The Mark II systems have been below a 10% failure rate per year for most of their deployment but their failure rate increased after the 5 years of deployment. Failures come from different sources than we expected when planning the project.

System power supply is the most common area for failures. The solar panels themselves which have not been the source of any failures. We discovered that sealed-lead-acid batteries suffer rapid degradation in performance after three years in the tropical environment, necessitating a 3 year battery replacement cycle. The solar charge regulators have an ongoing source of problems and in future system we will consider including redundant power supplies.

Sensors are another source of system failure. Their function limits the degree they can be protected from the environment. The dynamic microphones have been amazingly robust and have not failed directly but they are behind cloth screens which degrade eventually with extended UV exposure. On a number of occasions the failure of these screens has allowed insect penetration and microphone function has been lost as it filled with frass. The rainfall and temperature sensors have a high failure rate after several years of deployment.

Generally electronics have not been a significant source of failure except for vibration during transport causing pre-deployment problems. The Mark II electronics have been particularly reliable. Heat and humidity were both sources of much concern during system development. Internal temperatures peak at ~60C but this has not caused problems for the electronics. Humidity has also not been an issue. Systems are deployed with silica gel to remove initial moisture and heat from electronics apparently keeps canisters very dry.

Software bugs are obviously a risk when expecting a computer to operate untended for months. Despite defensive programming and extensive lab testing some software problems were not found until the first year software was fielded. The operating system (DOS) on the Mark I hardware had a Y2K bug not found in testing (because it occurred several weeks after the clock roll-over) which caused a number of failures. Linux, the operating system on the Mark II & III hardware, has been very reliable.

A dry season fire passes through each of our site roughly every two years. In these over fifty fires only one systems as suffered damage, a Mark III system (on a 2.5 m pole) was disabled by wiring from the solar panel being burnt. Flooding affects our sites probably with a similar frequency as fire. Two systems have been disabled by the electronics being submerged by flooding after extreme, in one case unprecedented, rainfall. In both cases the data collected prior to the flooding was recovered from the flash. Two cyclones have passed directly over our study sites. Neither affected the operation of any monitoring system.

Conclusions

We have deployed bioacoustic monitoring systems in a hostile environment and they have operated untended successfully between our yearly visits. It is already feasible to use automated recognition for long-term monitoring and the increasing availability of new technologies such as wireless broadband promises a bright future

Acknowledgements

We are grateful to CSIRO, Parks Australia North and Dept. Environment & Conservation WA for funding. Les Fletcher played a key part in the project: designing, constructing and assisting us deploy the Roper & Kakadu systems. Graeme Watson joined us for early field trips andprovided essential knowledge and recordings of NT frogs. Frank Seebacher, Peter Kind, Tim Jessop, Tim Schulz, Maurizio and Ida Bigazzi and Jan Grigg joined us in for the most arduous phase of the project. We are grateful to them, other volunteers and numerous parks staff who have assisted us in the field.

References

COVACEVICH, J, & M ARCHER (1975): The distribution of the cane toad, *Bufo marinus*, in Australia and its effects on indigenous vertebrates. - Memoirs of the Queensland Museum **17**: 305-310.

FROGGATT, W. W. (1936): The introduction of the great Mexican toad *Bufo marinus* into Australia. - Australian Naturalist **9**: 163 - 164.

GRIGG, G. (2000): Cane toads vs native frogs. - Nature Australia **26**: 32 - 41.

LEVER, C. (2001): The Cane Toad: The History and Ecology of a Successful Colonist. Westbury Academic and Scientific Publishing.

QUINLAN, J. R. (1993): C4.5: Programs for Machine Learning. Morgan Kaufmann.

TAYLOR, A., G. WATSON, G. GRIGG, & H. MCCALLUM (1996): Monitoring Frog Communities: An Application of Machine Learning. - *AAAI/IAAI* **2**:1564 - 1569.

TYLER, M. J., G. A. CROOK & M. DAVIES (1983): Reproductive Biology of the Frogs of the Magela Creek System, Northern Territory. - Rec. S. Aust. Mus. **18**: 415 - 440.

Wiggins, A. (2001): PLEB: a platform for portable and embedded systems research. Technical report UNSW-CSE-TR-0105 School of Computer Science and Engineering, University of NSW.

Automated Bioacoustic Identification of Insects for Phytosanitary and Ecological Applications

DAVID CHESMORE

Department of Electronics, University of York, Heslington, York, YO10 5DD, England
E-mail: edc1@ohm.york.ac.uk

Abstract. Many insect, animal and bird species produce sounds that can be used to detect their presence, and in many cases identify them. In recent years, there has been an increasing interest in the development of automated detection and identification systems for a variety of applications including species counting, biodiversity studies and pest detection. The rapid development of hand-held computers is leading to potentially commercially viable biodiversity assessment tools and pest identifiers. This paper discusses one particular approach applied to insects using time domain signal processing and artificial neural networks for the robust identification of taxa, concentrating on insects. The paper introduces the concept of time domain signal coding and then describes results for phytosanitary (plant health) applications for quarantine insect larvae in timber, identification of three species of Japanese cicada and species of British grasshopper. It concludes with a discussion on future directions and applications of automated bioacoustic identification systems.

Introduction

Automated Identification and Computer-aided Taxonomy

Species identification using acoustics has been employed for many years for species surveying and species counting, particularly for birds and mammals. However, this can be very time consuming, expensive and requires a high degree of skill. Surveys of this kind can only provide "snapshots" in time, often only a few hours in length and possibly one or two times per year (FISCHER ET AL. 1997). The availability of systems that can automatically identify species or taxonomic groups will be of great benefit, especially if they are capable of monitoring on a continuous basis (GARDINER ET AL. 2005, CHESMORE 2007a, CHESMORE 2007b).

Sound production in animals can be divided into deliberate (non-incidental) communication sounds and those produced as a by-product of activity such as flying and eating; these are known as incidental sounds. It is possible to use both types of sounds for automatically detecting and identifying taxa.

This paper discusses two application areas applied specifically to insects although the techniques described are applicable to any band limited signal and have been successfully demonstrated for heart sounds, faults in gearboxes, acoustic identification of vehicles and ECG monitoring.

The concept of automated species identification is not restricted to bioacoustics but can encompass image processing, radar and sonar systems as described in (CHESMORE 2007b). The other major work of automated species identification is in image processing as exemplified by Daisy (GASTON & O'NEILL 2004). Automated species identification is one part of computer aided taxonomy (CAT), and the other aspects being computer-based key systems, education and mathematical methods for, for example, genomics. The first serious consideration of all aspects of CAT was at the inaugural meeting of the Bionet-International group for computer aided taxonomy held at the University of Cardiff in July 1997 (CHESMORE 2000). This meeting was pioneering and brought together more than 30 researchers in the disciplines of biology, mycology, entomology, computing, film making and engineering.

Basic Structure of an Automated Identification System

An automated identification system is basically a form of pattern recognition and consists of the following four functional blocks as given in Figure 1 (CHESMORE 2004, CHESMORE 2007b):

a) Sensor. There are a variety of sensors depending on the application. For acoustic applications the sensors can be microphones (including ultrasonic microphones) or vibration detectors such as piezoelectric devices or accelerometers. For non-acoustic applications sensors include cameras, radar, sonar, SEM and it is possible to combine multiple sensor types in certain applications.

b) Preprocessor. In acoustic systems, signals are usually amplified and then filtered to remove unwanted signals by the preprocessor.

c) Feature Extractor. The extraction of features is the most important part of the identification system since any overlapping feature space between different taxa will cause reduced identification accuracy or misidentification. A large number of feature extractors exist and can be divided into frequency domain and time domain. For example, in the frequency domain the fast Fourier transform is very common with wavelet transforms becoming increasingly common. In the time domain, the time domain signal coding, autocorrelation, short time energy, etc are often used.

d) Classifier. The output of the feature extractor is fed into a classification system which may be one of a variety including artificial neural networks, expert systems, linear classifiers, etc.

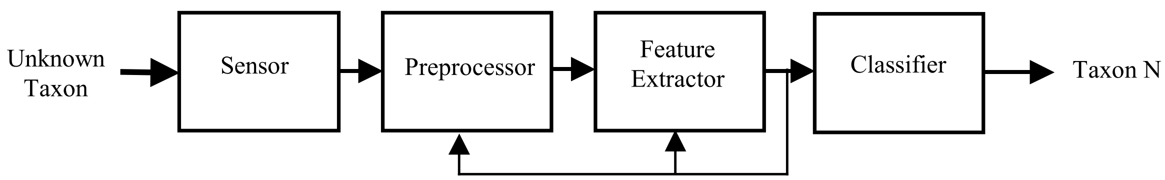


Figure 1: Schematic diagram of a typical automated taxon identification system.

The success of an identification system depends very strongly on the signals from each taxon being completely separable in the feature space. Overlapping features may cause misidentification. The term recognizable taxonomic unit (RTU) is commonly used to place taxa into similar groups in terms of morphology (RIEDE 1993). In an identification system, a RTU not likely to bear any resemblance to the features extracted, nor will the taxonomic grouping based on feature space be equivalent to that based on RTU. The author has coined the term "recognizable feature space unit" (RFSU) as illustrated in Figure 2. Here, the taxonomic space shown in a) is populated by a number of taxa which are then grouped in b) into four RTUs according to morphological features. Applying different feature extraction methods will cause the taxa to be grouped in a different way; for example, c) in Figure 2 shows them being grouped into two RFSUs (RFSU1 and RFSU2) whereas d) shows them being grouped into seven RFSUs, some of which correspond to individual taxa. It is evident that the latter feature extraction method provides a higher identification accuracy.

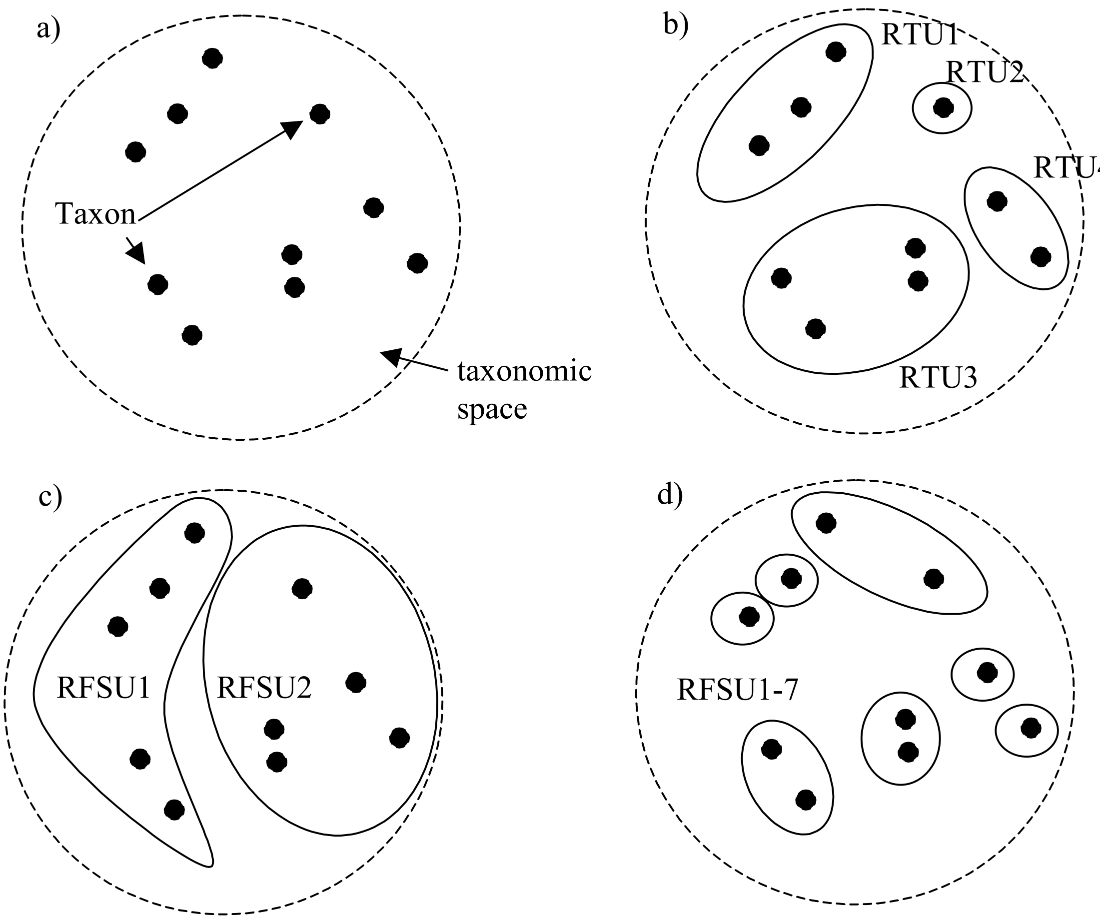


Figure 2: Representation of the relationship between Recognisable Taxonomic Unit (RTU), Recognizable Feature Space Unit (RFSU) and taxon. a) shows the taxonomic space populated by individual taxa; b) shows the taxa grouped into RTUs according to some morphological features; c) shows them grouped according to one set of features (RFSU1-2) and d) grouped according to different set of features (RFSU1-7).

Intelligent Bioacoustic Identification System (IBIS)

Over the past 10 years or so, work at York and previously at Hull University has been progressing towards the development of handheld devices for automated identification. The so-called intelligent bioacoustic identification system (IBIS) is still under development and is expected to be modular in that it can be used for different taxonomic groups.

IBIS uses a purely time domain approach to extract features for classification using one of a variety of artificial neural networks. The time domain approach, known as time domain signal coding (TDSC) is a computationally simple method. The technique has been tested on 25 species of British Orthoptera with 99% recognition accuracy (CHESMORE ET AL. 1997, CHESMORE 2000, Chesmore 2001, CHESMORE & NELLENBACH 2001) and 10 species of Japanese bird with 100% accuracy (CHESMORE 1999, CHESMORE 2001). These results were for high signal to noise ratio (SNR) and high quality signals and were somewhat unrealistic for " real" sounds. More recently, results for a smaller number of British Orthoptera recorded under field conditions indicate that the system is capable of accurate identification with lower signal to noise ratio and interfering signals as described in section 5.2. In all of the work described here, sounds have been recorded in .WAV format at 44.1kHz sampling rate (16 bit) on a variety of recording equipment including the use of external PC sound cards and hard disc recorders. It is important to note that PCs with internal sound cards generate significant electrical noise, especially disc drives, which can be a problem.

Time Domain Signal Coding

Time domain signal coding (TDSC) is a computationally simple method for describing the shape of a waveform between successive zero-crossings. The original concept was developed for low data rate speech communication and was called Time Encoded Speech (TES) (KING & GOSLING 1978). TDSC is a modified form of TES and is more generic. The basic concept of TDSC is illustrated in Figure 3 which shows two epochs (interval between successive zero-crossings) of duration 25 and 30 samples respectively. The shape of the waveform in each epoch can be defined in a number of ways, the simplest being the number of positive minima or negative maxima. Each epoch is therefore defined as a couplet (duration, shape) or (D, S); the example in Figure 1 has couplets (25, 1) and (30, 2). The range of values of (D, S) can be very large and is a function of the signal bandwidth and complexity. Reduction of the range of (D, S) can be achieved using a non-linear mapping into a codeword C using a codebook; however, a codebook has to be manually developed for each signal type by observation of (D, S).

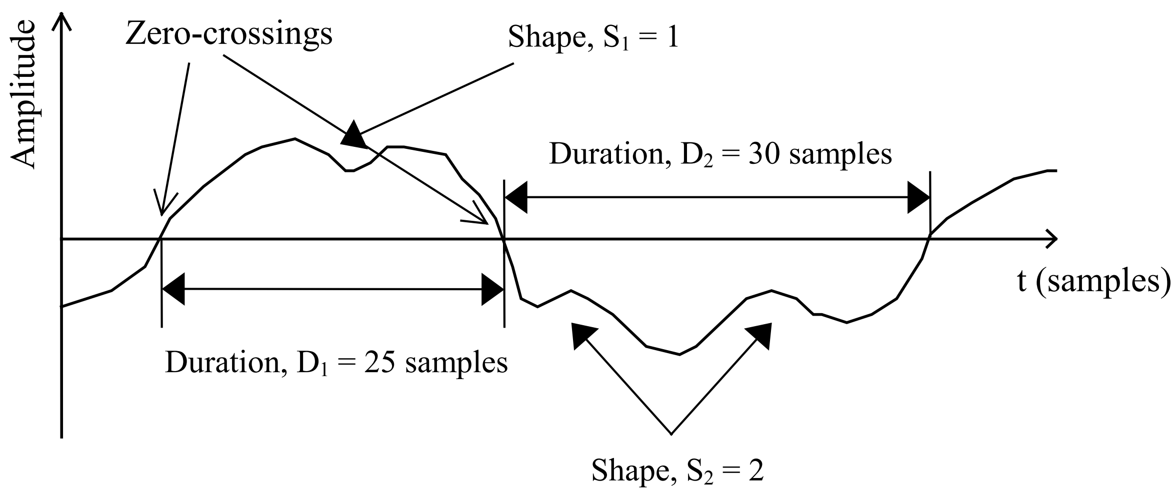


Figure 3: Time Domain Signal Coding of a Waveform.

The signal is further analysed by determining two matrices – the S-matrix and the A-matrix. The S-matrix is simply the frequency of occurrence of each codeword in the signal. The A-matrix is formed as the conditional probability of pairs of codewords and has been most often used as the feature space for classification.

TDSC in this form has a major drawback in that a codebook has to be produced for each application; generalised codebooks can be used but are sub-optimal. The author is investigating extensions to TDSC that do not involve a codebook. The most direct approach is to use (D, S) directly as inputs to an artificial neural network and this is currently being investigated. An alternative is to create a one dimensional vector, the D-matrix, by scaling duration by shape as follows (FARR 2007, FARR & CHESMORE 2007):

$$C = ((S \times S_F) + D)$$

where S = shape
D = duration
S_F = scaling factor

The scaling factor is determined by the maximum possible duration in the signal and S_F must be sufficiently large to avoid any overlaps. The use of D-matrices will be given in Section 4 where they have been successfully used for insect larva identification.

Pattern Classification

In this work, the majority of pattern classification has been carried out using a variety of artificial neural networks including multilayer perceptron (MLP), self organising map (SOM) and learning vector quantisation (LVQ). Early research used MLPs which have been successful for A-matrix-based identification as illustrated in the cicada and grasshopper identification system described in section 5. It has been shown that identification of insect larvae using their bites is improved by the use of LVQ as described in section 4.

Figure 4 shows the schematic diagram of the system known as Intelligent Bioacoustic Identification System (IBIS) in which pattern classification is carried out using an artificial neural network.

The research group at York is now investigating the use of syntactic pattern recognition for identifying temporally complex signals such as bird song. This work is in its early stages and there are no specific results to report. One problem that must be overcome is the lack of ability of most neural networks to be adapted to allow for retraining or addition of new taxa. Newer forms of artificial neural networks such as plastic self organising maps (PSOM) which adapt in real-time are worth consideration.

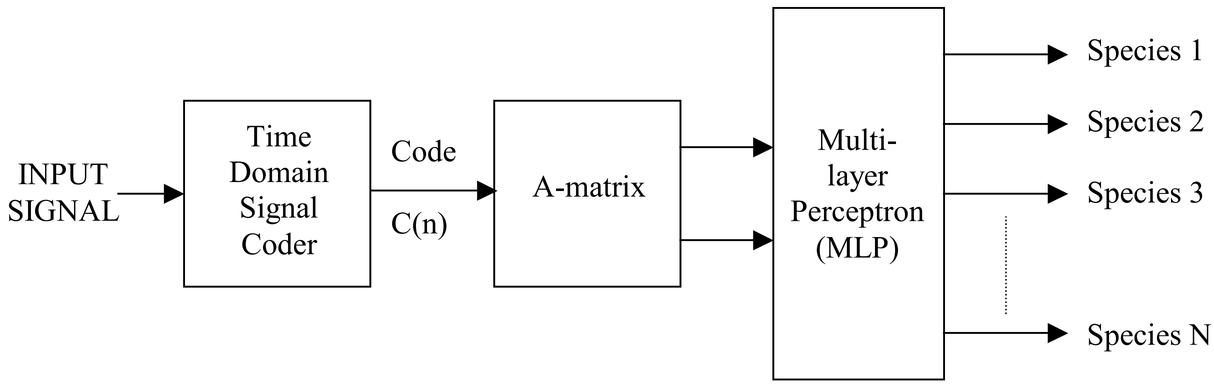


Figure 4: Schematic digaram of the Intelligent Bioacoustic Identification System (IBIS).

Identification of Insect Pests for Phytosanitary Applications

The acoustic detection of insects, particularly insect larvae within timber, is not new. For example, COLEBROOK described the design and construction of a device for the detection of destructive insects in timber in 1937 (COLEBROOK 1937). In recent years, the cost and availability of both suitable sensors and high-speed, low power processing is leading to the increased use of acoustic detection systems for a variety of applications including species counting, detection of pests (HAACK ET AL. 1997; HAGSTRUM ET AL. 1990; HICKLING ET AL. 2000; MANKIN ET AL. 2000; MANKIN & WEAVER 2000; SHUMAN ET AL. 1993, 1997) and education (OBA 2004). However, it must be stated that the majority of research is in detection only and not identification.

The number of imports and exports to many countries has increased rapidly in recent years leading to increased concerns about the potential import of serious of insect pests. For example, in the UK, thousands of tonnes of timber, plants and plant products are imported. The main problem with many insect species particularly beetles is that their larvae feed within wood and remain undetectable until they hatch of adults which may be several years. One example is the Asian Longhorn beetle (*Anoplophora glabripennis*) which is causing considerable damage to trees in the USA (MACLEOD ET AL. 2002) and is an EC A1 quarantine listed pest. In the UK it is the responsibility of the Department for Environment, Food and Rural Affairs (Defra) to maintain pest free status through the Plant Health and

Seed Inspectorate (PHSI). Despite the measures taken by PHSI there still remains the possibility of undetected larvae. The application described here aims to create a system that the PHSI inspectors can use for the detection and, more importantly, the identification of the larvae of wood-boring beetle species. Table 1 give a list of species recorded during the first Defra-funded project resulting in more than 80Gbytes of recordings. Two plant groups were investigated – soft plants and woody plants. Problems with poor acoustic coupling of sensors to the substrate in soft plants resulted in the project concentrating on woody material including live trees, wooden packing material and dunnage. Therefore only beetle larvae were extensively investigated.

Table 1: List of insect species recorded.

Insect Order	Insect Species
Lepidoptera (Moths)	*Acherontia atropos* (Death's-Head Hawkmoth)
Coleoptera (Beetles)	*Anoplophora chinensis* (Citrus longhorn)
Coleoptera (Beetles)	*Anobium punctatum* (Furniture Beetle)
Coleoptera (Beetles)	*Dorcus parallelopipedus* (Lesser Stag Beetle)
Coleoptera (Beetles)	*Hylobius abietis* (Pine Weevil)
Coleoptera (Beetles)	*Hylotrupes bajulus* (House Longhorn)
Coleoptera (Beetles)	*Leptinotarsa decemlineata* (Colorado Beetle)
Coleoptera (Beetles)	*Lucanus cervus* (Stag Beetles)
Coleoptera (Beetles)	*Prionus coriarius*
Coleoptera (Beetles)	*Rhagium bifasciatum*
Lepidoptera (Moths)	*Spodoptera exigua*
Lepidoptera (Moths)	*Spodoptera littoralis*
Coleoptera (Beetles)	*Trichoferus griseus*
Coleoptera (Beetles)	*Agrilus planipennis* (Emerald Ash Borer)

The main type signal generated by an insect larva is caused by biting of the wood fibres and is of short duration and impulsive nature as indicated in Figure 5. Different species will have characteristic bites that are a function of the jaw structure and the plant fibre structure. Some species such as the stag beetle (*Lucanus cervus*) produce a deliberate stridulating sound which is species-specific in character. It is not yet known why larvae produce this sound. Both types of sound can be used for identification purposes.

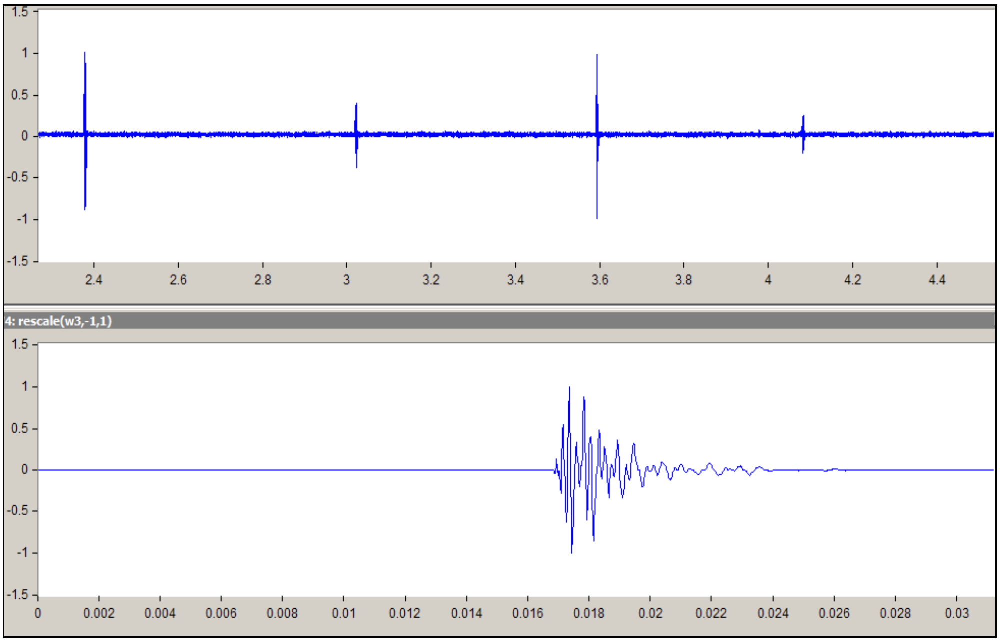


Figure 5: Example bite waveforms for *Hylotrupes bajulus*. Lower trace is a single bite in more detail. Horizontal axis is time in seconds. Signal amplitude is normalised.

Various types of sensors have been investigated optimal detection of biting sounds including microphones, bimorphs and piezoelectric sensors. Microphones do not couple well with would; bimorphs are very fragile and piezoelectric sensors whilst narrowband and resonant appear to be the most sensitive. A 5mm larva can easily be detected more than 2m away in pine timber using a low-cost piezoelectric sensor. An example of a waterproof piezoelectric sensor is shown in Figure 6 and Figure 7 shows a recent implementation using an ultra-portable PC and USB-based external sound card to reduce system noise.

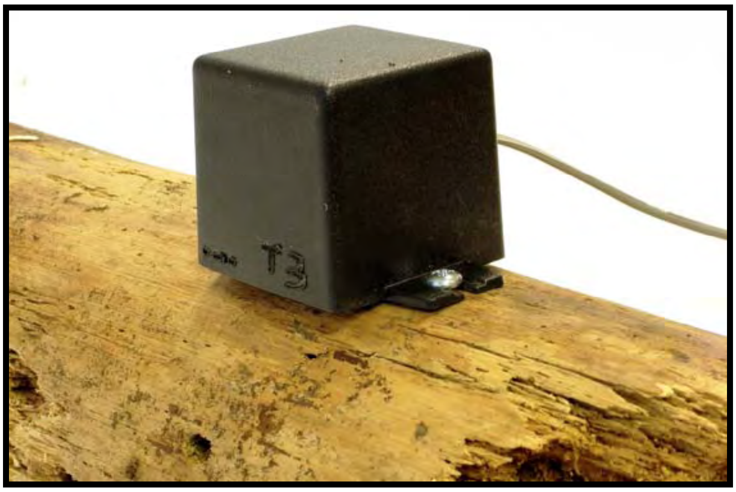


Figure 6: Sensor attached to branch.

Figure 7: Latest implementation of IBIS using Samsung Q1 ultra-portable PC and external sound card.

As noted in Section 3.2 the D-matrix is the feature set for input to the classification stage. Two types of artificial neural network were tested – a multilayer perceptron (MLP) and learning vector quantisation (LVQ) network. Each network was trained with a representative set of D-matrices from known species and then tested with unseen D-matrices from the species used in the training sets. LVQ outperforms MLP in most cases when the D-matrix is used; table 2 shows the misclassification identification matrix for two species using LVQ and a D-matrix. Results indicate that the species are unambiguously identified with accuracy no less than 97% (FARR & CHESMORE 2007). Table 3 is a similar matrix for three species using A-matrix features and MLP; here, identification is unambiguous and the accuracy approach 100%. Current work is under way to expand the species identification list.

Table 2: Mis-classification matrix for two species of beetle larvae using LVQ network and D-matrix.

	Hylotrupes bajulus	*Prionius coriarius*
Hylotrupes bajulus	98.70%	1.30%
Prionius coriarius	3.03%	96.97%

Table 3: Mis-classification matrix for three species of beetle larvae using MLP network and A-matrix.

	Hylotrupes bajulus	*Prionius coriarius*	*Rhagium bifasciatum*
Hylotrupes bajulus	0.86	0.0	0.14
Prionius coriarius	0.0	0.8	0.2
Rhagium bifasciatum	0.0	0.0	1.0

Applications in Ecology

Japanese Tibicen spp Cicadas

This section describes work carried out in collaboration with Dr Ohya of the biodiversity research group at the forestry and forest products research Institute in Morioka, Japan on the identification of three species of cicada of the *Tibicen* genus (*T. japonicus, T. bihamatus* and *T. flammatus*). The three species occur in the north of the Japanese main island of Honshu and are being used as indicators of diversity. The insects are usually seen high in the trees, making the capture difficult and therefore identification problematic. Dr Ohya's approach is to used principal component analysis using the peak and mean frequencies and the pulse rate of the songs. This work is described in (OHYA 2004) and Figure 8 is reproduced here with permission of Dr Ohya. The principal component analysis shows that of the 12 specimens recorded, 11 *T. bihamatus* and one closest to *T. japonicus*.

The same recordings were used in a TDSC system using a back propagation multilayer perceptron with inputs from an a matrix and with 10 neurons in the hidden layer. The results are shown in table 4 and are identical to the PCA analysis results. Results from a second series of recordings shown in table 5 indicate that *T. bihamatus* is accurately identified. Clearly more work needs to be done and more recordings analysed.

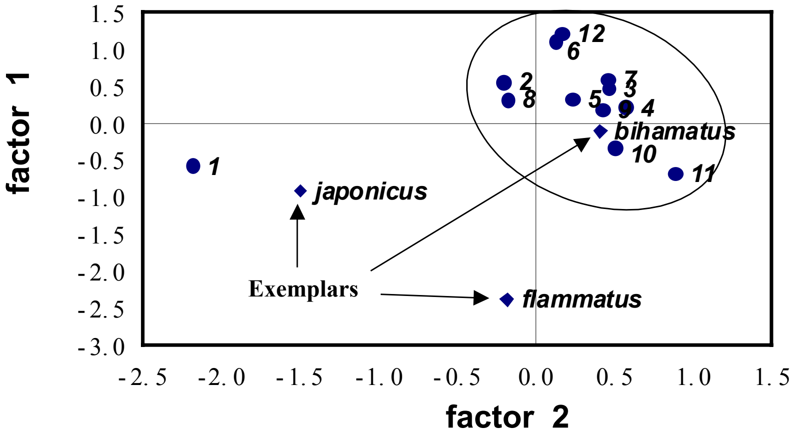


Figure 8: Results for PCA analysis of *Tibicen spp*. Adapted from (OHYA 2004). Sound file numbers correspond to those in table 4.

Table 4: TDSC identification of *Tibicen spp*. Sound file corresponds to the number in Fig. 8.

	Neuron Output (0.0 – 1.0)		
Sound File	***T. bihamatus***	***T. japonicus***	***T. flammatus***
1		0.98	
2	0.99		
3	0.99		
4	0.99		
5	0.99		
6	0.99		
7	0.99		
8	0.99		
9	0.99		
10	0.99		
11	0.99		

Table 5: Confusion matrix for the three *Tibicen* Species. The table should be read horizontally the species in the row being identified as a species in the column, e.g. *T. japonicus* is identified 50% of the time as *flammatus* and 50% of the time correctly.

	T. bihamatus	***T. flammatus***	***T. japonicus***
T. bihamatus	1.0	0.0	
T. flammatus	0.0	1.0	0.0
T. japonicus	0.0	0.5	0.5

British Grasshoppers

Work in the late 1990s show that it is possible to correctly identify 25 British species Orthoptera with up to 99% recognition accuracy (CHESMORE ET AL. 1997, CHESMORE, 2000, 2001; CHESMORE & NELLENBACH, 2001). However these were for high-quality signals with high signal to noise ratio and no interfering signals. In order to test the identification system under field conditions, recordings were made with Dr Ohya at local nature reserves in Yorkshire in 2002 using low-cost equipment (MiniDisc recorder and a stereo microphone). It was discovered that in addition to four species of grasshopper, there were many other interfering signals including birds vehicles and light aircraft. It was therefore decided to train

the system using exemplars of the four grasshopper species plus one blowfly sound, four bird sounds (alarm calls of three unknown species and the Chiffchaff), two vehicle sounds, one single engine light aircraft sound and one general background sound, making 13 sounds in total.
Analysis of the system's performance was carried out in two ways-performance for individual echemes and performance for whole song. Results for single echemes for each of the four grasshopper species are given in table 6, showing that *Omocestus viridulus* has 100% accuracy, *Chorthippus parallelus* 81%, *Myrmeleotettix maculatus* 90% and *Chorthippus albomarginatus* 85.7% for a threshold of 0.9. The output of each neuron has a range of 0.0 to 1.0 and a threshold can be applied to reject low neural output values. This has the effect of eliminating uncertain results and improving overall identification accuracy.

Table 6: Identification accuracy for single echeme for 4 species of grasshopper. The threshold is a value for the output neuron below which sounds are rejected.

Threshold		*O. viridulus*	*M. maculatus*	*C. parallelus*	*C. albomarginatus*
	sample size	34	17	14	16
0.5	**rejected**	2	4	1	1
	accuracy	100%	76.9% (10/13)	69.2% (9/13)	86.7% (13/15)
0.6	**rejected**	2	5	2	1
	accuracy	100%	75% (9/12)	75% (9/12)	86.7% (13/15)
0.7	**rejected**	3	5	2	1
	accuracy	100%	75% (9/12)	75% (9/12)	86.7% (13/15)
0.8	**rejected**	4	5	3	2
	accuracy	100%	75% (9/12)	81.8% (9/11)	85.7% (12/14)
0.9	**rejected**	4	7	3	2
	accuracy	100% (30/30)	90% (9/10)	81.8% (9/11)	85.7% (12/14)
None	**accuracy**	97% (33/34)	41.2% (7/17)	64.3% (9/14)	87.5% (14/16)

Figure 9 shows the results for recognition of a complete song, again using a threshold value. Best identification is again achieved with a threshold of 0.9, with identification accuracy between 80% and 100% for the four species. It must be remembered that the classifier has 13 outputs.

It is also possible to use the system in a different way by classifying sounds in a given period of time as illustrated in Figure 10 which shows 18 second sequence divided into two second blocks. The sound within each two second block is identified showing that, in this example, *Omocestus viridulus* is correctly identified in three blocks together with an aircraft sound and bird alarm calls. This approach leads to the possibility of identification on a continuous basis and has potential for more generalised sound mapping and identification. Full details of the results of the grasshopper identification work are given in (CHESMORE & OHYA 2004, CHESMORE 2007a).

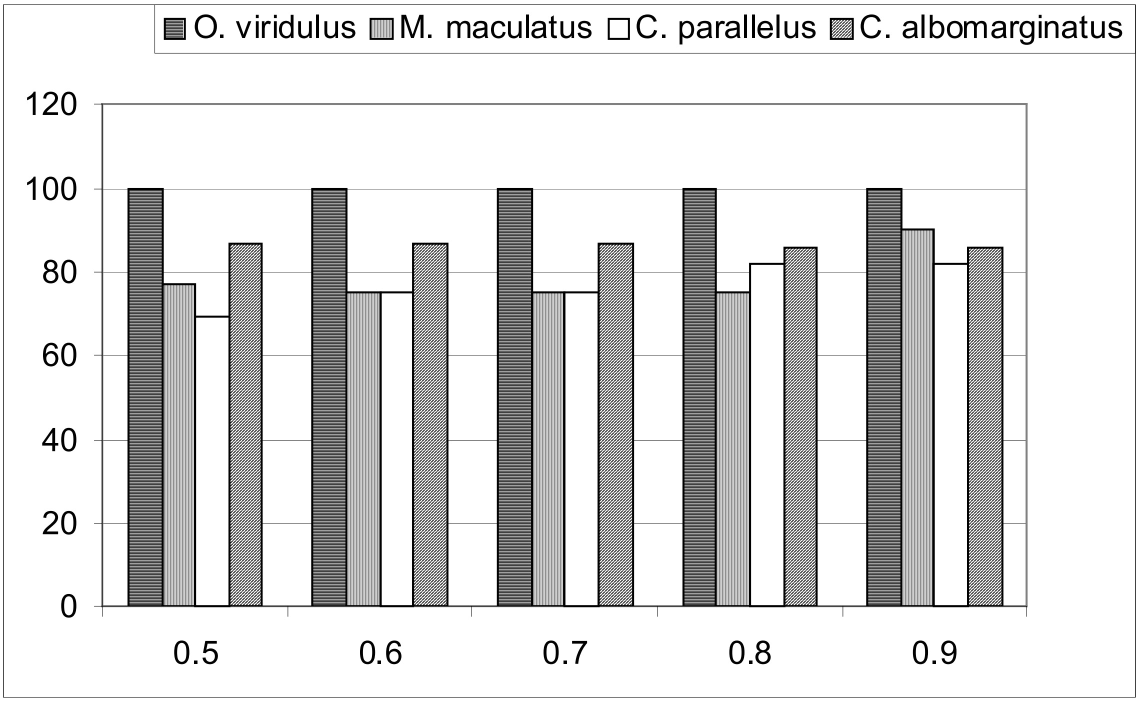
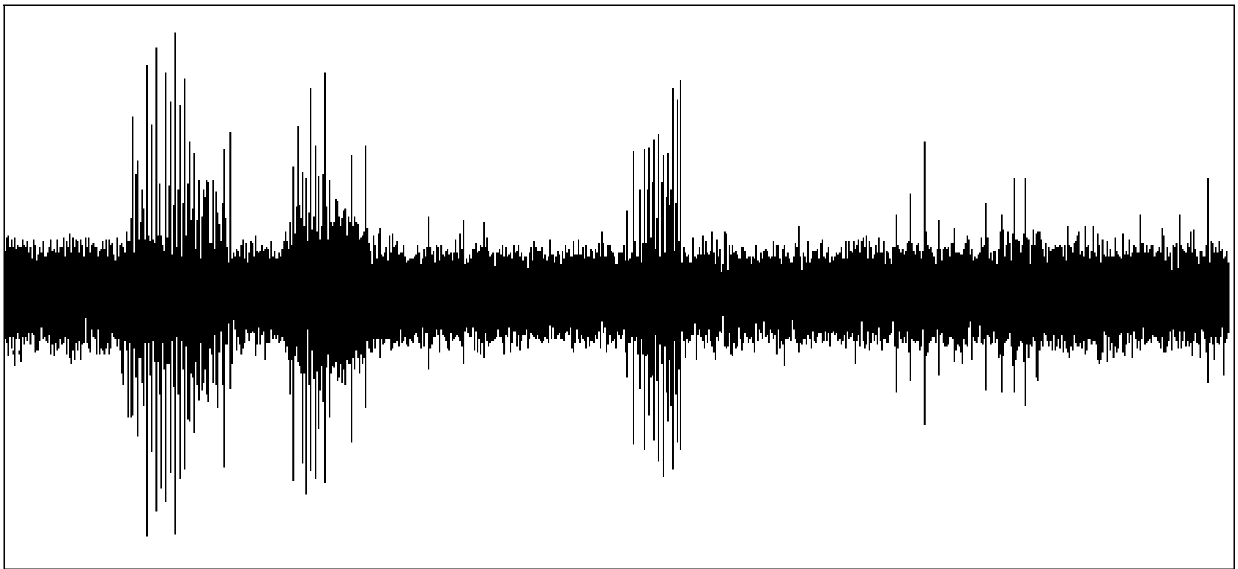


Figure 9: Identification accuracy for single echemes of four grasshoppers. The recognition accuracies are for four sounds out of a possible 13.



