Advances in Automatic Bird Species Recognition from Environmental Audio

Xueyan Dong*, Jingpeng Jia

1 Special Education College, Beijing Union University, Beijing, 100070, China
*Corresponding author’s e-mail: tjt.xueyan.dong@buu.edu.cn

Abstract. Bioacoustics has recently become one of the “big data” research topics since many bird monitoring projects have collected terabytes of audio using remote sensors. The challenge in recent years is to develop algorithms to realize fully automatic recognition of bird species through analysing environmental recordings. A number of approaches directly draw on experience of effective algorithms in signal processing and image processing areas. They seem working well for small data or lab data, however, the outcomes for large-scale environmental data shows a big gap between theoretical experiments and real applications. To provide possible clues for future research, we review the state-of-art development in automated bird species recognition, and identify wide range of algorithms on noise removal, bird call detection, feature extraction for classification. The significant software tools and publicly available datasets for the task are presented. This survey can be valuable for new researchers who are about to start the journey with birdsong analysis.

1. Introduction

Acoustic sensing techniques have been deployed to record vocal species to assist ecologists in bird studies, such as monitoring endangered species[1]and species richness[2]. The sensors operate continuously for long periods and the collected recordings provide a persistent and verifiable record of the soundscape. Furthermore, sensing technology is a cheap way for ecologists to study vocal species when combined with computer-aided analysis tools. However, the task of bioacoustics monitoring brings other challenges. It requires experts or birders to listen through all recordings or visually scan the corresponding visual representations of audio so as to provide ground truth data, and this is an arduous task. Therefore, automated tools are essential to process large collection of birdsong recordings with little annotation data. Besides, it is difficult to build accurate bird call recognisers on real-world birdsong recordings because environmental noise is undefined and calls can vary geographically and the audio content itself are not always clear to be identified.

While fully-automated analysis techniques can, in theory, be scaled up to process large volumes of audio data, in practice, their reliability and accuracy remain problematic. Most research effort to date has been on automated bird call recognition tasks to improve the recognition accuracy and increase the scalability[3-5]. Recently, many signal processing and image processing techniques have been used to automate the detection of bird calls. Many pattern recognition approaches have been explored for species classification. Commercial software programs, such as Raven[6], are available to detect and characterise bird calls. While these approaches and tools show promising results on automatic detection bird sounds from lab recordings, they have poor performance in real-world data.
In this paper, we review more than 200 papers related to automatic bird species recognition and report the state-of-art techniques used in the literature. We aim to provide a comprehensive survey to help readers to understand the recent developments in this field.

The rest of the paper is organised as follows: Section 2 lists the techniques used in bird species recognition. A summary of world-wide research is presented in Section 3 and a number of public datasets are given in Section 4. In the end, Section 5 makes conclusions and proposes future research directions.

2. Techniques used for automated bird species recognition

According to the reviewed literature, the approaches in automated species recognition usually contain three key steps: 1) noise removal, 2) bird call detection, 3) feature extraction for species classification and recognition. Background noise can be a significant issue in processing field recording. In this case, a noise reduction algorithm is often required to pre-process the raw audio. Detection aims to isolate individual bird calls or syllables from the continuous recording. To differentiate various bird calls or bird species, a multiple dimension of features are calculated. Based on the features, the classification task can identify a number of known species or bird calls, and recognition can find all positive calls.

2.1. Noise removal

Since bird vocalisations are recorded in the wild environment, they unavoidably contain background noises from wind, rain, other animal sound, and human speaking. In this context, a proper noise removal method is required. The methods based on the spectrogram aim to use filters to eliminate other unwanted signals but keep birdcalls of interest. However, it is not easy as the difficulties lie in three aspects. (1) the background noise tends to show various types. (2) some signals of interest are quite weak. (3) competing signals make the task even harder. Therefore, no single way can handle all types of noise, and be adaptive to all conditions.

