IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Bioacoustics Data Analysis – A Taxonomy, Survey and Open Challenges

RAMA RAO KVSN [1], JAMES MONTGOMERY[1] (Member, IEEE), SAURABH GARG [1] (Member, IEEE) and MICHAEL CHARLESTON[2]

[1]School of Technology Engineering and Design, University of Tasmania, Hobart, Australia
[2]School of Natural Sciences, University of Tasmania, Hobart, Australia

Corresponding author: RamaRao KVSN (e-mail: ramarao.kaluri@utas.edu.au).

This research was supported by an Australian Government Research Training Program (RTP) Scholarship.

ABSTRACT Biodiversity monitoring has become a critical task for governments and ecological research agencies for reducing significant loss of animal species. Existing monitoring methods are time-intensive and techniques such as tagging are also invasive and may adversely affect animals. Bioacoustics based monitoring is becoming an increasingly prominent non-invasive method, involving the passive recording of animal sounds. Bioacoustics analysis can provide deep insights into key environmental integrity issues such as biodiversity, density of individuals and present or absence of species. However, analysing environmental recordings is not a trivial task. In last decade several researchers have tried to apply machine learning methods to automatically extract insights from these recordings. To help current researchers and identify research gaps, this paper aims to summarise and classify these works in the form of a taxonomy of the various bioacoustics applications and analysis approaches. We also present a comprehensive survey of bioacoustics data analysis approaches with an emphasis on bird species identification. The survey first identifies common processing steps to analyse bioacoustics data. As bioacoustics monitoring has grown, so does the volume of raw acoustic data that must be processed. Accordingly, this survey examines how bioacoustics analysis techniques can be scaled to work with big data. We conclude with a review of open challenges in the bioacoustics domain, such as multiple species recognition, call interference and automatic selection of detectors.

INDEX TERMS bioacoustics, biodiversity, density estimation, species identification, features, syllables

I. INTRODUCTION

THROUGHOUT human history people have developed both mutually supporting and conflicting relationships with the natural world. Significant animal species loss has been observed in recent decades due to habitat destruction, which puts environmental integrity and biodiversity at risk [1], [2]. Imbalanced biodiversity may result in undesirable effects such as change in climatic conditions and pollution. Hence, assessing biodiversity and monitoring of individual species by ecologists and zoologists is increasingly important. However, zoologist's species estimates can vary significantly, with a high order of variation between their assessments [3]. Moreover, monitoring methods such as marking and tagging of animals, which primarily depend on visual characteristics, are invasive and may harm animals. Hence, to overcome the difficulties of invasive methods and manual data collection complications leading to unproductive analysis, there is need for a sophisticated non-invasive approach for monitoring animals. As animals indicate and communicate their presence through sounds, sound monitoring is a suitable non-invasive method. The sounds generated by animals can be used to understand their distribution and behaviour. The science related to studying sound is known as acoustics, and the subfield related to the study of biologically-produced sound is *bioacoustics*. Acoustic sounds can be recorded with ease, played, synthesized and analysed to recognize animal communication [4]. Acoustic monitoring reduces the need for invasive survey techniques [5] and offers a way to monitor remote locations in a cost effective manner. Moreover, acoustic signals provide significant information on environmental, seasonal and climatic effects on species [6].

Bioacoustics is used in a number of ecological applications including biodiversity assessment, density estimation and species identification as well as learning about tempo-spatial behaviour, ecology and communication. In order to obtain

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2978547, IEEE Access

significant information about these application areas using bioacoustics, a common series of processing steps is followed, and considerable research has been conducted in various aspects of this process flow. To the best of our knowledge, there is no broad survey in this domain discussing bioacoustic process flow activities and their application to particular environmental monitoring areas. Au and Hastings [7] and Zimmer [8] have presented a detailed discussion with a focus on marine bioacoustics. This paper reviews the current body of work in the field of bioacoustics, with an emphasis on bird acoustics and the problem of species identification. Based on this survey open challenges and research opportunities are identified. The key contributions of the paper are summarized as follows:

- Presents a holistic review of state-of-the-art research works in the bioacoustics domain.
- Summarizes bioacoustics workflow steps and comprehensively presents relevant detailed studies about each step, as many previous researchers have focused only on either a single step or single animal.
- Presents a taxonomy of bioacoustics applications such as density estimation and biodiversity, with a special focus on species identification.
- Reviews big data analytics for bioacoustics data to handle the massive data generated from different sources.
- Describes different bioacoustics specific software, identifying open issues and research targets for future bioacoustics research.

The review should benefit both ecologists looking to use bioacoustics analysis and researchers investigating analysis and processing techniques for this domain.

This paper is organized as follows. Section II formally introduces bioacoustics terms and important applications of bioacoustics analyses. Section III discusses collection technologies and pre-processing methods, while Section IV discusses different audio features that are widely used in bioacoustics analysis studies. We present a taxonomy of bioacoustics analysis techniques in Section V, summarising work related to density estimation, biodiversity and species identification. Section VI presents a discussion on several techniques that have been used to identify bird sounds automatically while Section VII discusses other animal species identification methods in order to provide a comprehensive view. Section VIII discusses different work that has been designed to deal with big data and various bioacoustics software that is available. Section IX discusses open research challenges in the bioacoustics field.

II. BIOACOUSTICS: OVERVIEW

Bioacoustics is a formal study that involves the production, transmission and reception of sound [9], [10] to gain insights about animal relationships with atmosphere. Before proceeding to discuss bioacoustics analysis in detail, this section provides a background to biacoustics:~~such as~~ its origins, applications, call categories and activities.

A. BACKGROUND

Bioacoustics has origins during the second world war [11] to monitor fish sounds. During World War II, acoustic devices like sonar were developed to detect acoustics in the ocean. The detected fish acoustics are primarily used for sensing, discovering and catching fish. At night time, using these acoustics, fisherman are able to detect the type of fish and their abundance. Dang and Andrews [11] also described work carried out in England and Russia on fish species identification. Initially it was done by correlating the response with swim bladder harmonics. Knowledge of fish school shape was used at a later stage to identify species. Based on this work, Dang and Andrews [11] concluded that:

- The sound of one species would be qualitatively different from another.
- Sounds may vary depending on the season and behavioural contexts.

Riede et al. [12] also supported these findings by experimenting on dog barking in two different behavioural contexts, healthy and unhealthy dogs. Barks of the dogs in these two different contexts are recorded. Riede et al. [12] established that there is considerable variation of sound in two different contexts. Hence, obtaining sounds of animals in different contexts, seasons and species may be useful for several bioacoustics analysis applications.

B. BIOACOUSTICS APPLICATIONS

Acoustics can be used to study and analyse several key phenomena such as:

Biodiversity: This concerns knowing about abundance and evenness of animal species living in particular surroundings or on the earth. Assessing and maintaining biodiversity will help us to maintain ecosystems, which is as essential as air and water. Bioacoustics serves as an important tool to assess biodiversity.

Species Identification: Species can be termed as a cluster of animals with similar characteristics. Each species can be classified based on biological characteristics specific to it. Bioacoustics can be applied to this field of research to assist the classification into genus, species and individual levels.

Density Estimation (abundance): In a defined location, some species may be commonly found but some may be rare. Using bioacoustics, we can find relative abundance such as the percentage of organisms found at a particular location.

Migrations: Most species relocate from one place to another place due to climatic conditions, foraging or for breeding. Using bioacoustics, we can monitor such relocation patterns.

Cryptic Species Detection: Some species which are not native to the surroundings, difficult to observe and rare species can also be detected using bioacoustics.

Animal Communication: This can be defined as exchange of information between sender and receiver animal that will affect and trigger change in the emotions and behaviours of the receiver animal.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2978547, IEEE Access

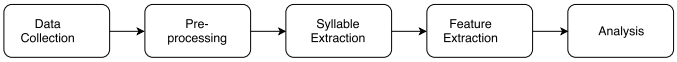

FIGURE 1. Common bioacoustics process flow.

C. TYPES OF ACOUSTICS

In the bioacoustics domain, typically used terms are active/passive acoustics and calls/songs. This section introduces such bioacoustic jargon.

1) Active versus Passive

Acoustic detection can be categorized into active and passive [13]. Active acoustics involves production of sound by a device that converts variations into electrical signals. To convert signals, transducer is typically used with several measures in active acoustics to understand or identify the species population. Species identification with this method requires more information. On the other hand, passive acoustics encompasses obtaining data, listening sounds and analysis.

2) Calls versus Songs

Bird vocalization activity is key for exchange of information. Each bird sound is different and varies in terms of pitch, rhythm and pattern. For vocalization, birds have evolved a unique primary sound-producing organ named the syrinx [14]. Functionally, the syrinx is very similar to vocal cords in human. In addition to the syrinx, bird sound is coordinated by several groups of muscles such as respiratory apparatus and upper vocal tract [15]. More detailed discussion of syrinx and call structure can be found in Bolhuis et al. [16] and Kershenbaum et al. [17].

The sounds made by birds can be broadly divided into two categories, Calls and Songs [18].

1) *Bird Calls*: Calls are very short and simple sounds. Calls fall into several categories such as mating calls, reproduction calls, feeding calls, distress calls, and excitement calls. These calls can be used to perform analytic activities such as species identification. For instance, Glaw and Vences [19] considered mating calls to identify species.
2) *Bird Songs*: Songs are of longer duration sounds than calls. Male and female birds produce songs of approximately the same length. Bird song can be best heard during Spring and are most likely to sing at dawn, as it is a favourable time for several activities [20]. These songs are to attract other birds or for defence.

In order to utilize these calls or songs for any bioacoustics application, several steps are involved constituting a process flow.

D. BIOACOUSTICS ANALYSIS PROCESS FLOW

Many bioacoustics analyses comprise the steps illustrated in Figure 1:

- *Data Collection*: This activity involves recording animal sounds in the field and obtaining the recorded audio data in digital format.
- *Pre-processing*: The collected raw audio data may contain noise. This step may be applied to remove noise and prepare the collected audio data ready for analysis.
- *Syllable Extraction*: These are like fingerprints of the audio signal that are analogous to phonetics for humans, which capture important segments of the signal.
- *Feature Extraction*: This activity captures the significant/informative properties of the signal for better understanding.
- *Analysis*: Various kinds of analysis can be performed using the above features.

A detailed discussion on these steps will be presented in the subsequent sections.

III. DATA COLLECTION AND PREPARATION

Data collection and preparation is the first bioacoustics analysis activity. It involves gathering the data and further making it ready for analysis through several pre-processing steps. This section primarily focuses on two aspects, recording and pre-processing.

A. FIELD RECORDINGS AND VISUALIZATION

For recording animal sounds, two devices are used, a microphone and a sound recorder. A detailed discussion on these two devices is presented below.

1) Microphones

Microphones convert sound into electrical signals with the help of a transducer. In addition to the transducer, several other parameters such as efficiency, self-noise, polar pattern and frequency also play a major role in recording quality. There are a variety of microphones including dynamic microphones, piezoelectric transducers, condenser microphones, solid-dielectric microphones, electret (capacitor-based)microphones, hydrophones and directional microphones [21].

1) *Dynamic Microphones*: These are robust, reliable and are particularly suitable for loud environments. A mechanical element in the device produces power by the process of electromagnetic induction.
2) *Piezo-electric Transducers*: These are particularly useful in detecting high frequencies(e.g., bats) and as microphones for musical instruments. Whenever acoustic waves are created, these transducers produce power which make them more suitable to detect ultrasonic sound.
3) *Condenser microphones*: These microphones have a diaphragm whose movement changes the capacitance in a condenser. This change in the condenser is transformed to electrical energy.
4) *Solid-dielectric microphones*: These have to be charged by external supply or by using internal batteries. These are best suited for bat detection but mechanical membranes are subject to humidity changes, giving scope for noise introduction.

5) *Electret microphones*: These consume low power requirements, being pre-charged. The recently introduced Micro-Electrical-Mechanical System (MEMS) comes under this category. These devices are small and very sensitive.
6) *Hydrophones*: These are particularly useful for detecting underwater sounds. The sensitivity issue in electret microphones is addressed in these by utilizing a piezoelectric element, which produces voltage when compressed by sound waves. These are omni-directional single transducer microphones, covering frequency range up to 100kHz. To monitor larger regions, an array of hydrophones can be used. A pack of hydrophones coupled with amplifiers and transmitter in a pressure resistant container is capable of recording even underwater sounds.
7) *Directional microphones*: These microphones are configured to focus on the sound originating from one direction or a single source. Directional microphone recording can be for two purposes. One purpose is for ambiance recording, which is used to understand detailed audio characteristics of a specific environment. All sounds in that particular environment will be recorded. The other purpose is for species recording, where sound is selectively recorded from a single source avoiding noise from other sources. For this purpose devices such as parabolic receivers are preferred, because they reduce unwanted noises coming from other directions and focus only on sound coming from the desired object.