B11 0.65	Ov 0.999	Ov 0.87	PL 0.54	Ov 0.98	PL 0.96	B11 0.999	B11 0.999	B11 0.999

Figure 10: Sounds correctly recognised by system: three short songs by *Omocestus viridulus* (Ov), a light aircraft (PL) and a bird alarm call (B11) (unknown species). The sound is an 18s sequence analysed on a 2s interval, recorded at Allerthorpe Common on 15 July 2002.

Conclusions and Future Directions

This paper has discussed the successful use of TDSC and artificial neural networks for the automated identification of a range of taxonomic groups. It is shown that TDSC is computationally efficient and provides a good feature set for signal classification. The range of applications of TDSC is not limited to bioacoustics but can be applied to any bandlimited signal. The research described here falls into two application areas-detection and identification of insect pests and ecological studies such as biodiversity assessment.

The future direction of this work lies primarily in five areas:

a) Improving the reliability of identification of individual taxa through the automated detection and identification of interfering signals. Progress towards this has been shown in the grasshopper research detailed in Section 5.2. More work on this is being carried out through an EPSRC research grant in collaboration with the Universities of Southampton and Newcastle to develop a real-time instrument for identification of sounds within a soundscape. The instrument uses a sound field microphone and will be capable of locating the source in three dimensions, leading to the possibility of active removal of interfering signals from other directions. Whilst this is at present focused on man-made sounds, natural sounds are included and it is will be feasible to train the system for specific taxonomic groups.

b) The development of real-time handheld species identification systems that can be reconfigured for different taxonomic groups. The trend towards lower cost tablet PCs with enhanced functionality (WiFi, cameras, etc) will make such systems more feasible and cost effective.

c) Increasing the size of the identification space. The author has several target taxonomic groups in mind and is currently writing a research proposal for the identification of all European bat species using echolocation calls. Preliminary work carried out at York university through an undergraduate project has indicated that at least 10 of 14 so British species can be identified (unpublished work). Identification of the *Myotis* species group is still problematical and will be the focus of the proposed research. Work is also underway to develop an identification system for all Japanese Orthoptera. Increased taxonomic group size is likely to lead to an increase in overlap in the feature space and work is under way to improve TDSC to overcome this and to also include other features.

d) Investigating methods for overcoming the problem of simultaneous singing taxa. This is a nontrivial problem and is likely to be the cause of the failure of most automated identification systems. Possible ways forward include 3-D sound localisation as mentioned in a) above and utilising knowledge of acoustic niche theory to provide partitioning of the frequency spectrum and therefore separation of some groups. However, this does not solve the problem of simultaneously singing taxa of the same species or those with significant overlap in frequency.

e) The development of sensor networks for, for example, the instrumentation of large areas such as forests or for “ ring fencing” of protected areas against the intrusion of pests.

In summary, the author believes that bioacoustic identification is now becoming a practical solution for many application areas, and is likely to make significant progress in the next few years.

One major issue that has been problematic over many years particularly in the UK is that of lack of funding. This type of research is considered to be multidisciplinary and until recently has been difficult to obtain funding from any of the research councils. Even though the funding situation is slowly changing, it is important to be able to access funding possibly on a European level in order to be able to develop demonstrators.

Acknowledgements

The author would like to thank Dr Ohya for his help in the grasshopper research and for giving permission to use Figure 8.

References

CHESMORE, E. D. (1999): Technology Transfer: Applications of Electronic Technology in Ecology and Entomology for Species Identification. - Natural History Research **5**: 111-126.

CHESMORE, E. D. (2000): Methodologies for automating the identification of species. – In: Chesmore, E.D., Yorke, L., Bridge, P. and Gallagher, S.(eds.) Proceedings of Inaugural Meeting of the BioNET - INTERNATIONAL Group for Computer-aided Taxonomy (BIGCAT), 1997: 3-12.

CHESMORE, E.D. (2001): Application of time domain signal coding and artificial neural networks to passive acoustical identification of animals. - Applied Acoustics **62**: 1359-1374.

CHESMORE, E.D. (2004): Automated bioacoustic identification of species. - Anais da Academia Brasileira de Ciencias **76**: 435-440.

CHESMORE, D. (2007a): Robust in-field automated bioacoustic identification of species for rapid biodiversity studies. - Proceedings of the 4th International Conference on Bioacoustics 2007, Institute of Acoustics, Loughborough, 29 (3), 10-12 April, 2007, University of Loughborough: 61-68.

CHESMORE, E. D. (2007b): The Automated Identification of Taxa: Concepts and Applications, in Automated Taxon Identification in Systematics. - CRC Press: 85-100.

CHESMORE, E.D. & NELLENBACH, C. (2001): Acoustic Methods for the Automated Detection and Identification of Insects. - Acta Horticulturae **562**: 223-231.

CHESMORE, E.D. & OHYA, E. (2004): Automated identification of field-recorded songs of four British grasshoppers using bioacoustic signal recognition. - Bulletin of Entomological Research **94**: 319-330.

CHESMORE, E. D., SWARBRICK, M. D. & FEMMINELLA, O. P. (1997): Automated analysis of insect sounds using TESPAR and expert systems - a new method for species identification. -In Information Technology, Plant Pathology and Biodiversity, Bridge, P. et al.), CAB International, Wallingford, UK: 273-287.

COLEBROOK, F.M. (1937): The aural detection of the larvae of insects in timber. - Journal of Scientific Instruments **14**: 119-121.

FARR, I.J. (2007): Automated bioacoustic identification of statutory quarantined insect pests. - PhD Thesis University of York.

FARR, I., CHESMORE, D., HARVEY, D., HAWES, C. & GANGE, A. (2005): Bioacoustic detection and recognition of Stag Beetle (*Lucanus cervus*) larvae underground using vibration sensors. -Royal Entomological Society National Meeting: Entomology 2005, 12-14 September 2005, University of Sussex.

FARR, I. & CHESMORE, D. (2007): Automated bioacoustic detection and identification of wood-boring insects for quarantine screening and insect ecology. - Proceedings of the 4th International Conference on Bioacoustics 2007, Institute of Acoustics, Loughborough, 29(3), 10-12 April, 2007, University of Loughborough: 201-208.

FISCHER, F.P., SCHULZ, U., SCHUBERT, H., KNAPP, P. & SCHMOGER, M. (1997): Quantitative assessment of grassland quality: acoustic determination of population sizes of orthopteran indicator species. - Ecological Applications **7**: 909-920.

GARDINER, T., HILL, J. & CHESMORE, D. (2005): Review of the methods frequently used to estimate the abundance of Orthoptera in grassland ecosystems. - Journal of Insect Conservation **9**: 151-174.

GASTON, K. & O'NEILL, M.A. (2004): Automated species identification -why not? - Philosophical Transactions of the Royal Society B **359** (1444): 655-667.

HAACK, R.A., LAW, K.R., MASTRO, V.C., OSSENBRUGGEN, H.S. & RAIMO, B.J. (1997): New York's battle with the Asian long-horned beetle. -Journal of Forestry **95**: 11-15.

HAGSTRUM, D.W., VICK, K.W. & WEBB, J.C. (1990): Acoustic monitoring of *Rhizopertha dominica* (Coleoptera: Bostrichidae) populations in stored wheat. - Journal of Economic Entomology **83**: 625-628.

HICKLING, R., LEE, P., VELEA, D., DENNEBY, T.J. & PATIN, A.I. (2000): Acoustic system for rapidly detecting and monitoring pink bollworm in cotton bolls. - Proceedings of the Beltwide Cotton Conference San Antonio: 1-10.

KING, R.A. & GOSLING, W. (1978): Time-encoded speech. - Electronics Letters **15**: 1456-1457.

MACLEOD, A. EVANS, H. F. & BAKER, R. H. A. (2002): An analysis of pest risk from an Asian longhorn beetle (*Anoplophora glabripennis*) to hardwood trees in the European community. - Crop Protection **21**: 635-645.

MANKIN, R.W. BRANDHURST-HUBBARD, J., FLANDERS, K.L., ZHANG, M., CROCKER, R.L., LAPOINTE, S.L., MCCOY, C.W., FISHER, J.R. & WEAVER, D.K. (2000): Eavesdropping on insects hidden in soil and interior structures of plants. - Journal of Economic Entomology **93**: 1173-1182.

MANKIN, R.W. & WEAVER, D.K. (2000): Acoustic detection and monitoring of *Cephus cinctus* sawfly larvae in wheat stems. - International Congress of Entomology, Brazil.

OBA, T. (2004): Application of automated bioacoustic identification in environmental education and assessment. -Anais da Academia Brasileira de Ciencias **76**: 445-451.

OHYA, E. (2004): Identification of *Tibicen* cicada species by a principal components analysis of their songs. -Anais da Academia Brasileira de Ciencias **76**: 441-444.

RIEDE, K. (1993): Monitoring biodiversity analysis of Amazonian rainforest sounds. - Ambio **22**: 546-548.

SHUMAN, D., COFFELT, J.A., VICK, K.W. & MANKIN, R.W. (1993): Quantitative acoustical detection of larvae feeding inside kernels of grain. - Journal of Economic Entomology **86**: 933-938.

SHUMAN, D. WEAVER, D.K. & MANKIN, R.W. (1997): Quantifying larval infestation with an acoustical sensor array and cluster analysis of cross-correlation outputs. - Applied Acoustics **50**: 279-296.

From Bird Species to Individual Songs Recognition: Automated Methods for Localization and Recognition in Real Habitats Using Wireless Sensor Networks

VLAD M. TRIFA[1], ALEXANDER KIRSCHEL[2], YUAN YAO[2], CHARLES TAYLOR[2] & LEWIS GIROD[3]

[1]Institute for Pervasive Computing, ETH Zürich, Haldeneggsteing, 8092 Zürich,SAP Research CEC Zürich, Kreuzplatz 20, 8008 Zürich,Switzerland, E-mail: vladounet@gmail.com
[2]Department of Ecology and Evolutionary Biology,University of California Los Angeles (UCLA), 621 Charles Young Drive South,90095 Los Angeles, CA,USA
[3]Computer Science and AI Lab,Massachusetts Institute of Technology (MIT),02139 Cambridge, MA,USA

Abstract. The recent advances in wireless networked sensing systems, combined with the ever increasing computational power of embedded systems has brought field biologists a large palette of opportunities for automated analysis of ecosystems. Now phenomena can be observed in real time and at their heart, where they happen. Our work presents such an attempt at combining wireless sensors with automated detection, localization, and classification of bird songs and we have successfully used Hidden Markov Models to identify tropical bird species in noisy environments with high success rates. In the context of behavioural studies, when social interactions between different birds are to be tracked, a very challenging goal is to be able to identify the individuals of the same species. We will give several examples from our recent progress, and discuss the difficulties that arise with such delicate processes. Additionally, we will provide some suggestions that could help increase the correct classification rates when using distributed sensor nodes.

For several decades, field biologists have attempted to design methods and tools that would help them collect data in real environments. More particularly, along with the growing computational power in computers, it became realistic to attempt automated detection and classification of bioacoustic signals. In many cases, the natural environments can be extremely noisy and many other sounds other than the target animals can be present at the same time.

Language is thought to be the key element in the development of the human intellectual properties. Throughout history, scientists and philosophers have always been interested in deciphering the mysteries of communication, and especially understanding how new languages emerge and evolve, while other disappear. Even if language *per se* is an exclusive characteristic of humans, many animals rely on communication for survival (CATCHPOLE & SLATER 1995). Countless examples support the hypothesis that animal communication is shaped by evolution (LEE ET AL. 2005), where only individuals that find an optimal communication strategy for a given environment are able to survive. For example, social interactions are detrimental in the bird song learning process (BEECHER & BURT 2004). Identification of the factors that influence the evolutionary process of communication, such as sexual and natural selection, could provide significant insights about the neural mechanisms underlying social behaviours, making this field of research very interesting for scientists in many domains including biology, linguistics, and neuroscience.

In particular, when the interest of biologists is focused on social behaviour of birds, it is not limited to the automation of detection and recognition of individual acoustic signals: real-time motion tracking provides important information about behaviour. Many methods for automated bird song analysis have been proposed (HÄRMA 2003, SOMERVUO & HÄRMA 2004). Unfortunately, few of them were able to cope with the unpredictability of real environments. For this purpose, we need new methods to acquire a more accurate representation of phenomena that is simply not possible using traditional recording techniques, such as manual recording and complex microphone arrays.

The National Science Foundation (NSF) sponsors an interdisciplinary project centered at UCLA that aims to create novel methodologies to help scientists study animal behaviour and communication, and to understand how communication evolves in correlation with factors such as environmental and social interactions. This project focuses mainly on birds, as they have been extensively studied and are relatively well known, so they provide an excellent test bed to analyze the transmission and perception of acoustic signals in noisy environments. Also, as a long-term goal it will be interesting to study how the structure of birds vocalizations could be used to design methods optimized to achieve these goals. Contrary to common belief that bird songs are frivolous, avian communication is the result of an impressive evolutionary process, where natural selection directly operates on the quality of males, and where only the most talented sopranos and the healthiest individuals will be selected by females for mating.

The goal of our project is to create a palette of tools that support field biologists in the study of birds' communication and behaviour in their natural habitats (TRIFA ET AL. 2007). Since human presence in the field introduces biases, non-intrusive techniques are critical for acquiring clean data. The reduced form factor and power consumption and enhanced capabilities of embedded sensor networks are opening a new world of possibilities for revolutionary environmental monitoring methods. However, many challenges remain in improving the robustness and adaptability of these devices and thus reducing human intervention and operational costs.

Tools and Methods

Because our focus is to provide field biologists a platform that can be used for the analysis of social interactions in birds, our tools must be placed at the heart of the phenomenon to observe, with minimal influence on the ecosystem to avoid biased results. For this type of application we require a tiny embedded computer that can sample and process data locally and exchange results wirelessly. To support high-quality audio sampling these systems need a sufficient amount of storage and processing power in order to buffer the data, process it on the fly, and either process it completely, forward it to a storage server, or archive it locally.

Acoustic EnsBox

The Center for Embedded Networked Sensing at UCLA has developed such a platform called Acoustic EnsBox. This platform is based on the Stargate processing module and a high-quality multichannel sound card with an external microphone array. The Stargate is a computational platform similar to a PDA, based on the 400 MHz PXA255 XScale processor with 64 MB of SDRAM, 32MB of flash memory. Audio sampling is done with a VXpocket 440 PCMCIA sound card, and a Compact Flash 802.11 card is used to support a wireless network. The VXpocket has 4 balanced mic/line analog inputs, provides 16-bit audio measurement at sampling frequencies ranging from 8kHz to 48 kHz in 100 Hz steps. We chose the LinearX M53 for our low noise, low distortion measurement microphone. Each M53 microphone is calibrated using a free-field comparison procedure with a laboratory grade reference microphone. The calibration produces a precision error response curve which can be used for correcting the response.

The Stargate runs a Linux 2.6 kernel with ALSA driver support for the VXPocket440. By layering our own user-space audio service over the ALSA API, our software provides a sampling interface that is accurately time-synchronized across several wirelessly connected nodes. On each node we achieve tight synchronization of the 4 local channels, and across nodes we achieve a synchronization accuracy of approximately 10 µs.

Development has continued on the ENSBox (GIROD ET AL. 2006), including a number of recent improvements to the packaging, the results of which are currently under review.

Emstar

Emstar is a comprehensive software environment for developing heterogeneous, distributed applications (GIROD ET AL. 2004, 2007). Emstar provides tools for simulation, emulation, and visualization of Emstar based distributed systems. It also provides many services such as networking and time synchronization across nodes. The strengths of Emstar lie in its message passing inter process communication (IPC) primitives. Each Emstar based system consists of multiple logically separable modules implemented as individual processes which enhances system robustness by allowing modules to independently fail and restart. Modules communicate with one another using message passing.

Full System Architecture

The whole project briefly presented here is based on the work of countless researchers and projects, however the initial attempt towards a unified system architecture has been developed during a master thesis and technical details can be found in TRIFA (2006). The general idea is to automate and chain the different processing stages involved in logging of individual bird songs in a single software workflow.

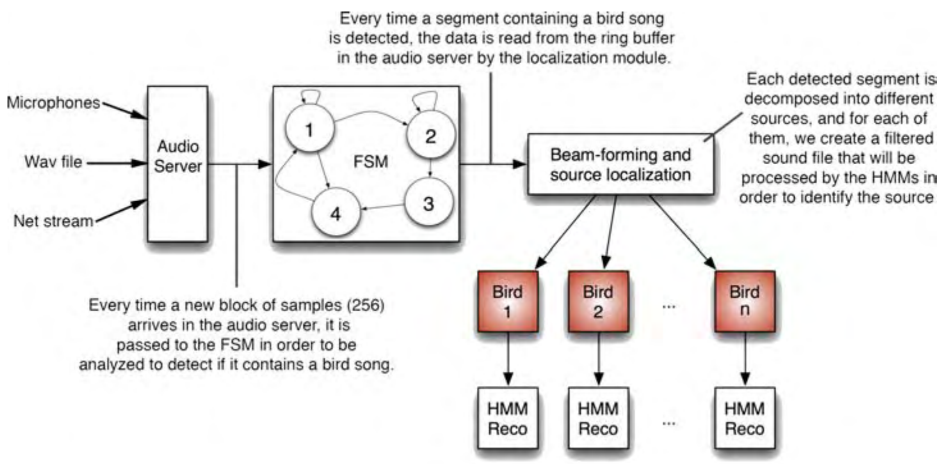


Figure 1: The workflow of the whole architecture developed during this project. The Audio Server task is to store all the incoming audio data (the data source is chosen when the server is started) into an indexed ring buffer. Then the incoming signals are processed to detection eventual bird songs. Every time a song is detected, it will be localized and beamformed, and each resulting sound file will be processed with HMM to recognize the species, and finally stored along with other metadata.

Experiments and Properties

In this section we describe two central aspects of this project. First, we briefly describe the collaborative localization algorithm to give an insight on how birds can be localized with improved accuracy when several sensor nodes are used to localize animals. Second, we give some results in the differences when HMMs are used to recognize species vs. individuals.

Collaborative Localization

Given that localization of sound sources is essential for fine-grained analysis of social interactions between birds, accurate localization has been of great interest for our project (WANG ET AL. 2004). The department of Electrical Engineering at UCLA has developed an efficient method that to cope with the localization of wide-band signals, even when they overlap in time and frequency (CHEN ET AL. 2006). The so-called Approximate Maximum Likelihood (AML) is a sound localization technique found to be very effective when it comes to separate each individual song from the many sources of noise in recordings from tropical rain forests. Furthermore, AML can be also used to perform beamforming (i.e. amplify all sound signals emanating from a particular spatial direction, while attenuating noise coming from other directions). Beamforming can greatly improve the quality and signal-to-noise ratio of recorded vocalizations.

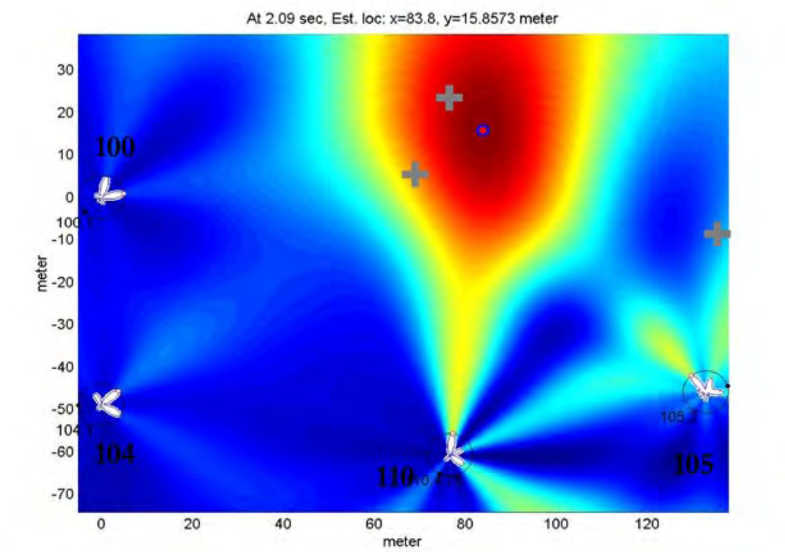


Figure 2: Results of the collaborative localization algorithm, presented as a 2D pseudo-likelihood map. One can see that the individual estimations of the angle of arrival (AOA) for each node (black lobes represent the likelihood for source AOA) are combined using their location as estimated by the self-calibration process.

The EnsBox has another very useful property that makes the whole system easy to deploy: the nodes are able to automatically localize each other (details of this process are to be found in GIROD ET AL. (2006)) to map the relative location of each node. Once each individual node has estimated the local Angle of Arrival (AOA) of an incoming sound signal, these estimates are recombined on the geographic map to derive the most likely actual location of the sound source relative to the nodes (ALI ET AL. 07).

Recognition of Bird Songs

We have run many recognition tests, with many types of songs, and in particular we have investigated the performance of hidden Markov models (HMM) with various parameters and data. The HMM were implemented using the free HTK toolkit from Cambridge University, an easy to install, yet powerful implementation of hidden markov models (YOUNG ET AL. 2002).
We have focused on mainly two types of experiments. In the first case we were interested only in recognizing the species of different songs. For that we have used several recordings of six types of tropical antbirds (Great Antshrike, Barred Antshrike, Dusky Antbird, Dot-winged Antwren, and the Mexican Antthrush), which were recorded in the Biosfera Monte Azules Natural Reservation in Chiapas, Mexico, during June 2005 and February 2006. For more information about the use of HMMs for species recognition, recent results are currently in revision (TRIFA ET AL. 2008).
The second type of experiment focuses on the recognition of the different individuals within the same social group of Acorn Woodpeckers (ACW). The acorn woodpeckers are found in California and the recordings were made in the Hastings Reserve, near Monterrey, CA. Results for the application of HMM for individual recognition are currently in preparation (YAO ET AL. in prep.). It is worth pointing out that the birds themselves are able to recognize the

identity of each caller based on different features contained in their songs (BLUMSTEIN & MUNOS 2005, NELSON 1989).

The challenges involved in these two types of experiment are actually quite different. Since we want to provide only a high-level overview in this article, the interested reader is invited to consult TRIFA ET AL. (2008) and YAO ET AL. (in prep.) for the technical details and detailed analysis. We also assume the reader to be familiar with the basics of HMMs; if not RABINER (1989) provides a very good introduction.

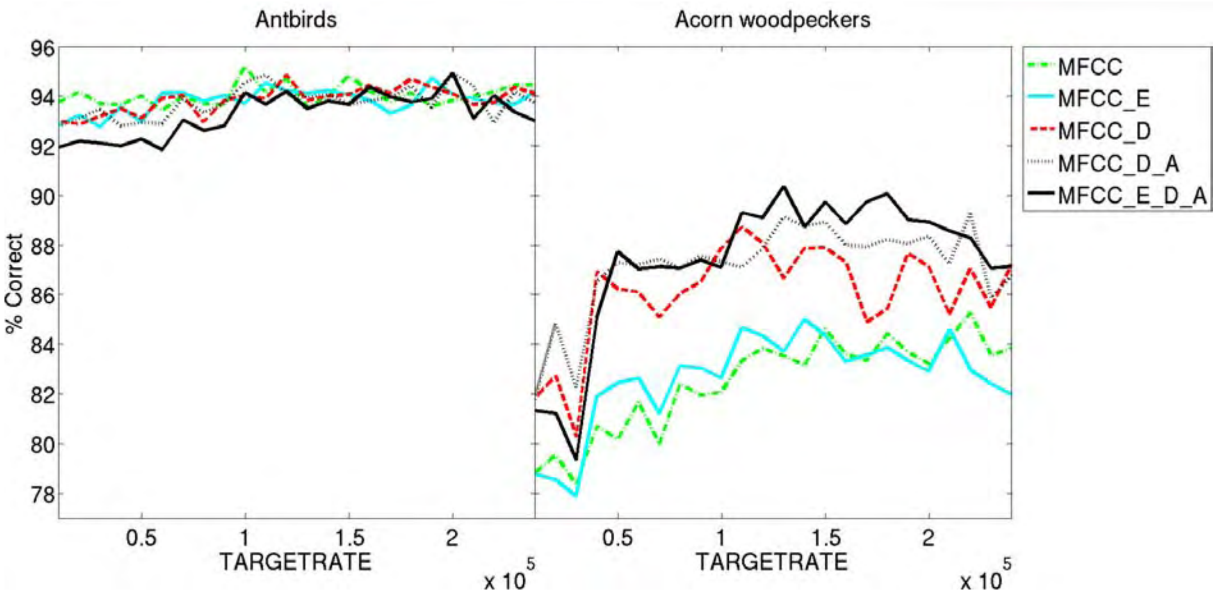


Figure 3: Comparison of recognition performance for different prefiltering methods. On the horizontal axis the overlapping between two consecutives frames used for feature extractionl. *Left:* average species recognition (% of correct classification). *Right:* the same but for several individuals. (Image: courtesy of Vlad Trifa, reprinted from TRIFA (2006)).

We have found that the in the case where only the species is to be recognized, the difference between different feature extraction methods is hardly noticeable, and is rather stable between 92% and 95% over a wide range of parameters. This result is especially impressive, considering that only 10 samples were used to train the models. The success of these techniques can be attributed to the many structural differences in the songs across species. The samples from each species are tightly clustered around different positions in the feature space, with a high separability between the clusters even when the parameters are estimated very roughly. Thus, changing the value of single parameter is not likely to have a large impact on the separability between these clusters. Also, adding other features such as the delta and the acceleration to the classification will not add discrimative power, as these features are likely to differ greatly between species, and does not bring much information to the classifier, but only increase the probability of correct classification. This hypothesis is partially supported by a quasi-perfect recognition of species, which shows the great separability between classes. Manual checking of the misclassified songs revealed that their SNR is very low and the bird songs in these recordings were barely perceptible to the human ear.

However, in the case of individuals of the same species, the structure of calls is very similar, and the separability between clusters is very small. In this case, slight changes to the value of a parameter can yield a great impact on the separability between different classes, and adding additional features can increase the discriminative power and thus the separability between the classes. It interesting to point out that adding the deltas improves the performance in the case of ACW, and the acceleration improves it even slightly more, while the energy does not seem to have an influence. This might give us some insight about what the way individuality is encoded in ACW calls. When the separability between classes is smaller, more training examples are required in order to correctly estimate the parameters of the HMMs and achieve accurate classification.

An analysis of instances of the misclassifications revealed some cases that a human expert could hardly distinguish on basis of their songs. The reason is that one bird was the parent of the other, and thus the similarity could be explained by the fact that children learn their songs by imitating the songs of their parents, though evidence that support this hypothesis is not given here.

The vertiginous drop of performance that can be seen for low values of target rate shows one of the main weakness of HMMs. Decreasing this parameter for a constant window size will increase the overlap between consecutive observations, thus also increasing the correlation between them. But, the Markov models are based on the Markov assumption where each observation is statistically independent from the previous one. Unfortunately, this assumption does not generally hold with real signals such as as human speech or bird songs, even without overlapping between the samples used to extract the observation vectors. However, we have shown that even when the Markov assumption is violated, classification with HMMs can be successful as long as the separability is sufficiently high.

We also ran tests with LPC methods and found out that these methods yield much lower performance than MFCC method, although these results have been omitted for brevity. LPC is less appropriate for bird song modelling, as it is based on more complex information as opposed to energy in different frequency ranges. By no means is MFCC a biologically relevant and accurate model of the animal auditory system, but several studies have shown that dominant frequency might be one of the most efficient acoustic cues used by animals in individual identification, and this could explain the impressive results we have achieved so far. There is also reason to believe that mechanisms to extract frequency information from signals are easier to implement biologically than those based on linear prediction coefficients, since neural pathways can act as natural delay lines to build frequency selective filters.

In summary, we have found that very good species recognition performance - over 95% of correct classification when 6 species are considered - can be obtained with HMMs, even when very few samples (15 songs) are used to train the models. This technique works well even when many of the samples used have a low SNR, and the bird song can hardly be distinguished from the background noise. However, we noticed that some conditions, such as incomplete songs, complicate the species recognition task. These promising results are most likely due to the structural differences in vocalizations across species, and thus effective clustering in the feature space.

It should be noted that the results presented here are only an initial indication of how recognition might be affected by several factors. The amount of data used to test species recognition (25 samples per species) is very small and statistically irrelevant of the real influence of these parameters upon global performance. To have a more precise idea of the performances of our system in real field experiments, more tests should be performed using a larger audio database with greater variation in the amount and nature of background noise.

Discussion

Research on embedded systems for signal processing is not a new research topic. Over the years, many powerful frameworks and tools have been invented for specification, modeling, simulation, verification, and code generation of signal processing embedded system. There are also many specially designed high-level programming languages to simplify digital signal processing. Space limitations preclude a thorough discussion of the other important work in embedded systems.

Nonetheless, many significant differences between the traditional embedded system for signal processing and our work on collaborative processing in sensor networks. Many such traditional design methodologies and tools often use specially designed platforms, while Emstar targets general purpose PDA-class platforms running Linux. The software used in this study ran in MATLAB, and thus was strictly for use in off-line processing. However, by porting the algorithms to C it could be possible to perform classification and recognition on nodes running in the field. This might yield advantages if it allowed the system to run for longer durations, or enabled additional sensing modalities to be engaged in the event that certain species or individuals were recognized. Integration of a DSP pre-processor to perform the signal processing, as is filtering, as well as other hardware improvements, would allow the main CPU to be used for more critical tasks and would greatly improve power efficiency, leading to a longer system lifetime.

We found out that individual recognition is by far a more delicate process than species recognition. In the best cases, no more than 90% of correct classification was obtained. This is due to the fact the structure of the calls among individuals are very similar, and the features extracted from the signals might not be the ones that convey individual identity. In contrast, even with only 10 samples to train the HMMs with the AB, we had over 97% of correct species classification. Also, we noticed that many more examples are needed to train the models in the case of individual recognition. Because the differences between songs are more subtle, more data is required to correctly estimate these slight differences. Increasing the overlap between consecutive blocks of data over which features are extracted might help to reduce the variability encountered in the experimental results. Increased overlap will increase the correlation among states and will reduce sensitivity to noise, which is the main source of error in automated recognition methods.

Further perspectives on this project include improvement of the performance of the different algorithms used, and a better interaction between the different processing stages, helped by a modular approach. Performance could be further improved using an appropriate design of the microphone array that would take into account the characteristics of bird songs.

We suggest that the standard feature extraction proposed by HTK (MFCC) is a very generic method and does not reflect at all the features that are actually used by birds to identify songs of callers. If one could design a feature extraction module based only on those relevant features that are used to encode individuality, performance could be significantly improved. Data mining techniques could give us information about what these relevant features are, and based on this information much more appropriate feature extraction methods that look only for the cues that differ significantly across individuals could be devised.

Conclusion

We have briefly described an architecture for automated detection, localization, and recognition of bird songs that can be used both with traditional recording systems (single directional microphone, microphone array, etc.) and also with distributed sensor networks. We have described how such novel computing paradigms can help to improve the accuracy and ease of use of existing technologies for environmental monitoring.

In the long run, to unlock the amazing potentials of embedded systems, one could focus on the development of swarm-intelligent algorithms that can take full advantage on the distributed nature of sensor networks in order to provide unmatched robustness and scalability to the system, while reducing the complexity of the algorithms used (TRIFA 2004).

Also, one could also investigate adaptive communication protocols that grounds sensor data into concepts that are shared via a symbolic adaptive language, which minimizes the data to transmit according to changes in the environment, exactly as birds do! But we need first to know how they do it. Therefore, understanding the neural mechanisms involved in human and animal language acquisition, production, and recognition would be a considerable milestone for the whole scientific community.

Acknowledgements

The authors would like to thank all the talented and helpful people that contributed to this ambitious project – they are way too many of them to be all quoted here. This research was supported by the National Science Foundation under Award Number 0410438.

References

ALI, A. M., K. YAO, T. C. COLLIER, C. E. TAYLOR, D. T. BLUMSTEIN, & L. GIROD (2007): An empirical study of collaborative acoustic source localization. - In IPSN '07: Proceedings of the 6th international conference on Information processing in sensor networks, pages 41–50, New York, NY, USA, ACM.

BEECHER M. D. & J. M. BURT (2004): The role of social interaction in bird song learning. - J. American Psychological Society, **13**: 224–228.

BLUMSTEIN D. T. & O. MUNOS (2005): Individual, age and sex specific information is contained in yellow-bellied marmot alarm calls. Animal behaviour **69**:353–361.

CATCHPOLE C. K. & P. J. B. SLATER (1995): Bird song: biological themes and variations. Cambridge University Press.

CHEN, C.-E., A. ALI, W. ASGARI, H. PARK, R. E. HUDSON, K. YAO & C. TAYLOR (2006): Design and testing of robust acoustic arrays for localization and enhancement of several bird sources. – In: Fifth International Conference on Information Processing in Sensor Networks.

GIROD, L., J. ELSON, A. CERPA, T. STATHOPOULOS, N. RAMANATHAN, & D. ESTRIN (2004): Emstar: a software environment for developing and deploying wireless sensor networks. - Proceedings of the 2004 USENIX Technical Conference.

GIROD, L.D., M. LUKAC, V. TRIFA, & D. ESTRIN (2006). The design and implementation of a self-calibrating distributed acoustic sensing platform. - In The 4th ACM Conference on Embedded Networked Sensor Systems (SENSYS 06), Boulder, CO, USA, 2006.

GIROD, L., N. RAMANATHAN, J. ELSON, M. LUKAC, D. ESTRIN (2007): EmStar: a software environment for Developing and Deploying Heterogeneous Sensor-Actuator Networks. - ACM Transactions on Sensor Networks, August 2007.

HÄRMA, A. (2003): Automatic identification of bird species based on sinusoidal modeling of syllables. - In IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP 2003).

LEE,Y., T. C. COLLIER, G. M. KOBELE, E. P. STABLER & C. E. TAYLOR (2005): Grammar structure and the dynamics of language evolution. In - Proc. of the 8th European Conference on Artificial Life, 2005.

NELSON, D. A. (1989): The importance of invariant and distinctive features in species recognition of bird song. - The Condor **91**:120–130.

RABINER, L. E. (1989): A tutorial on hidden markov models and selected applications in speech recognition. – In: IEEE (editor) Proceedings of the IEEE, volume 77, February 1989.

SOMERVUO, P. & A. HÄRMA (2004): Bird song recognition based on syllable pair histograms. – In: IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP 2004).

TRIFA V. (2004): Low-power sensor networks for collective hearing. Semester project, Ecole Polytechnique Fédérale de Lausanne.

TRIFA, V. (2006): A framework for bird songs detection, recognition and localization using acoustic sensor networks. - Master's thesis, University of California Los Angeles and Ecole Polytechnique Fédérale de Lausanne.

TRIFA, V., L. GIROD, T. COLLIER, D. BLUMSTEIN & C. TAYLOR (2007): Automated wildlife monitoring using selfconfiguring sensor networks deployed in natural habitats. In Proceedings of the International Symposium on Artificial Life and Robotics (AROB 12th 2007), Beppu, Japan.

TRIFA, V., A. KIRSCHEL, & C. E. TAYLOR (2008 in press): Automated species recognition of antbirds in a mexican rainforest using hidden markov models. - Journal of the Acoustical Society of America.

WANG, H., C. CHEN, A. ALI, S. ASGARI, R. HUDSON, K. YAO, D. ESTRIN, & C. TAYLOR (2004): Acoustic sensor networks for acoustic woodpecker localization. In SPIE conference on Advanced Signal Processing Algorithms, Architectures and Implementations, 2005.