Specifically, a low-pass filter was developed to excludes the noise generated from wind, waves and machines in the 0-6 kHz range[7]. This method is simple but leads to lower frequency signals being disregarded. As the spectrogram of acoustic data can be treated as an image, image processing approaches, such as edge detection and blur filters, are applied. Kittisuwann et al.[8] proposed a denoising algorithm based on wavelet decomposition. However, Hussein et al.[9] found the wavelet method was limited in its ability to reduce the noise generated from the Fourier transform. To address the issue, they explored a novel edge detection approach to eliminate the attached noise. However, the edges of overlapped signals and quieter acoustic sounds are more likely to be blurred and difficult to detect. Towsey and Planitz[10] found that the effects of noise tend to be reduced when frequency increases. Therefore, they initially applied a Wiener filter and then modified the adaptive level equalisation approach for noise reduction. In the study of [11], the morphological filtering was used to eliminate short bursts of noise and suppress weak competing signals. Lesseck[12] designed a median clipping approach, which set pixels to 1 when its pixel value is larger than 3 times of its corresponding frequency band as well as time frame. It resulted in a binary image which was further processed by closing, dilation and media filter to optimize the output.

2.2. Bird call detection

A more general definition on bird call detection is to find acoustic event in the audio. Its goal is to determine where the bird vocalisation begins and ends in the waveform or which particular region can be a bird call in the spectrogram. The detection target can be classified into five output examples[13]. Typical event detection methods include thresholding, frequency track detection, sinusoidal modelling, and template matching.

Thresholding is a widely used method, which is achieved by setting up a amplitude threshold on waveform or intensity threshold on spectrogram. Du and Troyer[14] reported the amplitude detector is inappropriate when bird calls show various amplitude levels. Therefore, rather than a fixed threshold, an adaptive angle threshold derived from the envelope data is used to determine the time boundaries of the signals. However, this segmentation method relies on the good quality of noise removal output.
Spectrogram blob detection is a method based on the energy of the signal in the spectrogram[15]. This method exploits an intensity threshold to pick up pixels that potentially belong to birdcalls. The selected pixels are grouped into an event through a eight directional searching algorithm. Since small events are less likely to be bird vocalisations, they are eliminated by a threshold. The method does not depend on a particular species; therefore, it can be used as a general method for acoustic event detection.

Brandes[16] assumed birds tend to make calls in a fixed narrow frequency band. Therefore, the spectral distribution intensity values over a specific frequency band change slowly. Applying a threshold filter to each frequency band proves useful for distinguishing among modulated animal sounds. In fact, many bird vocalisations occur across a wide frequency range, sometimes overlapping with those of other species. This contradicts Brandes’s assumption and makes his method unsuitable. Raven, a software package, provides a band limited energy detector[17]. The detector estimates the background noise of a signal and uses this estimation to find sections of signal that exceed a user-specified signal-to-noise ratio threshold in a region. The region is specified by frequency and time bounds. Duan et al.[5] examined the detector for bird component detection and found its performance worse in noisy recordings than in clean ones.

Härmä[18] believed that most birdcalls contain syllables showing sinusoidal characteristics, a sinusoidal model was developed to detect calls having syllables that show a spectral shape similar to sine waves. As reported by Somervuo et al.[19], many bird sounds do not have sinusoidal properties. Therefore, a spectral peak track detector was developed to segment bird calls consisting not only of pure tonal components, but also harmonic and inharmonic combinations of tones[3]. Co-vocalisations are common in field recordings, so a more robust segmentation scheme, called the frequency track method, was explored[20]. This method achieved promising results in removing the weak background vocalisations and keeping a dominant vocalisation.

Finding a way to detect all bird call structures can be difficult. Duan et al.[21] designed a generic method aiming to detect a wide range of bird calls. Since most Australian bird calls contain one or more of the following five commonly seen components: harmonics, whistles, clicks, slurs, and blocks, they developed a corresponding detector for each component with careful training. This method requires users to have prior knowledge on call structures to select which detector should be used. This makes the analysis semi-automatic rather than automatic.

Template matching is another method to detect bird calls from continuous recordings and has been studied by a number of researchers. The template was used to detect the end-time boundary from chaffinch songs[22]. The method is adequate for recognition of calls that show small variations because it is hard to cover all possible templates. A more effective method is through the similarity matching on points of interest from query call spectrograms[23]. Here points of interest refer to pixels which are more likely to belong to a bird call. This method allows some degree of variation in bird vocalisations. However, while this method performed well for bird call structures with edge characteristics such as click and whistle, it failed for the block shape of call structure. To address this problem, a generalised ridge detector was developed to compress the block shape in vertical or horizontal directions at a certain scale such that the block becomes a ridge structure[24]. Potamitis[25] constructed a dictionary of templates from reference recordings and then search through a large number of recordings to examine the underlying bioacoustic activity.