2) Sound Recorders

As discussed earlier, Microphones convert sounds to electrical signals and sound recorders subsequently record these generated electrical signals. Sound recorders can be of two types: analog or digital. Analog recorders such as cassettes suffer more signal degradation. Digital recorders have replaced analog by overcoming all the major disadvantages. Digital recorders usually store the audio files in 'uncompressed wave (.wav) format'. Such `.wav` files are stored digitally and the storage volume depends on available storage capacity. Hence long-term acoustic monitoring programs have to be planned properly with an automated recording system with sufficient systems deployed backed up with enough power and storage facilities. More detailed discussion on bioacoustics prerequisites, precautions while recording, hardware, power requirements and other information has been outlined by the Bioacoustic Diversity group [22]. The recorded sounds can be archived, similarly to any other digital data on a CD, DVD or BluRay disc, The recorded sound file documentation should be accompanied with metadata information such as locality, temperature and frequency.

Once audio data is obtained in digital format, it can be visualized using different tools.

3) Sound Visualization

Visualizing acoustic properties such as temporal and spectral characteristics is important for analysis. To visualize, play and edit audio recordings many tools are available. Two commonly used visualizations are oscillograms and spectrograms.

1) *Oscillogram*: This is a basic graphic display unit in terms of voltages and amplitudes revealing temporal changes of sound. It exposes the signal frequency composition data at a particular moment.
2) *Spectrogram*: This is the most widely used tool in bioacoustics for visualization as it visualizes in three dimensions: frequency, time and amplitude. Visualization is constructed by Fourier decomposition.

Once data are collected and optionally visualized, the next process is often to remove noise from the audio.

B. NOISE REMOVAL

In the context of bioacoustics, noise can be interpreted as an unwanted and unpleasant sound which is added to the desired sound involuntarily. These unwanted sounds may originate from several sources including airplanes, wind and rain [23]. Such noise should be processed and removed. To process noisy signals, there are several standard methods that can be applied to audio signals. This section discusses such noise processing methods.

1) Wiener filter methods

Lim et al. [24] pointed out that, relative to phase, short-time spectral amplitude is important for speech quality. Grounded on the concept of short-time spectral amplitude, noise removal techniques can be broadly classified into two categories [24]. The first category encompasses using the process of spectral subtraction for explicit estimation of the short-time spectral magnitude, which will remove noise and enhance the signal. The second category of speech signal enhancement is based on Wiener filtering. In this method, several filters are initially obtained from the degraded speech. Among these filters, an optimal filter will be found. To estimate the noise, this optimal filter is applied in the time or frequency domain. Since zero-phase frequency response is computed using the Wiener filter, spectral amplitude will be enhanced keeping the filtered speech phase similar to that of degraded speech.

2) Signal Subspace Methods

Ephraim and Van Trees has proposed the subspace method [25]. The basic principle lies in decomposing the noisy signal vector space into two subspaces, a noise subspace and a signal-plus-noise subspace, by using Karhunen-Loeve transform (KLT). On these sub spaces, linear estimation is applied on a frame by frame by using two criteria. The first estimator criterion is signal distortion power minimization on an average residual noise power. This is called a time domain constrained estimator. The second criterion

estimator is called spectral domain constrained estimator, again focused on signal distortion power minimization, but with a condition that the residual noise should be less than a threshold. KLT is applied to the two estimators. The KLT decomposition estimator is extracted by computing an eigen decomposition of the Toeplitz covariance estimate of noisy vector. By nullifying the KLT components of noise subspace, signal subspace KLT components are modified by using either of the estimators. While the spectral subtraction method uses DLT, the subspace method uses KLT for decomposing the noisy signal vector space. Ephraim and Van Trees [25] pointed out that sub space method performance was better than spectral subtraction method with less distortions, no residual noise, and noise being additive and white. These authors concluded that coloured noise could also be whitened. Lev-Ari and Ephraim [26] have extended the signal subspace approach to the colored noise process by proposing explicit forms of time and spectral domain estimators.

3) Statistical Methods

1) *Minimum Mean Square Error (MMSE) method*: Ephraim and Malah [27] stated that neither spectral subtraction nor Weiner filter are optimal spectral amplitude estimators. From noisy observations, these authors proposed to derive an optimal short time spectral amplitude (STSA) estimator by utilizing MMSE. To derive an MMSE STSA estimator, Fourier expansion coefficients of speech and noise probability distributions must be known. To obtain them, using either a statistical model or probability distribution measurement can be used. Since the probability distributions of samples is time varying, it would be complicated and impractical to measure them. Hence a statistical model utilizing asymptotic properties of Fourier expansion coefficients has been used by the authors. These Fourier coefficients are presumed to be independent Gaussian random variables. This estimator also considers the presence of speech signals in noisy observations. Since the complex exponential of MMSE does not affect STSA estimation, derivation is proposed from the noisy signal.
2) *Unified Approach*: Model based methods have become popular in speech enhancement. Ephraim [28] proposed a unified statistical framework to address speech enhancement issues such as quality, intelligibility, robustness of speech coders and recognition systems to noise. Vector quantization and hidden markov models (HMM) are two powerful statistical techniques that have found good application in speech enhancement domain in terms of signal estimation and classification of noisy signals. To design speech enhancement systems with HMM, the type of HMM used for signal and noise and distortion measures plays a key role. Ephraim [28] further discusses how HMMs can be extended in several ways to resolve noise problems.

However, all these methods (Weiner, signal sub space, statistical) assume that noise can be estimated from the first few frames determined by voice activity detectors (VAD) [23]. These authors discussed that it is not practical to assume that first few frames only contain noise. In bioacoustics recordings, bird calls without any gap in the first few frames are found in several recordings. In such situations noise estimation algorithms based on such an assumption will suffer. In addition, VADs do not work in low signal-noise-ratio (SNR). To address this, Cai et al.proposed a novel noise reduction algorithm which can estimate noise from every frame [23]. This algorithm also eliminates VAD.

After noise removal, next step is to extract features. Different audio features that are used by various studies in bioacoustic analyses are presented in the next section.

IV. AUDIO FEATURES USED FOR BIOACOUSTICS ANALYSIS

This section mainly discusses three different types of features that were used in bioacoustics studies during the process of analysis such as acoustic features, syllables and other features.

A. ACOUSTIC FEATURES

1) Spectral and Temporal

These are the features which are extracted from frequency and time plots of the signal using oscillograms or spectrograms. Commonly used features are Start Frequency, End Frequency, Maximum Frequency, Minimum Frequency, Middle Frequency, Maximum Intensity frequency and Duration.

2) Mel Frequency Cepstral Coefficients (MFCC)

MFCCs are extensively applied in human speech identification. Human and certain animal auditory systems possess similarities and their basilar membrane physical characteristics yield similar frequency characteristics [29]. Cai et al. [23] also discussed several similarities between human and bird with regard to hearing, vocal tract and auditory processing. These similarities enable MFCCs to be the useful for species identification across a diverse set of animals such as frogs, crickets [30]–[32] and birds [31], [33].

Neither human nor animal perception scale is linear [30], [31] with critical band frequencies very much influenced by energies. To resolve such linearity disturbances, MFCCs streamline the frequencies across the continuum. Further, MFCCs offer several advantages: they are simple, robust and computationally efficient. They have good accuracy, computation not requiring any performance tuning and exceptional recognition rates irrespective of call type [30], [32]. MFCC computation typically involves these steps:

1) Audio signal is split into short frames typically ranging between 20 ms to 30 ms
2) Since spectral content does not exist in every frame, segmentation irregularities exist which are rectified by

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2978547, IEEE Access

multiplying each frame with a Hamming window. This will minimize signal discontinuities and create frame overlapping with good resolution.
3) On this Hamming windowed signala Fourier transform is applied to calculate the power spectrum and identify the frames of concern.
4) The resulting spectrum is mapped to the mel-frequency scale.
5) Subsequently, a filter bank is applied to mel-scale mapped spectrum to identify each frequency region energy.
6) Logarithm is applied obtain energy log for each filter.
7) These log energies are transformed to cepstral domain by applying a discrete cosine transform(DCT).
8) The DCT will result in a 64-dimensional cepstral feature vector referred to as MFCC.

Classically, only the lower 12 DCT values out of 64 are retained to reduce the complexity and increase accuracy.

3) Entropy Based

To identify frog species, Han et al. [34] applied a hybrid spectral entropy method for extracting several features such as Spectral centroid, Shannon entropy and Renvi entropy. To enhance species identification, this study specifically used two types of entropies to create a hybrid feature system.

- Spectral centroid: This is very useful in pattern recognition. It represents the centre point of the spectrum where the sound is bright.
- Shannon entropy: This is used to measure the degree of the signal which quantifies the richness of sound. After the species richness index, this is the most preferred index as it increases the evenness in the relative richness in a habitat. This measure characterizes acoustic diversity even at time units with low probability amplitude mass function. Shannon entropy serves as an information content richness measure which quantifies the diversity.
- Renyi entropy: This has found its applications extensively in the fields of signal processing, data mining, classification and segmentation. In information theory, Renyi entropy is an estimate of the noise during signal transfer. In this study, authors used Renyi entropy to identify the noise content and its complexity.

Dayou et al. [35] has also used Shannon entropy, Renyi entropy with Tsallis entropy as features. Tsallis entropy, also referred to as q-entropy, is another generalization of Shannon entropy. This study utilized Tsallis entropy to measure the signal complexity.

B. SYLLABLES

The process of dividing audio signal into fragments is called segmentation. Bird vocalization hierarchies can be divided into notes, syllables, phrases and calls. Among these, for bird species recognition, syllables are ideal, as phrases and calls have more variations in terms of region and individuality [36]. Syllables are brief sounds which the species produces with the lungs in a single blow of air and hence can be considered similar to phonetic sounds in humans [29]. Somervuo et al. [37] termed syllables as organized sequence of brief sounds from a species-specific vocabulary.

The process of syllable segmentation primarily involves computing a threshold value of the signal. Using a threshold value and signal energy level, the syllables can be identified. The start of a syllable is recognized as the point where signal energy first exceeds the threshold and where the energy drops it is considered as the end point. The signal between the start and end point is referred as the syllable [30]. The number of syllables varies between 12 to 96 according to species [34].

The predominant way to perform this syllable identification is through spectrogram analysis and manual labeling. However, manual process is time intensive and subjective. Towards automating this, Harma [38] used a sinusoidal parameterization model to extract syllables. Each syllable is parameterized by the sinusoidal model by representing its frequency and amplitude trajectories. To improve recognition accuracy and characterize syllable harmonics, Somervuo and Harma [39] introduced four supplementary parameters. Kogan and Margoliash [33] have experimented with syllable extraction using HMM.

As discussed earlier, the bird acoustics hierarchy contains notes at the lowest level and syllables at the next higher level. Several works have been done based on syllables and their feature extraction. However, at the lower level below syllables there are notes, which may also be used for classification. Graciarena et al. [40] has created bird species models from note sequences. Note sequence models were produced using unsupervised clustering techniques. The note models are trained by performing vector quantization of acoustic features and two-pass index alignment training models. To create bird species models, a note n-gram model with support vector machine (SVM) is used to capture note sequence statistics. When compared with Guassian mixture models (GMM), note model is superior in several aspects such as multi-scale modeling, syllable modeling and modeling longer time dependencies. The features are obtained by computing note loop lattices, extracting n-gram statistics and normalizing them. SVM is used for training.

It is evident that any speech recognition system depends on the vocabulary and grammar of the language. Bird acoustics also have a vocabulary and grammar. These structured brief sounds are known as syllables. Lakshminarayanan et al. [41] developed probabilistic models to process these syllables for species identification. Inspired by document classification, syllables are treated as words and species as topic. A probabilistic model is built based on document classification. Lakshminarayanan et al. [41] introduced the Independent Syllable model in which the syllable frames are treated by two models, the Independent Frame Independent Syllable (IFIS) model and Markov Chain Frame Independent Syllable (MCFIS) model. Then a maximum *aposteriori* rule is derived for each model. These authors concluded through numerical evaluation that classifier performance is competitive.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2978547, IEEE Access

From the above, we can say that segmenting into pulses or syllables has become essential for most ecological machine learning problems. However, audio from recorders contains more noise, which compramise the classification accuracy. Hence for noisy signals syllable extraction, Neal et al. [42] proposed a supervised model for audio segmentation which transforms the input signals into a spectrogram representation and creates a binary masked label for each time-frequency slot as either bird sound or noise. This activity extracts individual song syllables of the bird, even when there are overlapping syllables. The binary mask is evaluated with manually-labelled ground truth using a metric of true positive rate (TPR) versus false positive rate (FPR), and a metric of energy-weighted TPR versus FPR. Experiments indicate that the method achieves 90.5% TPR for the first metric, and 93.6% TPR for the second, which suggests that the proposed method performs better than segmentation by energy thresholding method.

Most of the research in syllable extraction is call dependent. Cheng et al. [31] developed call independent identification of 10 passerine (perching birds) species using four machine learning methods, namely radial basis function networks (RBFN, a special kind of artificial neural network (ANN)), SVM, HMM and GMM. The models are trained using any type of syllables and used on the same or different types.