YOUNG, S., D. KERSHAW, J. ODELL, D. OLLASON, V. VALTCHEV, & P. WOODLAND (2002): HTK Book 3.2.1. Cambridge University Engineering department.

YAO, Y., Y. LIN, V. TRIFA, & C. E. TAYLOR (in prep.): Vocal individuality in acorn woodpecker (*Melanerpes formicivorus*) recognized by hidden markov models.

Advantages and Disadvantages of Acoustic Monitoring of Birds – Realistic Scenarios for Automated Bioacoustic Monitoring in a Densely Populated Region

KARL-HEINZ FROMMOLT[1], KLAUS-HENRY TAUCHERT[2] & MARTINA KOCH[2]

[1] Humboldt-Universität zu Berlin, Museum für Naturkunde, Invalidenstr. 43, 10115 Berlin, Germany
[2] Humboldt-Universität zu Berlin, Institut für Biologie, Invalidenstr. 43, 10115 Berlin, Germany
E-Mail of the corresponding author: karl-heinz.frommolt@rz.hu-berlin.de

Abstract. A wide variety of methods have been developed to evaluate breeding bird populations (territory mapping, point stop counts, line transects, etc.). These methods allow a good evaluation of changes in population density across a large number of species. However, these methods are often very time-consuming, and sometimes they still fail to yield reliable results. The application of acoustic monitoring methods can overcome some of these difficulties.
Specific advantages of bioacoustic monitoring (BM) are listed below:
- BM allows long-term recording in absence of an observer
- BM can be used in areas that are difficult to access (e.g., large reed habitats)
- BM observation data could be verified even after years
- BM ensures minimal subjectivity, due to independence from the skills of the observer
- BM tasks could be automated using pattern recognition software in the future

Based on our experience with bioacoustic monitoring, we describe some situations to highlight why this approach represents a powerful addition to the traditional methods. Specifically, we recommend using this methodology for counting owls and birds living in reed belts.

We live in a world where the impact of human activity on natural resources and climate is almost unavoidable. It's all but impossible to imagine a modern lifestyle that does not have detrimental effects on the natural environment. Urban development, industrial areas, widening of road networks – all these processes lead to loss of habitat suitable for wildlife, which has a huge impact on populations of wild animals.

In order to evaluate the impact of human activities on populations of wild animals and to decide on the most effective actions for nature conservation, we need fundamental information on the extent of changes in the living environment. Are populations of animals declining or increasing? Can we detect any trends in the changes in population size? Can we detect the fundamental causes for changes in population size?

To answer the questions mentioned above, we need an effective system for monitoring animal populations. According to HELLAWELL (1991), monitoring is defined as "intermittent (regular or irregular) surveillance carried out in order to ascertain the extent of compliance with a predetermined standard or the degree of deviation from an expected norm." For the purposes of nature conservation, we need standardized methods for periodic counting of animals in a specific region.

The progress in information technology in recent years, especially in the field of pattern recognition software, opens up important new perspectives for the automated bioacoustic monitoring of a multitude of animal species. In our paper, we will reconsider the application of acoustic pattern recognition algorithms for purposes of monitoring birds under realistic conditions in populated areas.

Using acoustic methods in densely populated areas, we have to deal with a wide range of different anthropogenic sound sources such as highways, railways, aircraft or construction sites. On the other hand, we already have a well-organised network of birdwatchers participating in different programs for monitoring bird populations in many countries.

In Germany, monitoring is conducted by amateur ornithologists coordinated by the Association of German Avifaunists (DDA). From 1989 to 2003, two main methods were used to monitor common bird species: stop counts and mapping of breeding territories. These methods were replaced by line mapping in 2004 (MITSCHKE ET AL. 2005). Line mapping is carried out during four surveys distributed over four time periods from March to June on study plots of 1 km². The observer slowly walks along a transect of 3 km length and maps all indications of breeding birds (singing males, visual observation of birds, bird's nests, etc.). The study plots for line mapping are randomly distributed over the entire country in consideration of the distribution of main habitats. In Germany, 1000 sites are planned. Currently 700 sites have been established. For monitoring purposes, results of line mapping were considered only when the same observer was mapping the plot for at least two consecutive years. At the end of each breeding season, species-specific maps of virtual breeding territories are created (Fig. 1.). For the determination of the virtual territories, only observation within the species-specific census periods were considered to minimize counting of migrating birds.

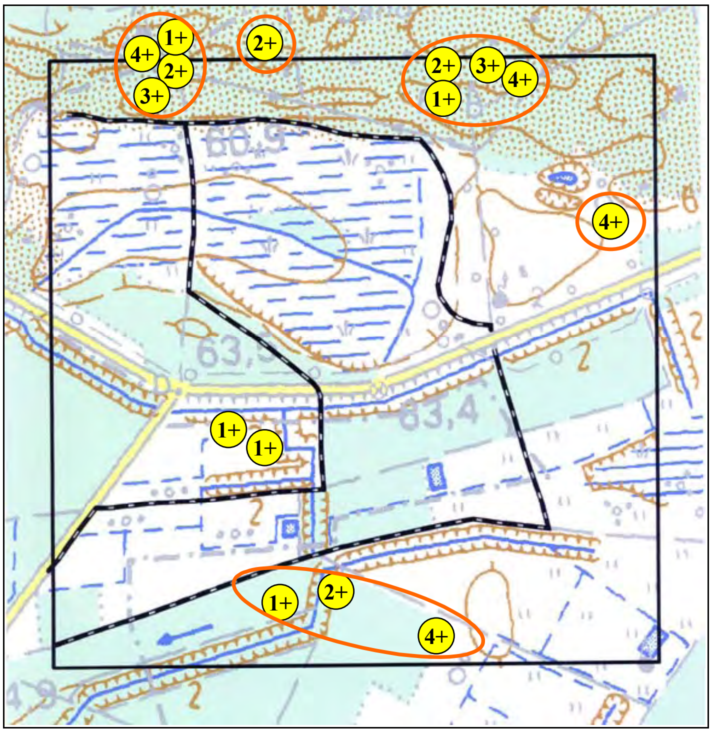


Figure 1: Estimation of virtual territories of woodlark (*Lullula arborea*) determined by line mapping in a study plot in the federal state Brandenburg. The numbers indicate the observation of an individual at a certain survey. The virtual territories were surrounded. The dashed line indicates the transect. For the creation of virtual territories for the woodlark, only observations concerning the periods 2 to 4 were considered.

In consideration of this very successful approach to monitoring breeding birds, we have to decide where bioacoustic methods should be used in addition to direct observations. The main advantage of a bioacoustic approach lies in the long-term recording in the absence of an observer. A recording device can be installed even in ecologically sensitive areas (e.g., nature reserves). Autonomous recording makes it possible to nocturnal animals and animals with low vocal activity. The method can be applied in areas that are difficult to access (e.g., large reed habitats). Another important advantage is keeping subjectivity to a minimum through independence from the skills of the observer. The raw material can be archived, and even years later the presence of species could be verified. Depending on the development of acoustic pattern recognition software, some tasks could be done automatically.
In the following section, we describe situations where bioacoustic monitoring could be a powerful tool in addition to direct observation. Places of realistic scenarios for an automated

bioacoustic monitoring of birds can be seen in reed zones of lakeshore, under low ambient noise conditions, in open landscapes and in GPS-supported mapping.

Acoustic Recording of Birds in Reed Zones of Lakeshore

Reed zones of lakeshores are very sensitive habitats and not easy to access. Thus, mapping of birds in reed habitats mostly occurs from the shore. In 2006 and 2007, we conducted a bioacoustic monitoring pilot study at Lake Parstein in North Eastern Brandenburg. We intended to verify the application of acoustic pattern recognition algorithms to real scenario acoustic recordings. Certain pattern recognition software was successfully evaluated on recordings held in sound archives. However, most of these recordings were done with directional microphones directed towards the caller. Most of the recordings have a good signal-to-noise ratio where the call or the song of the animal is prominent in relation to the acoustic environment. Our approach was to record with a four-channel stationary microphone array of cardioic microphones. The microphones were arranged in a cross configuration (Fig. 2). Our study was focused on nocturnal birds living in reed zones such as rails and bitterns. The following questions were raised:

- Where is the best place for data acquisition? Where should the microphones be placed?
- How far is the recording distance for certain signals?
- Which distortions occur during sound propagation?
- How good should the recordings be to meet the requirements for pattern recognition and sound localization?

Figure 2: Microphone array for acoustic monitoring. Four cardioic microphones (Sennheiser ME 64 or Beyerdynamic MC 930) were placed in cross configuration.

We took continuous recordings during nighttime from several positions on the lakeshore, from the lookout (Fig. 3) and from a boat. Two species of bitterns were found at the study site, the Great Bittern (*Botaurus stellaris*) and the Little Bittern (*Ixobrychus minutus*). Our first experiences with monitoring of nocturnal birds in reed zones have shown that noise level in the vicinity of lakes is very high during the night and in springtime. This is mainly due to calling amphibians, especially tree frogs (*Hyla arborea*). Calls of the Great Bittern could be recorded over distances of more than 1 km. The calls of the Little Bittern could be recorded over a distance of nearly 1 km. The latter finding was surprising since bibliographical references state that the Little Bittern has a very soft call, ranging over distances of no more than 50 m, or in rare cases up to 200 – 300 m (SÜDBECK ET AL. 2005).

Figure 3: Bird observation spot at Lake Parstein. The lookout was used for nocturnal registration of the soundscape.

We have seen that the best place to record birds in reed zones is from the lakeside (from a boat or from a platform). Good conditions for sound propagation (no obstacles, low transmission losses for sound propagating over a quite water surface) allow a large detection range. Consequently, in a second step we have arranged a continuous acoustic registration from a deep-seated boat using a solar driven recording device (Fig. 4). Two times per day, around sunset and sunrise, we recorded for four hours. Four-channel-recording at 48 kHz sampling rate and 16-bit data depth were done with four Sennheiser ME 64 cardioic microphones and an external audio interface (MOTU Traveller) using Avisoft Recorder software for triggering the recordings. The sound files were stored on an external hard disc of a notebook computer (JVC MP-XP 731).

Figure 4: Autonomous acoustic long-term recording at Lake Parstein (North Eastern Brandenburg). A solar driven recording device consisting of a notebook computer (JVC MP-XP731), a FireWire Audio-Interface (MOTU Traveller) and four cardioic microphones (Sennheiser ME 64).

The stereotyped calls of the Great Bittern were a good subject for testing (see also BARDELI ET AL. in this issue) pattern recognition software (Fig. 5). Applying a relative simple procedure, the booms of the Great Bittern could be visualised very well even in the presence of relatively loud anthropogenic noise (Fig. 6).

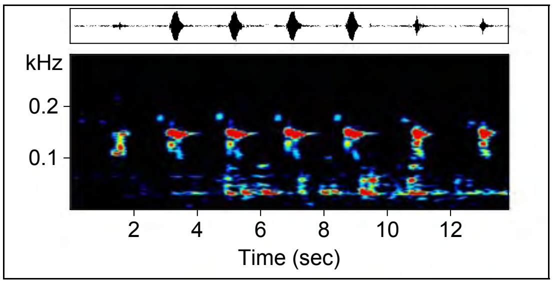


Figure 5: Spectrogram of a Great Bittern (*Botaurus stellaris*) booming. The call consists of a series of repeated elements at a low frequency near 150 Hz. The spectrogram was created using Avisoft SASLab Pro software.

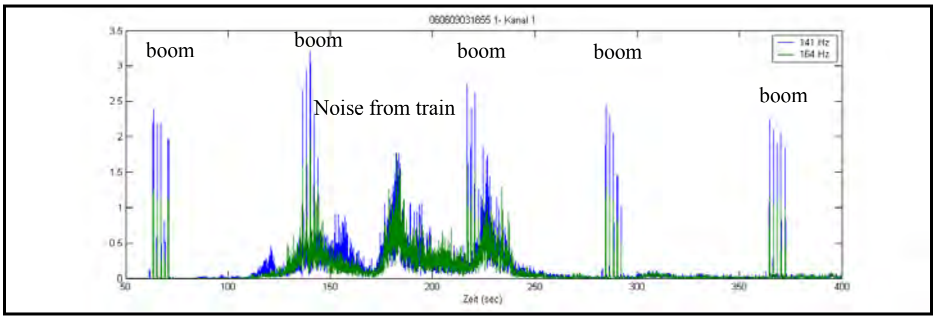


Figure 6: Temporal distribution of energy in two closely related narrow frequency bands (141 Hz and 164 Hz). The peaks in the curve indicate elements of bittern booming. The typical temporal structure of bittern booming is visible even in presence of very strong traffic noise (noise from a train).

Acoustic Survey of Birds Under Low Ambient Noise Conditions

As we have seen, simple structured signals can be detected even in the presence of strong background noise. Acoustic methods should have advantages for monitoring tasks where the surrounding noise level is low. This is due to the mating calls of owls. Most of the owl species in the temperate zone mate in late winter or early spring, and most of them are nocturnal. At this time of the year, we would hear almost no other animal sounds at night except the territorial calls of red foxes and calls of birds migrating at night.

We evaluated the use of bioacoustic approaches for the monitoring of owls in the Nature Reserve “Rochauer Heide” in the Southern part of the federal state Brandenburg. Three species of owls were detected at our study site, the boreal owl (*Aegolius funereus*), the tawny owl (*Strix aluco*) and the Pygmy Owl (*Glaucidium passerinum*). We conducted two- and four-channel sound recordings at stationary points as we did at Lake Parstein. Several observers simultaneously mapped the occurrence of owls in the Nature Reserve. Since owls call only sporadically, the mapping of owls is very time consuming. The goal of our study was to evaluate whether information obtained by sound recordings was as effective as the data obtained by listening at the site. Using a set of recorders placed at distances of approximately 300 to 500 m from each other, we could estimate that the calls of boreal and tawny owls could be recorded in coniferous forest over distances of at least 300 m. During parallel observations in the same spot, more sounds were recorded in general by the equipment than were heard by the observer.

By listening to the multi-channel recordings, the direction from which the owl calls came could be assessed. For all three species, we had indications of pair vocalizations (duetting, typical vocalizations during mating). This is clear evidence of breeding in contrast to high vocal activity of non-paired males. Therefore, by using an array of multi-channel recorders, owls could be mapped very effectively on the basis of their vocalizations. An example for mapping based on bioacoustic data is given in Fig. 7. By developing appropriate pattern recognition algorithms, the acoustic approach would be even more effective than nocturnal observations, which consist of listening to owls at the site.

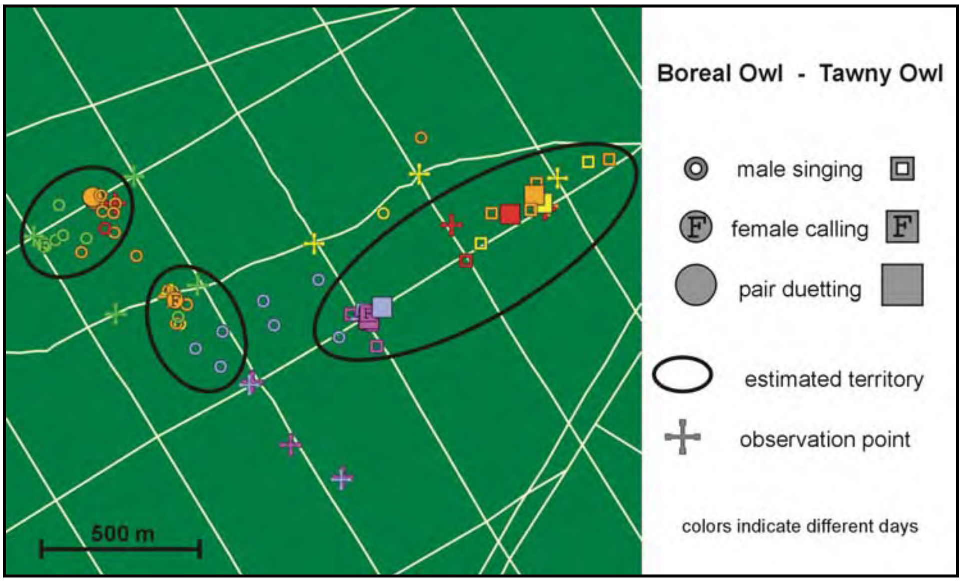


Figure 7: Territories of breeding boreal (*Aegolius funereus*) and tawny owls (*Strix aluco*) in the southern part of the nature reserve "Rochauer Heide" estimated by bioacoustic survey. The analysis of the sound recordings revealed in at least three cases clear indications of pair bonding (duetting). It is likely that the observations outside the marked territories are calls of unpaired males.

GPS-Supported Mapping of Birds

During our fieldwork on acoustic monitoring of birds in reed zones, we developed a method of GPS-supported acoustic mapping. As we have seen, the best place to record birds in reed zones is from the lakeside. In addition, large reed beds are often difficult to access. Applying the methodology of line mapping to an aquatic habitat, we recorded the Savi's warbler *(Locustella luscinoides*), a typical songbird inhabiting large reed beds, from a boat. A small boat equipped with an electromotor was driven at slow constant speed (approximately 30 m/min) along the reed belt, keeping a distance of ca. 20 m from the vegetation. During the boat trip, continuous sound recordings were taken using two directional microphones (Sennheiser MKH 70) arranged at a 90° angle (Fig. 8). The position of the boat was acquired continuously by a GPS device (Garmin Geko). In addition to the sound recordings, the song posts of Savi's warblers were mapped during the boat trip by an experienced observer. We carried out repeated mapping of Savi's warblers at two reed areas of the lake.

We evaluated the sound recordings using two approaches. In the first case, an experienced observer listened to the stereo recordings and marked the position of songs in the sound files. The stereo recording gave a spatial impression which allowed determination of the direction from which the song was heard, and allowed us to distinguish between neighbouring males. The localization of the song post was assessed using the GPS data. In the second approach, songs of Savi's warblers were detected by a pattern-recognition algorithm (see BARDELI ET AL. in this issue). Only pure songs were considered. The position of song posts was determined in analogue to the previous approach. We provided the determination of Savi's warblers territories in accordance with the criteria applied for line mapping. Comparing the three approaches, we found almost similar results (Fig. 9, Tab. 1).

Consequently, acoustic mapping could be successfully applied to the survey of the selected species. Since the combination of sound recordings and GPS tracks allows highly standardized data acquisition, this methodology could be used in the future even by inexperienced observers. However, the algorithms of pattern recognition need to be improved to allow discrimination of different birds by voice and not only by localization. In general, this approach could be used for other birds living in reed belts, too, particularly in warblers of the genus *Acrocephalus*.

Figure 8: Equipment for GPS-supported acoustic monitoring of Savi's warblers. A small boat equipped with an electromotor was driven at low speed along the border of the reed belt. During the movement, continuous acoustic recordings were taken by two directional microphones arranged at a 90 degree angle. The position of the boat was determined simultaneously by GPS.

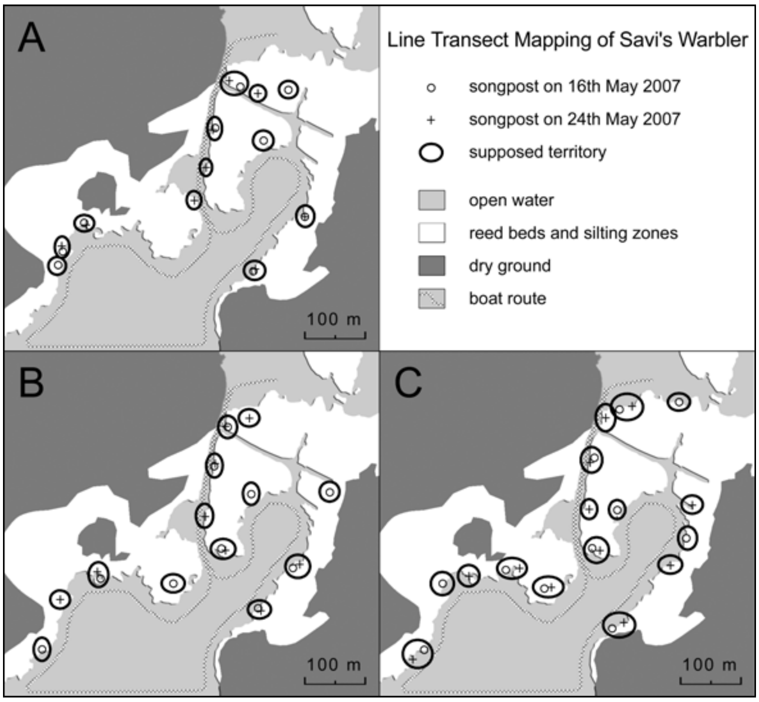


Figure 9: Estimation of breeding territories of Savi's warblers at Lake Parstein according to the criteria of line mapping (SÜDBECK ET AL. 2005). The position of the song posts were determined by A – mapping by an observer during the boat trip, B – listening to the recordings and determining the position of the boat on the base of the GPS track, C – recognizing songs of Savi's warblers on the recordings using pattern recognition software and determining the position by GPS data.

Table 1: Estimation of the number of breeding Savi's warblers at two reed belt areas at Lake Parstein, North Eastern Brandenburg

Method	Number of Territories in the Northeastern Part of Lake Parstein	Number of Territories in the Northwestern Part of Lake Parstein
„Mapping of territories“ by the observer in the boat (9 surveys)	17	33
Line mapping by the observer in the boat (3 surveys)	17	30
GPS-supported line mapping evaluated by listening to the records	17	25
GPS-supported line mapping evaluated by pattern recognition algorithms	19	24

Evaluation of Bioacoustic Survey by Line Mapping

One of the most crucial problems when applying bioacoustic approaches to bird censusing is the question of whether the estimation of numbers of species and specimens could be obtained from recordings. Therefore, we provided sound recordings in parallel with line mapping accomplished within the scope of monitoring of breeding birds of normal landscapes in Germany. We selected the northern part of a regular study plot with a size of approximately 500 m x 1000 m (Fig. 10). By line mapping, bird territories were estimated along a transect of almost 1.5 km. The recording device, consisting of four cardioic microphones in cross configuration and a four-channel recorder, were placed in an area of landscape (swampland) surrounded by deciduous and partly coniferous forest. The sound recordings were conducted at the same time as the observer was walking slowly along the selected transect. Applying species-specific criteria, the observer estimated the territories of breeding birds. For comparison with the acoustic survey, we included either all birds out of the northern side of the plot, or we restricted the data to the central region surrounded by the transect. We expected that the latter area should be adequate to the detection range of our acoustic equipment. The sound recordings were analysed by an expert. The multi-channel recordings allowed assessment of the number of specimens within certain accuracy.

In total, we detected 50 species, with 32 species registered both by line mapping and acoustic survey (Tab. 2). Five species were detected by line mapping only, where most of the locations were outside the range of the microphones. In contrast, 13 species were found only by listening to the recordings. Allowing for the estimation of the species composition, an acoustic approach is as accurate as line mapping. However, up to the time when we should have useful pattern-recognition software dealing with sounds in a very complex acoustic environment, estimation of species inventory would more effectively be provided by an skilled observer. We would recommend bioacoustic surveys in a complex acoustic environment only in areas with restricted access, such as core zones in nature reserves.

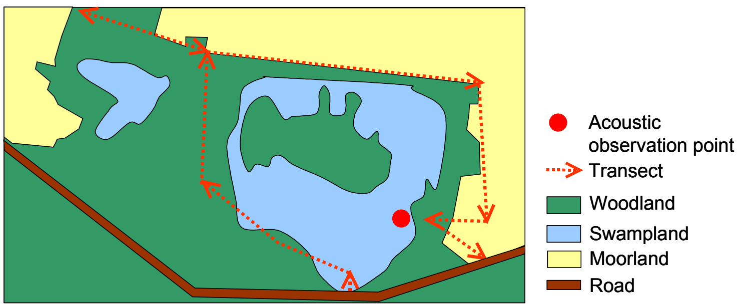


Figure 10: Evaluation of bioacoustic survey by means of line mapping. For four days during spring, birds were counted along a transect (line mapping). At the same time, acoustic recordings were provided from a central place in the line mapping area.

Table 2: Comparison of census data obtained by acoustic survey and by line mapping.

	I			II			III			IV		
Species	**AS**	**LM1**	**LM2**	**AS**	**LM1**	**LM2**	**AS**	**LM1**	**LM2**	**AS**	**LM1**	**LM2**
Aegithalos caudatus	2			2	1	1	1			1		
Alauda arvensis	1					1			1	1		1
Anthus trivialis						1	1	1	3	1	1	2
Certhia familiaris		1	2	1						1		
Coccothraustes coccothraustes	2			2		1	2			2		
Columba palumbus	2	1	2	2		3	3	3	4	3	2	3
Corvus corone	2	1	1	3	1	1	2			2		
Cuculus canorus	1						3		1	1		
Cyanistes caeruleus	1	1	1	2	1	2	2	2		2		1
Dendrocopos major	4	2	4	3	2	4	3					1
Emberiza citrinella	1		4	1		3				1		1
Erithacus rubecula	2	1	3	2	1	8	1		1			
Fringilla coelebs	2	2	6	2	1	3	1	2	6	1	1	4
Garrulus glandarius	2			1	1	1	1			2	1	1
Locustella fluviatilis							1	1	1	2	2	2
Lullula arborea	1		2	3		3	1		2	2		3
Luscinia megarhynchos							3	4	9	3	1	4
Oriolus oriolus				(1)			2		1	2	1	2
Parus major	2	2	12	2	2	10	3	1	6	3		2
Phasianus colchicus	2	2	3	3	1	2	4		1	1		
Phylloscopus collybita	1			2	2	3	2	1	2	2	1	4
Phylloscopus trochilus				2	1	2	2	1	1	1	1	1
Poecile montana	2	1	2	1					2	1		
Poecile palustris			2	1		1	(2)					
Streptopelia picturata				(1)						1	1	1
Sylvia atricapilla				3	4	9	3	3	9	3	2	8
Sylvia borin				1					1		1	5
Sylvia communis				1								3
Sylvia nisoria										1		1
Troglodytes troglodytes				(1)			1	1	2	1		1
Turdus merula	2			3	2	2	3	3	7	3		3
Turdus philomelos	3	2	7	2		1	1			2		1
Jynx torquilla									2			
Lophophanes cristatus									1			
Phylloscopus sibilatrix									1			
Sitta europaea						1				1		
Sturnus vulgaris		1	3			3			2			
Anas platyrhynchos	1			1			1					
Carduelis carduelis	1			1						1		
Carduelis chloris				1			2			2		
Corvus corax	1			2			1			2		
Cygnus olor	2			2								
Dryocopus martius							1			1		
Fulica atra				1								
Grus grus				1 (2)			1					
Hirundo rustica				1			1			1		
Lanius collurio				(1)			1 (2)			1		
Locustella naevia							1			(1)		
Scolopax rusticola	1			(1)						(1)		
Upupa epops				2						1		

I -IV - Days for data acquisition
I - March 23
II - April 15
III - May 13
IV - June 3
 - species specific census period
AS - Data obtained by acoustic survey
LM1 - Data obtained by line mapping from the surrounding of the microphones
LM2 - Data obtained by line mapping from the whole Northern part of the plot
 - Species detected both by acoustic survey and line mapping
 - Species detected by line mapping only
 - Species detected by acoustic survey only

In the above mentioned deliberations, we have given some examples of when a bioacoustic approach could be applied to monitoring tasks. We have seen that even given the current stage of development of pattern-recognition software, bioacoustic methods could be a powerful addition to conventional census approaches. The advantages of the method are obvious for detecting nocturnal birds and for assessing species inventories in areas where the presence of human observers is undesirable. The census of vocalizing animals could be carried out without any disturbance of breeding birds. However, at the current stage, the development of soft- and hardware would provide solutions for only a few species. More efforts in this direction are needed.

Acknowledgements

Our work was made possible by financial support of the Federal Agency of Nature Conservation and the foundation “Naturschutzfonds Brandenburg”. We are much obligated to all people supporting our research on site and/or taking part in data acquisition, especially Martin Flade, Stefan Kreft, Dana Hardrath, Jerzy Przemus and Reinhardt Möckel. Thanks to Julia Barfield for help with the manuscript.

References

FROMMOLT, K.-H. & K.-H. TAUCHERT (IN PRESS): Perspektiven eines auf bioakustischer Mustererkennung basierenden Monitorings. - Naturschutz und Biologische Vielfalt.

FROMMOLT, K.-H. & K.-H. TAUCHERT (2007): Anwendungsszenarien eines auf bioakustischer Mustererkennung basierenden Monitorings. - Verhandlungen der Gesellschaft für Ökologie **37**: 429.

HELLAWELL, J.M. (1991): Development of a rationale for monitoring. – In: Goldsmith, B. (ed.): Monitoring for conservation and ecology. London: Chapman & Hall, 1-14.

MITSCHKE, A., C. SUDFELDT, H. HEIDRICH-RISKE & R. DRÖSCHMEISTER (2005): Das neue Brutvogelmonitoring in der Normallandschaft Deutschlands - Untersuchungsgebiete, Erfassungsmethode und erste Ergebnisse. - Vogelwelt **126**: 127-140.

SÜDBECK, P., H. ANDRETZKE, S. FISCHER, K. GEDEON, T. SCHIKORE, K. SCHRÖDER & C. SUDFELDT, Eds. (2005): Methodenstandards zur Erfassung der Brutvögel Deutschlands. Radolfzell, Mugler.

Bird Song Recognition in Complex Audio Scenes

ROLF BARDELI, DANIEL WOLFF & MICHAEL CLAUSEN

Universität Bonn, Institut für Informatik III, Römerstr. 164, 53117 Bonn, Germany
E-Mail of the corresponding author: bardeli@iai.uni-bonn.de

Abstract. Unsupervised recordings of birds for wildlife monitoring usually contain a high number of sources and thus tend to be very complex. This makes automated recognition of bird species much more difficult than from dedicated recordings of individual birds using highly directed microphones. We present studies of bird song recognition algorithms for this case. On the one hand, we focus on features that can be detected reliably even from strongly distorted recordings. As a starting point, we examine bird songs of low complexity and variability. On the other hand, we study opportunities arising from multi-microphone recordings. In particular, we examine a method for source separation into less complex components.

The question whether computers can be used for the automated detection of animal presence in a given region by means of recognising their vocalisations is a strong stimulation for research in audio pattern recognition. But it is not merely interesting as a scientific stimulation. Monitoring the presence of animals is an invaluable tool for nature conservation. But it is also highly time-consuming when done by human listeners. In order to assess the animal population of larger areas, a great number of people are necessary, all well trained in the recognition of animals and their sounds. Thus, it would be of great help if they could be assisted by technical means.

During the year 2007, a project funded by the German Federal Agency for Nature Conservation (Bundesamt für Naturschutz) set off to investigate whether automated recognition of animal vocalisations might become a supporting tool for real world monitoring problems. An overview of the project from a biological perspective is given in the article FROMMOLT ET AL. in these proceedings. In the following, we describe its computer science and pattern recognition aspects.

First, we investigate methods for the automatic recognition of highly structured vocalisations characterised by blocks of repeated simple song elements. These can be found in the songs of many birds. Second, we show how less complex vocalisations of nocturnal birds can be detected, which is of great use for the monitoring of areas or times which are hard to cover by human listeners. Finally, we show that employing multiple microphones can help to extract simpler components from an audio scene which promise to be easier to analyse.

From Complex Audio Scenes to Structural Similarity

Audio recordings made for the purpose of monitoring an area for the presence of certain target species are strongly different from those dedicated to the recording of a single species, where highly directed microphones can be used and recordings can be made very close to the target. Figure 1 gives an example of how much these two settings differ.

Another source of complexity in the problem of recognising bird songs by pattern recognition techniques is the fact that most birds show a very high degree of variability in their songs.

One bird whose song shows high variability is the chaffinch. It is a well-known song bird with widespread occurrence. Many people can recognise its song in spite of its variability because its song is strongly characterised by its structure.

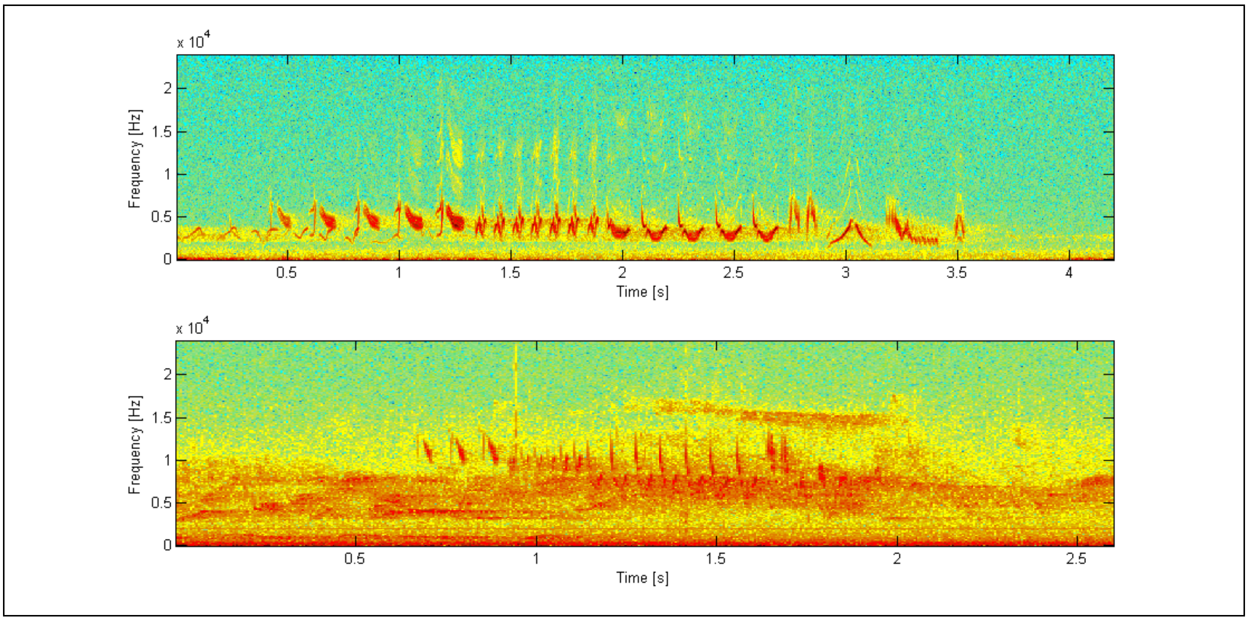


Figure 1: Top: The spectrum of a recording dedicated to capturing the song of the chaffinch. All elements of the song are easily discerned and little background noise is present. Bottom: The spectrum of a monitoring recording. The chaffinch's song is still discernable but embedded in a highly complex mixture of different audio sources.

Figure 2 shows some of the variability found in the songs of chaffinches. Note how the composition and form of the song elements varies strongly between songs. The general structure of the songs, however, is almost the same and can be used to describe an abstract model of the chaffinch song.

A chaffinch song typically consists of two to four segments in which one element is repeated, followed by an end segment. We therefore propose to detect chaffinch songs in three steps. First, all positions in a recording are found which are similar to typical end segments. Then, element repetition frequencies are estimated in a certain time window before the end segment. Finally, each candidate that has a combination of repeated element segments with parameters in a range typical of chaffinch songs is reported.

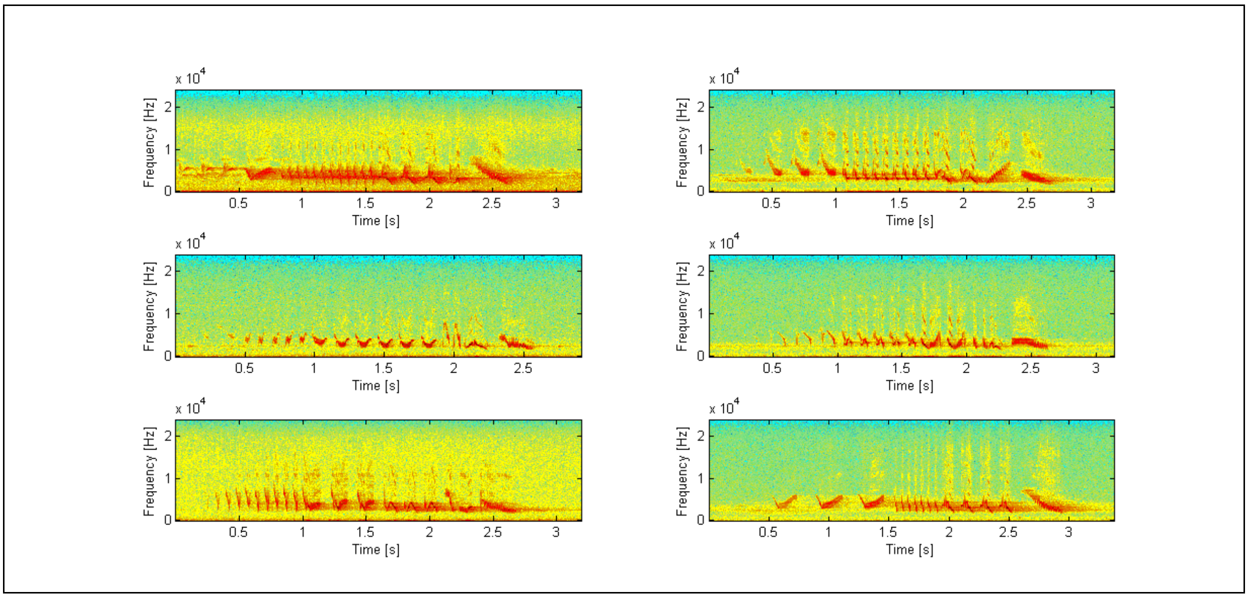


Figure 2: Variability in the chaffinch song. Note how the form and composition of song elements differs. The general structure of the song, however, is always the same.