2.3. Feature extraction for bird classification and recognition
As feature extraction is the key to automated species recognition, many studies have been conducted on this topic. The meaning of feature extraction lies in two points: (1) compress the detected call to a small amount of statistical values; (2) characterise the bird calls for differentiating one call/species from another.

2.3.1. Features from waveform. Audio signal can be easily converted into waveform, therefore many measures, such as short time energy and zero-crossing rate, are derived from the waveform to characterise bird calls. [26] measured zero-crossing rates to classify bird species. However, Wolff[27]
reported that these features alone are insufficient for distinguishing between bird species particularly when target acoustic signals are overlapped with background noise or other untargeted sounds.

2.3.2. Features from spectrogram. Audio signals can be transformed into a time-frequency representation (called spectrogram) by applying a Fourier transform. Based on the spectrogram, many features can be explored to represent bird sounds. These features include spectral features, temporal features, sinusoidal models, the shape descriptors in bird calls, and peak frequency track and intensity-based features. Spectral features (for example, call bandwidth and spectral flatness) can be extracted from spectrograms. Many bird calls consist of modulated tonal sounds, which can be modelled as time-varying sinusoids[18]. Härmä first found that syllables from birdsong structures in spectrograms exhibit modulated sinusoidal pulses, which are a set of frequency and amplitude values. To use this method, a strict requirement must be satisfied that the syllables in bird calls do not overlap in time or frequency. This limits its applications to non-overlapping signals.

In study[3], bird calls were detected as spectral peak tracks in spectrograms. These tracks were described by track frequencies, frequency differences, relative power, shape, and duration. These descriptors performed well in classifying bird species showing sinusoidal shape, harmonic and inharmonic structures. However, it was limited in representing calls containing rapidly modulated whistles and clicks. A different approach to characterising sinusoidal signals was developed[20]. The approach exploits the spectral magnitude shape and phase continuity to construct Gaussian mixture models (GMM).

2.3.3. Mel-frequency based features. Due to the successful application of mel-frequency cepstral coefficients (MFCCs) in human speech recognition, MFCCs were early used to represent bird sounds[19]. MFCCs are derived by transforming audio signals into the Mel scale. Kwan et al. (2006) characterised each frame of a birdsong using 13-dimension (13-d) MFCCs, such that the overall sound was described by a list of 13-d values. MFCCs show good performance in birdsong classification where bird calls have harmonic structure like human speech. However, such a representation results in a high dimension of the feature vector for a call containing multiple frames. This leads to a high computational cost. To reduce the dimensionality of the feature vector, the average MFCCs from the frames within a syllable are calculated[28]. To capture the dynamic information in bird sounds, MFCCs are often combined with delta MFCCs to represent bird call[29]. MFCC models offer a compact parametric representation of bird vocalisations with broadband characteristics and harmonics. However, as pointed out by Somervuo et al.[19], there is little evidence that MFCCs capture information optimal for bird species identification. Furthermore, cepstral coefficients misrepresent important pitch information and their suitability for many bird species must be questioned. However, as pointed by [30], MFCCs were originally designed to represent human speech whereas birds produce sounds differently from human speech. The failure of the MFCCs features is due to capturing information from the entire spectrum, which may be largely dominated by noise since bird sounds are often localised only in narrow frequency regions.

2.3.4. Wavelet transform based features. The wavelet transform(WT) is achieved by transforming audio signals to the spectral domain. The WT approach is useful for tracking changes of audio signals in various resolutions. Audio signals are decomposed into the wavelet coefficients using the wavelet packet decomposition. Selin et al.[31] proved that when audio signals are decomposed into 26 bands, the lower to middle parts of wavelet signals are relevant to bird species. Therefore, the coefficients from these parts are employed to calculate four features: maximum energy, position, spread, and width. Experimental results show that the features are appropriate in classifying bird species showing inharmonic and harmonic structures. This also means that other types of calls may be not applicable.

2.3.5. Features using image processing techniques. Spectrograms can be viewed as images, and a range of image processing techniques was applied to the problem of bird species recognition in
spectrograms. In the spectrogram, bird calls are represented by pixel clusters, each of which is characterised by central frequency, duration, and bandwidth\[16\]. These features are highly effective for detecting calls appearing within a fixed frequency band. However, the application to bird calls that cover variable frequencies needs to be investigated. Bardeli \[32\] used the structure tensor, an edge detection method, to detect local gradients in spectrograms. Each detected point is described by a 12-bit vector of the local feature description through the fast Fourier transform. These features are good for characterising curve-like birdsongs, but not for broadband bird calls. Another similar method called histograms of oriented gradients (HOG)\[33\] which combines with acoustic features but the contribution of the HOG features to the final result was not reported. A study of bird call retrieval\[34\] improved the HOG features and obtained promising results.

2.3.6. **Feature learning.** Many traditional feature extraction methods are based on prior knowledge. Different from these methods, feature learning was proposed by \[30\], which uses unsupervised machine learning approach. They believed that features learnt automatically from data outperform the manually designed features. The feature learning method shows great advantage in dealing with classification at a large scale. One thing needs to be noted, this method requires tons of training data to be effective.

In recent research, the dimensions of features are up to several thousands. How many features are really effective in classifying bird species remains problematic. In future, researchers may either do not need to consider the high dimensions of feature relying on cloud computing sources or a wise algorithm might be developed to select the most effective features with low dimensions.

3. **Software tools**

To put the existing contributions of automated birdsong recognition into practice, there are varieties of software tools used for automated detection of animal sounds. These programs are often easy to use and provide great tools for analysing a large amount of audio recordings. Table 1 gives a brief description of these tools.

All current software packages can achieve good performance in identifying a limited number of bird species, and they all rely on recordings with moderate SNR and the availability of sufficient training datasets. Their advantage is that they are ideally suited for supervised learning algorithms for analysis of different sounds. However, they cannot work well with low SNR audio data and when training data is insufficient. Therefore, more robust automatic toolkits need to be developed to overcome these limitations.
Table 1. Software tools for birdsong analysis.

| Title                | Software description and application                                                                                                                                                                                                 |
|----------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Audacity             | Audacity is a free, open-source software for editing sounds.                                                                                                                                                                            |
| SeeWave              | An R package for sound analysis.                                                                                                                                                                                                       |
| Kaleidoscope Analysis Software | An integrated suite of tools that make bioacoustic analysis easier, faster and more effective. The latest version, Kaleidoscope Pro, offers powerful tools including acoustic cluster analysis, viewer tools, classification tools, high-speed batch processing, and data export. For details, please go to the website (https://www.wildlifeacoustics.com/products/kaleidoscope-software-acoustic). |
| Raven                | It features several detectors and recognisers. It provides useful tools to perform band pass filtering (to remove background noise), and manually or semi-automatically segment syllables. It allows users to visualise the signal as a waveform and a spectrogram. For feature extraction, Raven mainly explores features including average power, centre frequency, bandwidth, entropy, and quality of syllable duration. It is useful for dealing with continuous recordings. |
| HTK                  | A toolkit for applying HMMs specially designed for speech recognition. Recently, it has been used for bird species identification.                                                                                                         |
| SoundRuler           | A tool for measuring and visualising sound and teaching. It is especially useful in the analysis of simple and repetitive signals.                                                                                                         |
| Avisoft              | Avisoft is a powerful Windows application for investigating animal acoustic communication. It provides a broad range of processing and analysis tools. It uses cross correlation for bird species recognition and is widely used in bioacoustic identification tasks. |
| SoundID              | SoundID provides hardware products for recording sounds in environments and a set of algorithms for sound recognition. It is good for simple and repetitive signals and has been used for bird species recognition with careful feature selection. |
| ChirpOMatic UK       | A top birdsong recognition app in 2015. It features a bird-safe mode and wind detection. The bird-safe mode allows a user to record birdsongs without disturbing other wildlife. The wind detection can identify the wind and remind the user to find a shelter. Once the recordings upload, they are sent to ecologists to help monitor changes in the distribution of species. It is currently only available for iPhones and iPads. |
| Warblr               | An automatic birdsong recognition app. It allows users to make a recording with an iPhone and upload for identifying the species included in the recording. This app contributed to the citizen science project. |