In the Multi-Instance Multi Label (MIML) framework developed by Briggs et al. [43], the classified objects constitute bags (audio recording) of instances (syllables) and class labels (set of species). This bag generator algorithm converts recordings into bag of instances.

Colonna et al. [44] presented a technique to treat syllables in real time by simply storing time series statistics instead of using sliding windows. The beginning and ending of each syllable is detected and the noise part is eliminated. Incremental versions of signal energy (E) and zero crossing rate (ZCR) equations are computed. The power of these two equations is that they remember past decisions even when new samples arrive. They also use few parameters thus consuming less memory and processing constraints. However they suffer from a high false positive rate. To handle the problem of false positives, a mode filter has been utilized which considers the most frequently observed value as precise. Further, the k-NN classifier is used for classification. Experiments demonstrated 37% improvement and is well suited for wireless sensor networks.

Following syllable identification, the next step is to extract various features from these syllables or directly from the signal. This is discussed in the next section.

C. OTHER FEATURES

1) Linear Predictive Coding (LPC)

LPC is robust in providing approximation of vocal track spectral envelope based on past samples. LPC exploits the relationship by converting them as a set of coefficients, called LPC coefficients. LPC coefficients can be computed by using Durbins method [45] and autocorrelation analysis.

2) Code Book of Frame Level Features

Any bird classification algorithm begins with extracting audio signal features. Such features that are extracted belong to single frame which are very short segments. However, Briggs et al. [46] pointed out that because single frame features are insufficient, it would be advantageous to extract features from multiple frames to represent sound in order to apply them on several standard classification algorithms. Features extracted from these multiple frames will form a fixed-length feature vector. From this vector, a common approach is to take an average of interesting frames. However, Briggs et al. [46] proposed a codebook concept, inspired by similar work in the fields of computer vision and music genre classification. In the codebook concept, the features from all frames are aggregated as a bag of code words through clustering.

3) MARSYAS Framework

The MARSYAS (Music Analysis, Retrieval and Synthesis for Audio Signals) framework was developed by Tzanetakis [47] for use in music information retrieval applications. This feature is well applied in music genre classification research work. Lopes et al. [48] has first made use of this feature set in the domain of bird species recognition. More details about this feature set can be found at the MARSYAS portal [47].

Once the features are extracted, the next step is to perform analysis on various bio-acoustic applications.

V. A TAXONOMY OF BIOACOUSTICS ANALYSES

Bioacoustics analysis provides us with deep insights into key environmental integrity issues such as biodiversity, density estimation and species identification. Each bioacoustics application has diverse analysis techniques and methods. For a comprehensive outlook of the techniques related to each bioacoustics application, we present a taxonomy illustrated in the Figure 2. This section focuses on presenting the details of taxonomy and work on Density Estimation, biodiversity and species identification.

A. DENSITY ESTIMATION

Density or abundance estimation is an approach to assess bird population size. This helps in bird conservation as it can assist in evaluating the impact of pollution and habitat loss on birds. After the 2002 World Summit on Sustainable Development, it has been globally agreed to reduce the extinction of animal species. This activated the investigation of approaches for monitoring bird population trends. To monitor population trends, there are two models, namely open and closed. Closed population models assume that births, deaths and immigrations do not occur while open population models relax these assumptions. Seber et al. [49] discussed several estimation methods for closed population and open populations. For closed population methods such as absolute density and relative density were proposed while Jolly-Seber

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2978547, IEEE Access

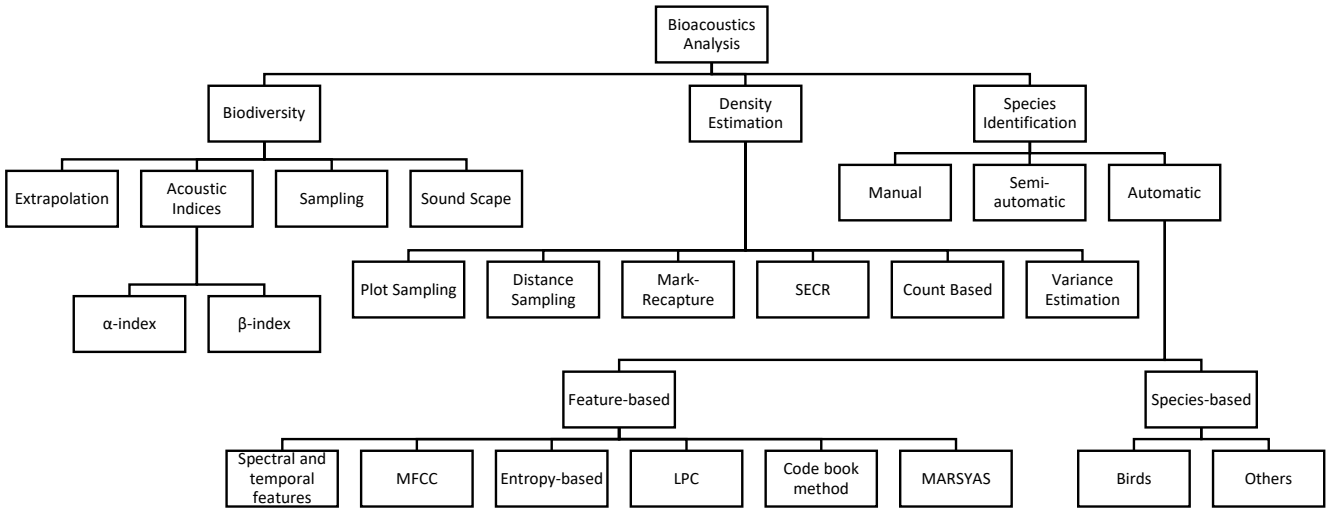

FIGURE 2. Analyses Taxonomy

method, regression and several methods are discussed for open population. Williams et al. [50] extensively discussed a framework for modelling, estimation and management of animal population. Some of the most frequently used density estimation methods are discussed below.

1) Plot sampling

Although a typical objective is to estimate total bird population, it is more practical to estimate the population in small areas and use these small area samples to estimate the total. Plot Sampling is one such method that estimates population based on samples from random plots. Reynolds et al. [51] developed a circular plot method that estimates the bird density in a given area. In their circular plot method, stations are established at scattered locations to monitor birds. Whenever a bird sound is heard, it is counted. The distance from each station to where they were found first is estimated. For each species, a plot is drawn for the areas where the observed bird species begin to decline. The density count is a sum of individual bird counts found within a specified radius.

Dawson et al. [52] discussed several issues with the plot methods and further pointed out that counting of one or other bird species and estimate their density may not be feasible until counts on sample plots are available and complete. This introduces reliability and detectability challenges which are addressed by using distance vector approach.

2) Distance sampling

Detecting an animal relies primarily on the distance between animal and sensor. Distance sampling methods use distance as a key parameter to estimate the probability of detection as a function of distance. To have a broad area coverage, a systematic design with a large number of distances will be recorded. Further, a detection function $g(y)$ should be defined to calculate animal detection probability by using the distance point values. Rosenstock et al. [53] made use of distance sampling method to count land birds. This method overcomes many drawbacks of index count. Burnham et al. [54] utilized a Fourier series method to estimate bird abundance.

Hiby and Ward [55] discussed a variation to distance sampling which is cue counting. In cue counting, cues produced by animals are counted rather direct animal observations. Cue counting was initially used in estimating whale density by using whale blows as cues. Since it is difficult to differentiate individual whale cues, cue density is estimated as the number of whale blows per unit area per unit time. The obtained value of cue density is divided by an estimated average cue rate to obtain whale density. Buckland and Handel [56] extended this technique to birds where a call or song from a bird is considered as a cue. Cue density is the number of bird calls per unit area per unit time and cue rate is the average number of calls per unit time. Using these two, bird density can be estimated.

3) Count based methods

Though spatial replication themes are generally used to count animals, it is difficult to estimate population size from them as they generate sparse count data. Royle [57] developed an N-mixture model to overcome this and estimate population from such sparse data. For a population size of N specific to the site, viewed as distributed poisson independent random variables. Carol and Lombard have developed estimators but they have limitations. This model has attempted to address this issue.

4) Variance estimation

To draw meaningful inferences, precision is considered to be an important parameter for density estimation. Precision estimation should be reliable. To obtain reliable precision measure, variance estimation method will be used. There are two approaches for variance estimation namely analytic

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2978547, IEEE Access

variance estimation and bootstrap variance estimation [58]. The first approach is analytic variance, which is estimated in terms of random (e.g. detection probability) and constant (e.g. number of sensors, recording time) components. The second approach, bootstrap variance estimation uses resampling of several units such as sensors to generate a boot strap dataset which is used to obtain estimate of density. The resampling is repeated many times which yields a density estimation set. Using the variance of these set of estimates, variance of original estimator can be approximated.

B. BIODIVERSITY

Conservation of biodiversity in a general sense aims at maintaining a balance of a large variety of interdependent biological creatures. These dependencies mean losing that balance threatens environmental health as a whole. One of the main threats to global biodiversity is tropical habitat loss due to human activities. Laiolo [4] has elaborately discussed the effects of human activities on animals and on environmental acoustics due to noise pollution, habitat fragmentation, chemical pollution, direct human disturbance, hunting, introduced diseases and food supplementation. This threatens biodiversity which initiates the dire need to survey. Acoustics will be an important input to analyse the biodiversity. Below we discuss biodiversity analysis research works.

1) Traditional Measures

The Shannon Wiener statistic (H) is a species diversity index. It is used to estimate species richness. Riede [3] has made use of this statistic to estimate crickets diversity of the Amazon forest. Sound generated by crickets were recorded at Amazon Rainforest for a duration of two weeks at 10 different points daily two times. With Condenser microphones, individual songs are recorded very closely. Spectrogram has shown frequencies ranging from 0 to 10 kHz. Mostly the frequencies below 3 kHz are identified to be produced by frogs, mammals, birds etc. Within a range of 4 kHz to 9 kHz, acoustics of crickets are identified. Bird wing movement causes resonance and generates sounds at a certain pulse rate and at a narrow frequency. These two parameters (pulse rate and carrier frequency) differ species to species and hence can be used to identify species. The frequency is dense at 5 kHz to 8 kHz with up to 80 pulse rate. We can get an understanding of species and their abundance by using recordings of ith species and total recordings by using shannon-wiener statistic. Based on authors experiment results, Shannon-Wiener statistic turns out to be $H = 2.789$. This value of H is low (typically should be 5.0), which indicates the cricket diversity is low.

2) Acoustic Indices

In earlier bioacoustic analyses, all the biodiversity estimation methods relied on species richness estimation. However, Sueur et al. [59] attempted to address biodiversity estimation by considering a community level. They derived α and β indices to analyse the animal sounds. The α index indicates species count in the area whereas β index designates different species in that community. The α index is derived from Acoustic Entropy Index (H) which is obtained by the product of temporal and spectral entropies. To obtain temporal entropy, Shannon index is used and for spectral entropy short time Fourier transform is used. Acoustic Entropy Index (H) will be zero for solo tone and tend to reach 1 as the number of tones increases. Using the value of H, species richness can be estimated. To derive β index, Acoustic dissimilarity Index (D) is computed as product of temporal and spectral dissimilarities. The value of D increases if there are different species in the community. Their results demonstrated that acoustic entropy index typically ranged between 0.3 to 0.9.

Particularly during spring, certain habitat's acoustics substantially affect these H index values. To address this, Depraetere et al. [60] developed an alternative to α index which is referred as Acoustic Richness (AR). Temporal entropy and amplitude of the signal are primarily used to develop this index. AR enhanced the acoustic clarity by addressing the noise component. Tests demonstrated that it is as good as human observation. Acoustic Richness revealed that there is good acoustic activity at small forests rather than in big forests.

Sueur et al. [10] has given an overview on α and β indexes developed during several years. Their survey indicates that, up to now 21 α indexes and seven β indices were proposed. Authors recommended the simultaneous use of these indexes to obtain more complementary information to build sophisticated mathematical tools for biodiversity monitoring.

However, technological advances such as sensors and other sophisticated devices have made it possible to collect huge amounts of acoustic recordings from several remote locations resulting in a surge of audio data. Processing such enormous quantities of data is a challenge. One approach to meet this challenge is to reduce the size of the audio recordings without loosing important information. Inline with this, Towsey [61] has stated that acoustic indices offer better solution in such scenario, which substantially reduce acoustic recording information and at the same time give a comprehensive understanding of the whole audio recording. The reduction is possible, as acoustic indices summarize the energy aspects of the recordings and yield a single value. Sankupellay et al. [62],Towsey et al. [61] and Towsey [63] have made use of several indices in their works and discussed calculation and application of acoustic indices for assessing biodiversity. The acoustic indices used in their work are Average Signal Amplitude, Background Noise, Signal to Noise Ratio (SNR), Acoustic Activity, Count of Acoustic Events, Average Duration of Acoustic Events, temporal entropy, Spectral Cover, Entropy of spectral maxima, spectral entropy, Entropy of the Variance Spectrum, Spectral Diversity and Spectral Persistence. Among these, authors concluded that, spectral entropy, Spectral Diversity and Spectral Persistence are the most useful indices for obtaining optimized samples.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2978547, IEEE Access

3) Soundscape

Although biodiversity can be estimated by species richness, Celis-Murillo et al. [64] has developed a soundscape recording system (SRS) to overcome the limitations of point count methods. The SRS design consists of four microphones positioned above ground level and capturing sound by placing microphones facing each other at 90 degrees to each other. On the collected acoustic data, a closed population capture-recapture approach is used to for species richness estimation. Further jackknife estimator is used to perform interpolation on estimates. Average detection probability is calculated which in turn helps to estimate species richness. To understand dissimilarities in species composition, a jaccard similarity index is used. Results have demonstrated that, SRS has given better average detection probability and with jackard index tending to 1.0 showing good similarity among species.