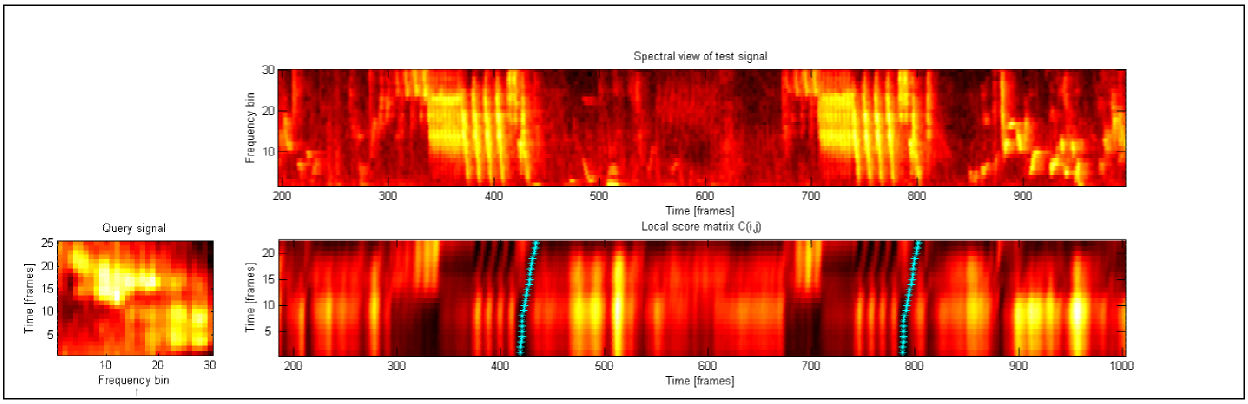


Figure 3: Finding the end segments of chaffinch songs by dynamic time warping.

We will now describe the steps of this algorithm in more detail. Starting from a collection of 20 templates of typical end segments, we first find possible occurrences of the chaffinch song by dynamic time warping (see, for example, DELLER, ET AL. 1993). Figure 3 shows a recording containing two chaffinch songs on the top and a template of an end segment on the left. In order to find potential occurrences of the template in the recording, a cost matrix C given by the difference of the template at position i and the recording at position j as matrix entry $C(i, j)$ is computed. Paths giving a low sum of matrix entries correspond to potential occurrences of the prototype.

After the detection of potential end segments, each such candidate undergoes a second analysis step. In this step, we estimate the repetition rate of elements.

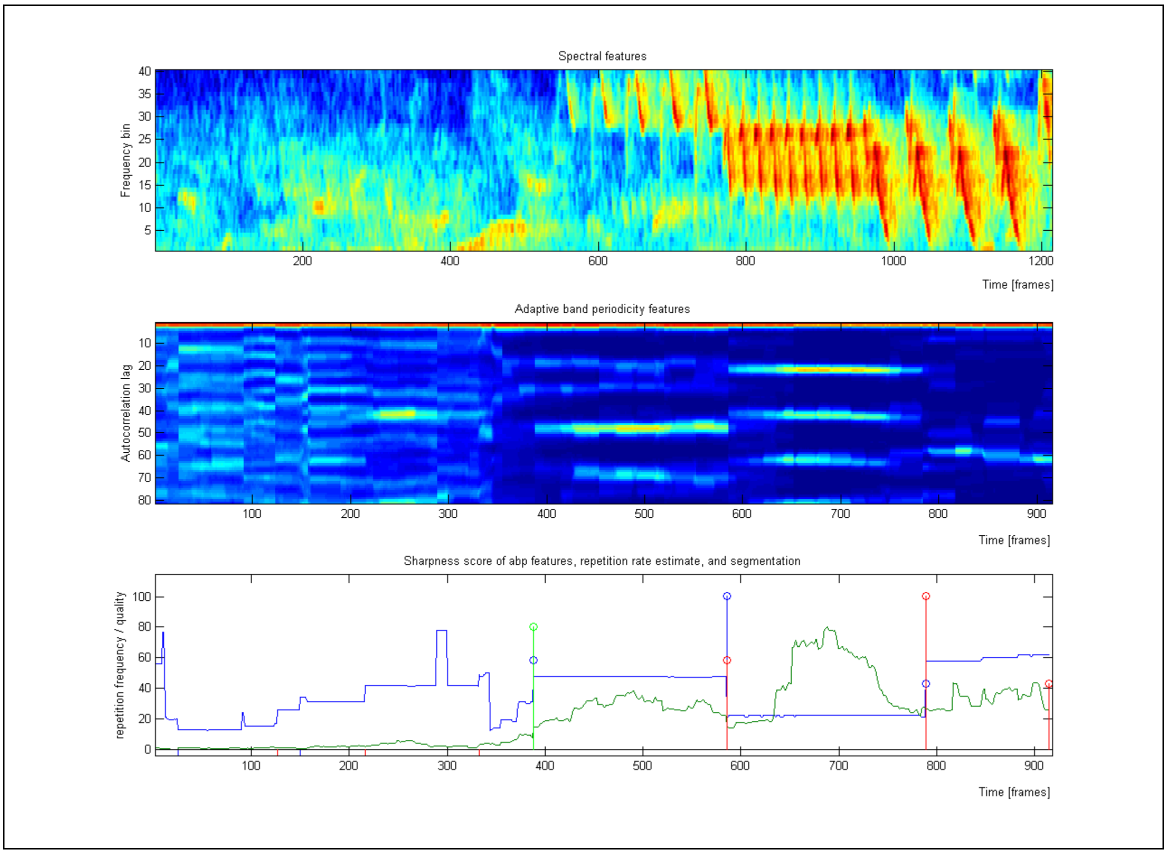


Figure 4: Autocorrelation based features for the recognition of repeated elements. Top: Spectral features in the subband containing the chaffinch song. Middle: Adaptive band periodicity (abp) features indicating repetition frequencies. Bottom: Estimates of repetition frequencies (blue) and a quality measure for the sharpness of the abp features.

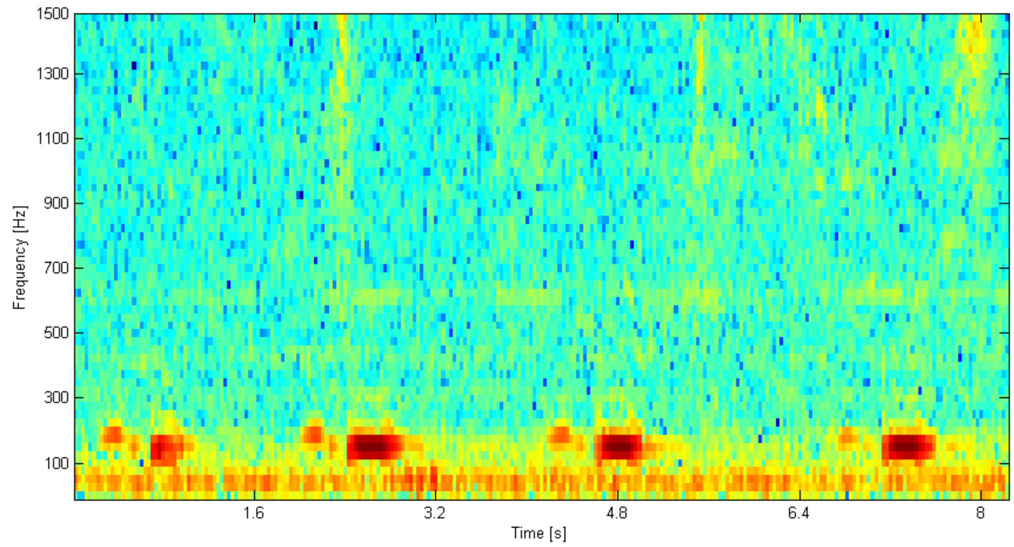


Figure 5: The booming call of the Eurasian bittern.

Figure 4 gives an overview of the features extracted from the part of a signal preceding an end segment. Computation starts by extracting a so-called novelty curve from five sub-bands of the frequency range typical for chaffinch songs. Novelty curves give a measure of change in the signal band over time. In particular, novelty curves show peaks where new elements begin. An autocorrelation curve is computed for each of these novelty curves.

At each time step, the autocorrelation curve is chosen from the sub-band giving the sharpest peaks in the autocorrelation. The resulting sequence of autocorrelation curves is called the adaptive band periodicity (abp) features of the signal. The repetition frequency of elements is then indicated by the peaks of the autocorrelation curves.

Now, grouping these curves by repetition frequency leads to a segmentation of song candidates into segments of constant element frequency. The final decision whether a song candidate is reported or not is based on the length of these segments and the repetition frequency of elements in each segment.

With this algorithm, chaffinch songs can be detected in fairly challenging signal-to-noise ratios. There are however certain cases when false negatives as well as false positives occur. By design, the algorithm cannot detect songs without an end segment or songs which are very short. The superposition of songs from other birds occasionally leads to false positive detections where one bird song resembles an end segment whereas the other consists of repeated elements.

Low Noise and Low Variability

While the chaffinch is a very useful study object for pattern recognition, its relevance for nature conservation is limited. As a starting point for more relevant tasks, the monitoring of birds in low noise environments promised to be a manageable problem. Additional simplification is found by restricting the task to birds with only one song type and little song variability.

The Eurasian bittern is a rare and threatened bird species living in large reed beds. The habitats are difficult to access and the most obvious indication of the presence of the

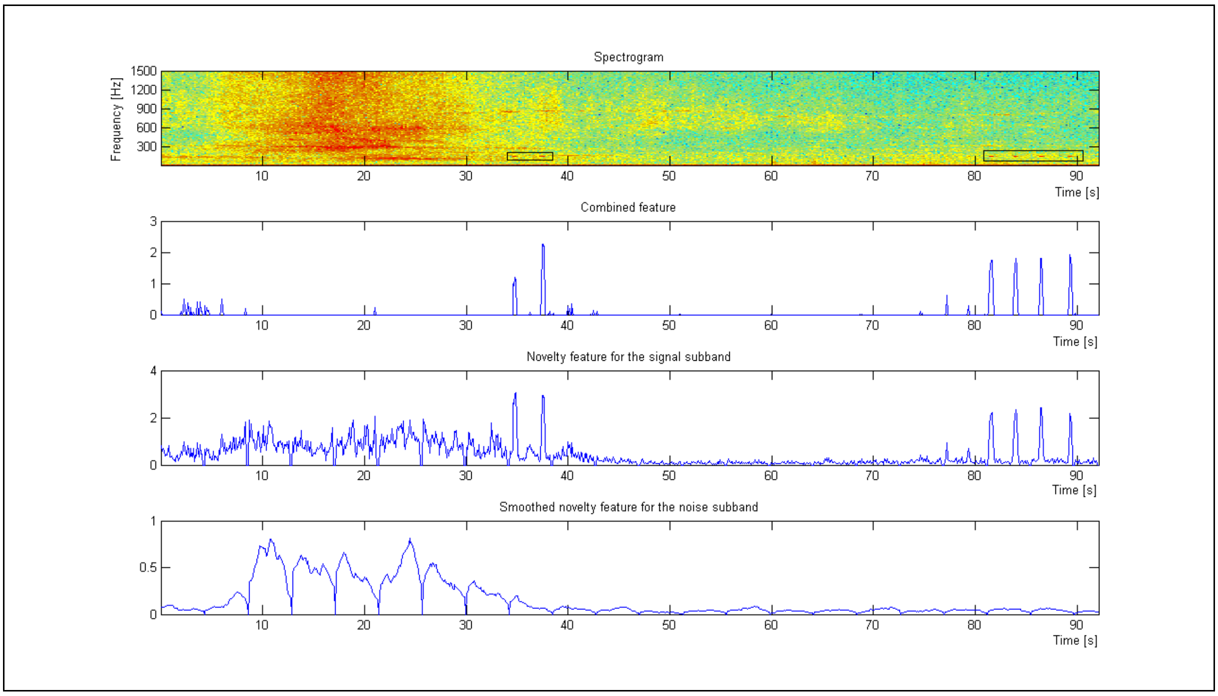


Figure 6: Removing broadband noise from the features used for the detection of Bittern calls.

Eurasian bittern is the booming vocalisation of the male. Acoustical monitoring allows for passive investigation of bittern activity.

The call of the Eurasian bittern is very simple. It is almost completely characterised by its frequency of about 150Hz (see Figure 5). Calls typically occur in call sequences with a characteristic repetition frequency. In low noise conditions, this call can be detected by finding energy peaks in the characteristic frequency band.

The main problem in detecting the bittern call, leading to false positive detections, is broadband noise overlapping the frequency band of the bittern call. This influence can be accounted for by estimating the noise level from a neighbouring frequency band. Figure 6 shows how broadband noise can be removed from the features by subtracting a low-pass filtered noise estimate.

This method alone still does not sufficiently reduce the number of false positive detections.

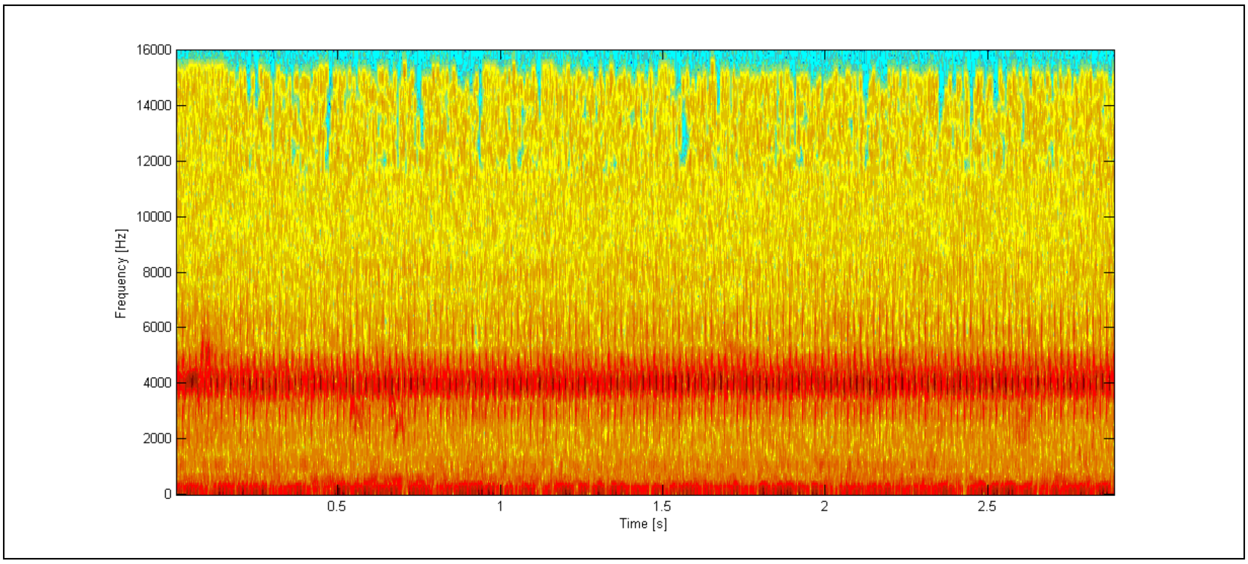


Figure 7: The characteristic song of Savi's Warbler.

Therefore, the characteristic repetition rate of the call is used as a second feature. Using this combination leads to a fairly reliable detector for bittern calls.

Another night active bird living in reed beds is Savi's Warbler. It has a very characteristic

song formed by the continuous repetition of simple song elements at a rate of roughly 50 repetitions per seconds. An example of its song is given in Figure 7. It is this composition of repeated elements which makes the application of the techniques described in Section 1 successful in this case.

The recognition of Savi's Warbler relies on the same features which were already used for the detection of the chaffinch. Starting from abp features, the repetition frequency of the basic elements of the song can be read off the Fourier transform of the features. Similar to the strategy followed in the detection of the Eurasian bittern, noise reduction of the abp features can be conducted by subtracting the abp features of a flanking frequency band.

Finally the decision whether a Savi’s Warbler is singing at a given time is found by deciding whether its characteristic element repetition frequency is present for a long enough time. This is combined with a second criterion, the sharpness of autocorrelation curves in the abp feature.

Evaluation of this algorithm has been performed on monitoring recordings from Lake Parstein. They consist of about 30 hours of audio material with frequent occurrence of the Savi’s Warbler’s song. The distance from the microphones is highly variable and the recordings contain a multitude of noise and background sounds.

We found that the detection of almost inaudible songs was possible. Moreover, false positive detections were very seldom.

Opportunities from Multiple Microphones: Source Separation

In Section 1, we gave an example of how complex natural audio scenes typically are. This is a very difficult problem for the detection of animal sounds. By combining multiple microphones, it is possible to extract simpler components by combining the recorded signals in such a way that songs from some directions are attenuated while those from other directions are amplified (see HYVÄRINEN ET AL. (2001) and VAN TREES (2002) for different approaches to array signal processing).

The central step in devising a source separation algorithm is the choice of a measure describing the complexity of an audio scene. Given such a measure, it is possible to evaluate it for several combinations of input sounds and choose the combination that gives the lowest complexity score.

The measure we use in our approach is the spectral flatness measure. It measures how much the energy at a given time is spread in the spectrum, giving a high value when the energy is equally distributed and a low value when the energy is concentrated in a small number of narrow frequency bands. The spectral flatness measure is computed from the spectrum as the geometric mean of the Fourier coefficients divided by the arithmetic mean.

The mixing coefficients of the source signals should be estimated from segments of constant mixing conditions. We assume that mixing conditions are locally constant and form a windowed spectral flatness measure by convolving short signal windows with a Hann window.

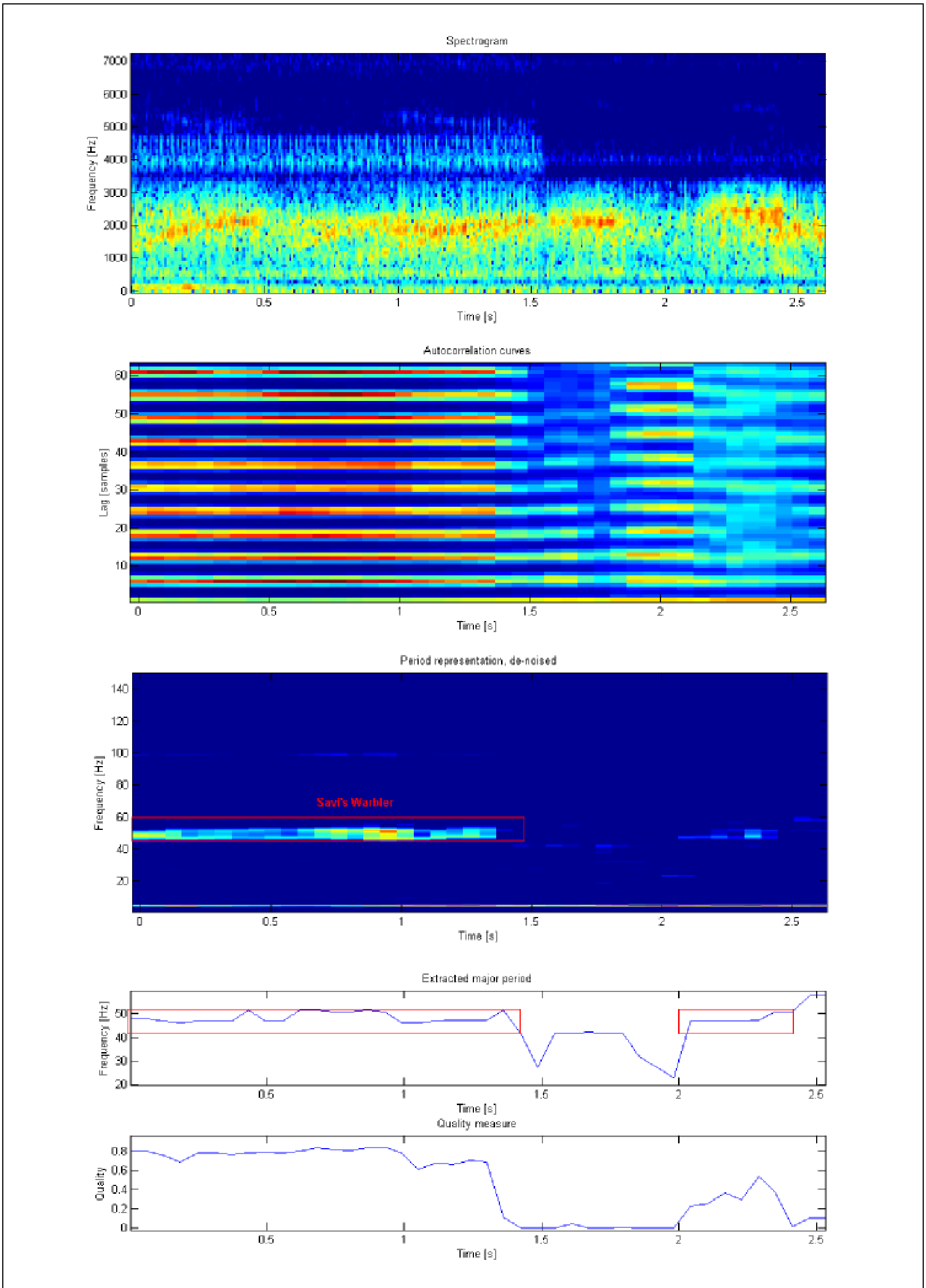


Figure 8: The recognition of the Savi's Warbler' song: Autocorrelation curves (second row) are extracted from the spectrum (first row). The repetition frequency of song elements can be read off the Fourier transform of the autocorrelation curves (third) row. The song is detected wherever the repetition frequency is in the expected range and a quality measure of the autocorrelation sharpness exceeds a threshold.

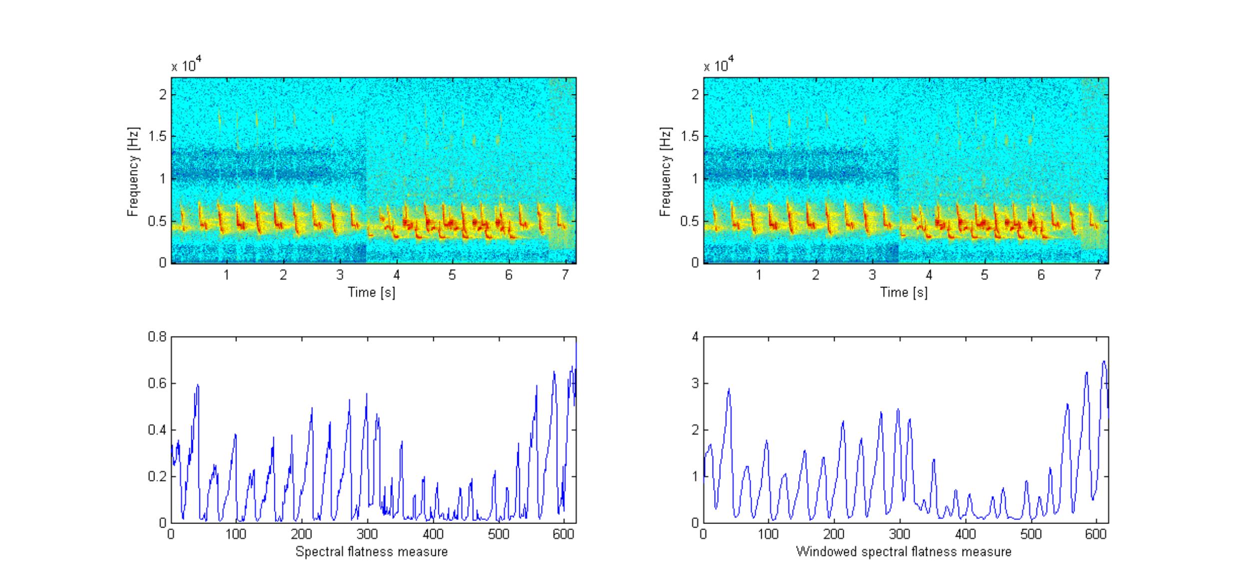


Figure 9: Left: The spectral flatness measure of a mixture of bird songs. Right: Windowed version of the measure which is used as an objective function for source separation.

The extraction of simpler components is achieved by optimising the spectral flatness measure for each window. This will lead to a vector of weights for each time window which describes how the input signals should be combined at the given time. For each window, a fixed number of hypotheses for this vector is generated. Dynamic programming allows combining the hypotheses from different windows such that a smooth variation of the mixing coefficients for the input signals is achieved wherever possible.

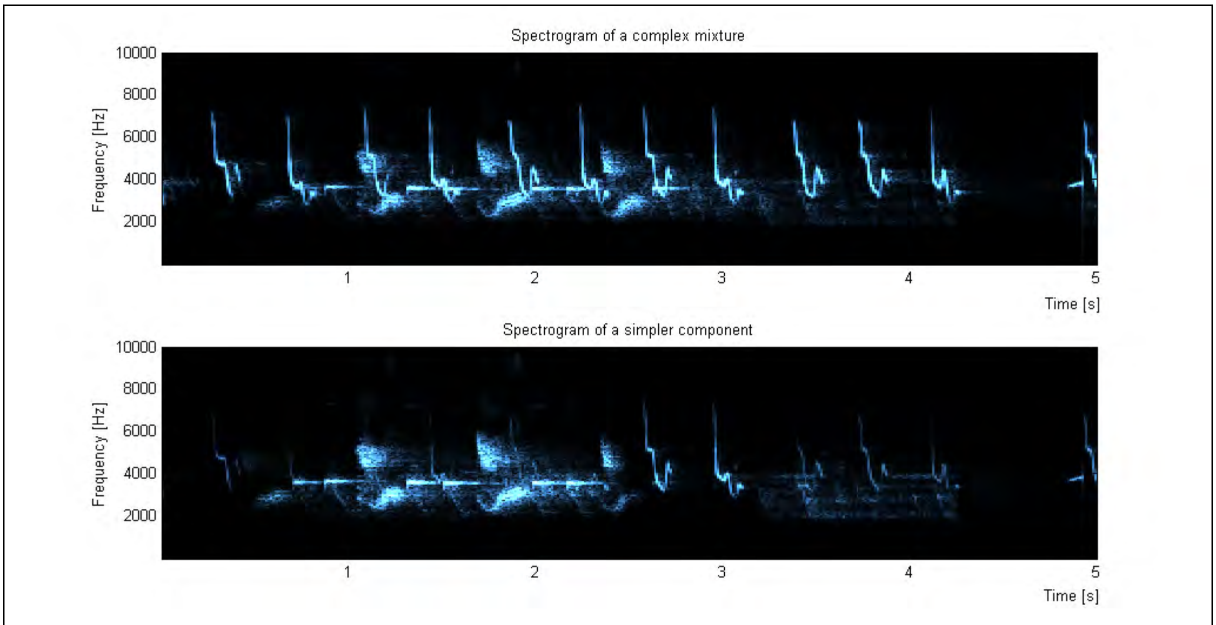


Figure 10: Extraction of a simpler component (bottom) from a mixture of two audio sources (top).

Figure 10 shows the extraction of a simpler component from the mixture of two bird songs. For better illustration this example was generated by artificially mixing two sources.

While component extraction from more complex natural audio scenes does not give perfect separation, it still leads to a strong attenuation of some sources while others are amplified. This promises to improve recognition rates.

Conclusion

In Sections 1 and 2, we have described tools for the recognition of vocalisations of various bird species. They are not restricted to these species but should be applicable to similar cases.

First, the recognition of low complexity songs and calls such as that of the Eurasian bittern can be used for the recognition of other simple sounds in silent environments such as the songs of owls.

It can also be the starting point of more complex recognition algorithms in the same way as the detection of end segments is a starting point for the detection of the chaffinch song.

Second, there are a large number of animals whose vocalisations are characterised by the repetition of simple elements like in the song of Savi's Warbler. This suggests, that methods like those described in Section 2 may not only allow recognising the vocalisations of different species of warblers but also of animals such as crickets, frogs and toads.

Third, many song birds show a highly structured song like that of the chaffinch. Thus, the techniques described in Section 1 may be applicable to other birds like the Blue Tit, the Coal Tit or the Wood Warbler.

Finally, while being in an experimental state, source separation techniques promise to become a valuable tool in breaking down the complexity of natural audio scenes into simpler components.

References

DELLER, J., J. PROAKIS & J. HANSON (1993): Discrete-time processing of speech signals. - New Jersey: Prentice Hall.

HYVÄRINEN, A., J. KARHUNEN, E. OJA (2001): Independent Component Analysis. – John Wiley & Sons.

VAN TREES, H. (2002): Optimum Array Processing. – John Wiley & Sons.

Techniques for Bioacoustic Signal Detection Using Image Processing

T. SCOTT BRANDES*

Conservation International, 2011 Crystal Drive, Suite 500, Arlington, VA 22202, USA
E-mail: tsbrandes@yahoo.com

Abstract. The use of spectrograms in sound analysis opens up the realm of sound in a way that provides visual detail of sonic nuances. Yet, this tool is underutilized in automatic call recognition (ACR) algorithms and is primarily used for manual detection. The use of image processing tools creates a path for developing ACR algorithms that can be used with great efficacy to detect a variety of call types. The two primary ways in which image processing is used are in filtering and feature extraction. In filtering, estimates of sound level in the prominent frequency bands are used to set threshold levels to filter out background sound, allowing the resulting spectrogram image to be used directly for the feature extraction of signals. This serves as the framework behind two ACR methods that are successful at detecting and classifying bioacoustic signals. The first method uses image filters that blur the spectrogram to help patterns stand out from background noise and a Bayesian classifier to identify the resulting signals. The other method uses a different set of features that result from the processed spectrogram with multi-layer hidden Markov models to identify signals. The blur filter method performs better with the honeycomb-structured calls, such as those made by many crickets, and the HMM approach works better with frequency modulated whistles, characteristic of many frogs and birds. Combining both allows additional call types to be better detected and classified.

The majority of bats, birds, frogs, marine mammals, many insects, and even some species of fish are easily detected by the sounds they make. Further, they each produce acoustic signatures, revealing not only their presence, but also their species type, and they can be recorded without impacting them or their environment. In this way, these sounds serve as an ideal way to asses and monitor environments and the ecosystems with which these species are a part.

The efficiency of acoustic surveying lends itself to both rapid assessment programs, which quickly assess the biodiversity of specific regions, as well as long term monitoring of biodiversity change (RIEDE 1998). This is largely because numerous animals are heard more often than seen or trapped. This translates into not only higher species counts, but also faster estimations of biodiversity. PARKER (1991) describes how in 7 days he recorded the vocalizations of 85% of the 287 species of avifauna his team of 7 ornithologists inventoried after 54 days of intensive field work within a 2km^2 area in Amazonian Bolivia, which included 36,804 mist-net hours. If this is representative of the advantages of monitoring birds acoustically, then the same is likely true for both stridulating insects and vocalizing anurans, particularly since they are even less visually conspicuous. Although no comparative numbers are provided, FISCHER ET AL. (1997) and RIEDE (1993, 1996, 1998) found it more fruitful to monitor insects acoustically than by physically collecting them. Likewise, PETERSON AND DORCAS (1992) as well as CROUCH III AND PATON (2002) strongly advocate acoustic monitoring to better study anuran activity and species presence. This trend also holds true for the field sampling of bats. O'FARRELL AND GANNON (1999) compared acoustic sampling of bats simultaneously with mist nets and double-frame harp traps in 57 locations. They found that 86.9% of the combined species present were detected acoustically, whereas only 63.5% of the species detected were physically captured.

There are many ways to monitor bioacoustic sounds. The least technical way is to use trained field personnel to perform systematic surveys throughout areas of monitoring interest. This is the most wide-spread approach, particularly with avian monitoring, since there are many skilled professionals capable of performing acoustic surveys of birds. This type of survey, however, is sensitive to observer bias in that the detection rate and accuracy are dependent on the skill level of the observer. While this is tolerable for short-term monitoring

* Current address: Signal Innovations Group, Inc., 1009 Slater Road., Suite 200, Durham, NC 27703, USA

in areas covered by the same technical staff, it makes long-term monitoring and comparisons among sites less meaningful if the personnel making the acoustic identifications are not the same for all of the field samples (ANGEHR ET AL. 2002).

One way to improve performance is to make audio recordings during the acoustic surveys to provide a record that can be reviewed or reanalyzed. Acoustic recording is particularly important for groups of species, such as orthopteroid insects (crickets, katydids, and relatives), that are not well documented, where audio recordings are needed to serve as a voucher. A further standardization can be achieved by automating the recordings themselves. In doing so, a great deal of work is transferred to the detection and classification of the bioacoustic sounds within the recordings.

Motivated with both the need for standardization and efficiency, a variety of approaches have been developed for automatic call recognition (ACR) of bioacoustic sounds. Early methods focus on techniques used in human speech recognition such as template matching with dynamic time-warping (ANDERSON ET AL. 1996), and modeling changes in ceptral coefficients with hidden Markov models (HMM) (KOGAN & MARGOLIASH 1998). These approaches work best with bird songs that have a rich harmonic structure recorded in low noise conditions. Other approaches have been successful with particular taxonomic groups. Neural networks have successfully been used to identify bat species by their echolocation (PARSONS 2001) (PARSONS & JONES 2000), as well as orthoptera and birds by using time-domain signal coding (CHESMORE 2001). In the presence of significant background noise, other approaches are needed. Successful methods of classifying calls given in high noise environments have focused on spectral intensity peaks (TAYLOR ET AL. 1996), (CHEN & MAHER 2006). These methods extract useful parameters from the spectral peaks which are used to classify the sounds. Focusing on spectral peaks in high noise environments enables the extraction of the dominant features of the loudest sounds. Through the use of image processing on spectrograms, a wider range of these sounds can be extracted from the background, leading to a wider range of sounds that can be identified from field recordings. The techniques to do so described here include threshold filtering, blur filtering, contour feature vectors modeled with HMMs, and call sequence modeling with HMMs.

Techniques

Preprocessing: Threshold Filtering

Spectrograms generated from field recordings often contain a great deal of background noise. This is particularly true of recordings made from autonomous recorders placed in forest canopies set to record sound at scheduled intervals of time. Recordings made with handheld microphones will often have a higher signal to noise ratio when the microphone is directly pointed at a singing individual, but even these recordings can have significant background noise when there are multiple singing organisms at any one time. We begin our approach to automatic call recognition (ACR) by generating spectrograms of the recordings. These spectrograms contain background noise that has different intensity levels as a function of spectral frequency, that are relatively consistent throughout the duration of the recording (usually 10 minutes or less for our recordings). Once the relationship for how the background noise levels vary across frequency is determined, a threshold filter is uniformly applied throughout the entire sound recording which preserves the signals louder than the background noise and drops out the remaining sound from the spectrogram image.

The process of creating this threshold filtering image has 2 steps: creating the relationship between the background noise and the spectral frequency, and applying the threshold level across the frequencies in the image. The noise level as a function of frequency is estimated by calculating the average intensity of each frequency within the spectrogram for some portion of the sound (e.g., the first minute of the recording). Once these averages are calculated, a set of frequency bands is created to group frequencies that have similar background sound levels. We do this by finding local minima of sound intensity throughout the frequency spectrum that bound local maxima of sound intensity, where all the local maxima are greater than the local minima that set the frequency bounds for them. In this way, we partition the spectrum into 5 or 6 frequency bands with which to apply separate threshold filters. By creating separate frequency bands, the threshold filters in bands with high noise do not suppress sounds in frequency bands with low noise, and allow us to create a spectrogram image of calls that are better separated from background noise (Fig. 1). By separating calls from the background noise, the pixel clusters of the calls can be measured directly from the spectrogram and classified.

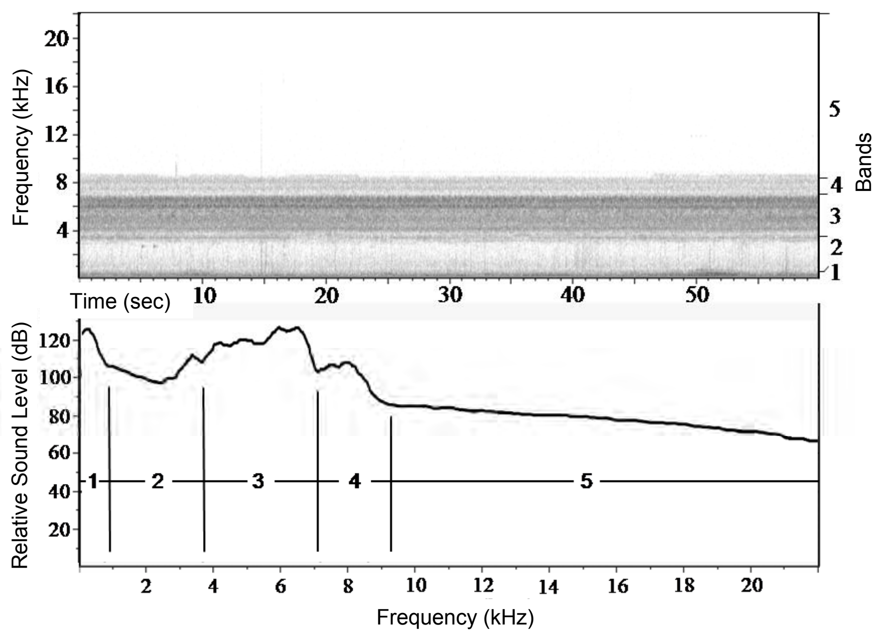


Figure 1: Dividing the spectrum into a set of frequency bands. A one minute segment of a recording is used. The average sound level as a function of frequency is calculated. The local maxima and minima are found. The band limits of the frequency bins are chosen as the local minima that bound local maxima, where all the maxima within the band are greater than the local minima bounds. In this figure, band 5 is created as a leftover chunk of the spectrum after the afore mentioned process. This prepares the spectrogram for threshold filters to be applied in each frequency band, improving the ability to extract calls from each frequency band independently of calls in the others.

ACR Method 1: Blur filters with a Bayesian classifier

Although the threshold filters work well to separate calls from background noise, additional image processing is necessary to detect calls recorded with a wider range of signal strengths. The quality of the spectrogram pixel cluster shapes of many calls degrades as signal strength declines, and the threshold filter by itself is effective only with singing individuals that are relatively close to the microphone. However, this pixel cluster degradation often takes the form of signals in the spectrogram developing rough edges and becoming fragmented. This is particularly true with cricket stridulations, which often consist of pulse trains with components that completely overlap in time when recorded with a high signal strength. Spectrogram images of the same calls recorded from a greater distance can appear fragmented. One of the most important features to use when classifying calls is their duration, and the duration of fragmented calls become difficult to measure. A useful way to combat this signal degradation is to apply a blur filter before applying the threshold filter. A particularly useful blur filter is an averaging blur filter, designed such that the pixel values on each side of every pixel in the image are averaged together with the pixel between them (as a group of 3 pixels) to become the value of that middle pixel for all the pixels in the new image. This sort of blurring distorts the image in time, but not in frequency. As such, it successfully joins pixel clusters in calls that would otherwise be fragmented. By applying such a blur filter, pixel clusters of calls from the same species will have a similar duration even when the call is not recorded very loudly (Fig. 2).