4. Published datasets
Public birdsong recordings and annotation data are important to conduct automatic birdsong research. They need to be well prepared for particular applications. Many studies described in this review used unpublished datasets for specific applications. While published datasets show great promise in boosting the development of automated birdsong analysis, only a small number of researchers use datasets from publicly available compact disks and the internet to conduct bird sound analysis. Table 2 lists the datasets publicly available for automated bird species analysis. According to the particular application, metadata (annotation data) are made differently. They can either be at syllable level (such as Bird-DB) or species level of bird sounds. That is, each syllable or birdsong in the bird sound recording has a label. The annotation data is often used for ground truth to validate the effectiveness of developed algorithms for classifiers recognisers.
Table 2. Available datasets for automated bird species recognition.

| Dataset               | Features                                                                 |
|-----------------------|--------------------------------------------------------------------------|
| Xeno-canto            | Xeno-canto provides a website (http://www.xeno-canto.org) for sharing bird sounds from all over the world. It has more than 221,547 recordings of birdsongs. The website has great advantages in discovering unknown species and geographic variability of same species in a crowdsourcing way. However, due to the diversity of involvements, the quality of audio and annotation data varies a lot. Some recordings are clear but some are hardly audible. |
| Berlin Museum für Naturkunde | The Animal Sound Archive at the Museum für Naturkunde in Berlin is one of the oldest and largest collections of animal sounds. Presently, the collection consists of about 120,000 bioacoustical recordings comprising 1,800 bird species (www.tierstimmenarchiv.de). HU-ASA database is a large archive of animal vocalisations annotated with the species and additional metadata, including 1,418 audio files available in MP3 encoding. The total recording length of the files was 20,423s (5h 40min 23 seconds). The majority of the available recordings consist of birds, mammals, and ‘Others’ including Sauropsida and Hexapoda. |
| Freefield1010         | Freefield1010 is an open dataset comprising of a wide range of sound sources, such as musical instruments, human speech, music, birds, trains and rain and etc. It was published by (Stowell & Plumley, 2013) in 2013. The URL to access the audio data and metadata can be found at http://c4dm.eecs.qmul.ac.uk/rdr/handle/123456789/35. In this dataset, 17,807 10-second segmented recordings from Freesound field recording were partitioned into 10 folders equally for the purpose of machine learning experiments. For each recording, a formatted metadata is provided with tag names, sample rate, channels, etc. |
| Bird-DB               | Bird-DB was provided by[35] has over 1000 recordings of 30 bird species in California and Western Australia regions; 428 of all were fully annotated at the phrase (syllable) level of bird songs. |
| Bioacoustic Unit      | The Bioacoustic Unit offers a growing data set for the study of biodiversity. The data are organised into various projects including terrestrial and wetland recordings for biodiversity study and recordings of particular vocal species, especially an endangered species, Yellow Rail for rare animal monitoring. |

Except for the datasets shown in Table 2, there are many other short-term competition. The 9th annual Machine Learning for Signal Processing (MLSP) competition was started in 2013. This competition provides a real-world dataset of bird sounds collected in field conditions. Its goal was to develop a classifier which predicts the set of bird species present in a given ten-second audio recording. Dan Stowell et al.[13] proposed a challenge on published datasets in 2016, which attracts many researchers participated. Another well-known project was the BirdCLEF task. This project offers a wide range of datasets and set yearly challenges to attract researchers worldwide participate in the task. Recently, the greatest challenges made focus on recognition of a large number of bird species over a large volume of acoustic recordings.

5. Conclusion

In this paper, we present the state-of-the-art algorithm developments in automatic bird species identification from recordings. The developed approaches for noise removal, bird call detection, and feature extraction are particularly discussed. Many methods in early years aimed at limited species recognition. After 2005, many researchers worked on recognising multiple species. Due to the increasing volume of collected audio recordings, some researchers sought a solution for large-scale analysis and algorithms for fully automated analysis.

Current techniques have achieved good performance in single species recognition and multi-label classification. In the future, new approaches are required for the recognition of all species singing in raw
recordings that can contain up to dozens even hundreds of birds singing simultaneously, such as dawn chorus. The difficulty lies in developing the segmentation algorithm that can effectively separate species from each other in the recording. With the development of classification techniques, feature extraction and recognition can be very effective and efficient but they require large amount of audio recordings for training the classifier algorithm. Therefore, citizen science is a trend to encourage more people to engage in not only collecting data but also annotating data at a large scale. A new trend to analyse the large volumes of audio data is to find the monthly or seasonally variations of species activity in a region. This brings collaboration between scholars from data science and digital arts areas and researchers in computer science field.

Acknowledgments
This paper is supported by two grants which are the 2019 Science and Technology Plan of Beijing Municipal Education Commission (Grant No. KM201911417005) and Beijing Municipal Natural Science Foundation under Grant 4202028.

References
[1] Bardeli, R., Wolff, D., Kurth, F., Koch, M., Tauchert, K.-H., & Frommolt, K.-H. (2010). Detecting bird sounds in a complex acoustic environment and application to bioacoustic monitoring. Pattern Recognition Letters, 31(12), 1524-1534.
[2] Wimmer, J., Towsey, M., Roe, P., & Williamson, I. (2013). Sampling environmental acoustic recordings to determine bird species richness. Ecological Applications, 23(6), 1419-1428.
[3] Chen, Z., & Maher, R. C. (2006). Semi-automatic classification of bird vocalizations using spectral peak tracks. The Journal of the Acoustical Society of America, 120(5), 2974-2984.
[4] Zhao, Z., Zhang, S. H., Xu, Z. Y., Bellisario, K., Dai, N. H., & Omrani, H., et al. (2017). Automated bird acoustic event detection and robust species classification. Ecological Informatics, 39, 99-108.
[5] Duan, S., Zhang, J., Roe, P., Wimmer, J., Dong, X., Truskinger, A., & Towsey, M. (2013). Timed Probabilistic Automaton: A Bridge between Raven and Song Scope for Automatic Species Recognition. Paper presented at the Twenty-Fifth IAAI Conference.
[6] CornellLab of Ornithology, C. L. (1999). Raven:interactive sound analysis software. Retrieved from http://www.birds.cornell.edu/brp/raven/ravenoverview.html
[7] Brandes. (2008a). Automated sound recording and analysis techniques for bird surveys and conservation. Bird Conservation International, 18(S1), S163-S173.
[8] Kittisuwan, P., Chanwimaluang, T., Marukatat, S., & Asdornwised, W. (2010). Image and audio-speech denoising based on higher-order statistical modeling of wavelet coefficients and local variance estimation. International Journal of Wavelets, Multiresolution and Information Processing, 8(06), 987-1017.
[9] Hussein, W., Hussein, M., & Becker, T. (2012). Spectrogram Enhancement By Edge Detection Approach Applied To Bioacoustics Calls Classification. Signal, 3.
[10] Towsey, & Planitz, B. (2011). Acoustic analysis of the natural environment.
[11] Oliveira, A. G. D. and T. M. Ventura, et al. (2015). "Bird acoustic activity detection based on morphological filtering of the spectrogram." Applied Acoustics 98: 34-42.
[12] Lasseck, M. (2013). Bird Song classification in Field Recordings: Winning Solution for NIPS4B 2013 Competition. 2013 Neural Information Processing Scaled for Bioacoustics (NIPS4B).
[13] Dan Stowell, Mike Wood, Yannis Stylianou, & Herve Glotin. (2016). Bird detection in audio: A survey and a challenge. 2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP).
[14] Du, P., & Troyer, T. W. (2006). A segmentation algorithm for zebra finch song at the note level. Neurocomputing, 69(10), 1375-1379.
[15] Towsey, M., Planitz, B., Nantes, A., Wimmer, J., & Roe, P. (2012). A toolbox for animal call recognition. Bioacoustics, 1-19. doi:10.1080/09524622.2011.648753.

[16] Brandes. (2008b). Techniques for bioacoustic signal detection using image processing. Computational bioacoustics for assessing biodiversity. Bundesamt fur Naturschutz, Bonn, Germany, 103-110.

[17] Mills, H. (2000). Geographically distributed acoustical monitoring of migrating birds. The Journal of the Acoustical Society of America, 108(5), 2582-2582.

[18] Härmä, A. (2003). Automatic identification of bird species based on sinusoidal modeling of syllables. Paper presented at the 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP’03).

[19] Somervuo, P., Härmä, A., & Fagerlund, S. (2006). Parametric Representations of Bird Sounds for Automatic Species Recognition. IEEE Transactions on Audio, Speech, and Language Processing, 14(6), 2252-2263.

[20] Jančovič, P., Köküer, M., & Russell, M. (2014). Bird species recognition from field recordings using HMM-based modelling of frequency tracks. Paper presented at the 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).

[21] Duan, S., Towsey, M., Zhang, J., Truskinger, A., Wimmer, J., & Roe, P. (2011). Acoustic component detection for automatic species recognition in environmental monitoring. Paper presented at the 2011 Seventh International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP).

[22] Bardeli, R., Wolff, D., & Clausen, M. (2007). Bird song recognition in complex audio scenes. Paper presented at the Computational bioacoustics for assessing biodiversity. Proc. International expert meeting on IT-based detection of bioacoustical patterns.

[23] Bardeli, R., Wolff, D., Kurth, F., Koch, M., Tauchert, K.-H., & Frommolt, K.-H. (2010). Detecting bird sounds in a complex acoustic environment and application to bioacoustic monitoring. Pattern Recognition Letters, 31(12), 1524-1534.

[24] Dong, X., Eichinski, P., Towsey, M., Zhang, J., & Roe, P. (2015). Birdcall retrieval from environmental acoustic recordings using image processing.

[25] Potamitis, I. (2015). "Unsupervised dictionary extraction of bird vocalisations and new tools on assessing and visualising bird activity." Ecological Informatics 26: 6-17.

[26] Fagerlund, S. (2007). Bird Species Recognition Using Support Vector Machines. EURASIP Journal on Advances in Signal Processing, 2007. doi:10.1155/2007/38637.

[27] Wolff, D. (2008). Detecting Bird Songs via Periodic Structures: A Robust Pattern Recognition Approach to Unsupervised Animal Monitoring. To be found at <http://www-mmbd.iai.uni-bonn.de/download/Diplomarbeiten/Diplomarbeit_Daniel_Wolff.pdf> [quoted 01.06.2011].

[28] Kwan, C., Ho, K., Mei, G., Li, Y., Ren, Z., Xu, R., . . . Stanford, V. (2006). An automated acoustic system to monitor and classify birds. EURASIP Journal on Applied Signal Processing, 2006, 52-52.

[29] Tsai, W.-H., Xu, Y.-Y., & Lin, W.-C. (2014). Bird Species Identification Based on Timbre and Pitch Features of Their Vocalization. Journal of Information Science and Engineering, 30(6), 1927-1944.

[30] Stowell, D., & Plumbley, M. D. (2014). Automatic large-scale classification of bird sounds is strongly improved by unsupervised feature learning. arXiv preprint arXiv:1405.6524.

[31] Selin, A., Turunen, J., & Tanttu, J. T. (2007). Wavelets in recognition of bird sounds. EURASIP Journal on Applied Signal Processing, 2007(1), 141-141.

[32] Bardeli, R. (2009). Similarity search in animal sound databases. IEEE Transactions on Multimedia, 11(1), 68-76.

[33] Briggs, F., Lakshminarayanan, B., Neal, L., Fern, X. Z., Raich, R., Hadley, S. J., . . . Betts, M. G. (2012). Acoustic classification of multiple simultaneous bird species: A multi-instance multi-label approach. The Journal of the Acoustical Society of America, 131(6), 4640-4650.
[34] Dong, X., Xie, J., Towsey, M., Zhang, J., & Roe, P. (2015). Generalised features for bird vocalisation retrieval in acoustic recordings. Paper presented at the Multimedia Signal Processing (MMSP), 2015 IEEE 17th International Workshop on.

[35] Arriaga, J. G., Cody, M. L., Vallejo, E. E., & Taylor, C. E. (2015). Bird-DB: A database for annotated bird song sequences. Ecological Informatics, 27, 21-25.