4) Sampling

Most bioacoustic analysis depends ideally on the quality of recorded acoustic data. However, some recordings may be made in less tha ideal circumustances such as bad weather. This should not affect the analysis. Bardeli et al. [65] has described algorithms which work even in bad recording conditions. Sounds of the Eurasian bittern species are used by authors to test the algorithms. The sounds are found to be short segments. Short segment calls with higher frequency are not suitable to apply pattern recognition. Hence input signals are down sampled and analysed by sliding window. This down sampled signal aids in obtaining energy weighted novelty measure. The peaks in this novelty measure curve indicates calls. But this may contain noise, which can be eliminated from a bittern call by subtracting with a low pass filter. Even then, this may contain several false positive detections. An auto correlation method is applied to reduce the false positives. This method is adequate for bittern calls, but if very low calls are to be detected, false positives will be a challenge. Also since the acoustic sensors record huge volumes of data over long periods, analysing such huge volume is complex and may be subjected to false positives and negatives.

Hence, Wimmer et al. [66] have made use of sampling methods. Data were collected at four locations for a period of five days by placing sensors at the center of each site. Surveyors critically analysed the calls manually and annotated. In the spectrogram, a specific marking is made and a tag is assigned to this selection. To simplify, the recordings are split into one minute segments. Five different samples (full day-random, dawn, dusk, dawn+dusk, systematic) are collected from these one minute segments. Samples drawn during the dawn period shown high detection rate, followed by dawn+dusk period samples. This sampling method can reduce the manual analysis burden. But removing noise and cryptic vocalizations will ease the analysis. Authors concluded that strengths of both manual and automated analysis makes biodiversity more feasible.

Thus far, we have discussed two important aspects of bioacoustics applications: density estimation and biodiversity. Another significant application area of bioacoustics is species identification, which is discussed in the next section.

C. SPECIES IDENTIFICATION

There are three broad categories of approaches for species identification: manual, automatic and semi-automatic.

1) Manual

Until a few decades ago, only ornithologists and ecologists used to identify birds using their expertise. Based on several factors such as bird origin, colour patterns, environment context (time, weather, habitat etc.) and its behaviour, experts could identify the species of the bird. However, this method has several disadvantages such as availability of few experts who can identify confidently and observer bias due to variations in their abilities. Moreover, manual observation is limited to smaller volumes.

2) Automatic

Overcoming drawbacks of the manual method, automatic species identification has taken this initiative to next level by using several Machine Learning Algorithms powered by numerous machine learning techniques. We would present automatic species identification methods used by various studies in case of birds and other animals.

3) Semi-automatic

Semi-automatic methods consist of using both manual and automatic notions and will be of immense use in real time scenarios. In real time, acoustic sensors, which are effective in monitoring and classification would capture huge amount of data. To analyse such big data, combination of manual and automatic methods i.e. semi-automatic methods were developed. This is because manual methods are not scalable while automatic methods suffer high false positives and false negative rate. Hence, to have a right balance of manual and automatic method benefits, Truskinger et al. [67] designed a semi-automatic analysis method (also known as crowd based analysis method) to analyse big data, which is a combination of manual and automatic methods. The central notion of this method is to source complex classification task to a set of participants (crowd) and use tools at certain places to perform analysis. Authors implemented this method for rapid scanning of spectrograms. Authors ascertain that, humans can distinguish regions of interest intelligently and effectively from the spectrograms. Beyond human capability, certain tools were used to analyze further. Experiment results show that analysis is sped up by twelve times. There are several other semi-automatic methods available in addition to crowd-based.

VI. BIRD IDENTIFICATION METHODS

As discussed earlier, bird sounds can be of call or song type, and both vary from species to species. Both types of sounds

are useful for species identification. To automatically identify species based on these calls or songs, a wide variety of sophisticated approaches have been proposed in the literature, discussed below.

A. NEURAL NETWORKS

Bird vocalization hierarchies can be divided into notes, syllables, phrases and calls. Among these, syllables (sets of notes) are considered to be the best for bird species recognition, as phrases and calls have more variation by region and individuals. To work on such syllables, Selouani et al. [36] developed a model with a combination of time delay neural networks and an autoregressive (AR) version of the back propagation algorithm. This combination increases context memorization capability and pattern matching capacity. For this combination model, during training, a multilayer perceptron is trained on sounds using standard back propagation techniques. Training is iteratively repeated until the error reaches a value which is less than a user-specified threshold. When used for prediction, a value of zero indicates no match while 1 indicates a strong match against the training data. Though recurrent networks perform quite well, to overcome difficulties when events are time shifted, a delay component is introduced to the network, which enhances its pattern recognition capacity even when encountering misalignments. Results demonstrated 83% accuracy with 16% of improvement over a base ANN system. However, this method had trouble processing large datasets. Seeking to address issues of scalability, Deecke et al. [68] developed a classification model that applies dynamic time-warping and an adaptive resonance theory (ART) neural network to categorize dolphins. This method classified the species as well as humans and was found to be effective in the case of large dataset analysis.

Juang and Chen [69] mentioned that bird sounds follow temporal patterns, which indicates the output not only depends on present input but also on past and future inputs. In such cases, rather than simply applying a back propagation feed forward neural network which increases the delay in input, creating a large network, it is advantageous to use a recurrent network. Towards this direction, Juang and Chen proposed a prediction based singleton-type recurrent fuzzy neural network model to address scaling and complexity issues. To apply the model data are gathered and syllables extracted. LPC is applied to calculate 12 coefficients. These 12 features are used to train a singleton-type recurrent fuzzy neural network, which achieved a high recognition rate.

Cai et al. [23] investigated bird species recognition with different set of features and pre-processing methods using neural networks. The authors established sensor networks to record sounds and images. They employed voice activity detectors (VAD) to estimate the noise level from noise-only frames, although these VADs may not work well with signals having low SNR. Hence, they developed a noise reduction algorithm. Using the output of this noise reduction algorithm, MFCC features are extracted. To model changes, delta MFCC features and delta-delta MFCC features were used in combination with time delay neural networks, which combine information both from past and current frames. Thirteen dimensional MFCC and 13 dynamic features are input to Time delay Neural Networks. Their results demonstrated 98.7% accuracy.

McIlraith et al. [70] tested classification of birds using a back propagation and multivariate statistics method. Initially, a windowed Fourier analysis was applied to acquire time frequency representation of the signal. Further, as a part of pre-processing, using power spectral densities statistical estimates of the spectrum are obtained and linear predictive coding (source-filter model) is applied. Data dimensionality was reduced using a discriminant analysis technique. To extract spectral features, a fast Fourier transform was applied to LPC wave forms. Next, a back propagation classifier with 10 inputs, 12 hidden nodes, and six outputs is applied. As a second method, after parsing and normalization, ANOVA is applied. Overall, the authors found that this combination of pre-processing and statistical methods helped in feature refinement and that the back propagation method can trade-off accuracy for computational efficiency.

In summary, machine learning techniques applied to bioacoustic data have been used extensively by ecologists. The reader is also referred to the 2008 review of machine learning applied to ecology in general by Olden et al. [71].

B. DEEP LEARNING METHODS

A relatively recent extension to ANN techniques are deep learning methods, which can be superior in identifying latent features in a dataset.

Fazekas et al. [72] has developed a deep learning based model for bird song identification. The input data for this model is two-fold: data collected from habitats as well as and its meta data. To clean the acoustic data and separate noise, several pre-processing steps are applied. The cleaned data is fed into a convolution neural network with four layers. The authors concluded that frequency features are more easily distinguishable than time features. Sankupellay et al. [73] have used a 50 layer neural network named ResNet-50 which is a deep learning approach used in residual learning. The authors concluded that model was able to achieve good accuracy for shorter input calls. Xia et al. [74] presented a survey outlining various deep learning based acoustic event detection studies.

C. PROBABLISTIC MODELS

Trawicki et al. [29] developed a model to assist in identifying species decline by using HMM and MFCC. The model is trained on data representing 115 species. Syllables are extracted and joined together to generate multiple song types, then divided into frames (of 25 ms size and 10 ms step size). Using frames with a hamming window, fast Fourier transforms(FFT) are computed. FFT is fed to filter bank to obtain log amplitudes. These are used to compute 12 MFCCs using a discrete cosine transform (DCT). To these

MFCCs, delta coefficients and log energy are also added as features. With this feature vector, the HMM was used for speaker identification comprising states, transitions and output likelihood through use of a GMM. Results indicated 83% to 95% accuracy in classification.

Many of the methods consider only a particular type of song. Similar to text independent recognition in humans, Fox et al. [32] developed a call independent identification method for birds. Recordings and divided into frames, then multiplied by a Hamming window. To these windowed frames, a Fourier transform is applied and multiplied by a mel-scale filter bank. To convert the filter bank energies to the cepstral domain, a DCT is applied. This generates the feature set. Once the classifier is trained with known signals, it is used to analyze the features from unknown signals to obtain a similarity score for identifying classes. Results indicate 67% to 97% accuracy. Cheng et al. [31] has also worked on a call independent recognition system using MFCC and Gaussian Mixture Models (GMM) for classification of passerine bird species. The GMM enables representation of bird-dependent spectral shapes and supports modeling any arbitrary densities.

Somervuo et al. [37] names syllables as "an organized sequence of brief sounds from a species-specific vocabulary". Feature extraction from these syllables is critical and at the same time forms the basis for classification. The authors compared three of the best parametric models, sinusoidal modelling, a mel-cepstrum model and a descriptive parameters model, to determine which provides superior features for classification. Initially, the recording data is divided into syllables using an iterative time-domain algorithm. Using a threshold, syllables that are 15 ms apart are grouped together, forming segmented regions. The segmented regions are parameterized using the three models. In the sinusoidal model, the analysis-by-synthesis method is used for parameter estimation to find the most significant frequency per frame. A frequency domain algorithm is also used to obtain the phase and magnitude of the sinusoidal pulse. Each leading sinusoid is analysed for harmonic structure by dividing into four classes. As the result of sinusoidal modelling, a sequence of triplets and the four harmonic classes are produced. However one issue observed with this model is that it resulted in a relatively large number of parameters, so DCT was then applied to reduce the number of parameters. The second model, a mel-cepstrum model, involves applying DCT to logarithmic mel-spectrum features to reduce dimensionality. However, the computed MFCC feature vector in this case may miss pitch information. This lost information can be obtained by the third method, which is descriptive parameters model. Seven spectral and five temporal features are identified by this method. Dynamic time wraping is applied to calculate the distances between syllables. To model probability density functions, GMMs were used and trained using the standard expectation maximization algorithm. Results indicated that the best model is MFCC combined with dynamic time warping. The optimal feature set can be obtained by combination of MFCC with descriptive parameters.

Lee et al. [75] proposed a two-dimensional MFCC model which collects both static and dynamic features. Instantaneous cepstrum represents static features, while temporal variations represent dynamic features. Static features are acquired by applying a 2D-DCT to logarithmic energies of Mel-scale bandpass filters to obtain a matrix. The first 15 rows and the first five columns of the matrix are chosen as preliminary sound features of a syllable, making 74 coefficients in total. Dynamic features are also obtained in a similar manner. These two types of features are combined to obtain a large vector. The feature vector is normalized, and then further PCA is applied for dimensionality reduction. To increase accuracy of feature vector, prototype vectors are generated using vector quantization (VQ) and a Gaussian mixture model. To improve further the discriminability between various brid species, linear discriminant analysis (LDA) was employed. A nearest neighbour classifier is used for classification, with 84% classification accuracy.

D. NAIVE BAYES AND DECISION TREES

Vaca-Castaño and Rodriguez [76] developed a database model of features that can save information as attributes and entities from various sources, which can be useful for later analyses. However, several analysis tools produce a huge number of attributes from the acoustic signal. In resource-constrained applications such as bird species identification using sensor networks, ascertaining the most important attributes plays a major role. Vilches et al. [77] has explored the use of data mining techniques to the problem of attribute dimensionality reduction in the case of bird species recognition. Three algorithms such as Decision tree based ID3 and J4.8, the probabilistic classifier naive bayes and vector quantization were applied to the dataset. In the first step, vector quantization is applied, which will convert a numeric vector to quantized vector values. The quantization calculates three values such as two intermediate vectors, partition and codebook. As a next step ID3, a decision tree algorithm is applied to the quantized data which calculates the attributes' entropy. In the resultant decision tree, the most significant attributes (leaves) are used for bird classification. This is applied in the J4.8 algorithm to gain the advantages of reduced error pruning and numeric data handling. The extracted attributes are then fed into a Naive Bayes classifier to overcome some drawbacks of decision trees (they may be unstable or complex). Results indicated that J4.8 algorithm is accurate up to 98.39% while Naive Bayes have given slightly better performance on reduced datasets (when the number of attributes is reduced from 71 to 47).

E. HIERARCHICAL CLASSIFICATION

Silla et al. [78] discussed the relatively new method of hierarchical classification of bird species from acoustic data. A dataset from the xeno-canto library was used. For feature extraction, they used the MARSYAS framework, producing 64 features. Hierarchical classification can be achieved us-

ing three approaches, a flat classification approach, a local model hierarchical classification approach (each non-leaf node trained and top down classification), and a global model hierarchical classification approach (which predicts class in hierarchy levels). In this work, the global model hierarchical classification approach was based on Naive Bayes. Hierarchical precision, hierarchical recall and the hierarchical F-measure were used as evaluation metrics. The hierarchical F-measure obtained from the experiments indicated that this approach outperforms flat and local models and was found to be useful when class counts are high and if class hierarchy exists.

Signal detection (feature extraction) and signal characterization (classification) are the two major activities of the species identification. Acevedo et al. [79] focused on signal characterization, for which they considered three supervised machine learning algorithms, linear discriminant analysis (LDA), decision tree and SVM. A dataset of 10,061 calls were used. LDA assumes features vectors of a class follow a Gaussian distribution. Decision trees iteratively partition until a condition is met and each partition represents a class. Further pruning is applied to avoid over fitting. SVM optimizes training and function complexity. Results indicate that SVM achieved 94.5% accuracy while decision trees and LDA achieved 89% and 71% accuracies respectively.

Many classification approaches implicitly make the simplifying assumption that only a single species is present in a recording (requiring multiple passes with different classifiers to identify all species present). Briggs et al. [43] developed a multi-instance multi label (MIML) supervised classification framework, which predicts multiple species present in a recording in a single analysis pass. In MIML, the classified objects constitute bags (audio recordings) of instances (syllables) and class labels (set of species). A bag generator algorithm converts the recording into a bag of instances. Random forest is used for segmentation. The MIML classifier is applied to identify the species in the recording using three underlying classification algorithms: MIMLSVM, MIMLRBF and MIMLkNN. Experiments demonstrated 96% accuracy.

Stowell and Plumbley [80] experimented with unsupervised feature learning as a surrogate to MFCC. Usually, MFCC features are obtained by applying short-time Fourier to audio, producing a Mel Spectrum which is transformed using cepstral analysis and preserving the last 13 coefficients. Rather than transformation, Stowell and Plumbley proposed automatic feature learning. For feature learning, initially a high pass filter and normalization are applied to spectrograms. Further spectral median noise reduction is performed. Subsequently, features are derived from the dataset using PCA. For feature learning, spherical k-means is used. To reduce the feature set further, mean and standard deviation, or maximum of each feature is tested. The authors used the random forest classifier. They observed that feature learning boosted performance, especially for single-label classification tasks, but for datasets with few annotations, the performance of this model is not promising.

F. METHOD COMPARISON WORKS

A common approach in audio classification is to use the average value of features calculated across several frames of the source audio. This approach was used by the SVM approach of Fagerlund [81]. In contrast, Briggs et al. [46] represented audio features using histograms, as in the codebook approach, instead of averaged frame level features. A codebook is collection of “words" where each word is a feature vector. The experiment was conducted by segmenting the signal into frames and applying noise reduction. Frames of interest were identified and only 10% with the highest magnitude retained. Then MFCCs are calculated and histograms constructed, resulting in a 5000 dimensional feature vector. The codebook is constructed using the k-means clustering algorithm. Several frame-level features are aggregated and applied with the classifiers nearest neighbour (with L1, L2, KL (Kullback-Leibler divergence) and Hellinger distances), Interval-IID MAP classifier, and SVMs. The interval-IID model was proposed by those authors to model feature dissemination, subsequently obtaining a MAP classifier. This classifier aggregates histograms of features and uses the KL-nearest neighbour algorithm for classification. The study indicated that frame histograms are better than average frame level features, nearest neighbour classifiers using KL and Hellinger distances outperformed SVM and rather than Euclidean distances, the KL and Hellinger distances metrics are appropriate for histograms.

Kampichler et al. [82] experimented with several classification algorithms including decision trees, ANN, SVM, random forest and fuzzy classifiers. Ocellated Turkey bird acoustics were chosen as the dataset for these five techniques. They concluded that the performance of neural networks, LDA and SVM was poor, while the joint use of decision trees and random forest is highly recommended to achieve a high level of accuracy, transparency and comprehensibility.

Lopes et al. [48] has also emphasised that the choice of machine learning algorithm and features plays a major role in classification performance. The authors used the MARSYAS framework to produce feature sets, which are mostly used in automated music genre classification problems. The feature set has 64 features which comprise 12 MFCCs, and means and variances of timbral features. In comparison with Inset-Onset Interval Histogram Coefficients (IOIHC) and the sound ruler feature sets, the authors claim that the MARSYAS feature set’s performance is better. Several classifiers were trained and tested using these features: Naive Bayes; k-NN with three clusters; the decision tree classifier J4.8; an MLP neural network trained with the back-propagation momentum algorithm; and the SVM classifier using the Platts Sequential Minimization Algorithm (SMO) implementation.Results indicated that, rather than using full bird song recordings, the use of pulses (sounds with high amplitudes) significantly increases classification performance. Using these pulses, the best results are obtained with a multilayer perceptron classifier and SMO classifier, with up to 95% accuracy.

Most of the research in species identification is call dependent. Cheng et al. [83] developed call independent identification of 10 passerine species using four machine learning methods, radial basis function networks (RBFN, a special kind of ANN), SVM, HMM and GMM. For these models, MFCCs and Linear Predictive Coefficients (LPCs) are chosen as feature sets. One feature vector per frame was extracted, which contains 13 LPCs and 24 MFCCs. Using these features, training and test data were prepared separately for the species identification purpose. They concluded that the LPC feature set with HMM classifier and MFCCs feature set with SVM were the best combinations.

Table 1 summarizes the discussion on bird species identification.

VII. OTHER ANIMAL IDENTIFICATION METHODS

While a major target for bioacoustics analysis is birds, many other animals emit identifiable sounds and have been the subject of bioacoustics-based ecological studies. This section summarizes such research on bats, cetaceans, terrestrial, ground-based animals, and insects.

A. BATS

1) Multivariate Analysis Technique

Vaughan et al. [84] used multivariate analysis of echolocation call parameters to identify bat species. Echolocation calls of around 536 bats were recorded from 15 known species in Great Britain. The recorded sounds were analysed using sonographs. From each sequence of calls, either the second or, i nthe case of noisy data, penultimate call was used. Calls were characterised by six features: duration, interpulse interval, peak, start, end and centre frequencies. Multivariate analysis was used, demonstrating good results. The authors concluded that, of the six features, the most informative are the duration, start and end frequencies. In related work, Farrell et al. [85] devised a qualitative identification approach using critical call parameters max and min frequencies, linearity, and slope. In a given time pass duration, each individual bat may make series of calls which is referred to as a sequence. The approach begins by establishing a library of recordings of known species. New, unidentified calls are compared with these archived calls visually using the min and max frequencies to qualitatively classify bats. The quality of the method depends on the library of identified recordings and visualization expertise of a human operator.

2) Pattern recognition approach using synergetic classifier

Within any species there will be call variation between individuals [86]. To address this, Obrist et al. [87] developed a pattern recognition approach using a synergetic classifier. Variation between individuals may mean that individuals of one species are misclassified as another when visually inspecting spectrograms of their calls. This is because spectrograms show characteristics in the form of frequency curve. This shape or curve may be better analysed by pattern recognition software to classify the signals. Using standard equipment, 643 sequences of calls were recorded and a high pass filter applied to cut single echolocation calls, with additional processing to identify the cleanest calls producing 14354 calls suitable for analysis. Calls were characterized by duration, high and low frequencies with the help of a discriminant function. The discriminant function is repeated for various percentages and these values are transformed for comparisons. Next synergetic algorithm is implemented which combines a class of training patterns in to a feature vector, allowing it to handle high dimensions. A training base was established by selecting 20 calls from each of 26 species and testing on the database. The database was successively refined, then within species variability was tested using coefficients of variation. Discriminant Function Analysis (DFA) is used to explore the classification strength. Finally, the authors concluded that this method has the strength of processing huge data sets in addition to classification accuracy and trustworthiness.

3) Discriminant Function Analysis (DFA)

Hughes et al. [88] tested the robustness of DFA on identifying Thai bat species. Recordings were made of free flying bats and spectrograms created to extract call structure information. Calls were divided into four categories based on their structure: broadband FM calls, narrowband FM calls, long multiharmonic calls and short multiharmonic calls. One call from each individual is considered for analysis with 10 features calculated: duration, frequency of maximum energy, start, end, max, and min frequencies, frequency range, number of harmonics, interharmonic distance, and pulse interval. The DFA technique with cross-validation was applied to classify the species. For calls in the broadband FM category, DFA achieved an accuracy up to 85.9%, while it achieved 70.4% for those in the narrowband FM category. For long harmonic calls DFA shows 84.4% accuracy, and 96.7% classification accuracy in the case of short harmonic calls. Hence for all call types DFA resulted in greater than 70% classification accuracy.

4) Artificial Neural Networks (ANN)

Although the DFA method has become a highly trusted approach for classification of bat species, other machine learning algorithms have also been used. Preatoni et al. [89] evaluated four different classification methods (DFA, cluster analysis, CART, and ANN) by collecting 3-second call sequences from 126 bats. Noise is eliminated by digital filtering and the cleanest ultrasonic clicks extracted. Each click was characterised by its pulse duration, max intensity frequency, and start, end, middle, min and max frequencies. Due to multicollinearity, only four features—duration, min and max frequencies, and frequency of maximum intensity—are considered for DFA and cluster analysis methods whereas all the seven parameters were used for ANN and CART. evaluate performance of networks during training. After comparison of the four methods, the authors concluded that DFA and ANN performed well, while cluster analysis was poor for

TABLE 1. Summary of articles addressing bird species identification

Study	Technique used	Features	Dataset	Accuracy
McIlraith [70] & Card	Back Propagation and Multi Variate statistics method.	Spectral and temporal features	Manitoba birds	82–93%
Trawicki et al. [29]	Hidden markov Models	12 MFCCs, log energy, delta coefficients	Norwegian Ortolan buntings	83–95%
Selouani et al. [36]	TDNN with back propagation	Front-end features using LPC	16 of 395 New Brunswick resident species	83%
Somervuo et al. [37]	GMM	sinusoidal modelling, mel-cepstrum model and descriptive parameters	14 common North-European passerine species	54–71%
Juang, & Chen [69]	singleton-type recurrent fuzzy neural network	12 LPC coefficients	10 species from Taiwan	98.7%
Fox et al. [32]	HMM	MFCC	Songs from seven individuals from each of three passerine species	67–97%
Cheng et al. [83]	GMM	MFCC	Four passerine species	89–92%
Lee et al. [75]	GMM, vector quantization, k-NN	2D MFCC	28 bird species	84%
Graciarena et al. [40]	Note N-Gram System using SVM and GMM	note loop lattices & n-gram statistics	Nine species database from Borror Lab, Ohio State University	n/a
Vilches et al. [77]	Decision trees ID3 and J4.8, Naïve Bayes, vector quantization	71 attributes of each pulse	Three species recordings from Cornell Macaulay Library	88–90%
Cai et al. [23]	Time Delay Neural Networks	13 MFCC and 13 dynamic features	Australian Bird Calls, Subtropical Rainforests	98.7%
Acevedo et al. [79]	LDA, SVM, Decision Tree (DT)	Min and max frequencies, call duration and maximum power	12 Species from 14 montane sites in Puerto Rico	SVM 94%; LDA 89%; DT 71%
Fagerlund [81]	Decision tree with SVM at each node	MFCC and descriptive signal parameters	14 species	85–100%
Briggs et al. [46]	SVM k-NN	Codebook approach frame histograms	6 species	90%
Lakshminarayanan et al. [41]	SVM, IFIS model and MCFIS model	Mean, frequency and bandwidth	6 species recordings from the Cornell Macaulay Library.	MCFIS 90.6%; IFIS 88.3%; SVM 86.2%
Kampichler et al. [82]	Random forest (RF), ANN and SVM	44 explanatory variables	Ocellated Turkey birds	RF 100%; ANN 69%; SVM 56%
Lopes et al. [48]	Naive Bayes, kNN, decision tree, MLP neural network, SVM	12 MFCCs, timbral features	73 bird species recordings from xeno-canto library	95% with MLP and SMO
Cheng et al. [31]	RBFN, SVM, HMM and GMM	13 LPCs and 24 MFCCs	10 passerine species	RBFN 57–96%; SVM 76–100%; HMM 69–93%; GMM 56–85%
Silla et al. [78]	Hierarchical Global Model Naive Bayes	MARSYAS framework, 64 features	74 bird species of South Atlantic Coast	F-Measure = 0.5

classification. CART had only moderate success separating the classes.

Jennings et al. [90] has conducted an experiment to evaluate the classification of bats by humans and an ANN. Recordings of 3–4 calls of 16 species were made. The recordings were given to 26 human participants working on bats (academicians, researchers, ecologists). They were asked to classify the calls using any method except statistical modelling. If a species is identified correctly, a score of 1 is assigned. On the other hand, from the same recorded calls dataset, several frequency parameters are measured and are fed into an ANN. Classification performance was assessed using sensitivity (% of known calls classified) and positive predictive power (% of unknown calls classified). Results indicated that the ANN performed 75% better than humans. The authors concluded that by improving the ANN and training data, this can be further improved.

Apart from DFA and ANN methods, support vector machines (SVMs) have also been used in the classification of bat species by Redgwell et al. [91]. The authors also tested ensemble neural networks (ENN) while keeping DFA as a canonical approach against which to measure performance. The experiment commenced by procuring a library of 713 calls from 14 species. A Butterworth high pass filter was used to remove noise and improve the signal. An algorithm from MATLAB was used for call extraction which will iteratively run until it finds a frequency with highest energy with 8 kHz difference from previous repetition or average signal to noise ratio (SNR) exceeds 0.01. Twelve parameters were extracted from each call. Five are considered to be “base parameters": duration, start frequency, end frequency, middle frequency and max energy frequency. The sixth and seventh are rate of change and bandwidth measurements, respectively. To obtain the next four parameters, each call was exposed to a Hilbert transform and multiplying by its conjugate to obtain the call’s energy distribution. The 12th parameter is a discrete variable that divides calls into various categories such as constant frequency, frequency modulated, quasi constant frequency

and kinkled (calls with unexpected weakness of frequency in last quarter). Untransformed data is used by DFA since it is robust to normality deviation. DFA showed 73% accuracy in species classification. Twenty SVMs were trained using a radial function with the step-varying range.All the classifiers were combined to identify calls, with the combined SVMs achieving 87% classification accuracy. The ENN approach achieved 98% accuracy. The authors concluded that SVM and ENN performed better than DFA and also that the five base parameters are most critical for classification.

Some research-based analysis techniques have been transformed into end-user tools. For instance, Walters et al. [92] developed an ANN based continental scale tool iBatsID, which assists in classifying any European bat calls, achieving 93% of group classification and 83% of species classification. This tool was made publicly available for monitoring bat acoustics in Europe.

In conclusion, for acoustic based bat species classification, methods such as DFA, SVM, and ANN have been extensively used. A summary is presented in Table 2.

B. OTHER TERRESTRIAL ANIMALS

1) Frogs

Bedoya et al. [30] has pointed out that acoustic-based identification of species has the advantage of being non-invasive, with advantages over marking procedures which may harm animals, particularly sensitive species. They used an unsupervised classification for anurans based on a fuzzy classifier and MFCCs. Datasets from two different sources were collected with 916 calls from 13 anuran species (from STRI) and 813 calls from 6 anuran species (from Antioquia) considered for analysis. The authors developed a four stage classification methodology:

1) Noise reduction: Threshold estimation is performed to identify noisy segments. A Fast Fourier transform (FFT) is applied on the windows of noisy segments, which gives a threshold estimate. During noise removal, the gain control is set to be in the limits of this threshold value. These gain controls are applied to the FFT, followed by an inverse FFT and finally a hamming window, which optimizes the signals.
2) Segmentation: The noise-reduced signal is divided into segments for easier analysis. A syllable of fixed length is formed, which is product of individual vocalizations. The start and end points of the calls are identified in comparison with the threshold value.
3) Feature extraction: The syllables are divided into frames. A Fourier transform is calculated for each frame to identify interesting frequency bands in the frame and to calculate the power spectrum, which is mapped to the mel frequency scale, from which MFCCs are calculated. The mean values of the MFCCs are used as the classification inputs.
4) Classification: The author's Learning Algorithm for Multivariate Data Analysis (LAMDA) is used, initializing cluster 0 as a non-information cluster (NIC). The first element goes unrecognized and is put into this NIC. A new cluster is created by calculating the mean and Global Adequacy Degree (GAD) values of the call. This process is repeated until all calls are analysed. To classify an unknown call, the GAD is calculated and the call is assigned to the species that exhibits the maximum GAD.

Results demonstrated 99% of accuracy in classification. Further this method is able to detect different species even they are not present in training stage.

Han et al. [34] introduced a spectral entropy approach for frog species identification. The method consists of the following three steps:

1) Syllable segmentation: Using the Raven software package (see Section VIII-A2), syllables are extracted and digitized. From 12 to 96 syllables can be extracted from a call of 3 seconds duration.
2) Feature extraction: Three different features are extracted from syllables:
   - Spectral centroid: A highly informative feature for machine learning representing the center point of the spectrum where the sound is "bright".
   - Shannon entropy: This quantifies the richness of sound.
   - Renyi entropy: It is used to identify the complexity of the sound, i.e., to identify the noise content.
3) Classification: The above entropies are given as input to a k-Nearest Neighbor classifier.

Results demonstrated that the use of entropy in the classifier has improved the accuracy.

Dayou et al. [35] also studied the use of an entropy-based approach for frog species identification. The work applied a k-NN classifier using Shannon entropy, Renyi entropy and Tsallis entropy as the features. The approach achieved 100% classification accuracy for seven species out of the nine present in their dataset. Table 3 summarizes frog species identification works.

2) Dogs

Bioacoustics relating to dogs has focused on identifying the state of the animal or situation they are in.

Yin and McCowan [93] analysed dog barks to examine classification in different contexts. Ten dogs ranging from 3 to 13 years of age, of both sexes, were considered for analysis. Recordings were made in disturbance, isolation and playing situations. Only 5% of the collected data was identified as noise and the remaining 95% used for analysis. Sixty frequency and 60 amplitude measures were taken, with a number of features derived from these. Spectrograms were generated using FFTs with a hamming window.To discriminate between different barks in several contexts DFA was used, producing a set of features useful for classification. Results indicated good recognition and classification accuracy.

Riede et al. [12] has also studied dog barking. They considered two dog categories, with a dataset comprising 10 dogs

TABLE 2. Summary of articles addressing bat species identification

Study	Technique used	Features	Dataset	Accuracy	Conclusion
Vaughan et al. [84]	multivariate analysis	call duration, interpulse interval, peak frequency, and start, end & centre frequencies	536 bats, 15 known UK species	67% (low duty cycle), 89% (intermediate duty cycle)	Duration, start & end frequencies are most important
O'Farrell et al. [85]	Anabat II detector using a zero-crossings analysis interface module (ZCAIM)	Max & min frequency, linearity, slope	59 locations in south-west US	62–97%	Unable to process up to 20% of the calls as they were too noisy
Obrist et al. [87]	Pattern recognition approach: synergetic algorithm (SC-MELT)	Duration, highest & lowest frequencies, frequency of main energy	643 sequence calls containing 362 hand-identified specimens	70–88%	High accuracy requires careful selection of training calls
Hughes et al. [88]	DFA	duration, the frequency of maximum energy , start frequency, terminal frequency , maximum frequency, minimum frequency and frequency range	510 calls from Thailand	70–96%	Bats emitting FM (broadband) calls should be analyzed carefully
Preatoni et al. [89]	DFA, cluster analysis, CART, ANN	pulse duration, max intensity frequency, and start, end, middle, min & max call frequencies	20 species from 126 hand released bats of northern Italy	DFA 77–100%; cluster analysis 30–100%; CART 41–100%; ANN 64–100%	DFA and ANN most effective, with DFA slightly better
Jennings et al. [90]	ANN	characteristic features of the call, the "rhythm" of calls	45 time expanded recordings of 16 species in the UK	92% (genus), 62% (species)	ANN performed 75.5% better than humans
Redgwell et al. [91]	DFA, SVM, ensembles of neural networks (ENN)	12 parameters: 5 base parameters, rate of change and bandwidth, 4 from Hilbert transform	713 calls from 14 species in the UK	ENN 98%; SVM 87%; DFA 73%	The five selected base parameters are essential features for classification

TABLE 3. Summary of articles addressing frog species identification

Study	Technique used	Features	Dataset	Accuracy	Conclusion
Bedoya, Isaza, Daza, & López [30]	Fuzzy classifier and LAMDA	MFCCs	916 calls (13 species) & 813 calls (6 species)	99%	Able to detect different species even if not present in training data
Han, Muniandy, & Dayou [34]	Hybrid spectral entropy approach and k-NN classifier	Spectral centroid, Shannon entropy, Renyi entropy	Nine species from Australian Microhylidae family	98%	Accuracy diminishes considerably at noise levels higher than -20 dB
Dayou, Han, Ahmad, Muniandy, & Dalimin [35]	k-NN	Shannon entropy, Renyi entropy and Tsallis entropy	Nine Australia frog species	100% (7 of 9 species only)	Model failed to identify two species due to similarity in entropy values

in good health and 10 dogs in unhealthy condition (e.g., undergoing treatment in a veterinary clinic). Barks of the dogs in these two different contexts were recorded. To identify the differences in the acoustics, harmonic to noise ratio (HNR) was calculated. The harmonics of healthy dogs' barks shows them to be regular, while the barks of unhealthy dogs are irregular. Due to illness, some noise will be introduced which also influences the acoustics. Recordings of all dogs over a 6-month period was used. The authors concluded that HNR-based classification is a good measure for quantifying noise in bioacoustics in the context of dogs. Table 4 summarizes dog species identification works.

3) Elephants

Clemins et al. [94] has made a study of one male and six female African elephants. Data were collected by fitting each animal with a microphone and radio frequency transmitter. Acoustic features are extracted from spectrograms using a hamming window. A 60 ms window is used for call classification and 300 ms window is used for speaker identification. To model transitions, a Hidden Markov Model is used. For classifying vocalization type, the Hidden Markov Model performed as well as humans. Further, speaker identification was performed in different contexts, one where the male was isolated from the females, and another was when the male was with four other elephants. The accuracy of the experiment was 82.5%.

TABLE 4. Summary of articles addressing dog species identification

Study	Technique used	Features	Dataset	Accuracy	Conclusion
Yin, & McCowan [93]	DFA and mixed-effects ANOVA	Frequency and amplitude measures	10 dogs (ages 3–13, in different contexts)	Up to 90%	Dog barks are context specific, affecting accuracy
Riede [12]	Harmonics-to-noise ratio and DFA (HNR)	60 acoustic parameters	10 healthy and 10 unhealthy dogs	83%	HNR can be useful in quantifying noise

4) Insects

Acoustics have also been used to study insects. Chesmore and Nellenbach [95] identified an automatic method for identification of grasshoppers and crickets. Using time domain signal processing and an ANN, they demonstrated 100% classification accuracy, and claimed that the approach can be used for other species as well.

Potamitis et al. [96] has highlighted that insects can be categorized based on their appearance and sound production. Trapping and identifying insects through appearance is difficult. Hence several forms of sounds made by insects in several instances such as eating, flying or locomotion can be used as a means of communication. The sounds produced include mating calls or sounds to warn others of danger. Species recognition is done in a two-step process:

1) Signal parameterization: the energy of fixed length frames is estimated, and then a discrete Fourier transform is applied. A linearly spaced filter bank is applied to the DFT and linear frequency cepstral coefficients are calculated. The first 24 of these are used for recognition. On all feature vectors dynamic normalization is applied.
2) Classification: These normalized features are used with probabilistic neural networks (PNN) and Gaussian mixture models (GMM) classifiers. For each target species, a model is built. The Bayesian rule is applied to finalize the class of the target predicted.

The models' accuracy was 90–99%. Table 5 summarizes works on insect species identification.

C. CETACEANS

Cetaceans includes whales, dolphins and related species. Soon after the Second World War it was identified that acoustics could be used with these for different kinds of analysis. By the 1970s the potential for species recognition was known, but could not proceed due to technology limitations. Advances in the digital technologies have paved the way for cetacean species identification using acoustics.

1) Statistical methods

DFA has been applied by Oswald et al. [97] to identify dolphin species. Whistles from four species were collected and only loud and clear whistles of nine decibels more than background noise were chosen randomly for analysis. Four different spectrograms with variable frequencies were created and eight measures extracted: start, end, min and max frequencies, duration, infection points, number of steps, and a harmonics indicator. Taking these measures as input, a DFA based on orthogonal linear functions classified to specified groups. Results indicated that, although the DFA classified accurately, recording quality and analysis bandwidth greatly influence the effectiveness of species classification. The authors concluded that a bandwidth of not less than 24 kHz is needed for accurate analysis and classification.

2) Real-time Odontocete Call Classification Algorithm (ROCCA)

Several large species like whales produce whistles that can be easily identified, but to recognize whistles of species like delphinds is more challenging. To address this, Oswald et al. [98] developed a tool named ROCCA (Real-time Odontocete Call Classification Algorithm) which classifies the species in real time. From the recordings, 50% (35 whistles per session) of whistles that were loud and clear were randomly selected. As in their previous work start, end, min and max frequencies, duration, infection points, number of steps, and a harmonics indicator were used as features, supplemented by slope of beginning and end sweeps. After normality tests and transformation, DFA was applied to these measures to classify them. Mahalanobis distance is calculated to identify the group centroid. Another classification method, CART, which creates a binary tree, was also applied. The jackknife method was used to obtain classification scores. DFA showed up to 63.5% accuracy while CART gave 57% accuracy.

3) Gaussian Mixture Model (GMM))

Roch et al. [99] used a Gaussian mixture model for species classification using the common analysis pattern of collection of the call data, call detection, feature extraction and finally classification steps. Recordings were obtained from the California coastline. Single species calls only were kept for further analysis. Using spectrograms, calls' start and end points were identified, and calls with good quality and above 18 dB, were considered for analysis. The identified calls are divided into frames of size 21 ms and a hamming window applied. A filter was applied to identify the poor calls below 5 kHz. To derive the classification features, cepstral features were identified by applying the discrete cosine transform on the filter. Since it is difficult to assess the values of a GMM, an initial Gaussian classifier was initialized using mean and covariance values. An iterative algorithm was applied which splits each single mixture into two mixtures with an offset value. The algorithm is iterated until the anticipated mixtures

TABLE 5. Summary of articles addressing insect species identification

Study	Technique used	Features	Dataset	Accuracy	Conclusion
Chesmore, & Nellenbach, [95]	ANN	Time domain signal processing	25 species of British Orthoptera	100%	Method works for other species (birds) as well
Potamitis, Ganchev, & Fakotakis [96]	Probabilistic neural networks (PNN) and GMM classifiers	Linear frequency cepstral coefficients	220 cicada & cricket species	90–99%	Method can be expanded to other insect species

are formed. Once the expectation is known with this algorithm, the model is trained. Then the posterior probability of each species was computed using Bayes' rule. The posterior values from each observation was summed up correctly to identify the species. The model achieved up to 75% classification accuracy. Table 6 summarizes the work on cetaceans.

From the studies summarised above (for birds, bats, other terrestrial animals, and cetaceans) one can see that various techniques have been used for different animal types, but with some commonalities. These are summarised in Table 7.

VIII. BIOACOUSTICS SOFTWARE AND BIG DATA

A. SYSTEMS AND SOFTWARE

As discussed earlier, bioacoustics analysis involves several activities such as data collection, visualization, preprocessing, feature extraction and analysis. Due to their high complexity, each activity requires support from sophisticated software tools to enable automatic processing. Several organizations have developed software which to support some or all of these activities. Initially we provide discussion on general audio processing software and, in subsequent sections, we discuss bioacoustics specific software and big data handling systems.

1) General Audio Processing Software

Software such as Sound Ruler, Audacity and seewave are designed for general audio processing purposes, but can also used in bioacoustics work.

1) *SoundRuler* is free and open source interactive visual tool to perform analysis tasks on acoustic data [101]. Bee [102] provides a comprehensive overview of the software's features and its operation. Its features include handling unlimited size audio files, adjustable filters, support for both manual and automatic analysis and generation of more than 50 acoustic measures. It can produce graphical summaries of acoustic data, such as spectrograms and oscillograms.
2) *Audacity* is sound processing software developed at Carnegie Mellon University by Dominic Mazzoni and Roger Dannenberg in 2000. It facilitates multiple source recording and several post-processing activities. It offers several features such as recording, editing, importing and exporting, as well as support for multiple channel modes, and spectrum analysis.
3) *Seewave* (typically rendered in all lower case) is an R package (available through CRAN) for performing sound analysis and synthesis. According to the seewave website [103], it offers a variety of functions for analysis, display, manipulation, editing and synthesis of audio. Further, this tool processes both 2D and 3D spectrograms, computes entropy, correlation and several other values.

2) Bioacoustics Specific Software

The software discussed below are specifically developed for bioacoustics applications.

1) *AviSoft-SASLab Pro* is a tool for performing various bioacoustics activities [104]. The software displays real time spectrograms with high quality output. It offers automated syllables classification with advanced metadata management capabilities. It is good platform for managing large number of audio files and is suitable for batch and real-time processing. It is available in both freeware and proprietary versions.
2) *Raven* [105] is designed by the Cornell Lab of Ornithology, which supports data uploading, visualization and acoustic analysis. Raven Lite is a freeware version that can used for research on birdsong recognition. It possess several features such as annotations, detection, correlation, support for various data acquisition software, multiple simultaneous windows and views, support for multiple audio formats, editing, filtering and amplifying facilities.
3) *Kaleidoscope Pro* [106] is produced by Wildlife Acoustics, a manufacturer of field recorders, and offers a variety of features, including detecting similar sounds and identifying them as clusters, sound visualization, and tools for editing and labelling, as well as intelligent classifiers that can perform species recognition. It also has support for batch processing and noise analysis.
4) *Bioacoustic Workbench:* Wimmer et al. [107] proposed a web-based workbench to address issues associated with the large volume of bioacoustics data being collected. They developed a framework which facilitates uploading, organisation and structure, visualization, recording and analysis through annotation facilities. Data upload is provided via a web service, which has the advantages of ease of access and support for data backup. Role-based access control mechanisms enable small to large project groups to be managed. The playback and visualization components facilitate splitting audio into segments. Users can perform man-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2978547, IEEE Access

TABLE 6. Summary of articles addressing cetacean species identification

Study	Technique used	Features	Dataset	Accuracy	Conclusion
Oswald, Rankin, & Barlow [97]	DFA	Start, end, min & max frequencies, duration infection points, number of steps, harmonics indicator	484 whistles (4 species)	30–42%	Recording and analysis bandwidth greatly influence accuracy; bandwidth $\geq$ 24 kHz should be used
Oswald, Rankin, Barlow, & Lammers [98]	CART DFA	Start, end, min & max frequencies, duration, infection points, number of steps, harmonics indicator, slope of beginning and end sweeps	Single-species recordings of nine delphind species	DFA up to 63.5%; CART up to 57%	Dataset characteristics strongly impact different classification algorithms; using more than one algorithm can result in higher accuracy
Roch, Soldevilla, Burtenshaw, Henderson, & Hildebrand [99]	Gaussian mixture model and Bayes rule	64 cepstral feature vector	Whistles (four dolphin species from southern California	75%	The cepstral feature space helps the system to capture the timbre of calls

TABLE 7. Similar techniques applied to the study of different animal groups

Technique	Target	Works
Discriminant Function Analysis	Bats	Hughes et al., [88] Preatoni et al., [89] Redgwell et al. [91]
	Cetaceans	Oswald et al., [97] Oswald, et al. [98]
	Dogs	Yin,& McCowan, [93]
Support Vector Machine	Birds	Graciarena et al., [40] Acevedo et al., [79] Fagerlund, Briggs et al., [81] Lakshminarayanan et al., [41] Kampichler, et al., [82] Lopes et al., [48] Cheng et al. [31]
	Bats	Redgwell et al. [91]
Artificial Neural Networks	Birds	Selouani et al., [36] Kampichler et al., [82] Cheng et al., [83] Cai et al., [23] Juang & Chen, Lopes, et al. [69]
	Bats	Preatoni et al., [89] Jennings et al., [90] Walters et al., [92] Redgwell et al. [91]
	Insects	Chesmore et al. [95]
CART	Bats	Preatoni et al. [89]
	Cetaceans	Oswald et al. [98]
Gaussian Mixture Model	Birds	Somervuo et al., [37] Cheng et al., [31] Lee et al., Graciarena et al., [40] Cheng et al. [83]
	Cetaceans	Roch et al. [99]
	Insects	Potamitis et al. [96]
Hidden Markov Model	Birds	Trawicki et al., [29] Fox et al., [32] Cheng et al. [83]
Multivariate Analysis	Birds	McIlraith et al. [70]
	Bats	Vaughan et al. [84]
	Frogs	Bedoya et al. [30]
k-Nearest Neighbors	Birds	Lee et al., [75] Briggs et al., [46] Lopes et al. [48]
	Frogs	Dayou et al., [35] Han et al. [34]
Naive Bayes	Bird	Vilches et al., [77] Lopes et al., [100] Silla et al. [78]
	Cetaceans	Roch et al. [99]
Back Propagation	Bird	McIlraith et al., [70] Selouani et al., [36] Kampichler et al., [82].,Lopes et al. [48]
Random Forest	Bird	Kampichler et al., [82] Stowell et al. [80]

ual analysis through marking annotations and search the acoustic database for matches.

5) *Automated Remote Biodiversity Monitoring Network (ARBIMON)*, developed by Aide [108] is a combination of research field project and web service supporting data hosting and species identification. The software component of the system generates spectrograms using short-time fourier transforms (STFT) and Hann windows. Signals are further processed to generate regions of interest. The interface for species identification has four components: a visualizer for inspection, listening and manual analysis; a species validation component that allows the user to specify the presence or absence of species vocalization in a recording; a model building component to train the data using ROIs (using HMM); and a model application component, which allows an entire dataset to be processed with a previously creaed model.

6) *Multi-layer framework model:* Zhang et al. [109] describe a big acoustic data management framework, analysis and visualization tool. Data is collected from sensors, recorders and handheld devices. This large scale data collection is exposed to issues such as large volume, and high variety and velocity. These three issues are expected to grow further in future as the monitoring and sensing area increases. To address this, the authors presented a Multi-layer framework model for acoustic data. The framework includes a data collection layer and data management layer for upload and management of audio, respectively. The third layer offers event processing which performs pre-processing, rapid scanning of spectrograms as discussed in Truskinger et al. [110] and identification of events of interest. The fourth layer performs knowledge discovery

(data mining). The system offers visualization using spectrograms, tag linking and 3D display. The authors concluded that their future work will focus on mining and extracting knowledge through tags, bird acoustics correlation analysis, their behaviours with respect to environment, time and location.

7) *High Performance Computer Acoustic Data Accelerator (HPC-ADA):* Also with a view to addressing big data issues in bioacoustics, Dugan et al. [111] proposed HPC-ADA, which operates on cloud infrastructure. authors designed software named Detection and Localization using Machine learning Algorithms (DeLMA) which can run across multiple cloud-based machines and run a variety of actual machine learning tools. Scalability is achieved by distribution of jobs, while use of a generic data format supports interoperability. MATLAB libraries are used in DeLMA to deal with algorithms efficiently across distributed computer architectures.
8) Finally, Reyes et al. [112] developed a prototype system based on the use of visual analytics to assists in tasks such as species identification. The experiment initiated by collecting bird audio sounds and extracting MFCC features. Further cosine distance function is used to identify similarity between audio records through its vector representations. To obtain a 2D representation and reduce dimensionalty, PCA is used. Based on 2D visualization and distance matrix, users can visually analyse bird species' sounds.

B. BIG DATA HANDLING SOFTWARE

Bioacoustics data is growing, as the data is received from different sources at different rates resulting in the generation of a massive volume of data. As a branch of ecology, bioacoustics has the task of identifying species, performing density estimation, tracking environmental changes, and so on, which requires data to be gathered for a long time periods and at appropriate, potentially huge scales. qualifies bioacoustics as a big data challenge. Al-Jarrah et al. [113] pointed out that current machine learning algorithms need to be scalable to deal with big data challenges. They reviewed the current state of research in sustainable data modelling such as ensemble models, Bayes nonparametric learning models, local learning strategy, semi-parametric approximation and deep learning authors also discussed several batch and stream processing big data computing methods. Their opinion was that these sustainable data models are capable of handling huge volumes of data with ease, in any field of science.

A number of research projects are working to address the big data challenges in this field. Kelling et al. [114] discussed the phases of a biodiversity research data science workflow. Initially observation data collected from various sources should be validated and organized. As an example of observational data, the authors developed an avian knowledge network containing 60 bird occurrences with descriptive metadata. Since the observation data is from a diverse set of sources and dispersed, a common data warehouse such as bird monitoring data exchange (BMDE) was developed. Then analysis can be performed on such common data by using bagged decision trees and several non-parametric tools.

Technology advances have enabled automated collection of larger volumes of acoustic data, which presents challenges for storage and organisation. Kasten et al. [115] developed an ensemble extraction model from sensor data streams. To support online and incremental learning from the sensor steams to detect bird species, the Multi-Element Self-Organizing tree (MESO) memory system. After filtering and transformations, spectrograms are plotted to represent each segment. Piecewise aggregate approximation (PAA) was used to decrease the dimensionality of the time series. Further, symbolic aggregate approximation (SAX) converts the PAA representation to symbols which helps in computing anomaly score of the signal. On this, a distributed stream processing pipeline named Dynamic River was created. Followed by this, MESO is used for classification and detection. A trained MESO instance returns either an exact matching pattern or similar pattern. The authors stated that the classification accuracies are promising. For these experiments, ensembles are extracted from data streams comprising a single signal. Ensembles extraction from multiple correlated data streams is left as future work.

Truskinger et al. [67] have presented their applied research on managing and analysing huge raw audio datasets. Their core contribution lies in the presentation of analysis methodologies for practical large scale acoustic data. authors categorized automated analysis activities into two broad groups. The first group comprises event detectors, which work on spectral components to identify regions of interest. The second group are acoustic indices, which calculate summary statistics of an audio stream. Initial exploration of results was performed in R, which was then transcoded to C#. Currently the project is migrated to cloud and their future work aims at developing a scalable architecture, enhancing analysis functions and spectrogram generation.

Peter et al. [116] has developed an efficient and scalable algorithm for data processing. The acoustic data-mining accelerator (ADA) algorithm uses distributed computing concepts, where sound is partitioned into blocks and each block is assigned to one worker. All results from the workers are collected and merged to provide analytical results. results demonstrated two times improvement in execution speeds. The authors' future work aims at working on multi-channel and multi-rate data.

IX. OPEN CHALLENGES

Even though there has been lot of work to address issues in analyzing bioacoustics data, there remain several open issues that have to be tackled before developing a fully automated analytical system.

1) *Adverse environmental conditions:* Animal recordings are obtained directly from their habitats and fields and, as such, are subject to environmental conditions.

For instance, it is very common that wind, rain or sounds of aircraft will be also recorded. Several works point out that this can hamper classification accuracy. Consequently there is a need for the development of effective pre-processing, feature extraction and classification methods that are robust to such interference.

2) *Automatic selection of detectors:* In general bioacoustics algorithms need to work in diverse and changing natural environments. Future bioacoustics analysis systems should be capable of automatically tuning based on these changes. This will also require data mining algorithms that are robust enough to detect changes and adapt the models accordingly. For example, deep learning and meta-algorithms need to be developed to automate detector selection for different weather conditions [117].
3) *Acoustic index optimization:* The space of alternative acoustic indices should be more widely examined, including their use as features in machine learning algorithms. When indices are calculated independently, biased sampling is another option to be explored [61]. Moreover, most current indices, even though they have wide applicability, also need to be specifically made applicable for a given context. In other words, new indices are needed that integrate the characteristics of the surrounding environment where the audio was captured or prior knowledge of the species found in that environment.
4) *Call interference*: Most classification methods that are applied for recognizing bird species are based on how precisely they can differentiate between different acoustic signals. Interference between different bird calls can make this difficult and can severely affect the accuracy of automatic recognition methods. Hence it is still a major challenge in the domain of bioacoustics.
5) *Multiple species recognition:* In a recording there may be several species present apart from noise. But classification methods typically assume that only one species is present in a segment of a recording [43]. The domain of multiple species detection should be explored by developing scalable models that can detect a set of all species present in a recording simultaneously.
6) *Scalable, streaming data analysis:* There is much scope for improvement in algorithms for efficient species detection from large scale multiple datasets. Current techniques are mostly tested with small datasets at a time, yet given the large scale deployment of the recorders processing algorithms should be scalable. Techniques for analysing multiple correlated data streams (from spatially-related recordings) are also an open area of research.

X. CONCLUSIONS

Modern bioacoustics has a nearly 70-year history, and has grown considerably in recent decades with advances in processing techniques and hardware. Most current bioacoustics research focuses on the application areas of biodiversity assessment using statistical techniques and acoustic indices, density estimation, using statistical analysis of detected calls, and species identification, using a variety of machine learning techniques. Growth in the use of acoustic monitoring has led to commensurate growth in the amount of raw audio needing to be analysed, necessitating the use of distributed computing tools, although the application of these tools in the bioacoustics domain is somewhat in its infancy. This review has identified a number of targets for future bioacoustics research.

REFERENCES

[1] Kevin R Crooks and M Sanjayan. Connectivity conservation, volume 14. Cambridge University Press, 2006.
[2] David S Wilcove, David Rothstein, Jason Dubow, Ali Phillips, and Elizabeth Losos. Quantifying threats to imperiled species in the united states assessing the relative importance of habitat destruction, alien species, pollution, overexploitation, and disease. BioScience, 48(8):607–615, 1998.
[3] Klaus Riede. Monitoring biodiversity: analysis of amazonian rainforest sounds. Ambio, pages 546–548, 1993.
[4] Paola Laiolo. The emerging significance of bioacoustics in animal species conservation. Biological Conservation, 143(7):1635–1645, 2010.
[5] Chloé Huetz and Thierry Aubin. Bioacoustics approaches to locate and identify animals in terrestrial environments. Sensors for ecology, towards integrated knowledge of ecosystems. JJ Le Galliard, JM Guarini, F. Gaill (eds.). CNRS, Paris, pages 83–96, 2002.
[6] Holger Klinck, Sharon L Nieukirk, David K Mellinger, Karolin Klinck, Haruyoshi Matsumoto, and Robert P Dziak. Seasonal presence of cetaceans and ambient noise levels in polar waters of the north atlantic. The Journal of the Acoustical Society of America, 132(3):EL176–EL181, 2012.
[7] Whitlow WL Au and Mardi C Hastings. Principles of marine bioacoustics. 2008.
[8] Walter MX Zimmer. Passive acoustic monitoring of cetaceans. 2011.
[9] David K Mellinger. Introduction to animal bioacoustics. The Journal of the Acoustical Society of America, 129(4):2406–2406, 2011.
[10] Jérôme Sueur, Almo Farina, Amandine Gasc, Nadia Pieretti, and Sandrine Pavoine. Acoustic indices for biodiversity assessment and landscape investigation. Acta Acustica united with Acustica, 100(4):772–781, 2014.
[11] Bhupinder Singh Dang and Frank Arland Andrews. A literature survey on the subject of the use of acoustics in fish catching and fish study. 1971.
[12] Tobias Riede, Hanspeter Herzel, Kurt Hammerschmidt, Leo Brunnberg, and Günter Tembrock. The harmonic-to-noise ratio applied to dog barks. The Journal of the Acoustical Society of America, 110(4):2191–2197, 2001.
[13] David A Mann, Anthony D Hawkins, and J Michael Jech. Active and passive acoustics to locate and study fish. In Fish bioacoustics, pages 279–309. Springer, 2008.
[14] AS King. Functional anatomy of the syrinx. Form and function in birds, 4:105–192, 1989.
[15] J Martin Wild. Neural pathways for the control of birdsong production. Journal of neurobiology, 33(5):653–670, 1997.
[16] Johan J Bolhuis and Martin Everaert. Birdsong, speech, and language: exploring the evolution of mind and brain. 2013.
[17] Arik Kershenbaum, Daniel T Blumstein, Marie A Roch, Çağlar Akçay, Gregory Backus, Mark A Bee, Kirsten Bohn, Yan Cao, Gerald Carter, Cristiane Caesar, et al. Acoustic sequences in non-human animals: a tutorial review and prospectus. Biological Reviews, 91(1):13–52, 2016.
[18] John R Krebs and Donald E Kroodsma. Repertoires and geographical variation in bird song. Advances in the Study of Behavior, 11:143–177, 1980.
[19] Frank Glaw and Miguel Vences. Bioacoustic differentiation in painted frogs (discoglossus). Amphibia-Reptilia, 12(4):385–394, 1991.
[20] Ruth Mace. The dawn chorus in the great tit paras major is directly related to female fertility. 1987.

[21] Martin K Obrist, Gianni Pavan, Jérôme Sueur, Klaus Riede, Diego Llusia, and Rafael Márquez. Bioacoustics approaches in biodiversity inventories. Abc Taxa, 8:68–99, 2010.

[22] Klaus Riede. Bioacoustic Diversity, 2010. http://www.bioacoustics.myspecies.info/en.

[23] Jinhai Cai, Dominic Ee, Binh Pham, Paul Roe, and Jinglan Zhang. Sensor network for the monitoring of ecosystem: Bird species recognition. In Intelligent Sensors, Sensor Networks and Information, 2007. ISSNIP 2007. 3rd International Conference on, pages 293–298. IEEE, 2007.

[24] Jae S Lim and Alan V Oppenheim. Enhancement and bandwidth compression of noisy speech. Proceedings of the IEEE, 67(12):1586–1604, 1979.

[25] Yariv Ephraim and Harry L Van Trees. A signal subspace approach for speech enhancement. IEEE Transactions on speech and audio processing, 3(4):251–266, 1995.

[26] Hanoch Lev-Ari and Yariv Ephraim. Extension of the signal subspace speech enhancement approach to colored noise. IEEE Signal Processing Letters, 10(4):104–106, 2003.

[27] Yariv Ephraim and David Malah. Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator. IEEE Transactions on Acoustics, Speech, and Signal Processing, 32(6):1109–1121, 1984.

[28] Yariv Ephraim. Statistical-model-based speech enhancement systems. Proceedings of the IEEE, 80(10):1526–1555, 1992.

[29] Marek B Trawicki, Michael T Johnson, and Tomasz S Osiejuk. Automatic song-type classification and speaker identification of norwegian ortolan bunting (emberiza hortulana) vocalizations. In Machine Learning for Signal Processing, 2005 IEEE Workshop on, pages 277–282. IEEE, 2005.

[30] Carol Bedoya, Claudia Isaza, Juan M Daza, and José D López. Automatic recognition of anuran species based on syllable identification. Ecological Informatics, 24:200–209, 2014.

[31] Jinkui Cheng, Yuehua Sun, and Liqiang Ji. A call-independent and automatic acoustic system for the individual recognition of animals: A novel model using four passerines. Pattern Recognition, 43(11):3846–3852, 2010.

[32] Elizabeth JS Fox, J Dale Roberts, and Mohammed Bennamoun. Call-independent individual identification in birds. Bioacoustics, 18(1):51–67, 2008.

[33] Joseph A Kogan and Daniel Margoliash. Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden markov models: A comparative study. The Journal of the Acoustical Society of America, 103(4):2185–2196, 1998.

[34] Ng Chee Han, Sithi V Muniandy, and Jedol Dayou. Acoustic classification of australian anurans based on hybrid spectral-entropy approach. Applied Acoustics, 72(9):639–645, 2011.

[35] Jedol Dayou, Ng Chee Han, Ho Chong Mun, Abdul Hamid Ahmad, Sithi V Muniandy, and Mohd Noh Dalimin. Classification and identification of frog sound based on entropy approach. In 2011 International Conference on Life Science and Technology, volume 3, pages 184–187, 2011.

[36] S-A Selouani, M Kardouchi, E Hervet, and D Roy. Automatic birdsong recognition based on autoregressive time-delay neural networks. In computational intelligence methods and applications, 2005 ICSC congress on, pages 6–pp. IEEE, 2005.

[37] Panu Somervuo, Aki Harma, and Seppo Fagerlund. Parametric representations of bird sounds for automatic species recognition. IEEE Transactions on Audio, Speech, and Language Processing, 14(6):2252–2263, 2006.

[38] Aki Harma. Automatic identification of bird species based on sinusoidal modeling of syllables. In Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on, volume 5, pages V–545. IEEE, 2003.

[39] Panu Somervuo and Aki Harma. Bird song recognition based on syllable pair histograms. In Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on, volume 5, pages V–825. IEEE, 2004.

[40] Martin Graciarena, Michelle Delplanche, Elizabeth Shriberg, and Andreas Stolcke. Bird species recognition combining acoustic and sequence modeling. In Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on, pages 341–344. IEEE, 2011.

[41] Balaji Lakshminarayanan, Raviv Raich, and Xiaoli Fern. A syllable-level probabilistic framework for bird species identification. In Machine Learning and Applications, 2009. ICMLA'09. International Conference on, pages 53–59. IEEE, 2009.

[42] Lawrence Neal, Forrest Briggs, Raviv Raich, and Xiaoli Z Fern. Time-frequency segmentation of bird song in noisy acoustic environments. In Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on, pages 2012–2015. IEEE, 2011.

[43] Forrest Briggs, Balaji Lakshminarayanan, Lawrence Neal, Xiaoli Z Fern, Raviv Raich, Sarah JK Hadley, Adam S Hadley, and Matthew G Betts. 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, 2012.

[44] Juan Gabriel Colonna, Marco Cristo, Mario Salvatierra, and Eduardo Freire Nakamura. An incremental technique for real-time bioacoustic signal segmentation. Expert Systems with Applications, 42(21):7367–7374, 2015.

[45] Lawrence R Rabiner and Biing-Hwang Juang. Fundamentals of speech recognition. 1993.

[46] Forrest Briggs, Raviv Raich, and Xiaoli Z Fern. Audio classification of bird species: A statistical manifold approach. In Data Mining, 2009. ICDM'09. Ninth IEEE International Conference on, pages 51–60. IEEE, 2009.

[47] Jakob Leben. MARSYAS: Music Analysis, