Pattern Recognition Letters 31 (2010) 1524–1534

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

# Detecting bird sounds in a complex acoustic environment and application to bioacoustic monitoring

R. Bardeli [a,*], D. Wolff [b], F. Kurth [c,d], M. Koch [e], K.-H. Tauchert [f], K.-H. Frommolt [f]

[a] Fraunhofer Institute Intelligent Analysis and Information Systems IAIS, Schloss Birlinghoven, 53754 Sankt Augustin, Germany
[b] Department of Musicology/Sound Studies, Institute for Communication Sciences at the University of Bonn, Adenauerallee 4-6, 53113 Bonn, Germany
[c] Department of Computer Science III, University of Bonn, Römerstraße 164, 53117 Bonn, Germany
[d] Fraunhofer-FKIE, 53343 Wachtberg, Germany
[e] Humboldt-Universität zu Berlin, Department of Biology, Invalidenstr. 43, 10115 Berlin, Germany
[f] Museum für Naturkunde, Leibniz Institute for Research on Evolution and Biodiversity at the Humboldt University Berlin, Invalidenstr. 43, 10115 Berlin, Germany

ARTICLE INFO

*Article history:*
Available online 18 September 2009

*Keywords:*
Bioacoustic monitoring
Animal sounds
Algorithmic bioacoustics

ABSTRACT

Trends in bird population sizes are an important indicator in nature conservation but measuring such sizes is a very difficult, labour intensive process. Enormous progress in audio signal processing and pattern recognition in recent years makes it possible to incorporate automated methods into the detection of bird vocalisations. These methods can be employed to support the census of population sizes. We report about a study testing the feasibility of bird monitoring supported by automatic bird song detection. In particular, we describe novel algorithms for the detection of the vocalisations of two endangered bird species and show how these can be used in automatic habitat mapping. These methods are based on detecting temporal patterns in a given frequency band typical for the species. Special effort is put into the suppression of the noise present in real-world audio scenes. Our results show that even in real-world recording conditions high recognition rates with a tolerable rate of false positive detections are possible.
© 2009 Elsevier B.V. All rights reserved.

## 1. Introduction

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. Birds are a good indicator for changes in biodiversity because they are distributed over a wide range of landscapes, are easy to detect in comparison to other animal groups and we have a good knowledge on the biology of most of the species. It is a fortunate fact that, at least in most European countries, we have huge associations of skilled and experienced birdwatchers, who willingly give their knowledge to non-profit service and support monitoring programs. Due to the activity of birdwatchers, data regarding trends in population sizes for certain European bird species has been recorded and made available since 1980 (Gregory et al., 2005). Different standardised methods for bird census have been developed (Bibby et al., 1992). Most of them are based on the mapping of singing males, assuming that the number of territorial males is equal to the number of breeding pairs. The most widely used method for estimating the number of breeding birds is based on point counts where all individuals heard and seen from stationary places are estimated (Klvaňová and Voříšek, 2007).

Complementing traditional approaches, vocalisations of birds serving for territory maintenance and mate attraction can be used for an automated acoustic monitoring of bird populations (Brandes, 2008; Frommolt et al., 2008). The main advantage of such an automated bioacoustic approach, as compared to previous methods, lies in the long-term recording in the absence of an observer. It allows to estimate bird numbers in ecologically sensitive areas (like nature reserves) or in areas that are difficult to access (for example large reed habitats). Even nocturnal birds and birds with low vocal activity could thus be effectively counted. In addition to the need of an applicable autonomous recording device, the greatest challenge is the development of appropriate pattern recognition algorithms giving reliable results even in complex acoustic environments. In order to apply acoustic methods for the monitoring of bird species, we have to solve two problems. We need pattern recognition algorithms for the automatic detection and identification of bird species and we need appropriate techniques for the estimation of the number of individuals.

Acoustic pattern recognition algorithms have been successfully applied to the study of nocturnal migrations of birds (Evan and Mellinger, 1999; Farnsworth et al., 2004; Farnsworth, 2005; Farnsworth

* Corresponding author. Fax: +49 22411443107.
*E-mail addresses:* rolf.bardeli@gmail.com (R. Bardeli), wolffd.mail@gmail.com (D. Wolff), frank@iai.uni-bonn.de (F. Kurth), koch_martina@gmx.de (M. Koch), klaus.tauchert@mfn-berlin.de (K.-H. Tauchert), karl-heinz.frommolt@mfn-berlin.de (K.-H. Frommolt).

0167-8655/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2009.09.014

and Russel, 2007; Hüppop et al., 2006; Mills, 2000; Hill and Hüppop, 2008; Schrama et al., 2008; Marcarini et al., 2008). However, these works mostly deal with relatively simple signals without background noise. The number of detected calls was used as a criterion to assess the occurrence of migrating birds. For noisy environments, hidden Markov models were successfully applied to frequency modulated sounds (Brandes, 2008; Trifa et al., 2008).

In this work we report about an approach using pattern recognition techniques for continuous bird monitoring. Our techniques are based on event detection and repetition rate estimation of bird song elements. Additionally, the algorithms use noise estimation from frequency bands known not to contain the bird's vocalisations and apply them to an effective noise reduction. They can be used in parallel with traditional monitoring methodology to yield methods with improved speed and reproducibility in those cases where reliable detectors for bird vocalisations are available. In contrast to other applications our algorithms have been designed specifically for certain endagered target species. This approach was chosen in order to improve recognition rates even under poor acoustic conditions and it leads to a larger range covered by a single sensor unit in comparison to human observers and algorithms requiring high signal-to-noise ratios. Moreover, we apply this methodology to a quantitative survey of a real bird population yielding automatic map generation of breeding territories for one of the considered bird species.

In Section 2 we summarise related work in the field of pattern recognition for animal sounds. Approaches for the detection of the vocalisations of two endangered bird species are developed in Section 3. These algorithms have been evaluated in a study described in Section 4, which also gives evaluation results. These results are sufficient for automatic map generation of breeding territories for one of the bird species investigated. Results of a corresponding study are presented in Section 5. Finally, conclusions are given in Section 6.

## 2. Pattern recognition for animal sounds

In comparison to other fields in pattern recognition, little work has been carried out regarding animal sound recognition. Nevertheless, a wide variety of methods and animal species have been examined. In previous studies, bird song recognition with hidden Markov models has been proven to be a useful tool in the recognition of bird song elements (Kogan et al., 1998). In this case, recordings were made under laboratory conditions with captive birds and microphones close to the cages.

The most obvious candidates for species recognition by sound are birds. But a lot of other species have also been subject to efforts in automated recognition, for example, crickets and grasshoppers (Schwenker et al., 2003; Farr and Chesmore, 2007), marine mammals like whales and dolphins (Deecke et al., 1999; Brown and Miller, 2007; Mellinger et al., 2007), frogs and bats ([Obrist et al., 2004). Methods like these are of interest in applications such as monitoring for the presence of certain species in an area, behavioural studies, assessing the impact of anthropogenic noise on animal vocalisations and many others.

Currently, there are no guidelines for the direct application of standard methods from machine learning to pattern recognition problems in time dependent data. The main problem here is how to decide which parts of a signal are to be used as input for such methods. Often, this problem is dealt with by applying segmentation algorithms. Unfortunately, this is equivalent to finding the starting and ending positions of animal vocalisations or their segments which is extremely difficult. This is particularly true for natural audio scenes where current solutions tend to be unreliable because of low signal-to-noise ratios. If, however, this gap is bridged in some way or the other, numerous classification algorithms are available for application. The classical dynamic time warping algorithm Deller et al., 1993 has found application in cases where animal vocalisations are not too variable and can thus be recognised by template matching (Brown and Miller, 2007). Neural networks present an obvious candidate for pattern classification. In addition to bird calls (Mills, 2000) and bird song elements (Nickerson et al., 2006), such networks have been applied to the recognition of animal species like marine mammals (Deecke et al., 1999) and crickets (Schwenker et al., 2003). Self-organising maps have also been used for various classification tasks concerning animal sounds (Mitsakakis et al., 1996; Somervuo and Härmä, 2003; Placer et al., 2006). Also, decision tree classifiers like C4.5 (Quinlan, 1993) have been used for bird song recognition (Taylor, 1995). When regarding spectrograms of animal vocalisations, the idea of applying image analysis methods springs to the mind, and several such approaches have been reported (Obrist et al., 2004; Brandes et al., 2006).

Several feature representations for animal sounds have been proposed. Because of the tonal quality of many bird songs, sinusoidal modelling is a promising feature extraction step for bird song recognition and has been studied extensively in this context (Härmä et al., 2003). Often, however, animal vocalisations are not very well represented using the Fourier transform. This is the case when vocalisations are noise-like in the sense that their energy spreads over a broad frequency range. Wavelets have been proposed as an alternative for analysing such sounds. They have been shown to concentrate energy in comparatively few wavelet coefficients for sounds which otherwise need many Fourier coefficients for their representation (Selin et al., 2007). In (Possart, 2002; Seekings and Potter, 2003), wavelets have been applied as feature extractors for the classification of bird and whale songs. In the latter case improvements over spectrogram matching techniques have been achieved. Here, wavelets were also introduced as a means to tackle the segmentation problem described above. In (Fagerlund and Härmä, 2005), other representations than wavelets have been proposed. For example, MFCCs have been found to give good representations in such cases. Moreover, some bird songs are rich in harmonic structure, a fact that can be used for their recognition (Härmä and Somervuo, 2004). More generally, several sets of features for the representation of bird songs have been compared in (Somervuo et al., 2006). Still, the recognition of bird calls in natural environments remains a great challenge (Tanttu et al., 2006). This is caused by the main source of complexity in natural audio scenes: the presence of multiple sound sources overlapping in time and frequency.

## 3. Bird song detectors

In this section, we describe algorithms designed for the purpose of detecting the presence of species specific vocalisations. The two target species, the Eurasian bittern (*Botaurus stellaris*) and the Savi's warbler (*Locustella luscinioides*), were chosen for their value for nature conservation. They are indicator species for extended reed beds. The two types of special purpose algorithms described in this section are tailored for different types of signals. The first algorithm to be described will be useful for detecting very simple spectral events in the presence of broadband noise. The second algorithm deals with signals characterised by the periodic repetition of simple elements which is often encountered in animal vocalisations.

### 3.1. Simple events: the Eurasian bittern

The most obvious indication of the presence of the 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 centre frequency of about 150 Hz. 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 a suitable frequency band. Fig. 1 shows a spectrogram of the bittern call. Each call begins with a short segment at a slightly higher frequency. This part, however, cannot be used for pattern recognition because it can no longer be detected reliably in the presence of noise or at larger distances from the animal.

In order to achieve sufficient frequency resolution in low frequency bands, input signals are downsampled to 6 kHz prior to further analysis. The signal is analysed with a sliding window of 21 ms length. Downsampling increases the effective window length to 170 ms.

Let $S(\omega,t)$ be the windowed power spectrum of a downsampled input signal $s$. The energy weighted novelty $N_{\ell,h}$ for a frequency range from $\ell$ to $h$ at time $t$ is defined by

$$N_{\ell,h}[S](t) := \sum_{\omega=\ell}^{h} S(\omega,t)(S(\omega,t) - S(\omega,t+1))^2. \quad (1)$$

A squared difference is chosen for this novelty measure in order to emphasise large changes in energy over small ones. Energy weighting is incorporated to make peak picking from this feature more robust in cases where silent noise produces sharp peaks in the novelty measure.

As indicated above, the calls to be detected are indicated by peaks in the novelty curve $N_{\ell,h}[S]$ for suitable values of $\ell$ and $h$. The main problem with this simple method, 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. Fig. 2 shows how broadband noise can be removed from the features by subtracting a low-pass filtered noise estimate.

From this, we derive a criterion $B[S](t)$ for the presence of bittern calls at time $t$ as follows:

$$B[S](t) := N_{\ell_b,h_b}[S](t) - \alpha(\phi * N_{\ell_n,h_n}[S])(t). \quad (2)$$

Here, $\phi$ is a given low-pass filter to be convolved with a noise estimate. This gives a smooth estimate of noise energy. For estimating the presence of the bittern call, bins $\ell_b$ to $h_b$ of the spectrogram are examined. Similarly, noise is estimated from a neighbouring frequency band given by the bins from $\ell_n$ to $h_n$. A fixed factor $\alpha$ controls how much influence the noise measure has in the combined criterion. Values for $\alpha, \ell_b, h_b, \ell_n$, and $h_n$ are found experimentally. In particular, the values for $\ell_b$ and $h_b$ are derived from a training set by spectrogram inspection. Suitable values for $\alpha$ are in a range from 5 to 10.

Using the feature $B[S]$ directly for finding bittern calls still leads to a high number of false positive detections due to noise. We can, however, use the fact that the bittern usually calls in sequences with almost constant length pauses between calls. Fig. 3a shows the features $B[S]$ for a 97-min recording from our study site (see Section 4.1). The recording is characterised by a high amount of noise caused by trains, wind, and water. Direct interpretation of the features would lead to a large number of false positive detections of the bittern call. An autocorrelation analysis of $B[S]$, however, allows to lower the number of false positives significantly. For this purpose, we calculate the windowed autocorrelation $A(\tau,t)$ of $B[S]$ (Fig. 3b), where the time $t$ gives the centre of the window and $\tau$ describes the autocorrelation lag. From this, we derive a feature sequence

$$\widetilde{A}(t) = \frac{1}{h}\left(\sum_{\tau=a}^{a+h-1} A(\tau,t)\right) - \frac{1}{k}\left(\sum_{\tau=b}^{b+k-1} A(\tau,t)\right). \quad (3)$$

This feature sequence measures the strength of the autocorrelation at lags $a \ldots a+h-1$ representing typical call repetition rates. We subtract the same measure for lags $b \ldots b+k-1$ indicating shorter repetition rates in order to remove the impact of noise events showing short repetition rates. Finally, candidate positions for the bittern call can be found from $\widetilde{A}$ by peak picking combined with thresholding.

### 3.2. Element repetition: Savi's warbler

The Savi's warbler has a very characteristic song formed by the continuous repetition of simple song elements at an almost constant rate of roughly 50 repetitions per second. Most of the warbler's song's energy is found in the frequency band between 3.8 and 8 kHz. When recording this song from a great distance, the lowest frequency part (around 4 kHz) is usually recognised best. An example of this song is given in Fig. 4.

The recognition of Savi's warbler relies on detecting the presence of repeated elements with the typical repetition rate of 50 Hz. Detecting this kind of repetition in the frequency band typical for the warbler, is considered an indicator for the bird's song.

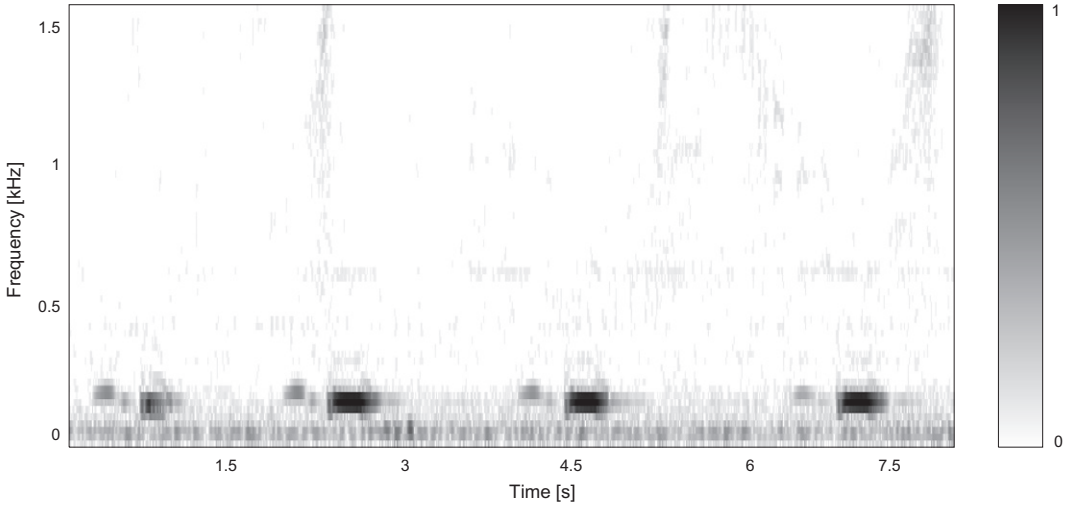


**Fig. 1.** Spectrogram of a call sequence of the Eurasian bittern. The call is characterised mainly by high energy in a narrow frequency band at about 150 Hz. (All subsequent figures encode intensity by the same normalised gray scale given in this figure.)

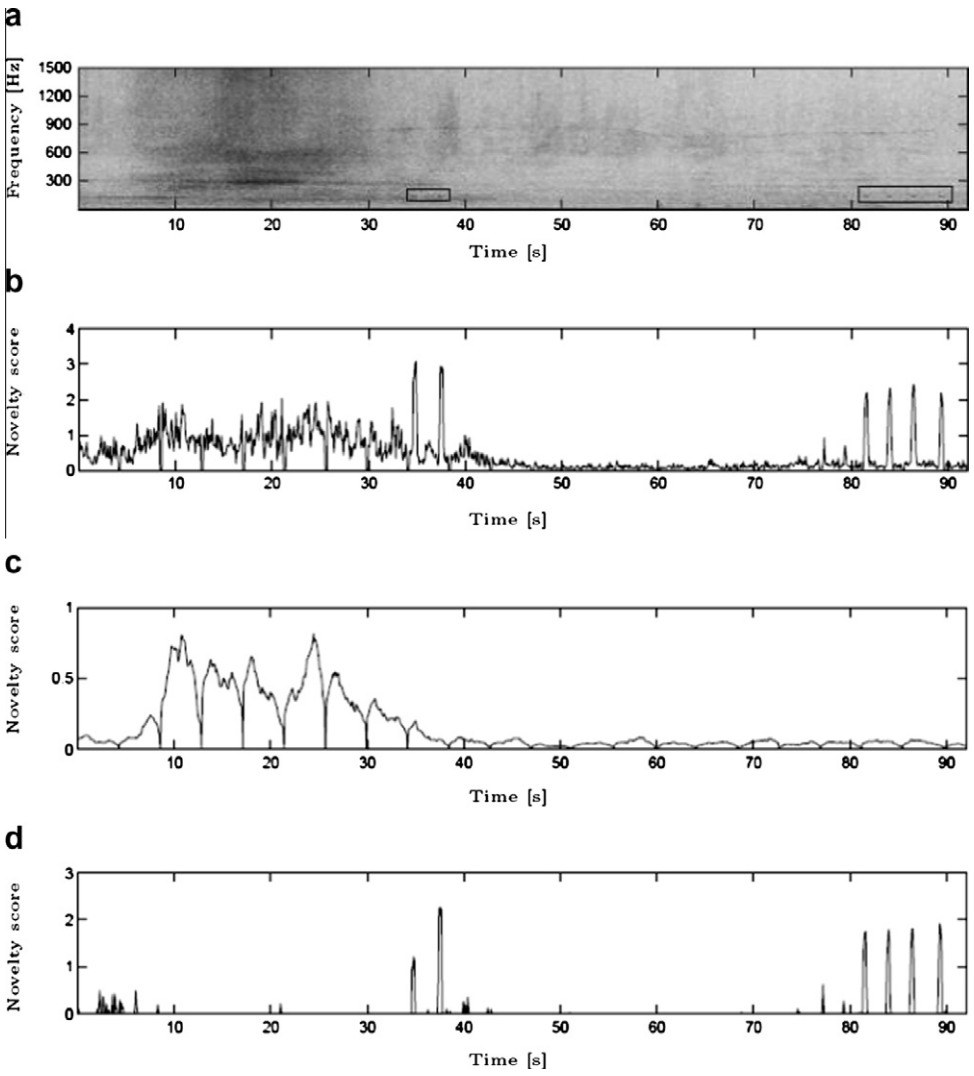


**Fig. 2.** Spectrogram $S$ of a recording containing the call of the Eurasian bittern (marked by boxes) as well as broadband noise (a). The impact of this noise on the novelty score $N_{\ell_b,h_b}[S]$ (b) is removed by subtracting the weighted low-pass filtered novelty $N_{\ell_n,h_n}[S]$ (c) estimated from a neighbouring frequency band, resulting in a combined feature $B[S]$ (d).

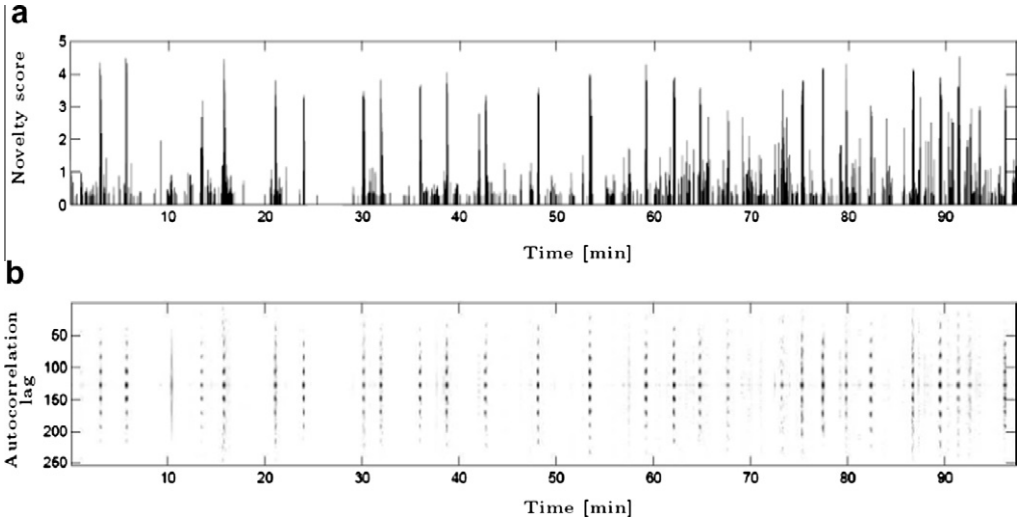


**Fig. 3.** Features $B[S]$ indicating activity in the frequency band characteristic for the Eurasian bittern after noise removal (a) and their windowed autocorrelation $A(\tau, t)$ (b) indicating call repetitions.

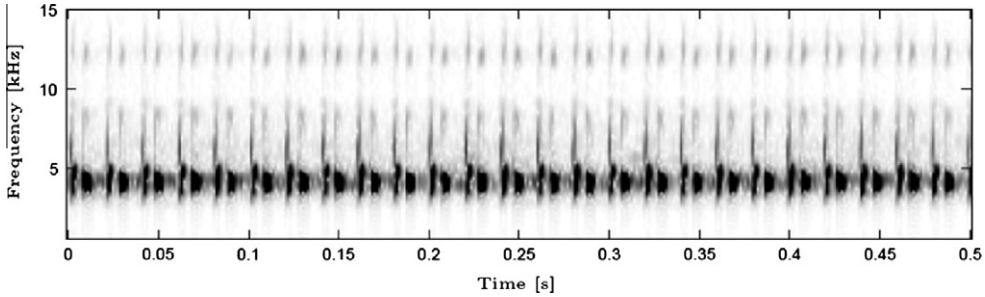


**Fig. 4.** Spectrogram of the Savi's warbler's song.

The first step in estimating repetition rates is to compute a novelty curve $\widetilde{N}_{\ell,h}$ for the frequency band from $\ell$ to $h$ similar to the one defined in the previous section. It is used to detect the onsets of song elements and is therefore designed to give peaks only when energy is rising in a given band. Here, parts of the frequency band typical for the warbler are chosen. The resulting version of the novelty curve is defined as:

$$\widetilde{N}_{\ell,h}[S](t) := \sum_{\omega=\ell}^{h} \max\{S(\omega, t+1) - S(\omega, t), 0\}. \tag{4}$$

Here, a different type of novelty feature than in the previous section is used because of the different signal characteristics. Now, repetition rates of song elements can be read off the autocorrelation function of the novelty curve. If the novelty curve has a period $\tau$, its autocorrelation $A_{\widetilde{N}}$ will show peaks at lags $n\tau$ for $n \in \mathbb{N}_0$.

In order to discard false detections generated by very noisy parts of the novelty curve, we assess the sharpness $\sigma[A_{\widetilde{N}}]$ of the autocorrelation function $A_{\widetilde{N}}$ which is measured using an inverted spectral flatness measure:

$$\sigma[A_{\widetilde{N}}](t) := 1 - \frac{\prod_{\tau=1}^{L}\left(A_{\widetilde{N}}(\tau,t)\right)^{\frac{1}{L}}}{\frac{1}{L}\sum_{\tau=1}^{L} A_{\widetilde{N}}(\tau,t)}. \tag{5}$$

Here, $A_{\widetilde{N}}(\tau,t)$ gives the windowed autocorrelation of the novelty curve $\widetilde{N}$ at window $t$ and time-lag $\tau$. For each $t$ the autocorrelation curve $A_{\widetilde{N}}(\cdot,t)$ is assumed to be evaluated for $L$ discrete time-lags. Sharp autocorrelation curves correspond to a dominant periodic signal being included in the associated part of the signal. Frames falling below a sharpness threshold of 0.2 are discarded from further processing at this stage.

How well song element onsets are reflected by the novelty feature is dependent on the actual frequency band that is used. Fig. 5 shows an artificial example illustrating this effect. In some frequency bands, song elements may be well separated whereas in other frequency bands the elements overlap and thus do not lead to clear peaks in the novelty curve. Therefore, the frequency band from 3.8 to 8 kHz known to contain the Savi's warbler's song is subdivided into five subbands. The novelty curve is computed for each subband and whenever the repetition rate we are looking for is detected in one of the subbands, this subband will be selected for feature computation.

The presence of a period fitting to the 50 Hz repetition rate typical of the Savi's warbler's song is most easily detected via the absolute value of the Fourier transform or power spectrum of the autocorrelation function. A strong peak in the frequency bin corresponding to the expected time lag indicates the correct period. Similar to the strategy followed in the detection of the Eurasian bittern, noise reduction of the Fourier transform features described above can be conducted by subtracting the same type of Fourier transformed features extracted from a flanking frequency band. Thereby, false detections due to periodic broadband signals like industrial noise are avoided.

By using a simple peak picking algorithm, it is assured that a peak at the Savi's warbler's typical element repetition frequency has significant energy. Given that case, the corresponding frame is marked as *candidate frame* along with a quality factor of the respective peak. Let $\widehat{A_{\widetilde{N}}}$ represent the power spectrum of a given autocorrelation curve. Furthermore, the frequency indices $p$, $q$ correspond to 44 and 58 Hz. The dominance of the warbler's periodic signal is defined by

$$d[\widehat{A_{\widetilde{N}}}] := 1 - \frac{\sum_{w\in\mathbb{N}_0}\widehat{A_{\widetilde{N}}}(w) - \sum_{w=p}^{q}\widehat{A_{\widetilde{N}}}(w)}{\sum_{w\in\mathbb{N}_0}\widehat{A_{\widetilde{N}}}(w)}. \tag{6}$$

An overview of this feature extraction process is given in Fig. 6.

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 in the analysed band for a long enough time, thereby accumulating a sufficient amount of domination.

When a Savi's warbler's song is detected, the direction of arrival for the song of this individual is estimated by comparing the domination values of all channels. As the Savi's warbler's song is usually perceived best by the directed microphone pointing towards the bird's songpost, the dominance value improves for channels being recorded with this microphone. The channel featuring the largest domination values is selected and output. This allows for a rough estimation of the direction from which a detected song is arriving at the array.

## 4. Results for a real-world monitoring scenario

In order to evaluate our pattern recognition algorithms, we applied it to recordings attained at Lake Parstein in the north-east of Germany (federal state Brandenburg) in 2006 and 2007. The northern part of the lake is surrounded by extended reed belts providing breeding habitats for the two target species. Recording was conducted in a project with the aim to examine the applicability of pattern recognition methods as a tool for monitoring bird vocalisations. A four-channel stationary microphone array of cardioid microphones (Beyerdynamic MC 930, Sennheiser ME 64) was used on several positions at the lake shore, from a lookout, and from a boat. The best recordings of vocalisations from the reed zone were achieved when the microphones were placed on a boat on the lake. In the best case, calls of the bittern could be recorded over distances of about 1 km. Acoustic data were acquired by an Edirol R4 field recorder at a sampling rate of 48 kHz and 16 bit accuracy. Recordings were performed at dusk and during night in order to cover audio scenes with a complexity level somewhere between the simple situation of a laboratory recording and the extremely complex situation of bird choirs at daytime.

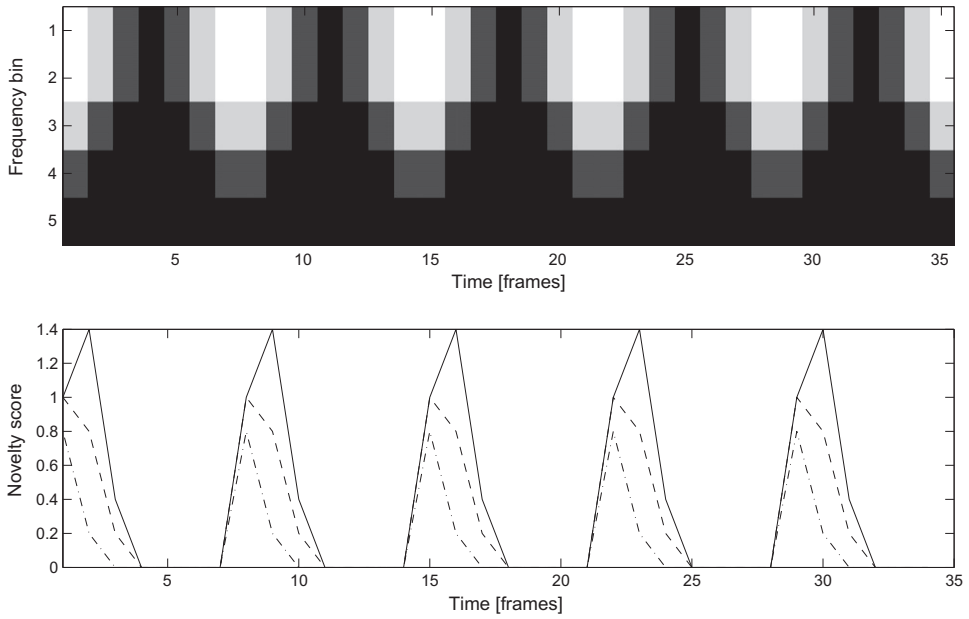


**Fig. 5.** The ability of the novelty features to reflect song element onsets depends on the subband. This is illustrated by an artificial example of a spectrogram (a). The sharpness of the novelty curves (b) depends on the frequency band from which they are extracted. The solid novelty curve is extracted from frequency bins 1–3, the dashed curve from bins 2–4, and the dot-dashed curve from bins 3–5.

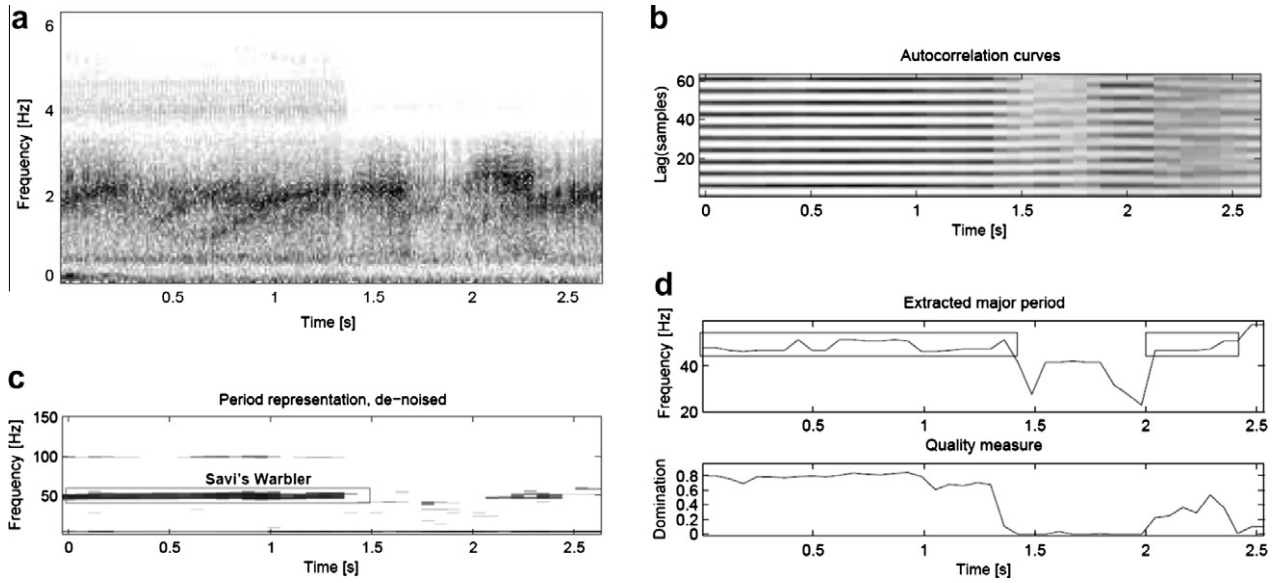


**Fig. 6.** The presence of elements repeated with a given repetition rate is estimated using autocorrelation-based features. (a) Spectrogram of the analysed signal. In the first half of the signal, a Savi's warbler's song is clearly visible. (b) Corresponding autocorrelation curves $A_{\sim_N}$ for successive frames. (c) Fourier transformed autocorrelation curves after subtraction of the flanking band's features and (d) Extracted element repetition periods and corresponding domination values. Frames indicate detected warbler songs.

### 4.1. Acoustic environment at the study site

At the study site, three kinds of noise influence the quality of the recordings and thus the reliability of pattern recognition algorithms. First, there are biogenic sound sources like the calls of amphibians. They differ considerably depending on the time of day. Even by night, the biogenic noise level is unexpectedly high. This is especially due to the calls of amphibians, as can be seen in the spectrogram in Fig. 7a. In contrast, in the morning the audio scene is dominated by the great number of birds that are audible (compare spectrograms of night and morning in Fig. 7).

Second, anthropogenic noise sources have a strong influence on the sound scene even in remote nature reserves. In particular, far reaching traffic noise from cars and trains has a broad-band influence disturbing nearly the whole frequency range of interest. Fig. 8a shows a spectrogram of an audio scene polluted by this kind of traffic noise.

Finally, wind plays a crucial role as its impact on the microphones cannot always be avoided. The effect of wind can be seen best in the results for the detector for the Eurasian bittern, see Section 4.2. A spectrogram showing this effect is shown in Fig. 8b.

### 4.2. Results for the Eurasian bittern

We have evaluated the algorithm of Section 3.1 for the detection of bittern calls using seven hours of recordings from Lake Par-

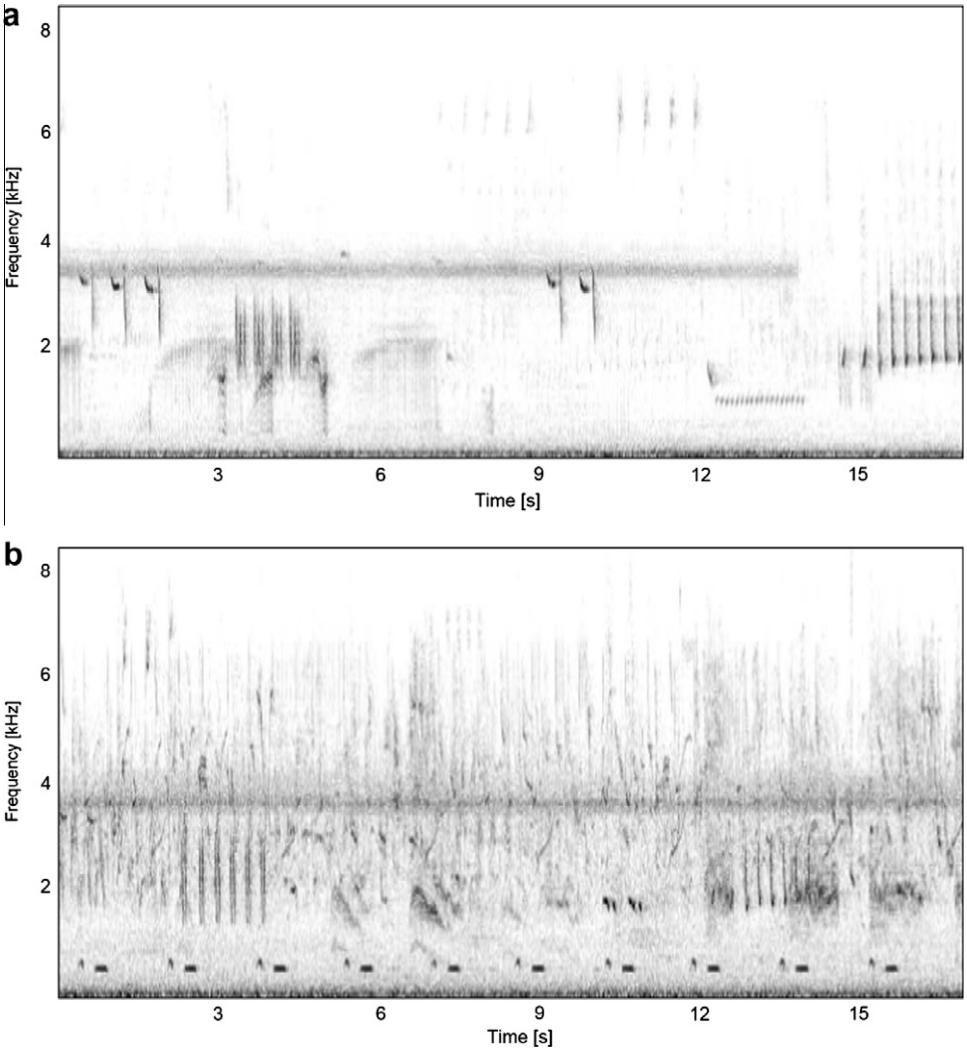


**Fig. 7.** Spectrograms visualising biogenic noise at the study site during the night (a) and in the morning (b). Both spectrograms show the song of Savi's warbler at just below 4 kHz.

stein. The recordings are from five different days, some of them were taken from a boat, others from the lakeshore. In Table 1, we give the number of time positions reported by the algorithm, the number of false positive detections (reported positions where no bittern activity is audible) and the number of false negative detections (positions, where bittern activity is audible, which were not reported by the algorithm).

The number of false positive detections is very much dependent on the acoustical situation. False positive rates were especially high on Day 3 and Day 5 for different reasons. On Day 3, a lot of speech is present in the recording. Especially in the beginning of the recording, a lengthy description of the recording equipment is spoken directly into the microphones and leads to a high number of false positives. While this kind of problem can often be eliminated before starting the analysis process, it still shows a typical problem met in long-term monitoring situations. When a lot of recording material from a field experiment is to be analysed, it is not always feasible to manually remove spoken annotations of the recording sessions. In that case it is a desirable property for detection algorithms to cope with this type of data.

On Day 5, the role of speech is replaced by wind rhythmically impacting on the microphone. Most of the false positives can be explained by these two effects. Some of the false positives may also be due to the fact that the autocorrelation feature $\widetilde{A}$ is computed from overlapping windows. This might result in finding one call sequence twice at different time positions.

False negatives are seldom found with our algorithm, which is a satisfying result. They only occur in two situations: first, extremely silent calls are sometimes dismissed, especially when there are few consecutive calls. Second, our method for noise reduction in the features sometimes leads to a masking effect. When the energy in the band used for estimating noise levels is significantly higher than the energy in the band used for detecting the bittern call, the bittern call can be removed from the features although the call is clearly visible in the spectrogram. Such masking effects also occur in human hearing (Zwicker and Fastl, 1999) and can be the cause for calls which are undetectable by human listeners.

Altogether, our algorithm is a very helpful tool in analysing the presence of the Eurasian bittern. Currently, the main problem is the high number of false positives from wind. How seriously this affects the utility of the algorithm depends on the task to be solved. For example, if the algorithm is used to demonstrate the presence of the Eurasian bittern in an area, most of the false positives can be disposed of by removing all reported positions with low values of $\widetilde{A}$. This would result in dropping detections of short sequences of low calls of the bittern which is bearable as long as some longer or louder calls are present. If detecting very low calls is crucial then an additional feature discriminating wind from bittern calls is

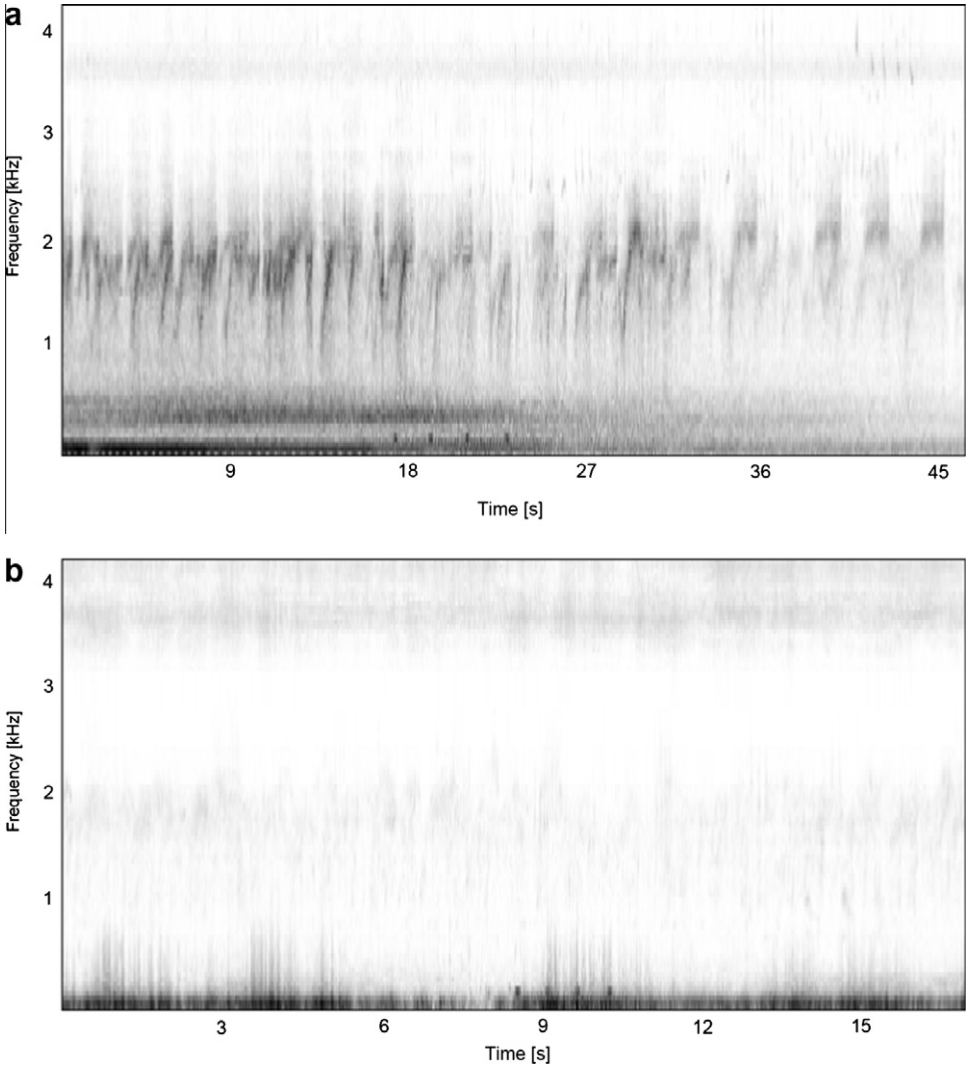


**Fig. 8.** Spectrograms visualising anthropogenic noise (a) and wind noise (b) at the study site. In this case noise is caused by a train passing by in the distance. Both spectrograms show calls of the Eurasian bittern in a noisy context ((a) from second 18 and (b) from second 9).

**Table 1**
Recognition results for the Eurasian bitter on five observation days.

| Recording | Duration (min:s) | Detections | False positives | False negatives |
|---|---|---|---|---|
| Day 1 | 15:21 | 3 | 0 | 0 |
| Day 2 | 15:41 | 9 | 0 | 0 |
| Day 3 | 116:51 | 74 | 20 | 7 |
| Day 4 | 174:30 | 77 | 12 | 2 |
| Day 5 | 97:15 | 83 | 52 | 1 |
| Sum | 419:38 | 246 | 84 | 10 |

needed. In this context, it is possible to detect calls that are all but inaudible to the human ear.

### 4.3. Results for Savi's warbler

The recordings used for evaluating the detector for Savi's warbler, have been attained during the sunrise and sunset periods. They contain a wide range of different birdsongs as well as a multitude of background noises. From a database containing several hundred hours of such recordings, an evaluation excerpt of 19 h has been composed to representatively cover the whole range of background noise met at the particular monitoring site. On this set, which has also been manually annotated by experts, an overall detection rate of 92% was achieved for the Savi's warbler detector.

For a more detailed evaluation of the impacts of different noise types, the signal excerpts were divided into 4 background noise classes. Signals featuring heavy wind or rain noise were bundled as well as those containing many birdsongs, quiet recordings and particularly clear warbler recordings.

As listed in Table 2, the detection rate is reasonably stable, falling to a minimum recall of 79.06% even when dealing with a set of very badly conditioned recordings.

For the group of quiet recordings, where a steady stream of amplification hiss and occasional noises (wind, remote airplanes or trains) dominate the signal's volume, we found that the detection of almost inaudible songs was still possible. Here, the selective properties of a periodicity-based analysis allow for the detection of birdsongs being recorded on a signal level equal to the accumulated amplification hiss. Furthermore, the final autocorrelation sharpness criterion defined in Section 3.2 filters out potential false detections which may otherwise occur in the precedent peak-picking step.

This robustness of our autocorrelation-based approach is also substantiated by the results for the class of very overloaded record-

**Table 2**
Evaluation of the Savi's warbler detector's precision, by signal categories. Row ***D*** gives the recording duration. Row ***R*** contains the recall values: the percentage of song time annotated manually also being found by the detector. ***FP*** measures the percentage of possibly falsely detected time spans.

| | Complete | Good SNR | Many birds | Quiet | RainWind |
|---|---|---|---|---|---|
| ***D*** | 19 h 1 min | 4 h 15 min | 4 h 29 min | 5 h 27 min | 4 h 30 min |
| ***R*** (%) | 92.95 | 99.59 | 97.28 | 79.06 | 80.55 |
| ***FP*** (%) | 1.22 | 0.71 | 1.12 | 1.42 | 1.60 |

**Table 3**
Number of breeding Savi's warblers in two different areas at the Lake Parstein estimated by two different methods.

| | Western reed bed | | Eastern reed bed | | Total | |
|---|---|---|---|---|---|---|
| Method | 2007 | 2008 | 2007 | 2008 | 2007 | 2008 |
| (A) Mapping by experienced observer | 15 | 14 | 9 | 15 | 24 | 29 |
| (B) Application of pattern recognition algorithm and GPS data | 14 | 16 | 10 | 15 | 24 | 31 |

ings performed during sunrise. Here, the chosen four-channel set-up helps to significantly improve the detection results. As the detection probability decreases with decreasing signal-to-noise ratio, the independent analysis of each channel allows the detection of an assumed warbler's song in a less distorted channel. Moreover, given the case of some birds singing simultaneously, detection

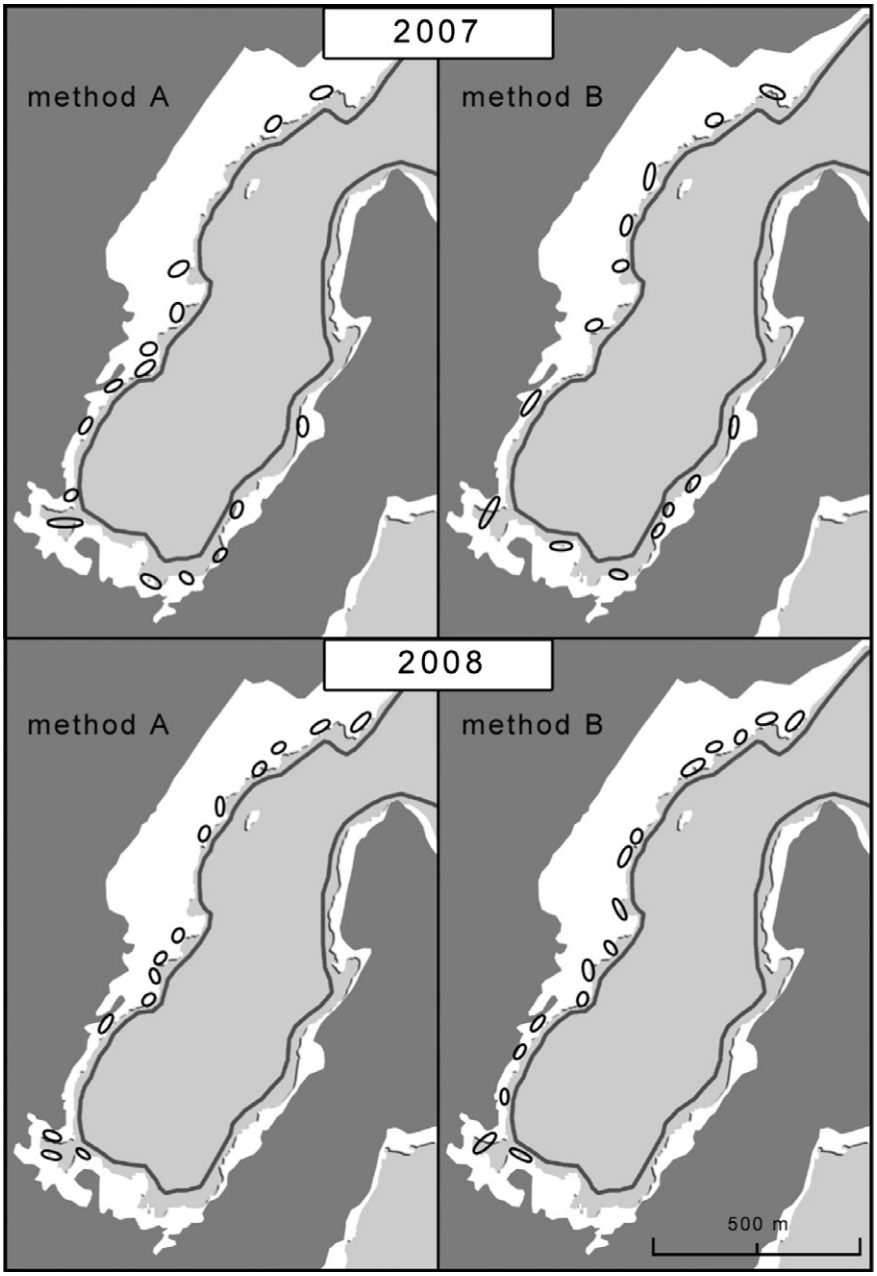


**Fig. 9.** Estimation of breeding territories of Savi's warblers along the Western reed beds at Lake Parstein in the years 2007 and 2008. A – results from mapping by an experienced observer; B – results from the analysis of sound recordings by pattern recognition algorithm and estimating the position by GPS data. The estimated breeding territories are indicated by black circles. Reed beds are displayed as white areas. The solid grey line shows the GPS track of the boat tours.

becomes much simpler if a separation of the songs is possible. Though not reflected in the above evaluation, individual localisation of multiple contributing warblers is also facilitated by the current recording setup.

## 5. Application of bird song detectors to the monitoring of Savi's warblers

In our work, we could already successfully apply the pattern recognition algorithm for the Savi's warbler described above to the monitoring of the population of this species at our study site at Lake Parstein. For this purpose, we examined songposts of the warblers at two different reed areas three times each year during the breeding seasons in the period from April to June in 2007 and 2008. In order to cover a more extended area, we decided to use a slow-moving small boat equipped with an electric motor. It was driven at constant speed (approximately 30 m/min) along the reed beds, keeping a distance of about 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 degree angle and facing towards the reeds. The position of the boat was continuously acquired by a GPS device (Garmin Geko).

On the sound recordings, songs of Savi's warbler were detected by the pattern recognition algorithm described in Section 2.2. Detected songs with a domination value exceeding a fixed threshold were considered for mapping. The location of song posts was determined concerning the following criteria. If the domination value was higher for the track recorded by the rear-facing microphone, a time mark was set to the beginning of the detected song because in this case the boat is moving away from the songpost and closest to it at the beginning of the song. If it was higher for the track recorded by the forward-facing microphone, the end of the detected song was marked respectively. If the total duration of a detected song sequence was longer than 160 s such long sequences could only be explained as overlapping songs from neighbouring birds (for 332 songs uttered during morning hours, 95% are shorter than 161 s in duration, most of them considerably). Therefore, additional time marks were added at a distance of 160 s in this sequence. Finally, the location of the birds was assessed on the basis of the synchronised GPS data. The expected song post positions were estimated by first finding the boat position from the GPS track using the time markers. This position was then projected into the reed bed.

In addition to the automated observations, an experienced supplementary observer on the boat marked every encountered animal in a map as a control test. According to the criteria established for the monitoring program for common breeding birds, we counted a warbler as territorial for both the automated and direct observation when it was observed at least two times a year at the same or nearly the same place.

Applying the two different approaches to two different reed areas in two subsequent years, we found a high consistency in the number of breeding pairs (Table 3). The maps of the spatial distribution of the territories estimated by the different methods reveal large similarities as well (Fig. 9). Even though the positions of song posts estimated by different ways did not match completely, the accuracy of the results is sufficient for purposes of long-term monitoring.

## 6. Conclusion

The computational evaluation of monitoring recordings introduces potent technology complementary to the existing means for assessing animal populations. Using robust feature extraction methods, the detectors introduced in the preceding sections feature a high detection precision even when used with badly conditioned recordings.

For less specific animal population surveys, the introduced features may be combined and used for the detection of a wider range of animal species. Several approaches to detecting multiple species using a general feature set have been published (Farnsworth, 2005; Brandes, 2008; Selin et al., 2007; Fagerlund, 2007). However, when working with unsupervised recordings performed in an acoustically unpredictable area, a great amount of overlap between different bird vocalisations and other noise sources may occur. In this case, the use of general feature sets may be problematic, particularly because the mix of different vocalisations is likely to be reflected in a complex mix in the feature space. As there is a high degree of variability of such mixtures when considering real scenarios, the proper training of corresponding classifiers may be a difficult task and can result in considerably less reliable detection outputs. In such cases, we therefore recommend the use of highly customised detectors which are tailored to the recognition of a small set of species only. An elaborate acoustic scene analysis, also detecting various noise sources, should be considered for further research on large-scale species recognition. In particular, the impact of wind on microphones could be detected by using bandwidth and entropy estimates for strong signal components. In the proposed detectors, the acoustic background is accounted for by using features from an individually shaped flanking band, used for filtering ambiguous events. This flanking band strategy is suitable for a wide range of animal sounds. Moreover, the feature extraction and classification methods developed in this project show further potential for successful application in the detection of a variety of other species. For example, the periodicity features used in the Savi's warbler detector also reflect elementary parameters of calls from frogs or crickets.

In combination with recorded GPS data tracks, as described in Section 5, robust localisation and mapping of Savi's warbler territories has been achieved. The synchronised audio and GPS tracks also enable the application of detectors for other species.

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 standardised data acquisition, this methodology is appropriate for future use even by observers with limited knowledge of bird songs. Moreover, we are currently improving the present pattern recognition algorithm in order to allow discrimination of neighbouring and simultaneously singing individuals to guarantee a more autonomous and automated analysis of the sound recordings. Thus, the population of the species discussed above could be assessed in a precise manner. In contrast to the monitoring of nocturnal migrating birds by flight calls where the number of calls was a rough measure for assessing the intensity of migration (Hill and Hüppop, 2008; Schrama et al., 2008), the actual number of territorial males and therefore the number of breeding pairs was estimated in our study. In general, this promising approach could also be used for other bird species living in reed beds.

For the monitoring of population sizes using stationary recording devices, the proposed algorithms could be combined with existing source separation techniques. This is planned for future research. In particular, the four-channel microphone setup was chosen for applying an acoustic beamforming routine previous to the detection step. Thereby, multiple signal sources should first be separated and then analysed separately. In addition to a simplified estimation of the number of individuals present, the signal-to-noise ratio will also be improved for each separated signal part. This should result in more precise detections and thus improve the accuracy of the derived population estimations.

## Acknowledgements

The study was supported by grants from the German Federal Agency for Nature Conservation (BfN, Grant 806 82 060 - K2) and the foundation NaturschutzFonds Brandenburg (Grant 557). During the accomplishment of the study, R. Bardeli and D. Wolff have been members of the Multimedia Signal Processing Group at the Computer Science Department of the University of Bonn headed by M. Clausen.

## References

Bibby, C., Burges, N., Hill, D., 1992. Bird Census Techniques. Academic Press, London.

Brandes, T., 2008. Automated sound recording and analysis techniques for bird survey and conservation. Bird Conserv. Internat. 18, 163–173.

Brandes, T., 2008. Feature vector selection and use of hidden markov models to identify frequency-modulated bioacoustic signals admidst noise. IEEE Trans. Audio, Speech, Language Process. 16, 1173–1180.

Brandes, T.S., Naskrecki, P., Figueroa, H.K., 2006. Using image processing to detect and classify narrow-band cricket and frog calls. J. Acoust. Soc. Amer. 120, 2950–2957.

Brown, J., Miller, P., 2007. Automatic classification of killer whale vocalizations using dynamic time warping. J. Acoust. Soc. Amer. 122, 1201–1207.

Deecke, V.B., Ford, J.K.B., Spong, P., 1999. Quantifying complex patterns of bioacoustic variation: Use of a neural network to compare killer whale (orcinus orca) dialects. J. Acoust. Soc. Amer. 105, 2499–2507.

Deller, J., Proakis, J., Hanson, J., 1993. Discrete-time Processing of Speech Signals. Prentice-Hall, New Jersey.

Evan, W., Mellinger, D., 1999. Monitoring grassland birds in nocturnal migration. Stud. Avian Biol. 19, 219–229.

Fagerlund, S., 2007. Bird species recognition using support vector machines. EURASIP J. Appl. Signal Process. (EURASIP JASP), 8 p.

Fagerlund, S., Härmä, A., 2005. Parametrization of inharmonic bird sounds for automatic recognition. In: 13th European Signal Processing Conf. (EUSIPCO 2005).

Farnsworth, A., 2005. Flight calls and their value for future ornithological studies and conservation research. Auk 122, 733–746.

Farnsworth, A., Russel, R., 2007. Monitoring flight calls of migrating birds from an oil platform in the northern gulf of mexico. J. Field Ornithol. 78, 279–289.

Farnsworth, A., Gauthreaux, S.J., Van Blaricom, D., 2004. A comparison of nocturnal call counts of migrating birds and reflectivity measurements on doppler radar (wsr-88d). J. Avian Biol. 35, 365–369.

Farr, I., Chesmore, D., 2007. Automated bioacoustic detection and identification of wood-boring insects for quarantine screening and insect ecology. In: 4th International Conf. on Bioacoustics. vol. 29, pp. 201–208.

Frommolt, K.-H., Tauchert, K.-H., Koch, M., 2008. Advantages and Disadvantages of Acoustic Monitoring of Birds – Realistic Scenarios for Automated Bioacoustic Monitoring in a Densely Populated Region. In: Computational Bioacoustics for Assessing Biodiversity. Proc. of the Internat. Expert Meeting on IT-based Detection of Bioacoustical Patterns. BfN-Skripten 234, pp. 83–92.

Gregory, R., van Strien, A., Vorisek, P., Meyling, A., Noble, D., Foppen, R., Gibbons, D., 2005. Developing indicators for european birds. Phil. Trans. Roy. Soc. B 360, 269.

Härmä, A., Somervuo, P., 2004. Classification of the harmonic structure in bird vocalization. In: Proc. of IEEE Internat. Conf. on Acoustics, Speech, and Signal Processing (ICASSP 2004).

Härmä, A., 2003. Automatic recognition of bird species based on sinusoidal modeling of syllables. In: Proc. of IEEE Internat. Conf. on Acoustics, Speech, and Signal Processing (ICASSP '03). vol. 5, pp. 545–548.

Hill, R., Hüppop, O., 2008. Birds and Bats: Automatic Recording of Flight Calls and their Value for the Study of Migration. In: Computational Bioacoustics for Assessing Biodiversity. Proc. of the Internat. Expert Meeting on IT-based Detection of Bioacoustical Patterns. BfN-Skripten 234. pp. 135–141.

Hüppop, O., Dierschke, J., Exo, K.-M., Fredrich, E., Hill, R., 2006. Bird migration studies and potential collision risk with offshore wind turbines. Ibis 148, 90–109.

Klvaňová, A., Voříšek, P., 2007. Review on large-scale generic population monitoring schemes in europe 2007. Bird Census News 20, 50–56.

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. J. Acoust. Soc. Amer. 103, 2185–2196.

Marcarini, M., Williamson, G., de Sisternes Garcia, L., 2008. Comparision of methods for automated recognition of avian nocturnal flight calls. In: ICASSP 2008. IEEE Internat. Conf. on Acoustics, Speech and Signal Processing. March 31–April 4, pp. 2029–2032.

Mellinger, D., Stafford, K., Moore, S., Dziak, R., Matsumoto, H., 2007. An overview of fixed passive acoustic observation methods for cetaceans. Oceanography 20 (4), 36–45.

Mills, H., 2000. Geographically distributed acoustical monitoring of migrating birds. J. Acoust. Soc. Amer. 108 (5), 2582.

Mitsakakis, N., Fisher, R., Walker, A., 1996. Classification of whale song units using a self-organizing feature mapping algorithm. J. Acoust. Soc. Amer. 100 (4), 2644.

Nickerson, C.M., Bloomfield, L.L., Dawson, M.R.W., Sturdy, C.B., 2006. Artificial neural network discrimination of black-capped chickadee (poecile atricapillus) call notes. Appl. Acoust. 67 (11–12), 1111–1117.

Obrist, M., Boesch, R., Flückiger, P., 2004. Variability in echolocation: Consequences, limits and options for automated field identification with a synergetic pattern recognition approach. Mammalia. 68, 307–322.

Placer, J., Slobodchikoff, C.N., Burns, J., Placer, J., Middleton, R., 2006. Using self-organizing maps to recognize acoustic units associated with information content in animal vocalizations. J. Acoust. Soc. Amer. 119, 3140–3146.

Possart, G., 2002. Signal classification of bird voices using multiscale methods and neural networks. Master's Thesis, University of Kaiserslautern.

Quinlan, J.R., 1993. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers.

Schrama, T., Poot, M., Robb, M., Slabbekoorn, H., 2008. Automated monitoring of avian flight calls during nocturnal migration. In: Computational Bioacoustics for Assessing Biodiversity. Proc. of the Internat. Expert Meeting on IT-based Detection of Bioacoustical Patterns. BfN-Skripten 234, pp. 131–134.

Schwenker, F., Dietrich, C., Kestler, H., Riede, K., Palm, G., 2003. Radial basis function neural networks and temporal fusion for the classification of bio acoustic time series. Neurocomputing 51, 265–275.

Seekings, P., Potter, J.R., 2003. Classification of marine acoustic signals using wavelets & neural networks. In: Proc. of 8th Western Pacific Acoustics Conf. (Wespac8).

Selin, A., Turunen, J., Tanttu, J.T., 2007. Wavelets in automatic recognition of bird sound. EURASIP J. Signal Process. Special Issue Multirate Syst. Appl., 9 p.

Somervuo, P., Härmä, A., 2003. Analyzing bird song syllables on the self-organizing map. In: Workshop on Self-Organizing Maps (WSOM'03).

Somervuo, P., Härmä, A., Fagerlund, S., 2006. Parametric representations of bird sounds for automatic species recognition. IEEE Trans. Speech Audio Process. 14 (6), 2252–2263.

Tanttu, J.T., Turunen, J., Selin, A., Ojanen, M., 2006. Automatic feature extraction and classification of crossbill (loxia spp.) flight calls. Bioacoustics 15, 251–269.

Taylor, A., 1995. Recognising biological sounds using machine learning. In: Proc. 8th Australian Joint Conf. on Artificial Intelligence. pp. 209–212.

Trifa, V., Kirschel, A., Taylor, C., 2008. Automated species regognition of antbirds in a mexican rainforest using hidden markov models. J. Acoust. Soc. Amer. 123, 2424–2431.

Zwicker, E., Fastl, H., 1999. Psychoacoustics: Facts and Models. Springer, Berlin.