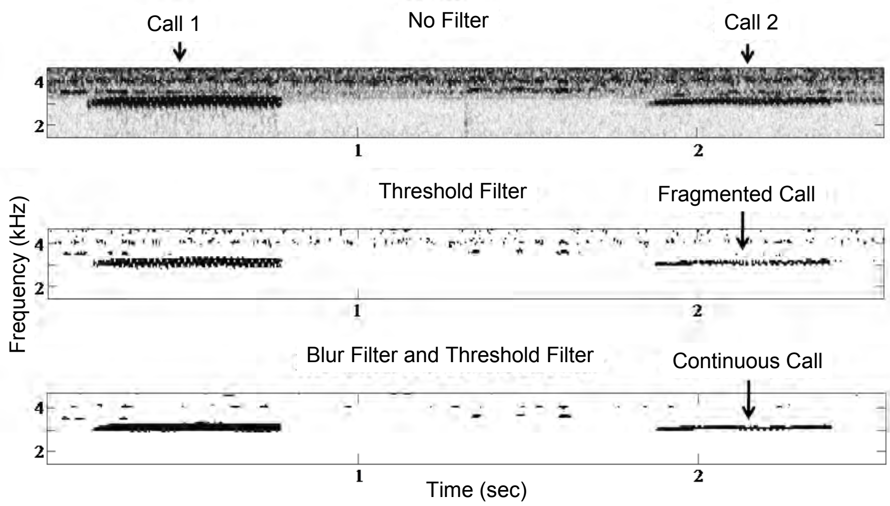


Figure 2: Applying a blur filter. Call 1 and 2 are from different individuals of the same species. Call 2 has less signal strength since the individual is further from the microphone. The top image is from the unaltered spectrogram. The middle image results from applying the threshold filter to the spectrogram. In this image, call 2 becomes a fragmented cluster of pixels. The bottom image results from applying the blur filter as well as the threshold filter. In this image, call 2 becomes a contiguous cluster of pixels and can be more readily matched to the pixel cluster shape of call 1, improving the detection and classification algorithm.

Once the blur filter and threshold filter are applied, pixel clusters from the resulting sounds are measured. A great many crickets and some frogs and birds have calls consisting of a constant frequency. This can be in the form of a stand-alone call or as a pulse train. Calls of this nature are particularly well suited for this sort of detection process. The central frequency, duration, and bandwidth of these pixel clusters can be automatically measured directly from the spectrogram and classified using a Bayesian classifier when there is little to no overlap of the species class boundaries in this feature space. The effectiveness of this approach has been shown with constant frequency calls from crickets and frogs. In one site in the lowland rainforest in Costa Rica, 20 species of crickets and 2 species of frogs are readily detected and classified this way (BRANDES ET AL. 2006).

ACR Method 2: Contour feature-vector with hidden Markov models

A much larger variety of bioacoustic sounds can be detected and classified using contour feature vectors and hidden Markov models (HMM). This is particularly true with calls that consist of frequency modulated whistles and sequences. When trying to determine the classification of a call, we want to choose the call classification that maximizes the following probability equation.

$$\hat{C} = \arg\max_{C} P(C)P(A \mid C)$$

Here, our estimate for call type, $\hat{C}$, is chosen as the call type, *C*, that maximizes the probability of collecting a set of signal measurements, *A*, when a particular call is given, multiplied with the probability of that call being given. The set of signal measurements is a sequential series in time, $A = \{a_1, a_2, \ldots, a_t\}$. The acoustic model, $P(A|C)$, is employed with the use of HMMs. It is often impractical to calculate this probability for each call type stored, so a viterbi search (VITERBI 1967) is used to efficiently find the most likely call type by looking for a maximizing state sequence.

Previous work where HMMs are used to classify particular bird songs rely on ceptral coefficients (KOGAN &MARGOLIASH 1998, KWAN ET AL. 2004). This is also true with the vast majority of human speech recognition work, which relies heavily on ceptral coefficients, and derivative thereof, as the feature vector to use to collect signal measurements. Ceptral coefficients are good at capturing the rich harmonic structure of human voice, however, they are particularly susceptible to noise (OPENSHAW & MASON 1994). A large number of bioacoustic sounds do not have any harmonics and they are often found in noise rich environments. A better performing feature vector to use is a contour feature vector that measures the peak frequency and bandwidth of a call. While these two parameters are important to model, the peak frequency can be susceptible to individual variation, and the bandwidth can be variable due to signal strength. For added robustness, it is also important to model the change in these parameters from the previous time step.

An additional level of pattern recognition that is also possible using HMMs is that of call sequence recognition. Here, we use a composite HMM, where the first layer consists of the HMM mentioned above. In the second layer, we model the sequences of the detected calls from the initial layer (Fig. 3). Here, our feature vector consists of the call type designation from the first HMM and its duration. The gaps between the calls are included in the sequence model as well. Both the single layer HMM and the composite HMM have shown effective at detecting frequency modulated whistles from birds, crickets, and frogs (BRANDES 2008).

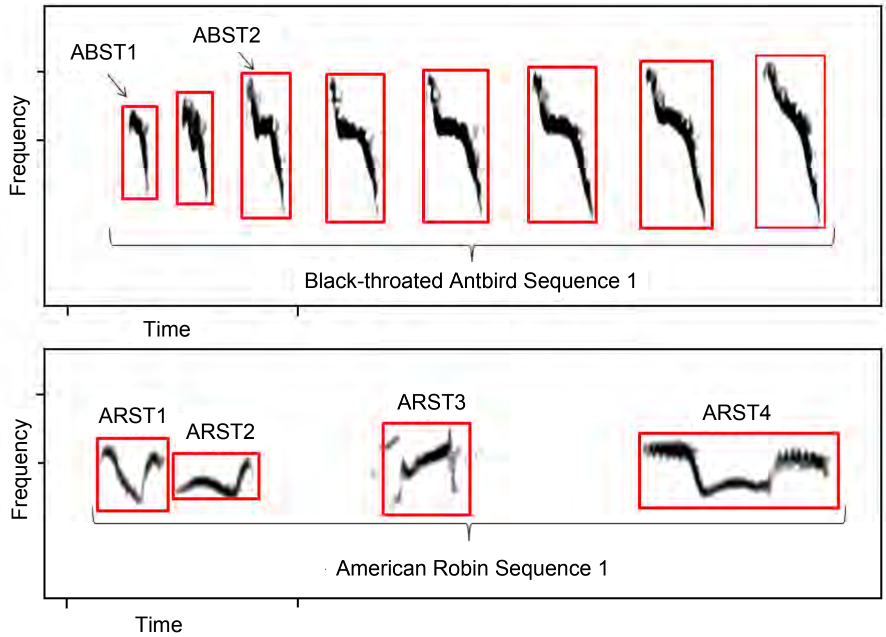


Figure 3: Sequence recognition with composite HMMs. The individual calls are found with the first layer HMM, and the sequence of these calls is classified with the second layer HMM.

Combining techniques and future work

A wider range of bioacoustic signal detection is possible by combining these techniques. One of the main shortcomings of the blur filter and Bayesian classifier approach is that it does not perform sequence modeling. After the threshold filtering, pixel clusters that represent noise are likely to be short in duration. This means that misclassification of calls that have a short duration are more likely to occur. With crickets in particular, the very short duration calls are likely to occur as elements of a pulse train. These pulse train sequences can easily be modeled by the second layer of the composite HMM mentioned previously, where the classification of the individual call elements can be found with the blur filter and Bayesian classifier method. This allows more confidence in the call type designations for short duration calls that are part of a sequence.

The main shortcoming of the contour feature vector method is encountered with calls that have either a honeycomb-shape or rough edges (Fig. 4). Calls in these categories tend to have variations in peak frequency and bandwidth that have a larger range of inconsistency among individuals of the same species than useful for identification. The contour feature vector picks up this high-resolution variation, and the HMM does not make identifications as readily. However, if these signals are first smoothed with a blur filter, then the variations are smoothed out and the overall trends of the calls are more discernable to the HMM classifier (Fig. 5).

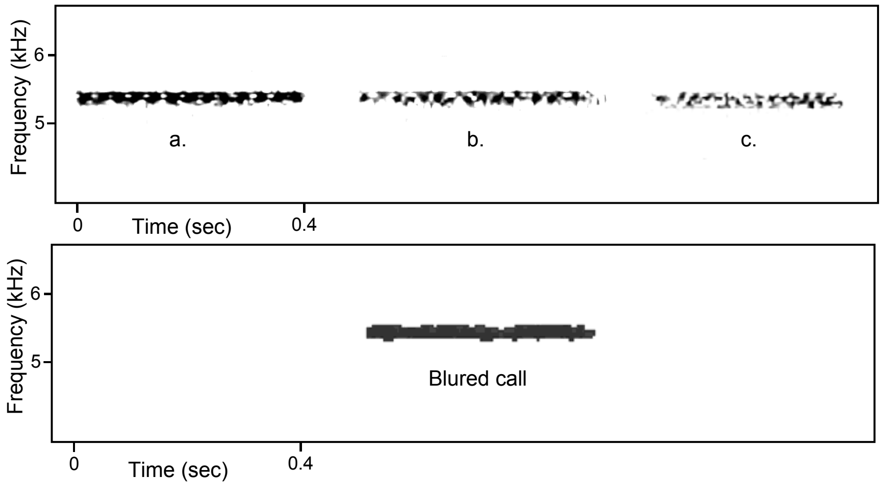


Figure 4: Improving detection of honeycomb-shaped calls with blur filtering and HMMs. In the top image, calls a, b, and c are each from the same species of cricket. Their uneven pixel cluster shapes result in poorer performance with the contour feature vector method. By first applying a blur filter, this unevenness becomes smoothed, and the calls become more readily classified with the contour feature vector method.

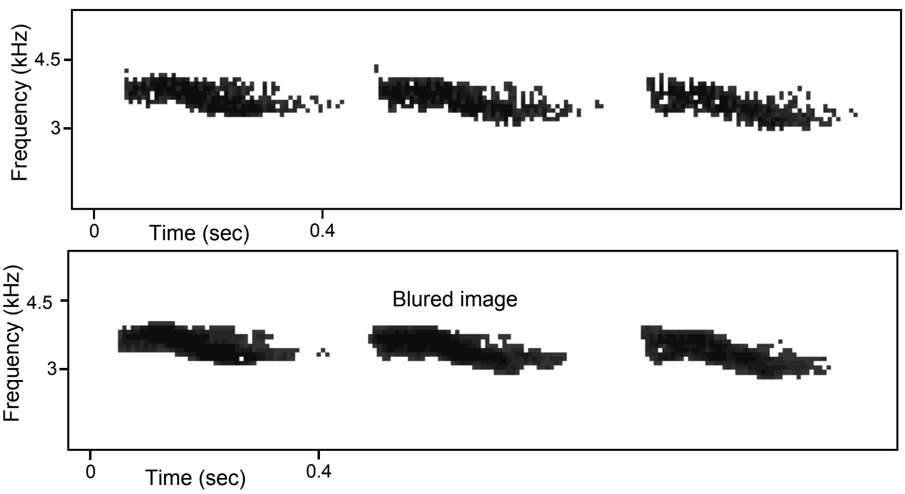


Figure 5: Improving detection of calls with uneven edges with blur filtering and HMMs. A sequence of calls is shown in the top image without applying the blur filter. Their rough edges creates difficulty for the contour feature vector. By smoothing the image with a blur filter, these edges become smoother and improve the performance of the contour feature vector.

The strength of the HMM classifier is in its ability to classify sequences that have a degree of stochastic variation among the samples. Their effectiveness is principally dependant on how well the feature vectors are chosen to capture the state changes within the sequence. The feature vectors described here are by no means exhaustive for bioacoustic signal detection. The best performance is achieved by choosing feature vectors appropriate to the patterns the model is designed to classify. By adding multiple levels of HMMs, additional levels of pattern sequences can be classified. This is a particularly attractive way to classify sequences of calls as described earlier. More work is needed in this area. Sequences of bat echolocations, for instance, are a good candidate for future work since they often consist of frequency modulated or constant frequency type calls. Likewise, calls with a clear harmonic structure can be looked at in a new way, by choosing feature vectors appropriate for their patterns. Cepstral coefficients, though useful in capturing the rich harmonic structure in signal, are susceptible to noise. Features extracted from spectrograms are less so, and provide a different way to explore classifying harmonically rich signals with composite HMMs that perhaps would improve performance of detection and classification models. Future work on additional features used to describe bioacoustic signals will lead to a greater range in signals that can be detected and will contribute to the scale and scope of acoustic monitoring possible around the world.

References

ANDERSON, S. E., DAVE, A. S. & MARGOLIASH, D. (1996): Template-based automatic recognition of birdsong syllables from continuous recordings. - Journal of the Acoustical Society of America **100**: 1209-1219.

ANGEHR, G. R., SIEGEL, J., ACCA, C., CHRISTIAN, D. G. & PEQUENO, T. (2002): An assessment and monitoring program for birds in the Lower Urubamba Region, Peru. - Environmental Monitoring and Assessment **76**: 69-87.

BRANDES, T. S. (2008): Feature vector selection and use with hidden Markov models to identify narrow-band frequency modulated bioacoustic sounds amidst noise. (*in review)*.

BRANDES, T. S., NASKRECKI, P. & FIGUEROA, H. K. (2006): Using image processing to detect and classify narrow-band cricket and frog calls. - Journal of the Acoustical Society of America **120**: 2950 – 2957.

CHEN, Z. & MAHER R. C. (2006): Semi-automatic classification of bird vocalizations using spectral peak tracks. - Journal of the Acoustical Society of America **120**: 2974-2984.

CHESMORE, E .D. (2001): Application of time domain signal coding and artificial neural networks to passive acoustical identification of animals. - Applied Acoustics **62**: 1359-1374.

CROUCH III, W. B. & PATON, P. W. C. (2002): Assessing the use of call surveys to monitor breeding anurans in Rhode Island. - Journal of Herpetology **36**: 185-192.

FISCHER, P. F., SCHULZ, U., SCHUBERT, H., KNAPP, P. & SCHMÖGER, M. (1997): Quantitative assessment of grassland quality: acoustic determination of population sizes of Orthopteran indicators species. - Ecological Applications. **7**: 909-920.

KOGAN, J. A. & MARGOLIASH, D. (1998): Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden Markov models: A comparative study. - Journal of the Acoustical Society of America **103**: 2185-2196.

KWAN, C., MEI, G., ZHAO, X., REN, Z., XU, R., STANDFORD, V., ROCHET, C., AUBE, J. & HO, K. C. (2004): Bird classification algorithms: Theory and experimental results. - IEEE Int. Conf. Acoust. Speech and Signal Processing (ICASSP'2004). V: 289-292.

O'FARRELL, M. J. & GANNON, W. L. (1999): A comparison on acoustic verses capture techniques for the inventory of bats. - Journal of Mammalogy **80**: 24-30.

OPENSHAW, J. P. & MASON, J. S. (1994): On the limitations of cepstral features in noise. IEEE Int. Conf. Acoust. Speech and Signal Processing (ICASSP'1994). II: 49-52.

PARKER III, T. A. (1991): On the use of tape recorders in avifaunal surveys. Auk. **108**: 443-444.

PARSONS, S. (2001): Identification of New Zealand bats (*Chalinolobus tuberculatus* and *Mystacina tuberculata*) in flight from analysis of echolocation calls by artificial neural networks. Zoological Society of London **253**: 447-456.

PARSONS, S. & JONES, G. (2000): Acoustic identification of twelve species of echolocating bat by discriminative function analysis and artificial neural networks. Journal of Experimental Biology **203**: 2641-2656.

PETERSON, C. R. & DORCAS, M. E. (1992): The use of automated data acquisition techniques in monitoring amphibian and reptile populations. Pages 369-378 in D. McCullough (ed.). Wildlife 2001: Populations. Elsevier Applied Science, New York.

RIEDE, K. (1993): Monitoring biodiversity: analysis of Amazonian rainforest sounds. - Ambio **22**: 546-548.

RIEDE, K. (1996): Diversity of sound-producing insects in a Bornean lowland rain forest. - in D. S. Edwards, W. E. Booth, and S. C. Choy (eds.). Tropical rainforest research – current issues. Kluwer Academic Publishers, Netherlands: 77-84.

RIEDE, K. (1998): Acoustic monitoring of Orthoptera and its potential for conservation. - J. Insect Conservation **2**: 217-223.

TAYLOR, A., WATSON, G., GRIGG, G. & MCCALLUM, H. (1996): Monitoring frog communities: An application of machine learning. - Proceedings of the 8th Innovative Applications of Artificial Intelligence. 1564 – 1569.

VITERBI, A. J. (1967): Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. - IEEE Transactions on Information Theory **13**: 260-269.

Bioacoustic Classifier System Design as a Knowledge Engineering Problem

SEBASTIAN HUEBNER

Pestalozzistraße 5, 14482 Potsdam, Germany, E-mail: sebastian@sejona.de

Abstract. The problem of programming classifier systems for bioacoustic signals is seen as a knowledge engineering problem. Advantages and disadvantages of traditional black box approaches are discussed. A new knowledge engineering approach for bioacoustic classifier system design is described and advantages of the approach are highlighted. The basic architecture of a general purpose knowledge engineering environment for the bioacoustics domain is outlined. Emphasis is placed on both expert knowledge and on the role of machine learning within the classifier system design process. Several examples illustrate how the approach can be used to create classifier systems for bioacoustic patterns. Examples include multi scale spectrographic visualization, subsymbolic classifiers for *Prunella collaris* (Alpine Accentor), a hierarchical symbolic classifier system for *Prunella collaris*, annotated whistle- and click-sounds of *Tursiops truncatus* (Bottlenose dolphin) and automatic extraction of symbolic whistle- and whistle type representations.

For the end user most systems for bioacoustic pattern recognition are black boxes. Biologists and field researchers usually cannot understand how these systems work, why classification decisions (annotations) are generated and why false decisions occur.

This is not surprising. Automated classification of bioacoustic patterns is a difficult computational task. Algorithms used in this domain are just as complicated as in other fields of pattern recognition, e.g. in natural language- or in picture processing. At least basic knowledge in higher mathematics and in computer science is necessary to understand how such systems work.

However, as even the most advanced methods in bioacoustic pattern recognition are far from being perfect and as we do not have better classification algorithms to our immediate disposition, we can look for a practical alternative that help biologists to use existing bioacoustic classification algorithms in a more satisfactory way than up to now.

Such an alternative can be a knowledge engineering environment for computer aided bioacoustic classifier system design. The purpose of such an environment is to facilitate the modeling, implementation, test and application of bioacoustic classifier systems and to hide from the end user the complexity of all involved digital signal processing and pattern recognition algorithms.

In this article the fundamental architecture of a general purpose knowledge engineering environment for bioacoustic classifier system design is described and several examples are shown that highlight important aspects of the classifier system design process. In order to keep this article concise only the most basic topics are discussed and more advanced features, e. g. the resolution of logic classifier systems, are omitted.

Knowledge Engineering Bioacoustic Classifier Systems

In order to understand the reasons that speak for a knowledge engineering approach in bioacoustic pattern recognition it is best to first have a look at the advantages and disadvantages of the traditional or "black box" approach. Afterwards, it will become evident that the knowledge engineering approach is superior in this highly knowledge-intensive domain.

The Black Box Approach

The task of a black box classifier system (Fig. 1) is to process bioacoustic signals and to generate annotations (classification decisions) that indicate the presence of certain acoustic event types at certain points of time in the audio data.

In many cases it is the explicit wish of both the computer scientists (the constructors of the black box) and the experts (the biologists and field researchers) that classification is done unsupervised by the black box.

Usually, after the box has been implemented, the job of the computer scientists is done. Expert's control on the system is restricted to the input (the audio signals) and possibly to a few general parameters that configure algorithms inside the black box. The output (the annotations) can neither be influenced nor be understood properly by the experts. This is especially awkward if within an unsupervised classification process many false classification decisions occur for unknown reasons. Experts do not have the slightest chance to correct this type of ill behavior even if the underlying problem actually is trivial.

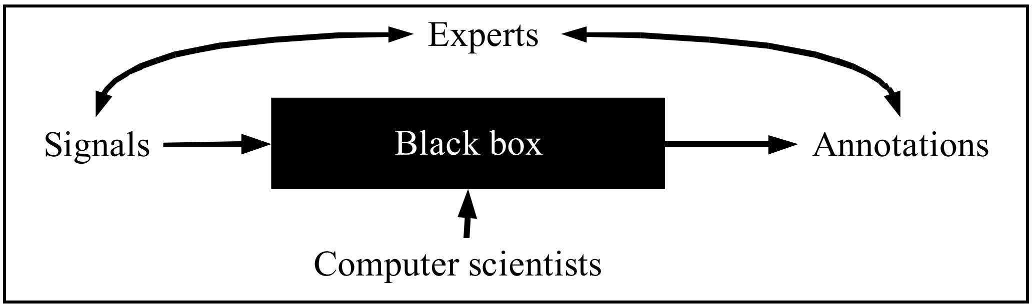


Figure 1: Black box approach.

It may be said that this approach fails the more often, the more complex the classification task is. It may also be said that the more complex the classification task is, the more expertise in bioacoustics and biology is needed to build the black box itself. Indeed, to solve complex bioacoustic classification problems within a black box approach a close cooperation of both computer scientists and biologists is inevitable.

Furthermore, in many cases experts start understanding the nature of bioacoustic sounds only after having intensively worked with them for a while. Expertise grows slowly but is needed to build and modify the black-box. Unfortunately, the modification of the box requires each time a close cooperation with the computer scientists. In order to take new important insights into account, the computer scientists have to be engaged for one more time – if the experts still can afford it.

It should be mentioned that the traditional black box approach has also some advantages. A black box can be a very compact solution efficiently implemented by computer specialists. Bioacousticians do not have to bother with computational details and may entirely rely on the qualifications of the programmers. However, the above mentioned circumstances lead to a whole series of severe disadvantages characteristic for this approach:

- Experts have low influence on how their expertise is implemented.
- Expertise incorporated in classifier systems is not sharable among experts.
- The work of the computer scientists is difficult to re-use.
- Experts can't understand why the system behaves the way it does.
- Experts have little chance to correct an ill-behaving system.
- The incorporation of new insights into the nature of bioacoustic signals requires a re-implementation of parts of the system.
- General inflexibility and expensiveness.

In many other expertise-intensive problem domains the art and science of knowledge engineering has proved to be much advantageous compared to black box solutions. How this approach can be understood in the bioacoustic classifier system design domain is described in the next section.

The Knowledge Engineering Approach

Knowledge engineering is the art and science to transfer domain specific expertise into a computer system in such a way that the expertise can both be computed by the system in a reasonable way and still be understood by experts without to much effort.

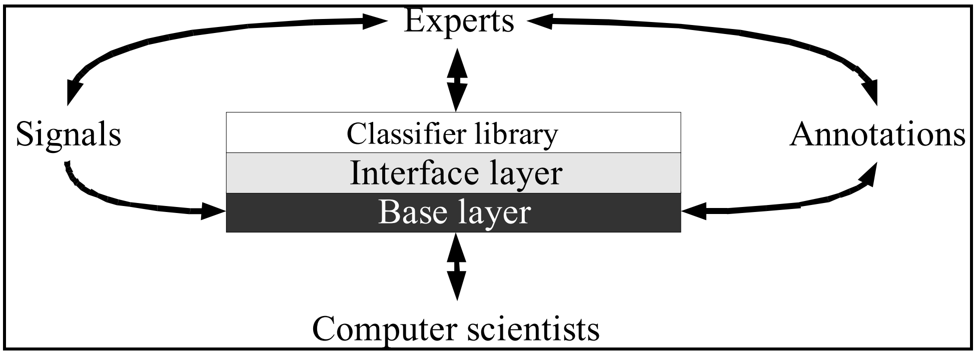


Figure 2: Knowledge engineering approach.

The claims of computability and at the same time comprehensibility require an entirely new conception of bioacoustic pattern recognition systems. A knowledge engineering environment (KEE) for bioacoustic classifier system design (Fig. 2) is very much different from the above described black box solutions:

1. The KEE can annotate audio material only if equipped with a library of appropriate knowledge based classifiers systems. *A priori* this library is not part of the system.
2. The KEE is designed to give experts (non-technicians) the opportunity to create classifier system libraries by making intense use of their individual expertise. Note that the creation of such libraries is a programming task comprising all typical steps of software engineering: modeling, implementation, test and application.
3. The KEE is equipped with a special knowledge design interface. This interface provides all means to model, test and apply classifier systems for bioacoustic data and at the same time hides the complexity of the machinery that carries out all kinds of difficult computations.
4. The KEE guarantees full transparency on all important levels to the experts. Subject of computation are not nebulous concepts, obscure networks or sophisticated but incomprehensible systems but *well defined* signatures of acoustic event types, *comprehensible* similarity measures and *clear cut* decision making algorithms.
5. The KEE provides immediate visual feedback within all steps of the classifier system design process. Visual feedback concerns the structure of signatures of acoustic events, the behavior of similarity measuring functions and classifier decisions generated by the system.

It is easy to see, that such an environment must be far more complex than just a black box classification engine made to respond to certain patterns in audio signals. Especially the interface layer can be a very extensive and complicated piece of software though at first glance it appears trivial as it's main purpose is to solve mere pragmatic problems.

However, given a state of the art knowledge design interface the architecture shown in figure

2 opens new possibilities. High performance pattern recognition algorithms may be linked into the base layer with ease. Experts may use them interactively and independently in order to create and modify classifier systems in accordance with their specific needs. New important bioacoustic insights may be taken into account immediately by the experts – without having to engage computer scientists for one more time.

The knowledge engineering approach has also some disadvantages that should not be left unmentioned here. The system itself is far more complex than black boxes are and experts have to do specific knowledge design work on their own. Also the needs of full transparency and comprehensibility within all steps of the classifier system design process may cause problems with certain types of existing sophisticated classification algorithms. However, the above described architecture leads to a whole series of advantages characteristic for this approach:

- Experts have full control on the flexible and interactive classifier system design process.
- Experts can understand *what* is computed and *why* annotations are generated.
- Expertise incorporated in classifier systems is sharable worldwide.
- The work of the computer scientists (the KEE) is suitable for a virtually infinite number of different tasks.
- Experts can easily correct an ill-behaving system (even if the underlying problem is not trivial).
- The incorporation of new insights into the nature of bioacoustic signals does not necessarily require re-implementations of parts of the system.
- Knowledge discovery and data mining instruments easily fit into the architecture.
- General flexibility and cheapness in the long term.

It can be concluded that the knowledge engineering approach does not only not suffer from the disadvantages of the black box approach but that it also has several additional advantages.

General Architecture

A KEE for bioacoustic classifier system design comprises a large variety of more or less independent tools, instruments and controls. The most essential of them may be grouped into three modules or *layers* as depicted in figure 2. In figure 3 these layers are explained in more detail. Small helper-tools are not included in the following description though they also play an important role in practical classifier system design tasks.

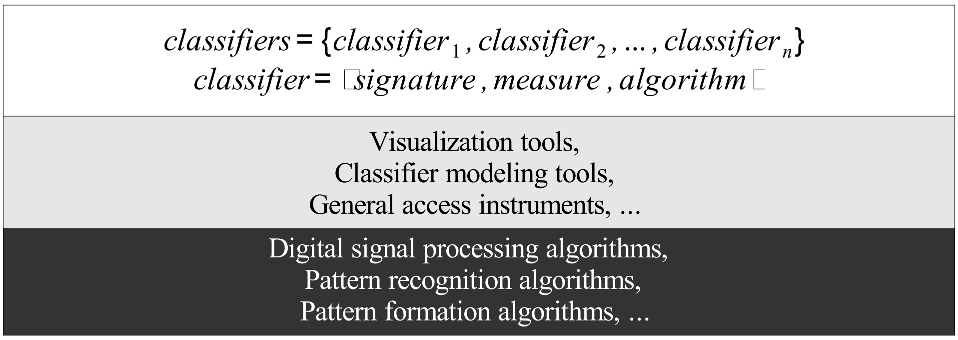


Figure 3: The three layers of a bioacoustic knowledge engineering environment.

The Classifier System Library

Transparency and comprehensibility are the two most important features of the classifiers in the classifier library. To achieve this, a classifier can be defined as a triple consisting of a signature, a similarity measure and an annotation generating algorithm:

- The *signature* is a well defined formal descriptor of the class of acoustic events that the classifier is supposed to detect in audio signals. Note that expertise is contained mainly in the signature - not in the other two constituents of the classifier. Signatures may be either *subsymbolic* or *symbolic*.
- The *similarity measure* is a well defined function that is suitable to compare a signature with a piece of audio data of exactly the same duration. By systematically applying signature and measure to a longer audio signal it is possible to compute the degree of similarity between the signature and the signal for each possible point of time.
- The *annotation generating algorithm* systematically evaluates an audio signal by making use of *only* signature and similarity measure. The algorithm creates annotations if certain constraints are satisfied. An annotation optionally may include a set of automatically extracted measurements such as the degree of computed similarity at positive classification decision time.

The above described classifier architecture lies at the heart of a comprehensible KEE for bioacoustic classifier system design. Shapeliness of both signature and similarity measure as well as purity of the annotation generating algorithm are decisive for scientific quality of both design and classification processes. Furthermore, properties and behavior of all three classifier constituents may easily be visualized and thus understood immediately by experts.

The Knowledge Engineering Interface

Classifiers for the library have to be designed in accordance with the requirements of the study or application they are needed for. The knowledge engineering interface gives experts the opportunity to model classifier systems independently. Tools in this layer may be grouped into data visualization, classifier modeling and general access instruments.

- *Visualization* tools serve three main purposes: (1) Visualization of audio data (e.g. in form of spectrograms), (2) visualization of classifiers (including signatures, numerical, nominal and hierarchical properties) and (3) visualization of classifier decisions and of the behavior of similarity functions. Visualization tools need to be intuitive and precise.
- *Classifier modeling* tools serve all tasks necessary to initialize and modify classifier systems and their constituents (e.g.. subsymbolic and symbolic signatures). Interactive manual editing tools known from picture editors (e.g. 'rubbers') belong to

this group as well as automatic signature extraction, merge and clustering instruments necessary for data mining of signatures.

- *General access* instruments are necessary to control all data flows within the KEE. They serve the access to audio file collections, sets of annotations as well as the configuration and conduction of classification tasks. Access instruments often have the form of *wizards* that guide through the configuration of digital signal processing and pattern recognition tasks.

Future knowledge engineering interfaces may with ease include dozens of different such tools. For example, simple subsymbolic signature extraction may be based on dozens of different time-frequency or time-energy based visualizations and the number of thinkable signature modeling tools is limited only by the number of algorithms linked into the base layer.

The Base Layer

The foundations of a KEE are contained in the base layer. It is invisible to experts and implemented by computer scientists. For experts it is a black box but it's algorithms may be freely accessed through the knowledge engineering interface. Tools in this layer may be grouped into digital signal processing (DSP), pattern recognition and pattern formation algorithms.

- *Digital signal processing* algorithms are necessary for all kinds of computation typical for audio data processing. Included in this group are time-energy, time-frequency an wavelet transforms as well as filtering, signal generating and mapping algorithms.
- *Pattern recognition* algorithms are used to compute annotations by searching patterns in audio signals. Annotation generating algorithms are physically represented in this layer, not inside the classifiers which actually contain only a symbolic algorithm identifier and a set of related parameters.
- *Pattern formation* algorithms are necessary for advanced data mining functionality. They are especially useful for inductive programming of signatures either on subsymbolic or symbolic levels. Note that inductive programming of signatures is one of the most promising applications withing knowledge engineering based classifier system design. New and interesting acoustic patterns may be revealed by such algorithms.

Today, we can choose from a wide variety of high performance algorithms for the first two subgroups of the base layer. Most advanced are low level DSP standard algorithms like DFT or Wavelet-transform computing routines. Selected pattern recognition algorithms may be adopted from natural language and picture processing. Suitable pattern formation algorithms, however, are rare especially in the field of subsymbolic signature formation. Symbolic data mining methods can be used to work with sets of annotations.

The Classifier System Design Process

Knowledge engineering simple classifier systems is a straightforward programming task. In such cases the classifier system design process fits into the waterfall-scheme known from traditional software engineering. This scheme however is applicable only if both the structure of the acoustic events to be classified is highly deterministic and all types of events are *a priori* known by the expert who programs the classifiers.

Knowledge engineering of complex classifier systems is a knowledge discovery task. In such cases the classifier system design process should be conducted within the general knowledge discovery in databases (KDD) procedure (SHAPIRO & FRAWLEY 1991, FRAWLEY & SHAPIRO 1992). This scheme is applicable if either the structure of acoustic events is not deterministic or not all types of acoustic events are *a priori* known by the expert.

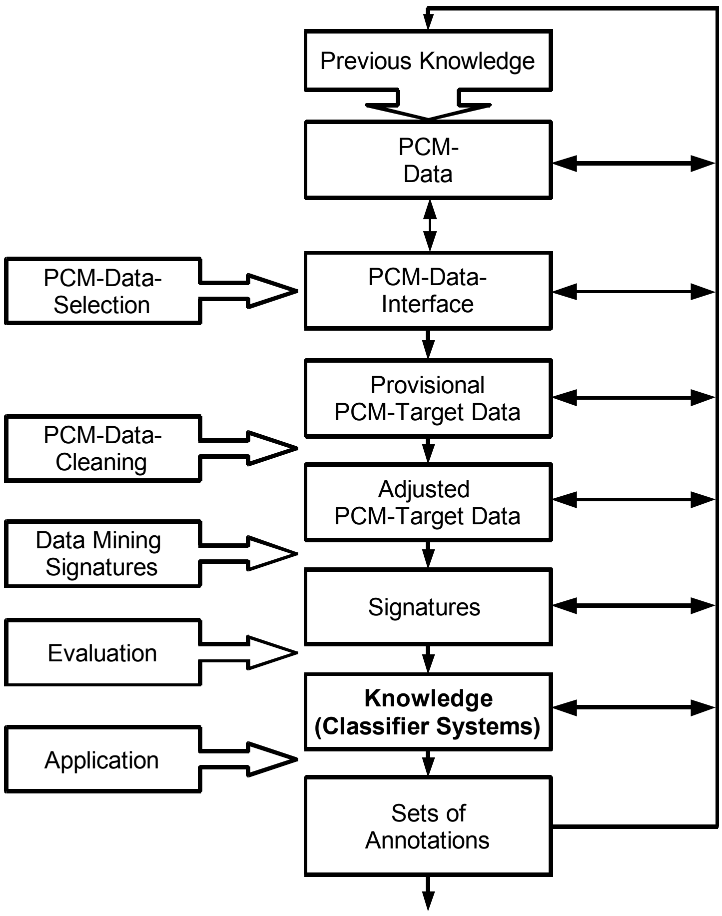


Figure 4: KDD for bioacoustic classifier system design.

The conduction of a KDD process on audio file collections can lead not only to better audio signal classifier system libraries but also to the discovery of new and interesting acoustic patterns. Such patterns can become relevant to scientific discoveries.

Examples

It has been mentioned above, that knowledge engineering of bioacoustic classifier systems is a dynamic and interactive process. It may resemble computer gaming much more than dry source code programming. Unfortunately this dynamics cannot be transferred to two dimensional paper. Thus we are limited to provide a few examples of visualizations of classifiers and annotations that represent typical results of properly conducted knowledge engineering processes. All classifiers and graphics were created with the acoustic knowledge engineering environment DSProlog (HUEBNER 2007a, 2007b, http://www.sejona.de/dsprolog.php).

Alpine Accentor *(Prunella collaris)*

Standard Spectrographic Visualization

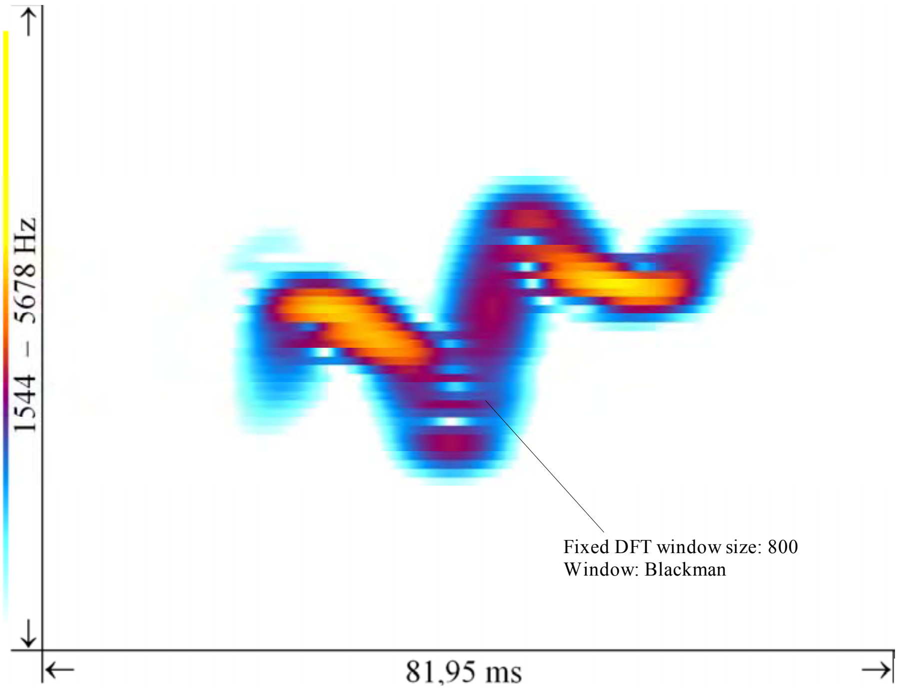


Figure 5: Standard spectrogram of a syllable of Prunella collaris. Frequency resolution is fine grained but time resolution is poor.

Improved Spectrographic Visualization

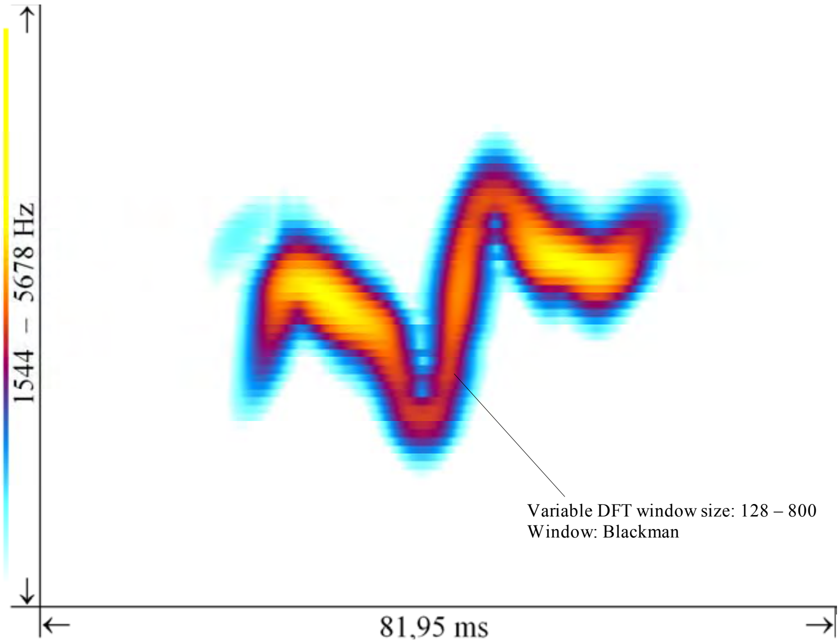


Figure 6: Multi scale spectrogram of a syllable of *Prunella collaris*. Both time- and frequency resolutions are fine grained.

Template Based Subsymbolic Signature

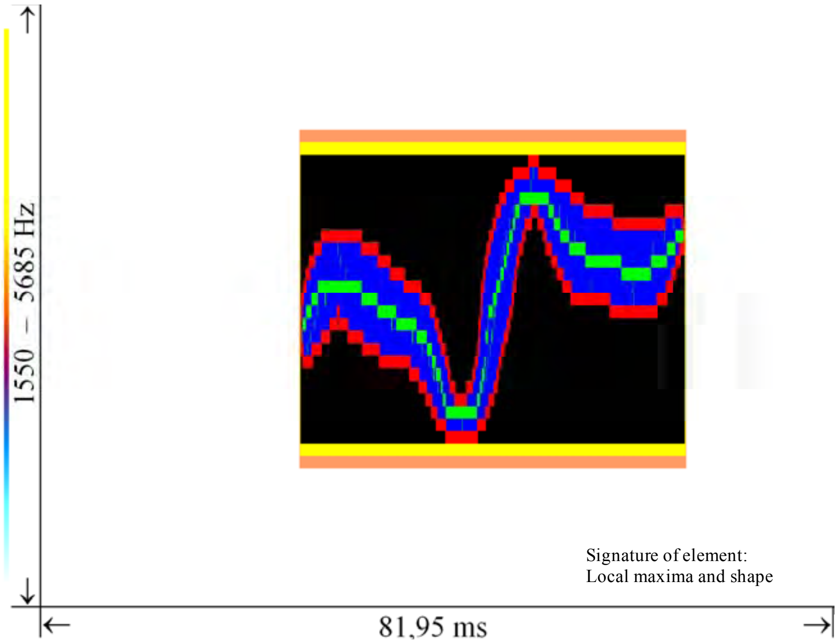


Figure 7: Template based signature in highlight mode. The signature was derived manually from one single example by applying a template to the spectrogram in figure 3. Green are local maxima of Fourier coefficients. Red is the shape of the acoustic event. Blue are above threshold coefficients important for pattern formation algorithms.

Induced Subsymbolic Signature

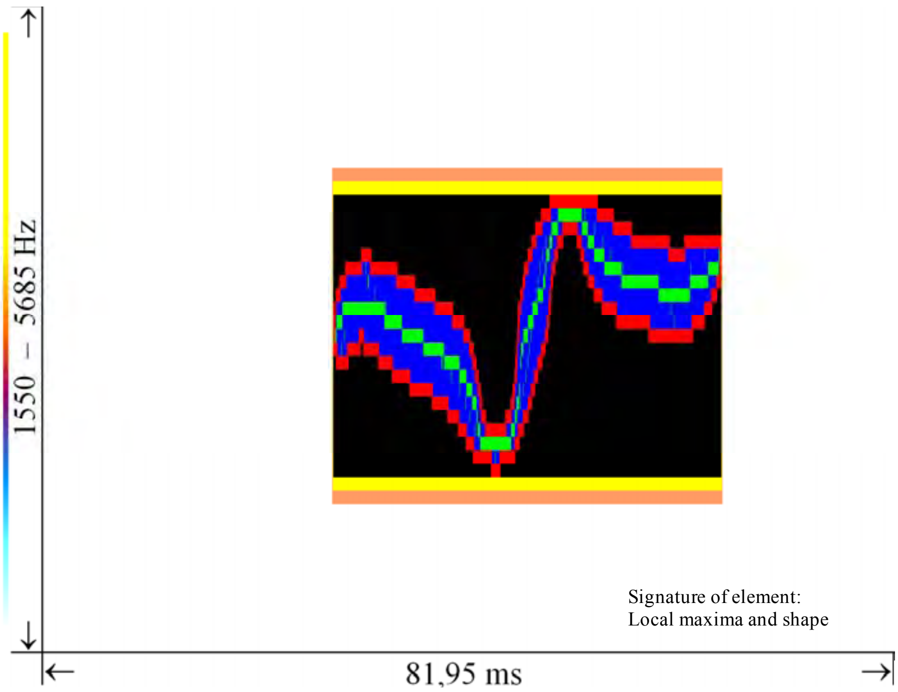


Figure 8: Induced signature in highlight mode. The signature was induced by merging twelve similar patterns found by a classifier using the signature in figure 4. Induced signatures usually perform better than template based signatures though they look similar at first sight.

Visual Feedback - Classifier Decisions and Similarity Measures

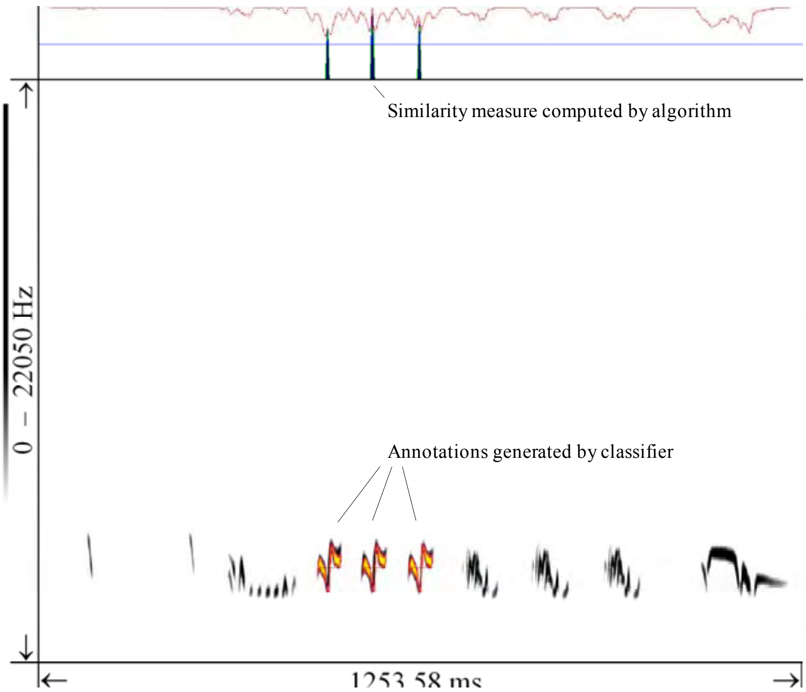


Figure 9: Visual feedback - classifier decisions and similarity measures. This diagram shows how the classifier in figure 5 performs with respect to a short strophe of *Prunella collaris*. Colored visualizations of classifier decisions are mapped onto the black and white background of the spectrogram. Similarity vectors computed by the annotation generating algorithm are shown on top.

A Hierarchical Classifier

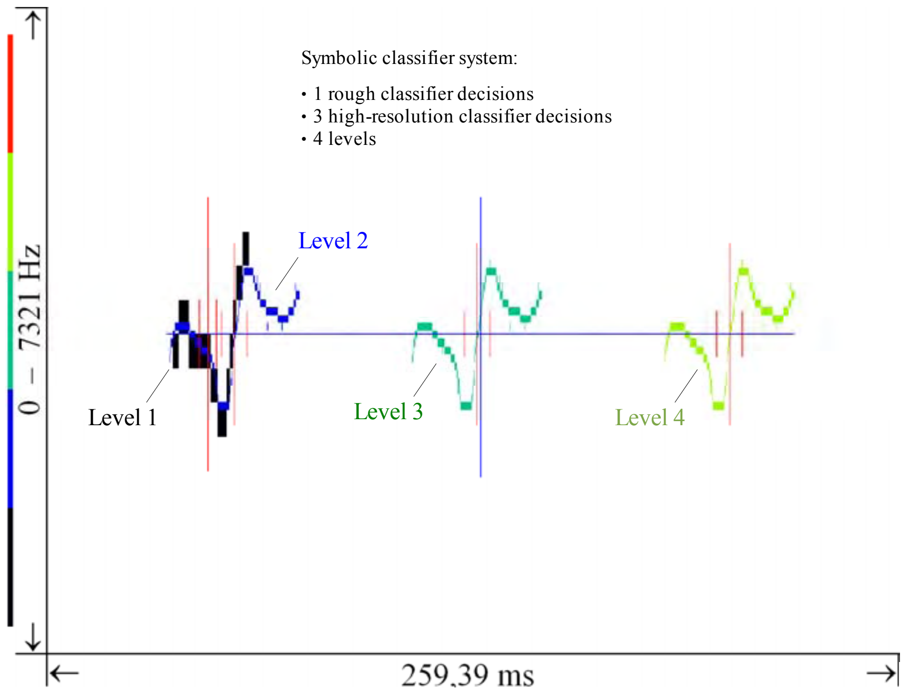


Figure 10: Symbolic hierarchical classifier system for a sequence of syllables. Hierarchical symbolic classifier systems can be quicker and more accurate than simple classifiers. They can be designed to recognize entire sequences of elements. Tolerance parameters enable fuzzy pattern recognition. The levels of the internal hierarchy are color coded.

Bottlenose Dolphin *(Tursiops truncatus)*

Annotated Whistle

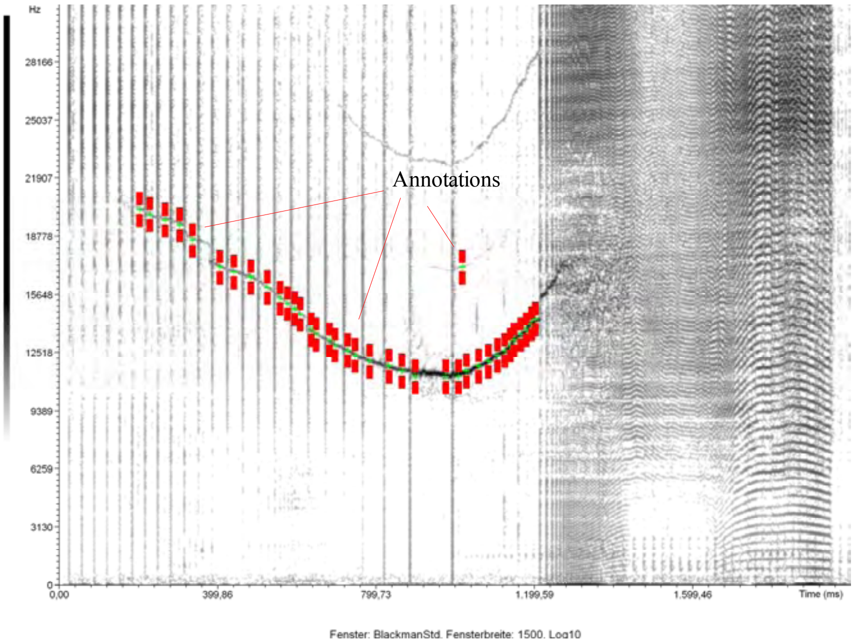


Figure 11: Annotated whistle of *Tursiops truncatus*. Annotations were generated by a simple classifier for very short sine-like events. The whistle is overlapped by loud click sounds. Classification decisions are generated only between the clicks of the train. The sequence of annotations represents the contour shape of the whistle.

Annotated Click Trains

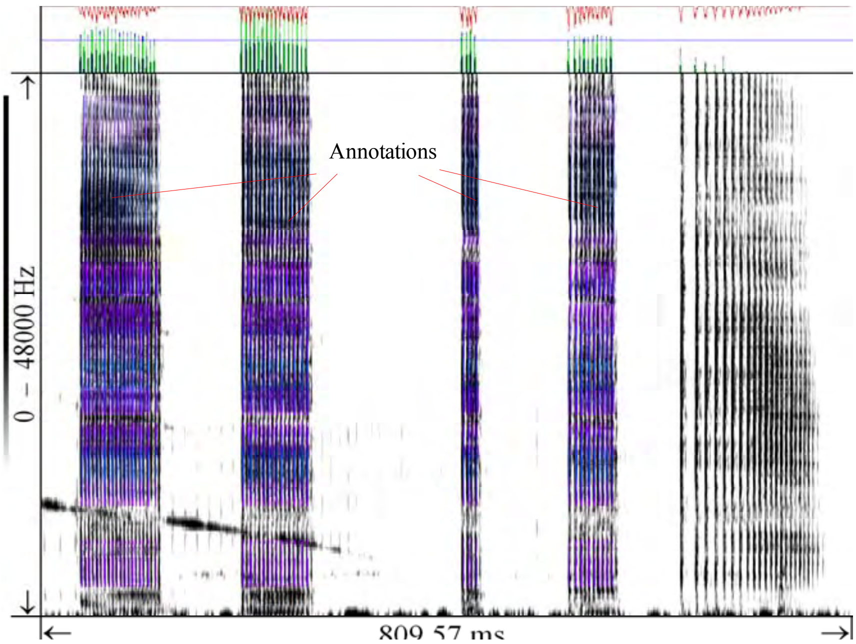


Figure 12: Annotated click trains of *Tursiops truncatus*. Classifiers respond only to clicks with an certain internal time-frequency shape. The clicks in the last train have a different shape and are ignored by the classifier. Similarity vectors computed by the annotation generating algorithm are shown on top.

Automatically Extracted Symbolic Whistle Representations

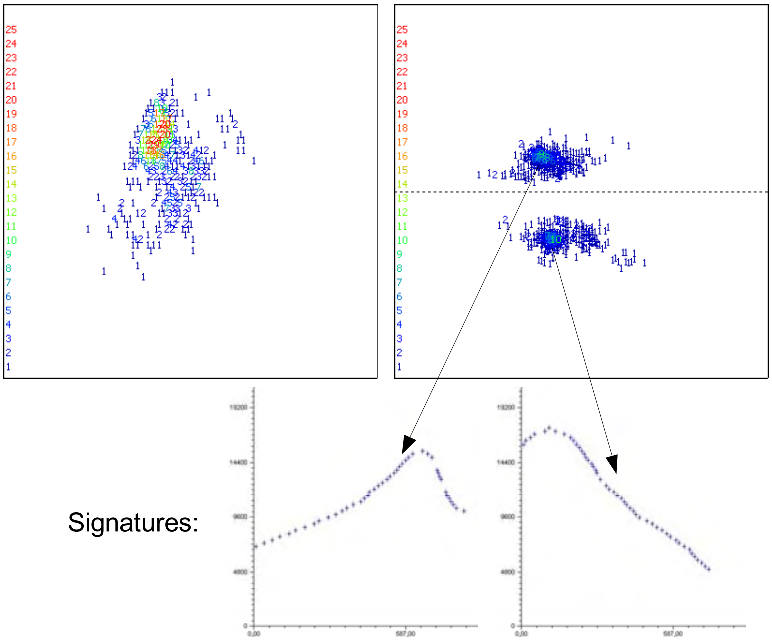


Figure 13: Scatter plot of 1072 long symbolic whistle signatures. Left: center frequency to bandwidth. Right: center frequency to gradient. Below: two automatically extracted symbolic signatures from the center of the clusters.

Induced Symbolic Whistle Representations

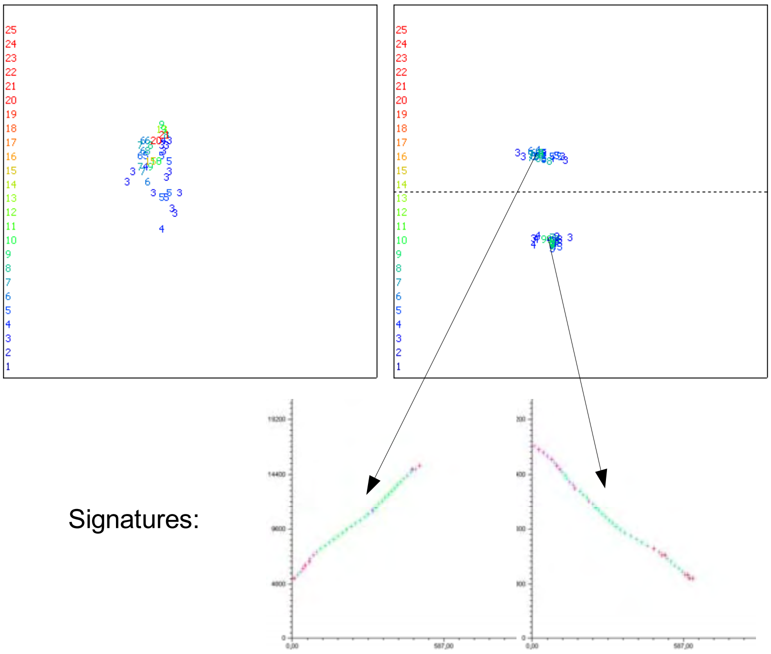


Figure 14: Scatter plot of 52 induced symbolic whistle-type signatures generated from 1072 symbolic signatures. Left: center frequency to bandwidth. Right: center frequency to gradient. Below: two symbolic signatures from the center of the clusters. Colors indicate the degree of computed support of single symbolic elements.

Conclusions

Knowledge engineering bioacoustic classifier systems is a new approach that has the potential to solve many severe problems of traditional black box approaches and to open new horizons in bioacoustics:

- Biologists and field researchers will be given the tools they need to independently create classifier systems for their specific needs. It does not matter whether they work with birds, fish, frogs, insects or marine mammals.
- Data mining bioacoustic file collections may lead to the finding of new interesting acoustic patterns. From other fields of science (e.g. from genetics) it is known that such findings can lead to new scientific discoveries.
- Classifier libraries are independent from the knowledge engineering environment and can be shared worldwide through the Internet. Sharing expertise based classifiers not only facilitates day-to-day work but also has the potential to make bioacoustic processes tractable on a large scale.

Today, knowledge engineering bioacoustic classifier systems is at it's very beginning. Much theoretical, methodological and practical work still has to be done to create reliable knowledge engineering standards valid for the bioacoustic community. However, this work is worth to be done and certainly will pay.

References

FRAWLEY, W., SHAPIRO, G. & C. MATHEUS (1992): Knowledge discovery in databases - an overview. AI Magazine, pages 1-34.

HUEBNER, S (2007a): Wissensbasierte Modellierung von Audio-Signal-Klassifikatoren. Zur Bioakustik von *Tursiops truncatus*. Dissertation, University of Potsdam.

HUEBNER, S. (2007b): Wissensbasierte Modellierung von Klassifikatoren für Zeit-Frequenz-Muster in PCM-Daten. Logos: Berlin.

SHAPIRO, G & W. FRAWLEY, EDS. (1991): Knowledge Discovery in Databases. AAAI/MIT Press, Cambridge, MA.

Computational Methods in Analysis of Bird Song Complexity

JUHA T. TANTTU & JARI TURUNEN

Tampere University of Technology, Pori, P.O. Box 300, FIN-28101 Pori, Finland
E-mail: juha.tanttu@tut.fi, jari.j.turunen@tut.fi

Abstract. We will present algorithms that have been developed for the complexity analysis of the male Pied Flycatcher territorial songs. The songs were recorded in Ruissalo, Turku, Finland, during May 5th - 17th, 2006, by Dr. Toni Laaksonen from University of Turku with his students. Because of the environment and partly because of the inexperience of the students the 14 hours of recordings are extremely noisy and contain many unwanted sounds. So the analysis of the recordings has been quite a challenging task.

The computational bird sound analysis has been an active research area over several decades. Probably the most common tool in use has been the song spectrogram. However, usually the use of spectrogram has been combined with human made visual assessment of the results. The advancements in signal processing and pattern recognition algorithms and constantly increasing computer power offer, at least in principle, possibilities to automate the process even further. In this paper we discuss algorithms for measuring the complexity of the song of male Pied Flycatcher (*Ficedula hypoleuca*) and the lessons learned during the process. Our aim was to reduce the human and manual work as much as possible because the amount of data was quite large and our resources were quite limited.

A widely used measure of the song complexity is the repertoire. There are two types of repertoires: a *song repertoire* where the male sings several different song types but individual song types do not vary much and a *syllable repertoire* where a number of syllables are recombined to produce different songs (COLLINS 2004). In both cases it is essential to find the basic elements (sometimes called notes) of the song spectrogram.

LAMPE AND SÆTRE (1995) have shown that high quality male Pied Flycatchers (defined by brighter plumage, better condition and more experience) have larger song repertoires. LAMPE AND ESPMARK (2003) propose that female Pied Flycatchers use the song to find the high quality males and their territories. Also the song similarity seems to predict hybridisation of the sympatric species pied and collared (*F. albicollis*) flycatcher (QVARNSTÖM ET AL. 2006). So considerable biological interest for objective measures of similarity and diversity of Pied Flycatcher territorial song exists.

“However, sometimes the research ideas and harsh reality will collide very roughly”. In the development of automatic pattern recognition methods large data sets of sounds are needed. Our experience has been that the bird species that are interesting for the researchers, are usually boring for the professional natural sound recordists. Recording sound samples during mating season from several individuals of common bird species is not a fascinating task since it is usually the rare species that are the first target for an ambitious recordist. Our solution to obtain sufficiently many recordings of songs of Pied Flycatcher males was to use student resources with professional biologist / recorder guidance. Since several students made the recordings differences in the recording techniques posed an extra challenge for sound analyst and especially the normalization of different recordings was quite difficult.

The question - does the complexity of the song of the male Pied Flycatcher correlate with its mating success – is quite complex, because of several uncertainties. How faithful the male is? What does the female hear, and especially prefer, from the male song, is it pitch, complexity or something else? Are we sure that the male from nestbox number 101, is always singing near the nestbox 101?

Sound analysis can not answer these questions, because the only information is in the recording metadata (date, nest, etc), and thus the sound analyst must rely on the recordist’s skills and ability to identify the individual males. Usually these assumptions are correct in our case because professional biologist work together with the student recorders.

Ruissalo island is located in the city of Turku, Finland and has plenty of good habitats for Pied Flycatchers and many other bird species as well but it is also a popular recreational area for Turku citizens. Thus the recordings from the spring 2006 contain besides many overlapping songs of different bird species, wind and sea noise also lot of anthropomorphic noise including local ferry traffic, harbour noise, cars and talking and walking people.

Methods

The recorded sounds were analyzed in Matlab environment. First the amplitudes of the recordings were normalized into the range [-1, 1] and averaged to zero. Typically, the male Pied Flycatcher sound frequency is in the range 3500 – 6000 Hz with occasionally drops down to 2000 Hz and sometimes up to 10000 Hz. So it is safe to use 1500 Hz highpass filtering to remove the low frequency wind noise. The wideband noise was reduced from the recordings by using a filter bank based reduction method. The sound was divided into eight frequency regions. The standard deviation of the highest frequency band was multiplied by three and this number was selected as noise threshold. This noise threshold was subtracted from amplitude values of each frequency band, and values below the threshold were zeroed out. The subtracted frequency bands were summed together to produce a ‘cleaned’ sound.

The segmentation was performed in two phases. A Matlab algorithm sought energy concentrations from the recordings, but the human analyst’s task was to confirm whether the found sound segment was the song of a male Pied Flycatcher or not. The accepted songs were then stored for analysis and the segments containing many simultaneously singing Pied Flycatcher individuals, sounds heavily disturbed with background noise or songs from other species were discarded.

The accepted songs were then split manually to elements. The element is the smallest continuous unit that can be identified from the spectrogram. In this phase the misidentified or too noisy sounds were discarded. After the position of each element in each song was determined the data was ready for automated complexity analysis.

Pied Flycatcher is a tonal singer, and the elements of the male song have a very powerful first harmonic component in the spectrogram. This feature is very easy to estimate in time domain using for example a high order (20^{th} order in our case) autoregressive model (KAILATH ET AL. 2000) . The autoregressive model is computationally easy to estimate directly from the waveform and it reduces the effects of remaining background noise (Fig. 1).

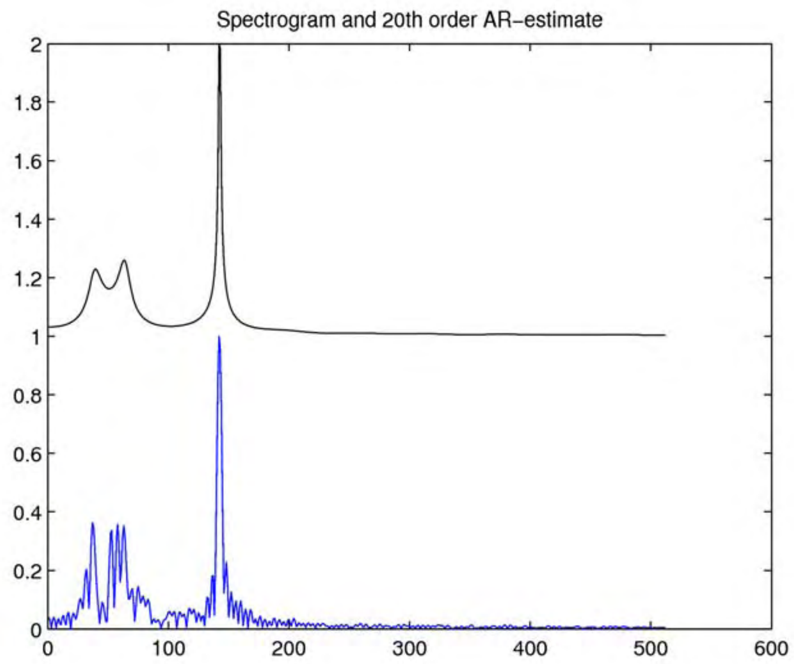


Figure 1: Fourier spectrum of a part of the song element (blue line) and its 20th order autoregressive estimate (black line). Both curves are normalized and the AR estimate is biased for visualization.

The vocabulary building process, i.e. element identification, proceeds as follows: Each new element is compared to the earlier labelled elements. If the similarity is below an experimental threshold, the element in question is copied to same directory where the similar element resides. Otherwise, if none of the elements in current directories show similarity, new directory is created and the element in question is copied there.

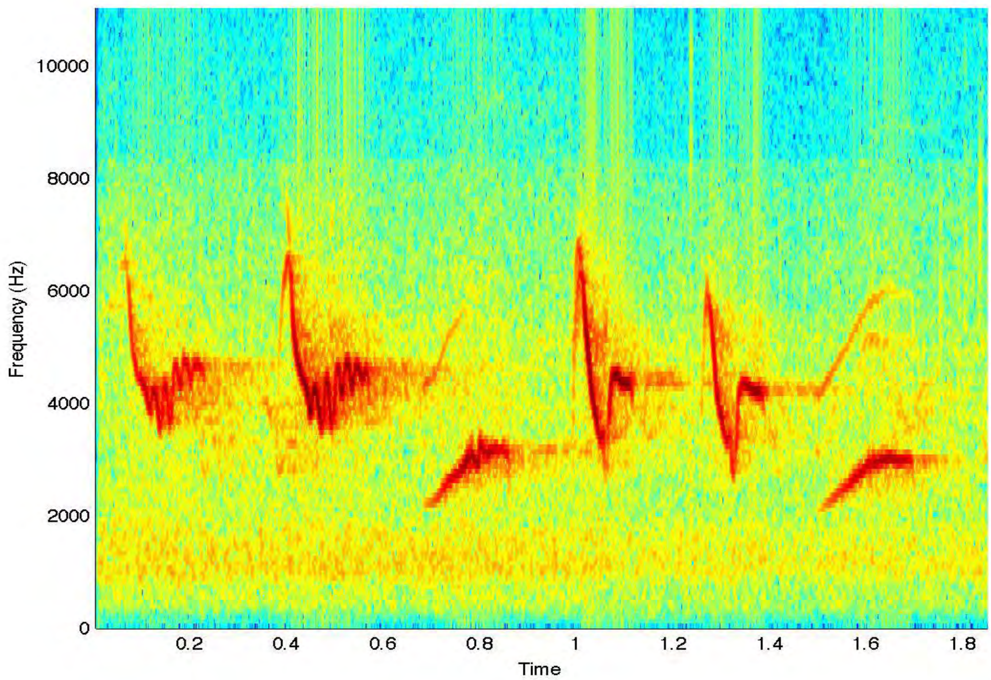


Figure 2.: An example of element diversity in Pied Flycatcher song.

The element similarity assessment is also a difficult task, because in the male Pied Flycatcher song the element size and shape vary over time (Fig. 2).

The similarity of the first harmonic components was measured using a modified dynamic time warping (DTW) algorithm (KRUSKAL & LIBERMAN 1983). The original DTW searches the optimal path through a matrix, that can be used to convert one item to the other. The point-to-point similarity distance is measured using Euclidian distance measure. However, the path information and the cumulative cost based on Euclidian measure may give false results, because any item can be forced to be another using the warping path. Instead of the path and the cost, we used another metric which measures the warping path distance from the diagonal. If the minimum cost warping path goes through the diagonal, the two items are identical and thus the distance measure is zero. The amount of computation is also reduced by comparing the lengths of the first harmonic components. If the lengths of the first harmonics differ more than 20 percent, they probably do not belong to same class and they are not compared. Otherwise the DTW-similarity measure was performed. After the DTW-similarity measure the element directories and labels are ready for further analysis.

Results

A total of 2237 songs, (867 rejected and 1370 accepted) were analyzed. The accepted 1370 songs were then manually split into elements. The number of elements varies from 2 to 20 per song with the mode of the distribution being 9 elements per song. The distribution of the number of song elements is depicted in Figure 3.

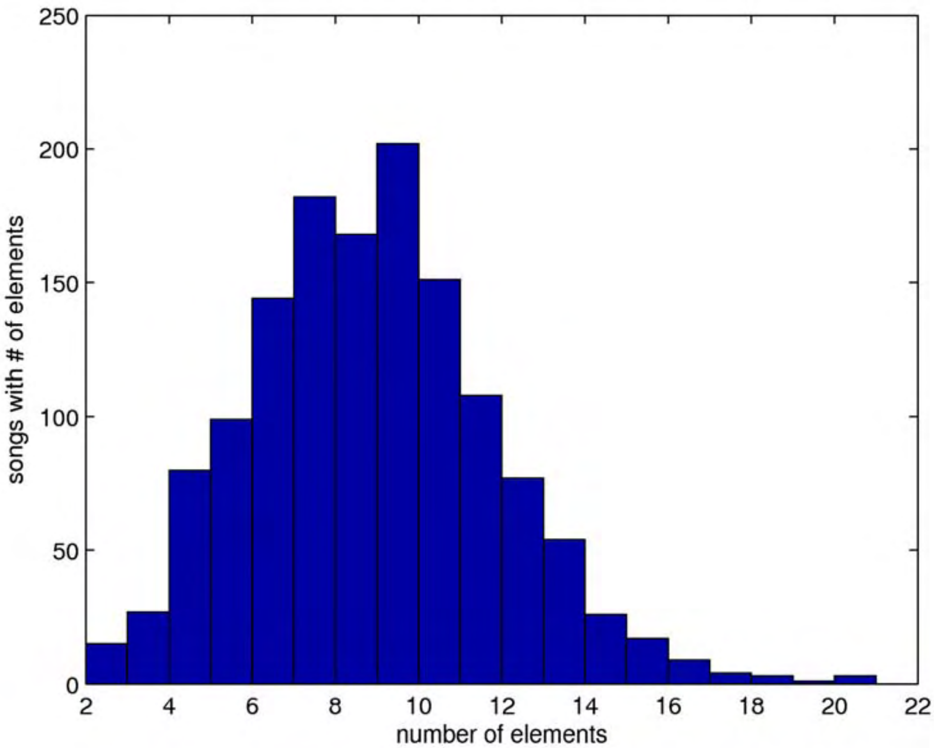


Figure 3: The distribution of the number of elements per song.

Discussion

The song structure identification is basically a straightforward task. First build a vocabulary and then label every element in each song. But the interpretation of the song structure and whether it affects the attractivity of the song to female are more difficult questions. There are several acoustic features that can be used for measuring the sound parameters, but which are the right ones? The only way to clarify this problem is to trace the biological fingerprints of the male to the juveniles using for example DNA sequences, number of juveniles per female etc. At this moment the DNA sequence identification is currently going on by Dr. Laaksonen's research group in University of Turku.

In signal processing, the Fourier transform based spectrogram is a widely used tool for sound analysis. However, there are several other candidates for the analysis of nonlinear signals. To mention few, Choi-Williams distribution, Wigner-Ville Distribution, Hilbert-Huang Transform and many others are available. Most of these methods can reveal nonlinear structures that Fourier transform is unable to reveal (QUATIERI 2002).

In statistical bird sound analysis there is always the concern about the number and quality of the recordings. How many song samples are sufficient to reliably estimate the complexity (or attractivity) of the sound? It is obvious that the more sound samples we have the more reliable our analysis will be. However, gathering large amounts of sound data requires plenty of time or many recordists. Several different recordists will record using several different ways. So in order to minimize the variation in the recordings, the recordists should be trained to use the equipment as similar ways as possible.

Large number of recordings will need an organized storage and retrieval system. In other words, the storing and managing the data needs a decent database system. Also the manual work that is required in the analysis of the recordings will become more difficult when the number of recordings is very large. This basically means that there is a need to automate the sound analysis algorithms as much as possible. A fully automated method is usually very difficult to develop because of the diversity of the nature recordings. In one record there might be several different species singing at the same time, animal sounds, human made sounds, wind, noise etc. In most of these cases, an experienced human can easily identify bird sounds from the spectrogram, while the fully automated methods will face huge problems. Perhaps the specific tools that will minimize the manual work are the best approach at the moment.

References

COLLINS, S. (2004): Vocal fighting and flirting: the functions of birdsong. - In P. Marler & H. Salbbekoorn (eds), Nature's Music The Science of Bird song. Elsevier Academic Press: 39-79.

KAILATH, T. , A. H. SAYED & B. HASSIBI (2000): Linear Estimation, Prentice Hall, New Jersey.

KRUSKAL J. B. & M. LIBERMAN (1983): The symmetric time-warping problem: from continuous to discrete. - In D. Sanko & J. B. Kruskal (eds.), Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparisons. Addison-Wesley, Reading, Massachusetts, 1983.

LAMPE, H. M. AND G.-P. SÆTRE (1995): Female pied flycatchers prefer males with larger song repertoires. - Proc. R. Soc. Lond. B, **262**: pp. 163 – 167.

LAMPE, H. M. & Y. O. ESPMARK (2003): Mate choice in Pied Flycatchers *Ficedula Hypoleuca*: can females use song to find high-quality males and territories? - Ibis **145**: E24 – E33.

QUATIERI, T. F. (2002): Discrete-Time Speech Signal Processing: Principles and Practise. Prentice Hall, New Jersey.

QVARNSTRÖM, A., J. HAAVIE, S. A. SÆTHER, D. ERIKSSON, AND T. PÄRT (2006): Song similarity predicts hybridization in flycatchers. - Journal Compilation **19**: 1202 – 1209.

Automated Monitoring of Avian Flight Calls During Nocturnal Migration

THIJS SCHRAMA[1], MARTIN POOT[2], MAGNUS ROBB[3] & HANS SLABBEKOORN[1]

[1] Institute of Biology Leiden, Leiden University, PO Box 9516, 2300 RA Leiden, The Netherlands
[2] Bureau Waardenburg, Consultants for Environm. & Ecology, PO Box 365, 4100 AJ Culemborg, The Netherlands
[3] The Sound Approach, Baarsjesweg 237/3, 1058 AA Amsterdam, The Netherlands
E-Mail of the corresponding author: TSchrama@FSW.leidenuniv.nl

As part of a project involving monitoring bird migration over windmill farms in the North-Sea (KRIJGSVELD ET AL. 2003), we initiated the development of an automatic monitoring system based on avian flight call detection and identification. Our goal is to develop a computerized acoustic monitoring system for assessing number and species of birds passing by on nocturnal migration. Such an automated monitoring system has several advantages over human observation such as the ability to 1) make observations 24 hours a day, 2) collect data in environments which are uncomfortable to human observers, 3) perform objectively and consistently over time and 4) accumulate comparable data in multiple units spread out over the geographical area of interest. Many bird species give flight calls during nocturnal migration, especially waterbirds and songbirds. These calls can be identified to the level of species or at least species group by a trained ear and offer a way to quantify migration in the lower air layers during hours of darkness (EVANS & ROSENBERG 1999). Using these nocturnal flight calls, an automatic bird call recording, detection and identification system for North-western European species has been developed. First an algorithm was developed to detect Acoustic Events on continuous recordings, subsequently a bird species is associated with an Acoustic Event by finding a match within a pre-established flight call library based on Dynamic Time Warping and a Euclidian Distance Algorithm.

Aim and Methods

The principle aim of the current study was to calibrate and compare the automatic flight call detection and identification system to a human observer. In October 2005 we made a 12 hour pilot-recording during nocturnal bird migration on the Oosterscheldekering, in the South-West of the Netherlands (51°37'48.13"N, 3°42'19.82"E) (Fig. 1). We made both field observations as well as automatic recordings of flight calls, using a Sennheiser Shotgun microphone ME67 connected to a custom built battery powered preamp and a digital 20 GB hard disk recording unit (Creative ® Jukebox III).
The software program was written on the Matlab® and LabVIEW® platforms to detect so-called Acoustic Events in continuous recordings. We developed a Maximum Normalized Narrowband Amplitude algorithm to detect the faint bird calls in continuous recordings. The procedure involves a step of normalizing frequency bands by a low-pass filtered version of the same frequency bands. For each instant of time the maximum amplitude from all normalized frequency bands is selected. An Acoustic Event will be detected when this amplitude exceeds the average Maximum Normalized Narrowband Amplitude. This method is simple to code and extremely fast in execution: 30 to 300 times faster then real-time. It has been designed to be sensitive to faint bird calls while maintaining a robustness against false positives due to environment noise such as car traffic, airplanes, or sea waves.

Figure 1: Installation of the continuous sound recording set-up in October 2005 on the Oosterscheldekering, in the South-West of the Netherlands for the audio recordings of avian nocturnal flight calls. Simultaneously bird migration was studied with radar and auditive observations were made by human observers (photo Bureau Waardenburg 2005).

Two Classification Algorithms

Our first classification algorithm was based on a Euclidian Distance Algorithm to associate a bird species with the Acoustic Events. The algorithm searches for the optimal match within a pre-established flight call library using a set of seven acoustic parameters: call duration, highest frequency, lowest frequency, loudest frequency, average bandwidth, maximum bandwidth and average frequency slope. The Euclidian distance was based on the similarity in these seven parameters between the Acoustic Events and the reference set for 12 species in the library (based on a total of 574 calls).

The second, more advanced algorithm makes use of Dynamic Time Warping (BROWN & MILLER 2007). The Dynamic Time Warping algorithm seeks the optimal alignment of the vector, representing the frequency contour, with each of the 12 vectors representing the frequency contours of the same 12 library species. Dynamic Time Warping allows for small variation within species to be ignored, reducing false negative results. The algorithm associates a bird species with the Acoustic Events by finding the best match within the pre-established flight call library.

System Performance

We compared the two automatic identification algorithms, Euclidian Distance and Dynamic Time Warping, with the species identification of an experienced observer scoring by ear. Actual field recordings of nocturnal migration were played back to both the automated identification system and the human observer. Due to the inherent low signal to noise ratio of the distant bird call recordings, the more advanced Dynamic Time Warping algorithm does not always perform better compared to the Euclidian Distance applied to the same recordings (Fig. 2).

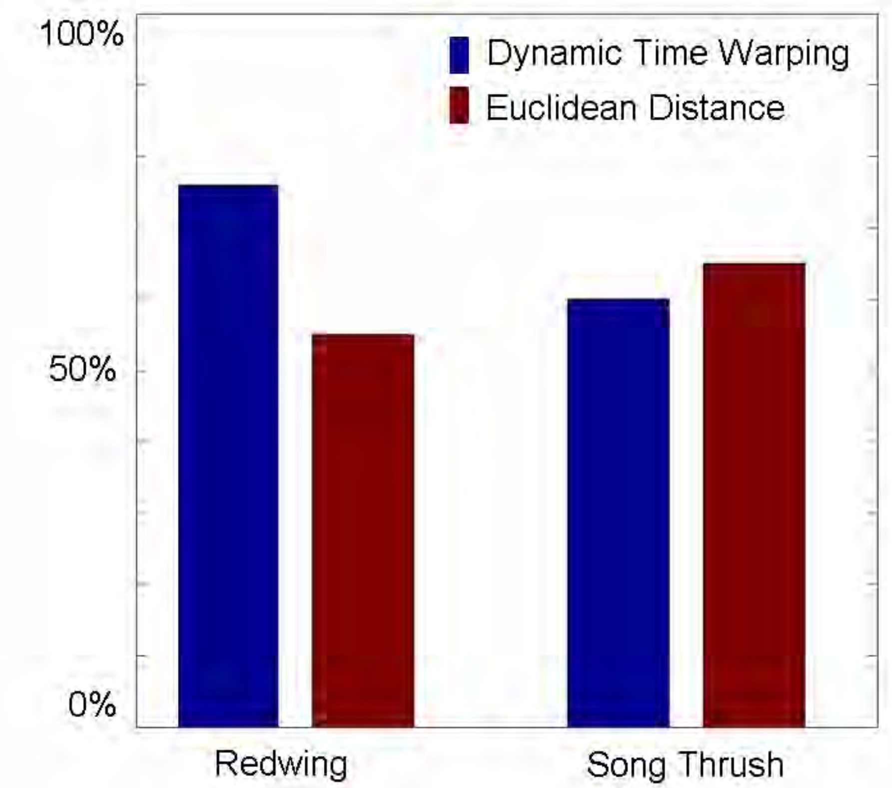


Figure 2: Performance of the two recognition algorithms as evaluated by an experienced human reference listener (MP).

The automated system detects less bird calls than a experienced user (Table 1). This will likely depend on the migrating altitude of the Birds. Nevertheless the system is capable of giving an objective measure for call density which can be used to infer bird migration intensity.

Table 1: Performance of the computerized flight call detection and identification system on actual field recordings made on the Oosterscheldekering, as evaluated by simultaneous observations by an experienced bird observer.

	Human field observations	Automated field observations	ratio auto/human
redwing	158	35	22%
song thrush	112	39	35%
total	270	74	27%

Conclusion and Future Work

Due to the high altitude of the migrating birds, the recordings have an inherent low signal to noise ratio. Applying a more advanced Dynamic Time Warping classification algorithm did not improve the classification of these noisy recordings. Improvement of the classification performance is more likely to be obtained by applying a more advanced denoising algorithm, not by a applying a more advanced classification algorithm. The comparison of the results of an automatic classification system with the species identification by an experienced human observer showed that for European species like redwing the automatic identification looks promising. The system is capable of giving an objective measure of species-specific call density, which will be related to the number of migrating individuals that passed by for the set of species involved. However, the automatic detection and identification system detected only about 27% of the number of flight calls of low altitude migration detected by a human field observer. Here a limitation is set by the capacity of the microphone, likely caused by the narrowness of the sampled air volume in comparison with the wider reach of human ears.

Aim is to further optimize the system, including optimizing the microphone and adding more Northwestern European species to the identification library. The Maximum Normalized Narrowband Amplitude method makes the system especially suitable to employ in noisy surroundings like offshore recording sites and cities. Currently new recordings are being made simultaneously with radar recording at a North-Sea windmill farm to evaluate the systems performance in an off-shore situation (Fig. 3).

Figure 3: The new recording site with windmills in the background in the North Sea along the Dutch coast (photo Bureau Waardenburg 2007).

References

BROWN J. C & P. J. O. MILLER (2007): Automatic Classification of Killer whale vocalizations using dynamic time warping. - JASA **122**: 1201-1207.

EVANS, W.R. & K.V. ROSENBERG (1999): Acoustic monitoring of night-migrating birds:a progress report. - In: R.E. Bonney. Jr., D.N. Pashley and R. Cooper (eds.). Strategies for bird conservation: the Partners in Flight planning process. Cornell Lab of Ornithology, Ithaca, NY.

KRIJGSVELD, K.L., LIESHOUT, S. M. J. VAN , SCHEKKERMAN, H. , LENSINK, R., POOT, M. J. M. & S. DIRKSEN (2003): Base line studies North Sea wind farms: strategy of approach for flying birds. Culemborg : Bureau Waardenburg bv.

Birds and Bats: Automatic Recording of Flight Calls and their Value for the Study of Migration

REINHOLD HILL & OMMO HÜPPOP,

Institute of Avian Research „Vogelwarte Helgoland“, Island Station, P.O. Box 1220, D-27494 Helgoland, Germany
E-mail: reinhold.hill@avitec-research.de, ommo.hueppop@ifv.terramare.de

Abstract. Within the framework of a study on potential impacts of offshore wind-farms on migrating birds and bats we developed and applied, besides radar and thermal imaging, systems for the automatic registration of bird and bat calls at a research platform in the south-eastern North Sea some 45 km off the coast. Because of the 'noisy environment' special software had to be developed for the automatic detection and registration of calls. For a few species an automatic identification by software was also realized. The automatic registration throughout the years 2004 to 2007 revealed clear time patterns of migration for many species, both seasonal and daily. Calling intensity can be related to weather parameters, namely wind direction and speed. The results also allow conclusions on the collision risk of single species.

Many bird species, though by far not all, utter calls while they fly. In the old world namely geese, some duck species, many waders, thrushes and several other passerines call frequently and can be identified by their flight calls (DIERSCHKE 1989, FARNSWORTH 2005). Bats produce species specific ultrasonic calls for echolocation and communication. Within the framework of a study on potential impacts of offshore wind-farms on migrating birds and bats we developed and applied, besides radar and thermal imaging, systems for the automatic registration of bird and bat calls at the research platform FINO 1 (54° 01‘ N, 6° 35‘ E) in the south-eastern North Sea some 45 km north the Eastfrisian island Borkum. Although it is sometimes difficult to distinguish between real changes in migration intensity and mere changes in calling activity, e.g. due to bad visibility or light attraction, the registration of bird and bat calls can provide reliable data on migration intensity on a species level.

Methods

Acoustic detection of bird calls has worked satisfactorily for many decades and has been recently standardized (FARNSWORTH 2005, KUNZ et al. 2007 a). Recordings at sea are often degraded by strong wind noises, but may be automated with certain restrictions (see DIERSCHKE 1989). It has to be noted that sole acoustic data collection is not suitable for the quantification of bird migration, as some species of birds utter no calls during migration, while others increase their call activity during poor visibility conditions or when they are attracted to light (ALERSTAM 1990). Because of ‘environmental noises’ like wind, rain and waves, a special software had to be developed for the automatic detection and registration of calls. Bird calls close to the platform were detected and recorded automatically by a directional microphone (Sennheiser ME67) with windshield and a mic muff (Fig. 1). Our self-developed capturing software, AROMA (‘Acoustic Recording of Migrating Aves’ based on the audio-processing-toolkit ‘Snack’ for Tcl/Tk: www.speech.kth.se/snack), automatically recognizes birdcalls by their characteristic narrow sound spectrum and filters out wind and rain noises to a large extent (see also HILL & HÜPPOP 2006). Fig. 2 shows a screenshot of the software. Recordings are stored as WAV-files (16 bit, 22 kHz, mono).

For a few species, e.g. the Redwing (*Turdus iliacus*), an automatic identification by software was also implemented by analysing formants with Praat 4.6, a free software for speech and other acoustic analyses (www.praat.org). A formant is a peak in an acoustic frequency spectrum which results from the resonant frequencies of any acoustic system. First a bandpass filter (6200 - 8500 Hz) was used to improve the detection probability by reducing noise at other frequencies. After that, we searched for formants with a bandwidth of less than

50 Hz with a duration of at least 130 ms and linear decrease of the frequency (r < -0.8). The detection rate of Redwing calls was around 80% compared to the manual analyzation of the sound files and, sometimes calls were detected which had not been recognised by the human ear before.

Figure 1: Bat detector and microphone with windshield and mic muff on FINO 1.

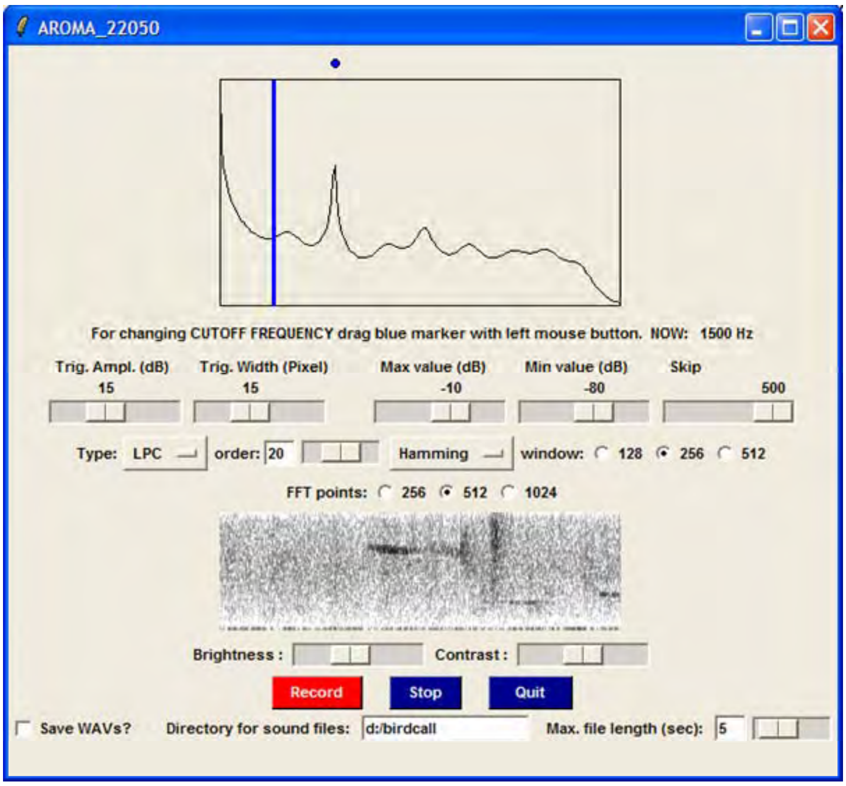


Figure 2: Screenshot of the software AROMA.

In order to automatically register ultrasounds we use a Pettersson D-230 bat detector (Fig. 1) with a heterodyne (frequency mixer, set to 45 kHz) and 1:10 frequency division detector (see SKIBA 2003 for details). The detector is connected to a computer (line-in of the sound-card). One recording channel is used for the heterodyne, the other one for the divider. After threshold detection with the software RecALL 2.4 (www.sagebrush.com) all relevant sounds are recorded as WAV-files (16 bit, 44 kHz). Automatic detection of bat calls is realized by pitch–analysis with the software Praat 4.6 (see above). All sounds with a maximum pitch-frequency > 19 kHz are automatically copied for a later manual species identification at a desktop computer. Fig. 3 illustrates this analysis procedure.

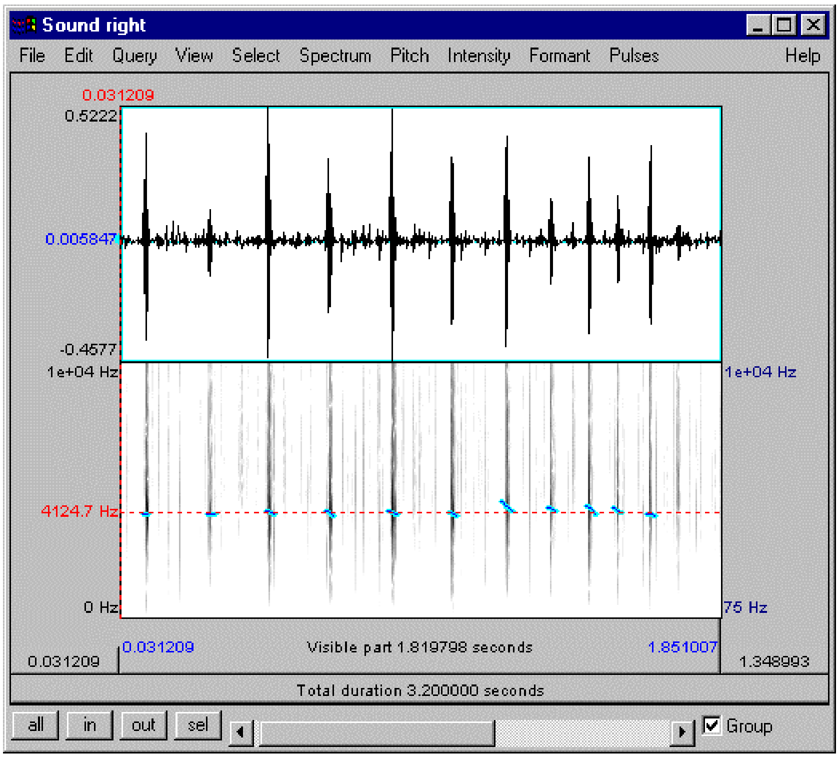


Figure 3: Only the right channel of the bat detector (divide-by-ten) is used for pitch analyses. The screenshot shows oscillogram (top, intensity vs. time) and spectrogram (bottom half) with pitch (blue). The dominating frequency of 4125 Hz x 10 ~ 41 kHz identifies this indiviual as a Nathusius's Pipistrelle - *Pipistrellus nathusii.*

Results

The call intensities of the most frequently recorded species on the FINO 1 show a very characteristic migratory pattern, which largely corresponds to studies of birds captured and ringed on the island of Helgoland (HÜPPOP & HÜPPOP 2004). In addition to the already well-known intensity of migratory movements in spring and autumn, it has become quite clear that migration is concentrated to just a few „migratory waves“ per migration period, and that the greater part of migration takes place at night. Intensive call activities during the day occur for the most part in the summertime (HÜPPOP & HILL 2007). Nights with substantial amounts of bird calls or nights with masses of disoriented birds are at the same time potential collision nights.The weather-conditioned causes that go along with the latter are not yet entirely clarified and thus still hardly predictable. Results of the measurements on FINO 1 are described in HÜPPOP et al. (2005, 2006a and 2006b).

Audio recordings of species-specific migratory calls prove to be especially helpful in migration surveys (DIERSCHKE 1989, FARNSWORTH 2005). The three most common thrush species (Turdidae) in Northern Germany and over the open North Sea often call during migration. Their calls were mainly recorded at night, with a few calls during the day presumably from birds resting on the platform (for the Redwing as an example see figure 4). There appears to be a clear temporal differentiation between the three species in terms of seasonal phenology, which is supported by migration patterns determined by captures on the island of Helgoland, 90 km to the east. This emphasizes the suitability of standardized recording of migratory calls for the study of migration phenology in frequently calling birds (see also discussion in FARNSWORTH 2005). The vernal migration of the Blackbird (*Turdus merula*) was nevertheless only occasionally observed on the FINO 1, in contrast to the trappings on Helgoland. The majority of the migratory calls of thrushes were registered in the second half of the night, above all with an observed intensity just before dawn on some nights. This information corresponds to various observations of songbirds in North America (FARNSWORTH 2005), but contradicts the observations of VLEUGEL (1960) and other authors on European thrushes. The differences may be explained by the situation of the platform located far out at sea.

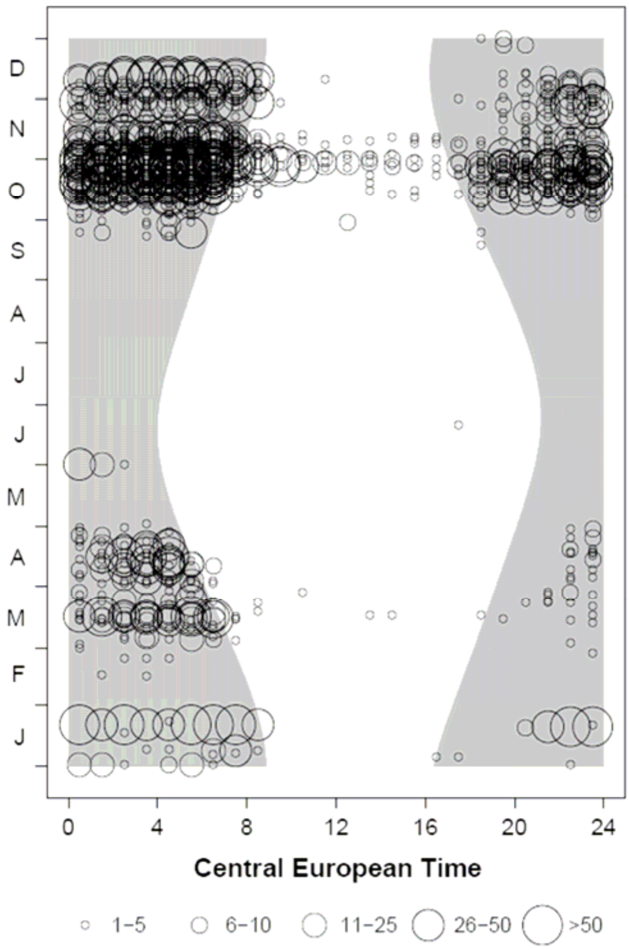


Figure 4: Daily distribution of Redwing calls throughout the year. Dark areas represent the night, light areas the day. The size of circles illustrate the amount of recorded calls (ind./h).

Calls of Common Starlings (*Sturnus vulgaris*) were recorded both in spring and even more in autumn. The Starlings were recorded mainly at night, similar to Common Redshanks (*Tringa totanus*), which were almost exclusively recorded during their autumnal migration at night. In contrast, calls of Sandwich Tern (*Sterna sandvicensis*) and Common Tern (*Sterna hirundo*) were nearly only recorded by day (see figure 5). This was apparent particularly after the breeding period, in which intensive migration was registered in the area around the FINO 1 platform and/or resting birds were observed on the platform itself (terns only rarely swim). In individual cases some breeding birds of the East Frisian Islands may also forage in this area (see GARTHE & FLORE 2007). The frequent appearance of Sandwich Terns on the FINO 1 platform indicates that special attention must be dedicated to the further observation of this species during and after the construction of the pilot wind farm. Recent studies show that terns are apparently at greater risk of collision with wind turbines (EVERAERT & STIENEN 2007), even though they are almost exclusively diurnal. As on the research platform terns presumably also like to rest on structures in wind farms. Germany carries a high international responsibility for the protection and conservation of Sandwich Terns (GARTHE & FLORE 2007).

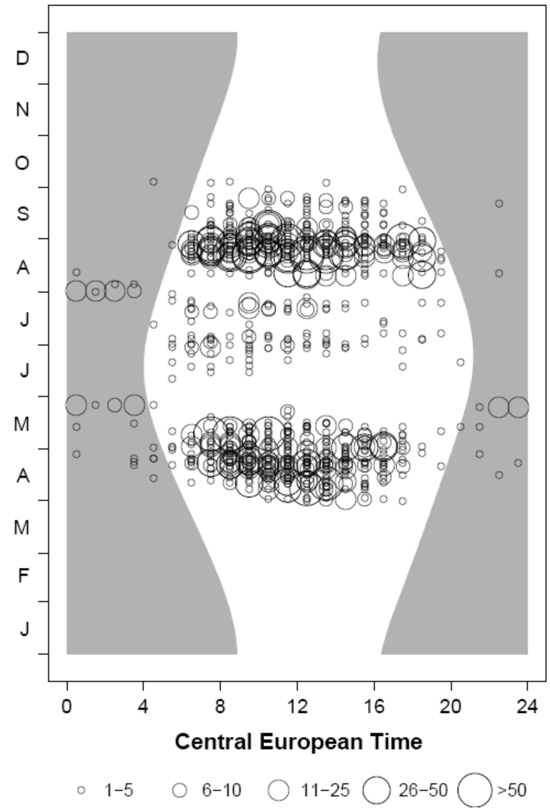


Figure 5: Daily distribution of Sandwich Tern calls throughout the year. Dark areas represent the night, light areas the day. The size of circles illustrate the amount of recorded calls (ind./h).

The acoustic monitoring system for bat calls on FINO 1 registered eight single bats between August 2004 and the end of 2007 (with a break due to technical problems between August 2006 and February 2007): seven Nathusius' Pipistrelles *Pipistrellus nathusii,* a species that is frequently recorded during migration on the German offshore island Helgoland (HÜPPOP in prep.), and one Serotine Bat *Eptesicus serotinus* (rarely recorded on Helgoland). Additionally, on Helgoland Common Pipistrelles *Pipistrellus pipistrellus* (regularly) and Common Noctules *Nyctalus noctula* (rarely) are recorded during migration with the same type of recording system, but there is so far no evidence on FINO 1. The numbers of bats at FINO 1 indicate however regular migration of bats over the open North Sea. It has to be noted here that migrating bats presumably „switch off“ their ultrasonic echolocation system temporarily making them undetectable by our system but susceptible to deadly collisions with stationary objects (VAN GELDER 1956, CRAWFORD & BAKER 1981). From the recorded ultrasounds it can be concluded that at least some individuals obviously hunted insects on FINO 1, which were possibly attracted by the lights for ship and aircraft safety of the platform. This could additionally increase the risk of collision for bats.

Discussion

The automatic registration throughout the years 2004 to 2007 with 73,506 recorded files of bird calls revealed clear time patterns of migration for many species, both seasonal and daily. High calling intensity can often be related to weather parameters, namely wind direction and speed, rain, cloud cover and visibilty. The results also allow conclusions on the collision risk of the single species. A network of acoustic devices at research platforms and later at offshore windfarms in the German Bight could provide detailed information of the temporal and spatial pattern of migrating birds on a species level. This would provide data for modelling and forecasting mass migration events, at least for common species that utter flight calls (see also FARNSWORTH & RUSSEL 2007).

The results of an automatic registration of ultrasound on FINO 1 show that bats migrate offshore and are also at risk to collide with wind turbines, as we know from onshore wind farms (e.g. KUNZ et al. 2007b). Intensities are unknown due to limited range of the ultrasound detectors. Improvements of the hardware and a closer network of monitoring sites would enormously extend the knowledge of migrating bats over the open North Sea.

Acknowledgements

We would like to thank our co-workers and colleagues from two projects for their cooperation and support. Above all we would like to thank the German Federal Ministry for the Environment, Nature Conservation and Nuclear Safety (BMU, FKZ 0327526 and 0329983) for financial support.

References

ALERSTAM, T. (1990): Bird migration. Cambridge University Press.

CRAWFORD, R.L.; BAKER, W.W. (1981): Bats killed at a north Florida television tower: a 25-year record. J. Mammal. **62**:651–652.

DIERSCHKE, V. (1989): Automatisch-akustische Erfassung des nächtlichen Vogelzuges bei Helgoland im Sommer 1987. Vogelwarte **35**: 115-131.

EVERAERT, J. & STIENEN E.W.M. (2007): Impact of wind turbines on birds in Zeebrugge (Belgium) Significant effect on breeding tern colony due to collisions. Biodivers Conserv **16**: 3345-3359.

FARNSWORTH, A. (2005): Flight calls and their value for future ornithological studies and onservation research. Auk **122**: 733-746.

FARNSWORTH, A. & RUSSEL, R.W. (2007): Monitoring flight calls of migrating birds from an oil platform in the northern Gulf of Mexico. J. Field Ornithol. **78**: 279-289.

GARTHE, S. & FLORE, B.-O. (2007): Population trend over 100 years and conservation needs of breeding sandwich terns (Sterna sandvicensis) on the German North Sea coast. J. Ornithol. **148**: 215-227.

HILL, R. & HÜPPOP, O. (2006): Techniken zur Erfassung des „unsichtbaren Vogelzugs“ über See. Jber. Institut Vogelforschung **7**: 21-22.

HÜPPOP, K. & HÜPPOP, O. (2004): Atlas zur Vogelberingung auf Helgoland Teil 2: Phänologie im Fanggarten von 1961 bis 2000. Vogelwarte **42**: 285-343.

HÜPPOP, O., DIERSCHKE, J., EXO, M., FREDRICH, E. & HILL, R. (2005): AP1 Auswirkungen auf den Vogelzug. In: OREJAS C., JOSCHKO, T., SCHRÖDER, A., DIERSCHKE, J., EXO, M., FREDRICH, E., HILL, R., HÜPPOP, O., POLLEHNE, F., ZETTLER, M.L., BOCHERT, R.: Ökologische Begleitforschung zur Windenergienutzung im Offshore-Bereich auf Forschungsplattformen in der Nord- und Ostsee (BeoFINO) - Endbericht Juni 2005, Bremerhaven: 7-160.

HÜPPOP, O., DIERSCHKE, J., EXO, K.-M., FREDRICH, E. & HILL, R. (2006a): Bird migration studies and potential collision risk with offshore wind turbines. Ibis **148**: 90-109.

HÜPPOP, O., DIERSCHKE, J., EXO, K.-M., FREDRICH, E. & HILL, R. (2006b): Chap. 9 Bird Migration and Offshore Wind Turbines. In: KÖLLER, J., KÖPPEL, J. & PETERS, W.: Offshore Wind Energy – Research on Environmental Impacts, Berlin, Heidelberg: 90-116.

HÜPPOP, O. & HILL, R. (2007): Bird migration over the North Sea. In: Morkel, L., Toland, A., Wende, W., Köppel, J. (Eds.): 2nd Scientific Conference on the Use of Offshore Wind Energy by the Federal Ministry for the Environment 20. and 21. February 2007 in Berlin Conference Proceedings: 35-40.

KUNZ, T.H., ARNETT, E.B., COOPER, B.M., ERICKSON, W.P., LARKIN, R.P., MABEE, T., MORRISON, M.L., STRICKLAND, M.D. & SZEWCZAK, J.M. (2007 a): Assessing Impacts of Wind-Energy Development on Nocturnally Active Birds and Bats: A Guidance Document. J. Wildl. Managem. **71**: 2449-2486.

KUNZ, T.H., ARNETT, E.B., ERICKSON, .P., HOAR, A.R., JOHNSON, G.D., LARKIN, R.P., STRICKLAND, M.D., THRESHER, R.W. & TUTTLE, M-D. (2007 b): Ecological impacts of wind energy development on bats: questions, research needs, and hypotheses. Front Ecol Environ **5**: 315–324.

SKIBA, R. (2003): Europäische Fledermäuse. Neue Brehm-Bücherei 648. Westarp Wissenschaften. Hohenwarsleben. 212 S.

VAN GELDER, R.G. (1956): Echo-location failure in migratory bats. Trans. Kansas Acad. Sci. **59**: 220-222.

VLEUGEL, D.A. (1960): On the temporal pattern of nocturnal migration in thrushes. Auk **77**: 10-18.

XBAT: An Open-Source Extensible Platform for Bioacoustic Research and Monitoring

HAROLD FIGUEROA & MATTHEW ROBBINS

Cornell Laboratory of Ornithology, Cornell University, 159 Sapsucker Woods Rd., Ithaca, NY 14850, USA
E-mail of the corresponding author: harold.figueroa@gmail.com

Abstract. The Extensible Bioacoustic Tool (XBAT) is a sound analysis application designed for extensibility along with a framework and toolbox for the development of sound analysis applications focused on the demands of bioacoustic research and monitoring. XBAT was developed in close collaboration with experts in the fields of conservation-monitoring and animal-communication under real-world project pressures to support: human evaluation and annotation of recordings, the development and use of automatic techniques for processing recordings, and a seamless integration of these two modalities. XBAT development makes use of various software engineering strategies of such as extensible software development (WILSON 2004) and domain-driven design (EVANS 2004) to promote the sustainable development of a project with this scope. XBAT has been publicly released as free and open-source software in the hope of promoting more effective collaboration and faster development in the field of algorithmic and software tools for bioacoustic research and monitoring.

The Extensible Bioacoustic Tool (XBAT) was initially developed by the first author at the Bioacoustics Research Program (BRP) at the Cornell Lab of Ornithology (CLO) in 2002 to support algorithm and software developers as well as biologists in a range of conservation-focused bioacoustic-monitoring projects (Fig. 1). It was and continues to be developed in close collaboration with experts in the fields of conservation-monitoring and animal-communication under real-world project pressures to support: human evaluation and annotation of recordings, the development and use of automatic techniques for processing recordings, and a seamless integration of these two modalities. Importantly, through the seamless integration of human evaluation and automated processing XBAT creates an effective and powerful communication channel between signal-processing or pattern-recognition experts and biological or field experts. This communication is critical as technological advancement makes automated processing and sophisticated pattern-recognition more and more possible, necessary for, and essential to bioacoustic-monitoring.

Currently, the software supports data management, human evaluation and annotation, and the automated processing efforts for most bioacoustic-monitoring projects at BRP, as well as a number of animal-communication studies. Further and importantly, it also supports a small group of researchers that continue to develop techniques adapted to the varying analysis demands and data conditions of each of these projects. Many of these new and adapted techniques are implemented as extensions to the XBAT system, and this implementation enables their quick integration into the data-processing workflow at BRP as well as their reuse in other projects.

These various BRP projects involved have included monitoring of both terrestrial and marine environments with diverse analysis goals including but not limited to: signal detection and species identification for presence/absence or abundance censusing, analysis of long-term noise conditions, and the interaction of the above. They have also included the localization of sound sources using array-processing techniques.

Starting in 2005 the second author joined in the development of XBAT, in particular to assist development that would promote the preparation of a public release. Beginning in 2006 preview versions of XBAT have been available through http://xbat.org as free and open-source software licensed under the GNU Public License (GPL). Since making the code publicly available various other efforts have been initiated to support the development of XBAT as a successful open-source community project. A first non-preview release with sufficiently tested and stable programming interfaces is projected for the end of 2008 or the start of 2009.

Figure 1: BRP-developed autonomous recording unit (ARU) in the Big Woods of Arkansas during a large-scale survey seeking supporting evidence for the presence of the Ivory-Billed Woodpecker, a challenging project that used XBAT to support human evaluation and annotation and automated processing of the collected recordings.

The following discussion provides a brief description of the process, goals, and efforts of XBAT development. These are first presented from a software engineering perspective, and then very briefly from a community support and development perspective. The discussion will not include particular strategies, algorithms, and related functionality in the areas of signal detection and measurement developed and currently implemented as part of XBAT; these will be the subject of future publications.

Software Perspective

A basic goal of the XBAT development project is to create an **extensible** application that supports the kinds of **interactive** visualization and **large-scale** automated sound analysis typically required in bioacoustic research and monitoring. Interactive analysis and visualization tools support our familiarization with the data, the process of pattern discovery, and the development of new automated techniques as well as the evaluation of their results. The large-scale automated analysis tools allow us to scale various processes typical in the bioacoustic-monitoring data-processing scenario well beyond unassisted human capacities. Typical data-processing scenarios include, but are not limited to:

- Scanning a long-term field-recording to detect events of interest (events typically means transient acoustic events of biological origin) and

- The measurement and classification of large collections of such events into biological categories, where the categories vary in meaning and degrees of granularity depending on the application.

Finally, an extensible application offers the opportunity to adapt. It provides mechanisms to implement new algorithms, tools, and strategies that support both human and automated processing and evaluation.

Extensibility creates a second basic focus for XBAT development. This focus of XBAT development is to create tools that support algorithm and software development for bioacoustic research and monitoring more generally. A primary goal here is to extend the MATLAB language and environment as a platform for the development of interactive visualization and large-scale automated sound analysis tools required for bioacoustic research and monitoring. MATLAB is an excellent learning and prototyping environment for research in signal-processing, numerical computation, and other fields relevant to bioacoustic tool development. For the above reasons it is already a commonly used as a development environment for bioacoustic tools; XBAT development aims to continue to facilitate these efforts. Some activities related to this aspect of development include:

- Ample use of C language MEX extension interface (THE MATHWORKS 2007) to implement relevant computationally-intensive functions, as well as to access relevant open-source C libraries for efficient improved sound file access, spectrogram computation, and database interfacing.
- The development of the XBAT workflow-oriented programming interfaces to facilitate the implementation of various typical automated processing tasks, such as event detection, measurement, and classification.
- The development of a GUI generation framework to support the agile creation of rich and uniform interfaces for newly developed tools, and
- Along with provision of the basic XBAT programming interfaces, MATLAB command-line and menu-based tools to assist in the development of new extensions using these programming interfaces. These tools also enable easy exploration of existing extensions, so that these serve as living documentation for the interfaces.

XBAT software and development represent an ongoing realization of the previously stated goals. Through current versions of the software users and developers can:

- Easily and efficiently access short or long-term recordings (with single or multiple-channels and in a variety of storage formats); navigate such recordings using familiar spectrogram and waveform amplitude visualizations, as well as listen to the data in various ways.
- Manually browse for acoustic events of interest, as well as automatically search for these events and log their occurrence in persistent database documents accessible through a number of software interfaces. Further, the logged events can be visualized, used as a basis for navigation through the data, and annotated and measured in a growing number of ways.
- Easily extend XBAT in a variety of functionally meaningful ways using MATLAB and the provided programming interfaces, for example to support new sound-feature visualizations or event detection or measurement strategies. Then, easily create integrated and familiar interfaces to the new functionality that allow effective exploration of configuration during research and development and also make it easy to share and work with non-developers.

(Aspects of the above, as implemented in current development versions of the software, are shown in Figures 2, 3 and 4)

The initial construction and current development of XBAT, a project of considerable scale and scope, have relied on the application of a number of software engineering and management strategies (HUNT 2000, BERKUN 2005). Continued sustainable development of XBAT in an open-source environment will rely even more on the effective application of such strategies. Three of these are discussed here: extensible software development (Wilson 2004), domain-driven design (EVANS 2004), and the creation of language-workbenches for domain-specific languages (FOWLER 2005).

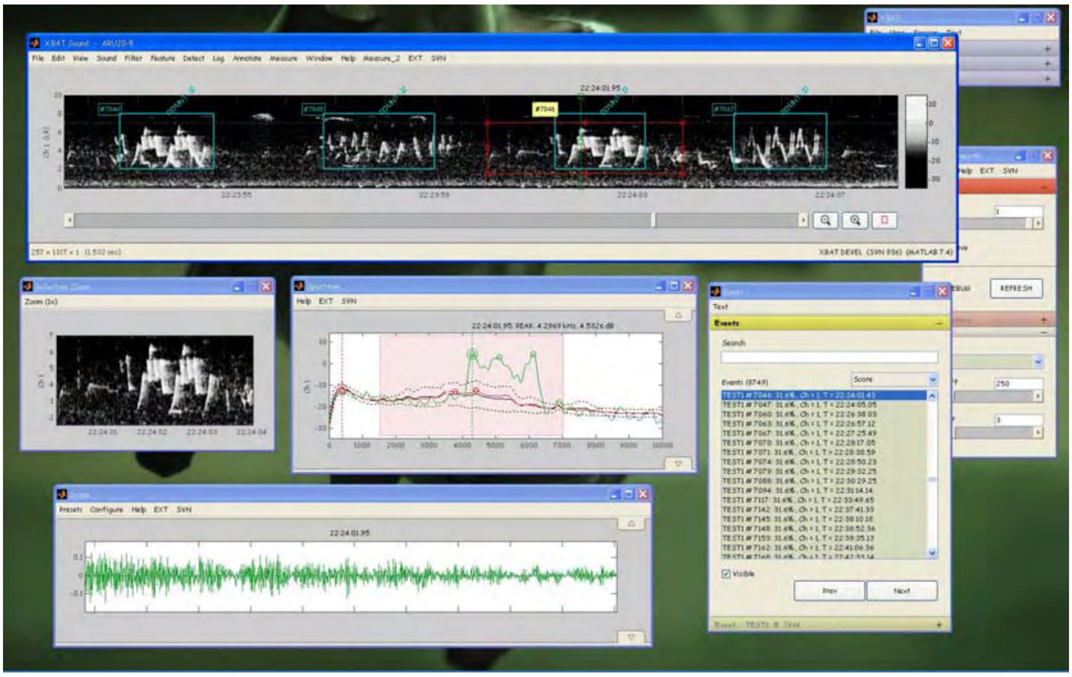


Figure 2: XBAT display demonstrating the automatic detection, visualization, and event-based navigation, as well as various other interactive displays. It shows the results of an automatic detection scan of a 30 hour field recording of a forest using a BRP-developed autonomous recording unit (ARU). The scan resulted in the detection of 9000 songs instances from of various songbirds including the Common Yellowthroat, Cerulean Warbler, and Indigo Bunting among others.

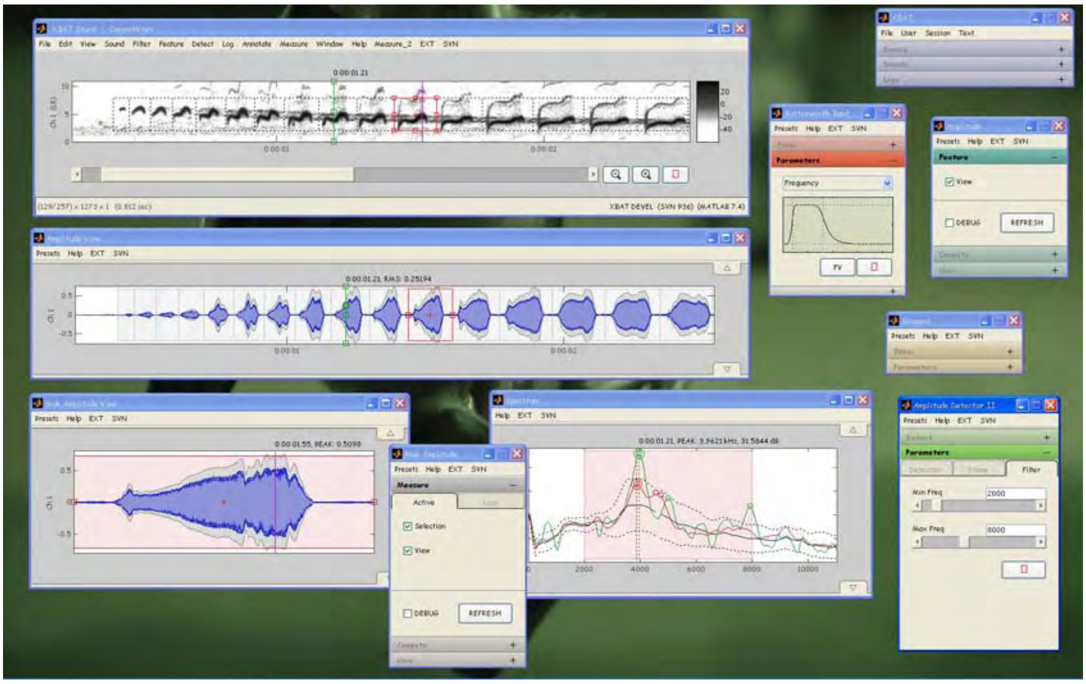


Figure 3: A sample XBAT desktop demonstrating multiple linked interactive and configurable feature views, pre-selection through “active detection”, and “active measurement” of current selection. The various configuration palette interfaces are generated from simple descriptions by the GUI generation framework with no graphical programming effort required.

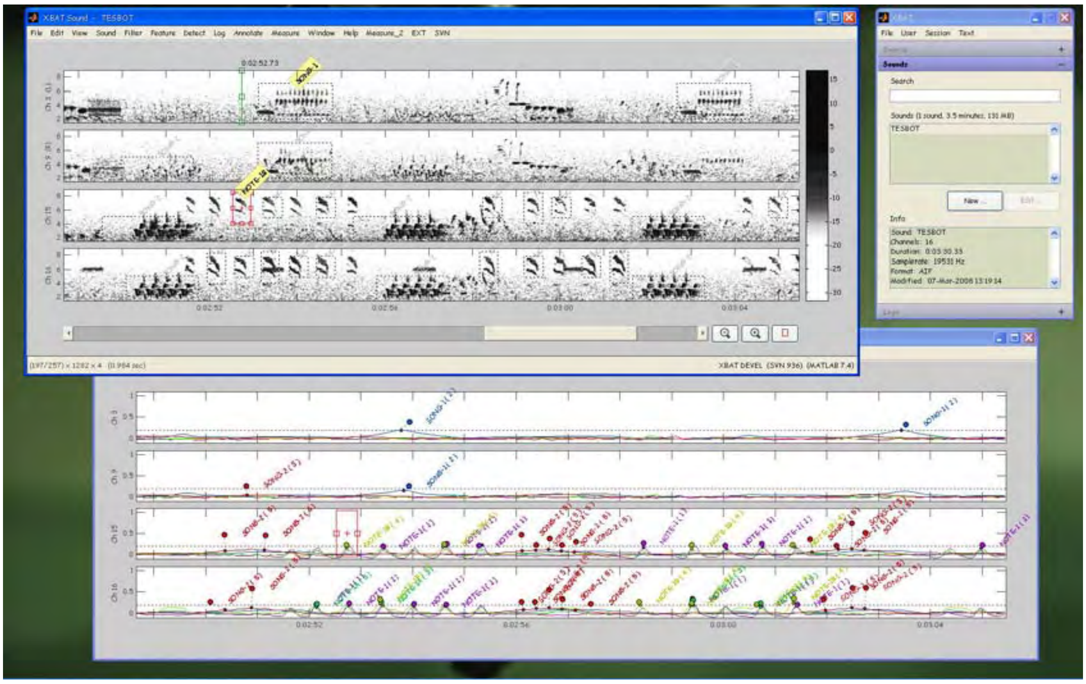


Figure 4: An XBAT "active detection" display for the interactive exploration of an automated detection strategy and configuration. Active detection allows the user to configure the detection strategy (in this case by indicating a few examples along with some tags, and setting a threshold). This capacity is not limited to a particular detection strategy, but part of the programming interfaces to support such behavior for any developed extension. Any detector extension can express its "view" of the data, and reveal as much as it wants about its decision process. We can also see the multiple-channel access and browsing capabilities, the displayed browser shows a subset of a 16-channel recording.

Extensible Software Development

In software development in general, and in the particular case of bioacoustic software, the need for both common function and case-specific adaptation stress the tool development process. Mismanagement of this stress can lead to various wasteful software development scenarios. In some of these for example, we can observe divergent versions of software containing common functionality along with duplicate and often incoherent development of new features that could benefit multiple projects. Bugs and inefficiencies are also addressed in non-uniform ways in such a scenario. Both effects result in software of overall lower quality.

The use of version control systems such as Subversion (http://subversion.tigris.org, COLLINS-SUSSMAN 2005) that support branching and tagging, can ameliorate the effect of such mismanagement for a particular development project, but they are not a solution. In particular, these problems are observed in development efforts at all scales, from a single developer to the development efforts of a community that depends on software for shared needs. Version control systems do not address the community case directly. A possible exception to this last statement is provided by distributed version control systems, for example the decentralized workflows supported by the Bazaar version control system (http://bazaar-vcs.org), however even these do may not constitute a good solution (CLATWORTHY 2007).

Extensible software development resolves the tension created by the need for common function and case-specific adaptation by providing a core of needed common infrastructure along with appropriate programming interfaces to allow for required adaptation (WILSON 2004). Achieving an appropriate decomposition of the problem into a core of common functionality along with appropriate and well-designed programming interfaces for adaptation are the challenges here. The decomposition problem may be called the **modeling** problem; the provision of appropriate programming interfaces is the problem of API development and design, or simply **API design**. These two problems should be considered separately, but they are not independent.

The modeling problem is largely an empirical one, of knowing what the core should contain, of expertly knowing the domain. In the case of XBAT this knowledge comes from working in close collaboration with experts in the fields of conservation-monitoring and animal-communication under real-world project pressures to support: the human evaluation and annotation of recordings, the development and use of automatic techniques for processing recordings, and a seamless integration of these two modalities. These experiences suggest for example, that core needs include, but are not limited to:

- Comprehensive and efficient access to sound data collections of arbitrary size. For example, we should be able to access a recording stored in any of a variety of audio storage formats, even a custom one. Or a long-term recording that may be meaningfully represented as a single file or frequently as a collection of files.
- Efficient computation of various fundamental operations and representations typical in sound analysis such as frequency-dependent filtering and spectrograms.
- Supple visualization and navigation through the data that includes the capacity to explore the data at a range of temporal scales, from very small to very large.
- Access to current and reliable signal-processing and numerical algorithms from a range of relevant related fields.
- The need to create tools simply usable by non-developer biological or field experts by providing familiar and intuitive graphical interfaces.
- Support for enlightening visualization of automated processing strategies to bridge the gap in understanding between signal-processing or pattern-recognition experts and non-experts.

XBAT development addresses these listed common needs and many more such needs in very effective ways and offers these as core functionality. XBAT offerings with respect to the last two listed items are especially effective and interesting. These reflect the very important commitment to the integration of human evaluation and automated processing, as well as to the provision of an effective communication channel between experts from different fields.

Beyond empirically gained knowledge, the decomposition problem can be assisted by more formal approaches. One particularly effective approach in this respect used to support XBAT development is domain-driven design (Evans 2004). Initial XBAT development followed many of the patterns and principles proposed by domain-driven design intuitively. However, learning the more formal perspective allowed more focused development, making better use of the appropriate related practices and principles.

The API design and development problem is considered difficult in general and of great importance to the success of any software development effort (BLOCH 2005, HENNING 2007). It is of particular importance in the context of open-source software development. API design *a priori* constrains the capacity of the developer using the API to express their solution to a particular problem. API constraints can have a positive or negative effect, they can clarify, focus, or muddle a solution strategy, and they can make a solution trivially-easy or nearly-impossible to express and implement. Further, the clarity and expressiveness of the resulting code is also affected. In the context of open-source community development, both of these effects influence the likelihood of participation of would-be developers as well as the effectiveness of those who participate.

XBAT makes use of domain-driven design principles along with dynamic and reflexive qualities of the MATLAB language to provide simple, meaningful, and effective programming interfaces. For example, parts of the interfaces that represent a parametrized computation use essentially a parametrized mathematical function model: Y = F(X; P). Represented in code as,

```
[result, context] = compute(data, parameter, context);
```

Where a further "context" input is often required to develop the full meaning of the remaining inputs, and may also be used to maintain a global storage mechanism between calls to the functions.

At the current state of development these have been used to develop between 100-200 extensions of various types, by a handful of developers without the availability of extensive API documentation. However, work on the programming interfaces is not done. Not all proposed extension-type programming interfaces have been sufficiently exercised and some documentation will be required. This is one of the main reasons dissuading a first non-preview release of XBAT; because API changes after public release become much more problematic.

Domain-Driven Design

A first step and final goal in the application of domain-driven design to a particular domain is to create a **ubiquitous language** that can be used by both domain experts (restricted during this part of the discussion to biological or field experts) and developers to describe and talk about their questions and activities (EVANS 2004). Software development should then strive to offer a faithful representation of this ubiquitous language in the code. The idea is that a stubborn if not strict adherence to this goal will lead to a software system and human language that are essentially equivalent. Through the use of this language both domain experts and developers can effectively communicate currently known problems and solutions in a way that is immediately relevant to development and readily expressible in code. Further, this language will sustainably support communication between domain experts and developers, through time and through the consideration of ever more complex problems. The closing of this communication circuit has a very powerful positive effect on development. However, we should not only consider software development, as this effect is a superficial observation of a deeper scientific implication of this exercise. Closing the communication gap through the ubiquitous language exercise helps formalize our natural language understanding of problems; it heralds a mathematical formalization and understanding of our problems.

XBAT development strongly adheres to these ideas and tries to prominently represent a number of concepts typical in bioacoustics research and monitoring through its data structures, programming interfaces, and graphical user interfaces. The following paragraphs help to define and describe some of the basic language offered by XBAT to support user and developer thinking and communication, and of course software development (some of these are also represented in Figure 5):

> A human becomes a **user**, and may create or subscribe to any number of **libraries**.
> A **library** contains a collection of **sounds**, along with all the associated annotation **logs**.
> A **sound** is a source of audio samples. A sound may have a number of associated auxiliary **attributes** that assist our interpretation and computation with the samples.

Sounds can be annotated by storing **events** of interest in **logs**. Logs contain a collection of **events** and can be edited both by humans, and by automated processes of various types.
Automated processes can edit **logs** in the following ways: **detectors** can add new events to a log, **measurements** can append measurement values (the result of repeatable computations) to the existing **events**, and **classifiers** can update event **tags** or **annotation** information based on the measurement values.
Human **users** can also create new **events** and edit the time and/or frequency boundaries, **tags,** and **annotation** information. Many other operations are possible through the open-ended extension **action** interfaces.
(That both human annotation and automated processes share a common storage format is a key design element that supports the seamless integration of manual and automated annotation modalities.)
An **event** is a reference into a sound, containing time position, channel position, and frequency band of interest. Events can be **annotated**, **tagged**, and **measured**. At the moment the concept of **measurements** is reserved for repeatable, data-derived computations.
An **annotation** is a collection of essentially qualitative observations about an object, their structure is specified through a protocol of arbitrary complexity, and they implement controlled annotation vocabularies.
Tags are a collection of space separated strings that allow us to quickly record essentially qualitative observations, perhaps idiosyncratic, or related to developing categorizations. Tags implement annotation with an open vocabulary.
A **measurement** is the result of computation, based on data, and repeatable given access to the data and any required configuration parameters. The result of a measurement can be of arbitrary complexity.

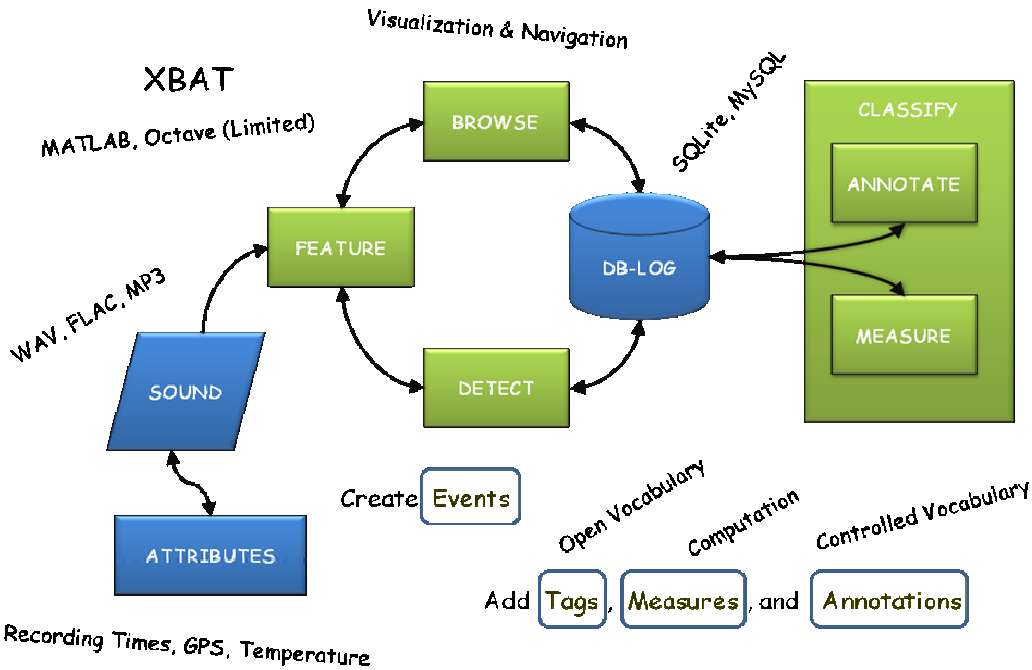


Figure 5: Simplified and annotated diagram of the domain-driven XBAT architecture. Most prominent in the representation are the workflow-oriented programming interfaces and the basic **sound** and **log** object concepts. Also noted are the various complementary annotation categories and some key implementation technologies. All block elements represent extension types, available programming interfaces. Blue block elements represent data, metadata, and annotation stores, and green block elements actions typical in the bioacoustic-monitoring data processing workflow. Again, these block elements and their relations are concepts prominently represented in the XBAT workflow-oriented programming interfaces and fundamental domain data structures.

Provided with such language we can describe our activities or desires in ways that share and retain meaning across domain experts, developers, and code. Here are some examples: (Note that all boldface words in the following paragraphs are effectively represented as data structures and/or programming interfaces, and typically reflected or visualized through the graphical interfaces in XBAT.)

> A recording consists of a **sound** with **attributes** particular to the recording system and which are relevant to various forms of **measurement** on this sound or derived sound **selections**. For example, sounds generated by an array recording system require a sensor geometry attribute to enable the computation of source locations.
>
> Often we apply a **signal-filter** to a sound to emphasize a particular band, reduce-noise, or remove a bias as a pre-processing step for further computation.
>
> We may want to visualize and navigate the raw or filtered sound through a particular **feature** view such as the spectrogram, zero-crossing rate, or output-power from a filter-bank.
>
> In the context of bioacoustic-monitoring we are typically interested in the occurrence of bioacoustic **events** of interest, and a collection of these events can be recorded in a **log** or set of logs for future reference and further consideration. We may create multiple logs for a given sound depending on the focus of our study and our approach to organization; we may be interested in separate logging of different species or perhaps call types.
>
> To make possible the logging of new events of interest in very long-term field recordings or in real-time by autonomous systems, we can automatically **detect** these **events** through the evaluation of indicator sound **features**, and perhaps **classify** of the resulting events through more refined **measurement** of these events.
>
> We may want to quickly and informally **tag** events (to indicate their relation to various developing categories), or carefully **annotate** them according to a pre-specified protocol. Tags and annotations are stored in the logs along with the events.
>
> We may also want to **measure** acoustic properties of these events. The results can be used to **cluster** events and discover latent categories. Or we can use the measurement results to try to explain or harmonize tags or annotation information in an effort to create an automated **classifier**.

All the hypothetical examples in the previous paragraphs either currently exist as part of the XBAT system or are easily implemented with relatively little code.

From a simplified perspective there are two overarching concept categories extracted through the domain-modeling process: “nouns” that typically become data structures, objects, or data stores in XBAT, and “verbs” or perhaps verb categories which are represented as programming interfaces. So it may seem that this domain-driven perspective is not doing much. However, the previous perspective is too simple as many apparent “nouns” have effective dual/representations as both data structures and interfaces, and sometimes even as a collection of interfaces. To illustrate, we note that the **sound** concept appears in the XBAT model as a data structure or object, and at least one programming interface, the **sound-format** interface. In fact, there are a couple more interfaces required to complete the representation of the concept of **sound**, the **sound-file-format** and **sound-attribute** interfaces. It turns out that the construction of the more complex representations can be assisted by the formal perspective. In this case it reflects the implementation of a layered-architecture that isolates infrastructure interfaces, related to aspects of the I/O systems, from higher-level domain interfaces.

This layered-architecture approach is reflected more generally in the fact that the provided XBAT programming interfaces support the following range of adaptation activities:

- High-level workflow-oriented interfaces designed to support operations typical in bioacoustic research and monitoring, these support for example, the development of **detection**, **measurement**, **annotation**, and **classification** strategies for events of interest
- Lower-level programming interfaces to support the integration of I/O technologies and processing primitives, examples of these are the **sound-file-format, log-format**, and **signal-filter** programming interfaces, and finally
- Open-ended extension of the system, through the **action** and **extension-type** programming interfaces. The latter allows for the creation of new extensions types, altogether new programming interfaces.

Together these provide a very effective and meaningful way to apply an extensible software development strategy.

Language-Workbenches for Domain-Specific Languages

The application of an extensible software development strategy guided through domain-driven design, along with a persistent effort to extend the MATLAB language to provide ever simpler and more complete support for the typical operations required during sound analysis as well as the construction of interactive visualization and large-scale analysis results in the development of an internal domain-specific language (DSL). Tools to support development in the domain-specific language complete the idea of a language-workbench (FOWLER 2005). So a way to describe XBAT development is as an effort to create an effective domain-specific language and workbench (FOWLER 2005) for bioacoustic tool development. It provides an adequate and concise answer to the question “What is XBAT?”. It is interesting to observe that MATLAB serves as an excellent base for this kind of development as it is in itself a domain-specific language and workbench for numerical computation, signal-processing, and many other technical fields! It is perhaps, one of the best known and most successful examples of this approach to development in general.

Related terms used to describe this approach to or level of development are “intentional software” (ROSENBERG 2007) and “software factories” (GREENFIELD 2004). In the end they all point to reaching a point in language development and tooling where a developer can “imaginate”, defined as “to instantiate into reality by pure will of imagination” (HUNT 2004).

At the current state of development, this seems to accurately describe some aspects of extension development, namely the creation of new extensions either using command-line or menu-based tools related to the one described and represented in Figure 6. A working skeleton of a new extension for the system can be generated instantly. To the extent that MATLAB, and the XBAT provided functions, data structures, and programming interface abstractions allows us to add to this skeleton the behaviors we intend using concise code largely readable by non-developers, the statement that we can “imaginate” extensions to the system seems true in general. This second step however is not an end but an ongoing process.

Finally we should note that previous discussions are not a gratuitous theoretical rehashing of development. This effort to develop an effective domain-specific language for bioacoustic tool development along with the set of tools to facilitate the use of this language did not stem from a theoretical exercise, but rather developed from very practical considerations and benefited considerably from theoretical reflection.

Quick prototyping from user specifications motivated the domain-model objects previously described. User proposed workflows and algorithm development pressures motivated the development of programming interfaces.

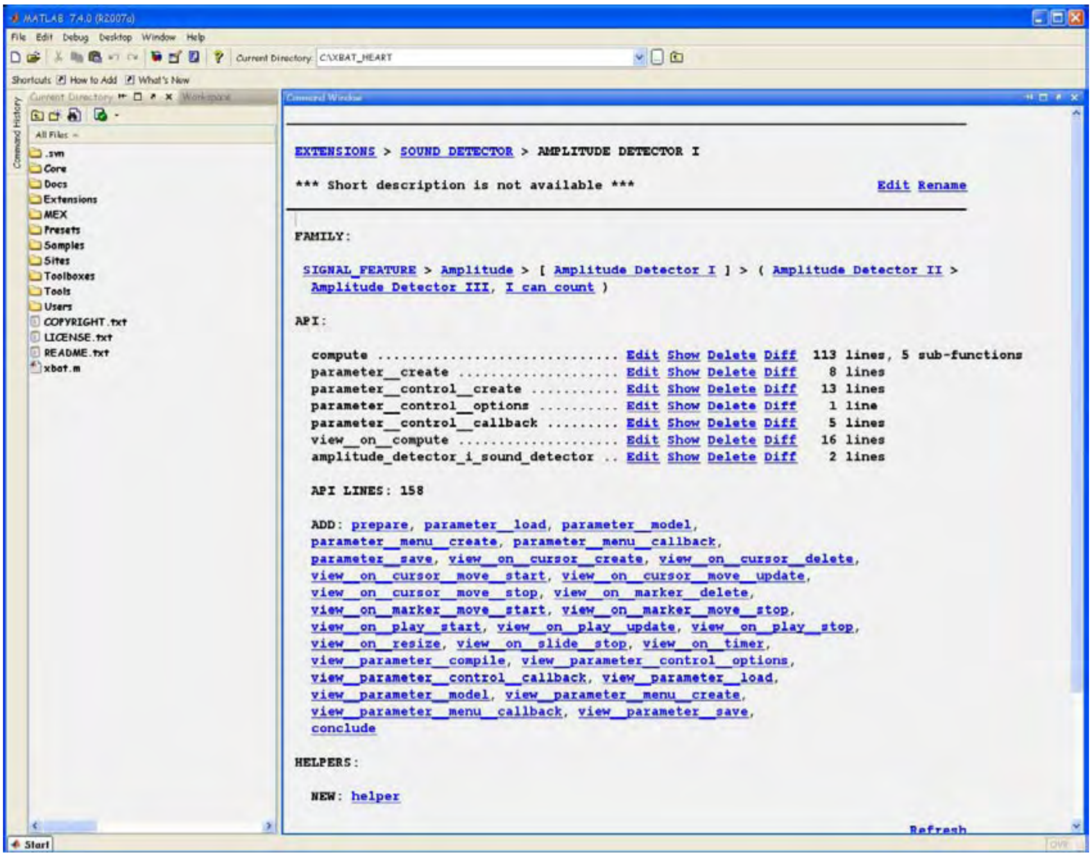


Figure 6: XBAT offers MATLAB command-line tools for the development of new extensions and the exploration of existing ones. The blue text is hyper-linked to perform various expected operations, such as navigating to a given API category and seeing what is available, navigating to a related extension, operating on existing files (including Subversion integration), and adding new implementations of a method. This tool serves as active documentation and support for the various API definitions. Intuitive and repeated API patterns and naming conventions also support the developer. A related menu-based system is also accessible for developers from the same extension interfaces available to users for extension configuration.

The fact that many of these tools were to be operated by biologists during analysis of their data also required that the domain-specific language allow the agile development of graphical user interfaces. These quick code and tool development pressures also led to use of design patterns and metaphors in both arenas of interface design to support simpler code, faster learning, and error reduction in interface use (FREEMAN & FREEMAN 2004, RASKIN 2000, TIDWELL 2006). A large part of this is simply achieved through the reduction of modal interfaces which is apparent in the dominant use of palette-based graphical interfaces in XBAT in contrast to dialog-based graphical interfaces.

The creation of the tools to support the domain-specific language and interfaces also emerged intuitively and practically in response to the pressure to quickly adapt and develop new techniques and software tools for many sound analysis and monitoring projects, while supporting multiple developers and dealing with a nearly complete lack of documentations for the programming interfaces provided.

Community Perspective

This aspect of XBAT development is recent and largely unknown, and therefore this is a very brief perspective. This is also true as some community related considerations also appeared very naturally from a software perspective. This is nevertheless a very exciting and important aspect of XBAT development, and we believe the potential positive effect of developing such a community resource is considerable.

From a community perspective, the XBAT development project aims at the creation of a community resource for the efficient and effective communication of ideas as concise and understandable running code. This free and open-source software resource can offer:

- Support to students and researchers in the learning, evaluation, and reproduction of published research (BUCKHEIT & DONOHO 1995, TREFETHEN 2005),
- Help in providing tool access to non-developers to recently developed techniques, by helping developers create intuitive and familiar interfaces with very little effort, and
- The incremental improvement of tools and techniques by a community of users and developers.

We believe that progress in any single one of these would be worth the development efforts incurred in the development of XBAT, but believe and hope much more is possible.

From a more mechanical perspective in preview versions of XBAT have been available starting in 2006 have been available through http://xbat.org as free and open-source software licensed under the GNU Public License (GPL). A first non-preview release with sufficiently tested and stable programming interfaces is projected for the end of 2008 or the start of 2009.

Various other efforts have been initiated to support the development of XBAT as a successful open-source community project, these include:

- Along with the website and the stable preview release downloads, a version controlled source code repository for XBAT development is available through Google Project Hosting at http://code.google.com/p/xbat-devel/, and
- Various Google-Groups to support announcements, user, and developer discussions.

More information on how to access these various resources is available from the XBAT website at http://xbat.org. The described efforts constitute merely essential elements and continued work is underway to promote the success of the open-source XBAT development project.

Acknowledgements

The XBAT development owes its existence to the initial support, financial and otherwise, of Christopher Clark and Kurt Fristrup. Since then, it has been funded through various grants from sources such as CI, NIH, ONR, and SERDP. Its success within our group is owed to the open communication environment that has existed in the Bioacoustics Research Program (tremendously critical during the early stages of development when documentation is sparse or simply non-existent).

References

BERKUN, S. (2005): The Art of Project Management, O'Reilly

BLOCH, J. (2005): How to Design a Good API and Why it Matters, keynote address at Library Centric Software-Design conference, electronically available through http://lcsd05.cs.tamu.edu/

BUCKHEIT, J. B. & DONOHO, D. L. (1995): Wavelab and Reproducible Research, available at http://www-stat.stanford.edu/~wavelab/Wavelab_850/wavelab.pdf

CLATWORTHY, I. (2007): Distributed Version Control Systems – Why and How, Canonical, available electronically at http://ianclatworthy.files.wordpress.com/2007/10/dvcs-why-and-how3.pdf

EATON, J. W. (2002): GNU Octave Manual, Network Theory Limited, project home at http://www.octave.org

EVANS, E. (2004): Domain-Driven Design, Addison-Wesley, summary available at http://domaindrivendesign.org/books/PatternSummariesUnderCreativeCommons.doc

COLLINS-SUSSMAN, B., FITZPATRICK, B. W. & PILATO C. M. (2005): Version Control with Subversion, O'Reilly, also available in electronically from http://svnbook.red-bean.com

FOGEL, K. (2006): Producing Open Source Software, O'Reilly, also available electronically from http://producingoss.com/

FOWLER, M. (2005): Language Workbenches: The Killer-App for Domain Specific Languages?, available at http://martinfowler.com/articles/languageWorkbench.html

FREEMAN, E. & FREEMAN, E. (2004): Head First Design Patterns, O'Reilly

GREENFIELD, J. (2004): Software Factories: Assembling Applications with Patterns, Models, and Tools, Microsoft Corporation, available electronically at http://msdn2.microsoft.com/en-us/library/ms954811.aspx

HUNT, A. & THOMAS, D. (2000): The Pragmatic Programmer, Addison-Wesley

HUNT, A. & THOMAS, D. (2004): Imaginate, IEEE Software 9/10, 96-97, also available electronically at http://media.pragprog.com/articles/sep_04_imaginate.pdf

THE MATHWORKS (2007): MATLAB 7 External Interfaces, The Mathworks, current version available at http://www.mathworks.com/access/helpdesk/help/pdf_doc/matlab/apiext.pdf

RASKIN, J. (2000): The Humane Interface, Addison-Wesley

ROSENBERG, S. (2007): "Anything You Can Do, I Can Do Meta", Technology Review, also available electronically at http://www.technologyreview.com/Infotech/18047/

TIDWELL, J. (2006): Designing Interfaces, O'Reilly

TREFETHEN, N. (2005): Ten digit algorithms, http://web.comlab.ox.ac.uk/oucl/work/nick.trefethen/tda.html

WILSON, G. V. (2004): Extensible programming for the 21 century, ACM Queue 2, 9, 48-57, electronic reference http://doi.acm.org/10.1145/1039511.1039534

International Expert meeting on IT-based detection of bioacoustical patterns

December 7th until December 10th, 2007

at the International Academy for Nature Conservation (INA), Isle of Vilm, Germany

Agenda

Friday, Dec. 7th

Arrival

Dinner, Welcome party

Saturday, Dec. 8th

8:00	**Breakfast**
9:00	**Opening of the expert meeting**
9:30	**Large-scale, Spatio-temporal Acoustic Mapping: Insights into Acoustic Ecology, Individual Fitness and Population Health in Whales and Elephants** Christopher W. Clark
10:15	**Short Term and Long Term Bioacoustic Monitoring of the Marine Environment. Results from NEMO ONDE Experiment and Way Ahead** Gianni Pavan
11:00	**Coffee Break**
11:15	**An Perennial Acoustic Observatory in the Antarctic Ocean** Lars Kindermann
12:00	**Probabilistic Evaluation of Synergetic Ultrasound Pattern Recognition for Large Scale Bat Surveys** Martin K. Obrist & Ruedi Boesch
12:45	**Lunch**
13:30	**Excursion to the Nature Reserve at the Isle of Vilm**
15:00	**Coffee break**
15:30	**A Case Study: Acoustic Monitoring of Anura in the Iberian Peninsula** Rafael Marquez & Diego Llusia
16:15	**A Decade of Autonomous Monitoring of Tropical Frog Communities** Andrew Taylor, Gordon Grigg & Hamish McCallum
16:45	**Automated Bioacoustic Identification of Insects for Phytosanitary, Quarantine and Ecological Applications** David Chesmore
18:00	**Meeting Dinner**

Sunday, December 9th

8:00	**Breakfast**
9:00	**XBAT: An Open-Source Extensible Platform for Bioacoustic Research and Monitoring** Harold Figueroa
9:45	**From Bird Species to Individual Songs Recognition: Challenges Involved and Comparison of Identification Methods** Vlad Trifa
10:30	**Coffee break**
10:45	**Advantages and Disadvantages of Acoustic Monitoring of Birds – Realistic Scenarios for Automated Bioacoustic Monitoring in a Densely Populated Region** Karl-Heinz Frommolt, Klaus-Henry Tauchert & Martina Koch
11:30	**Bird Song Recognition in Complex Audio Scenes** Rolf Bardeli & Daniel Wolff
12:30	**Lunch**
13:15	**Techniques for Bioacoustic Signal Detection Using Image Processing** T. Scott Brandes
14:00	**Bioacoustic Classifier System Design as a Knowledge Engineering Problem** Sebastian Hübner
14:45	**Computational Methods in Analysis of Bird Song Complexity** Juha T. Tanttu & Jari Turunen
15:30	**Coffee break**
15:45	**Automated Monitoring of Avian Flight Calls During Nocturnal Migration** Thijs Schrama, Martin Poot, Magnus Robb & Hans Slabbekoorn
16:30	**Birds and Bats: Automatic Recording of Flight Calls and their Value for the Study of Migration** Reinhold Hill & Ommo Hüppop
17:15	**General Discussion and Conclusion of the Meeting**
18:30 – 19:30	**Dinner**

Monday, Dec. 10th

Breakfast, Departure

Participants of the International Expert Meeting on IT-based Detection of Bioacoustical Patterns

Bardeli, Rolf: Universität Bonn, Institut für Informatik III, Römerstr. 164, D - 53117 Bonn, Germany

Brandes, T. Scott: Signal Innovations Group, Inc., 1009 Slater Road, NC 27703 Durham, USA

Chesmore, David: University of York, Department of Electronics, YO10 5DD, York, UK

Clark, Christopher W.: Cornell University, Cornell Lab of Ornithology, 159 Sapsucker Woods Road, NY 14850 Ithaca, USA

Clausen, Michael: Universität Bonn, Institut für Informatik III, Römerstr. 164, D - 53117 Bonn, Germany

Figueroa, Harold: Cornell University, Cornell Lab of Ornithology, 159 Sapsucker Woods Road, NY 14850 Ithaca, USA

Freiberg, Horst: Bundesamt für Naturschutz, Konstantinstraße 110, D - 53179 Bonn, Germany

Frommolt, Karl-Heinz: Humboldt-Universität zu Berlin, Museum für Naturkunde, Invalidenstr. 43, D - 10115 Berlin, Germany

Hill, Reinhold: Vogelwarte Helgoland, PF 1220, D - 27494 Helgoland, Germany

Hübner, Sebastian: Pestalozzistr. 5, D - 14482 Potsdam, Germany

Kindermann, Lars: Alfred-Wegener-Institute for Polar and Marine Research, Am Alten Hafen 26, D - 27568, Bremerhaven, Germany

Koch, Martina: Humboldt-Universität zu Berlin, Institut für Biologie, Invalidenstr. 43, D -10115 Berlin, Germany

Koeppel, Hans Werner: Robert- Koch-Str. 38, D - 53115 Bonn, Germany

Márquez, Rafael: Museo Nacional de Ciencias Naturales , Dept. de Biodiversidad y Biologia Evolutiva, Jose Gutierrez Abascal 2, E - 28006 Madrid, Spain

Meffert, Beate: Humboldt-Universität zu Berlin, Institut für Informatik, Rudower Chaussee 25, D - 12489 Berlin, Germany

Obrist, Martin K.: Eidg. Forschungsanstalt WSL, Zürcherstr.111, CH – 08903 Birmendorf, Switzerland

Pavan, Gianni: Universitá degli Studi di Pavia, Centro Interdisciplinare di Bioacustica e Ricerche Ambientali, Via Taramelli 24, I - 27100 Pavia, Italy

Schrama, Thijs: Leiden University, Institute of Biology Leiden, PO Box 9516, NL - 02300 Leiden, The Netherlands

Tanttu, Juha T.: Tampere University of Technology, P.O. Box 300, FIN - 28101 Pori, Finland

Tauchert, Klaus-Henry: Humboldt-Universität zu Berlin, Institut für Biologie, Invalidenstr. 43, D - 10115 Berlin, Germany

Taylor, Andrew: University of New South Wales, Computer Science & Enginering, AUS - 02052 Sydney NSW, Australia

Trifa, Vlad M.: ETH Zentrum, Institute for Pervasive Computing, Clausiusstr. 59, CH - 08092 Zürich, Switzerland

Turunen, Jari: Tampere University of Technology, P.O. Box 300, FIN - 28101 Pori, Finland

Wolff, Daniel: Universität Bonn, Institut für Informatik III, Römerstr. 164, D - 53117 Bonn, Germany

Participants of the international expert meeting on IT-based detection of bioacoustical patterns in front of the conference hall of the International Academy for Nature Conservation on the Isle of Vilm (photo K.-H. Frommolt).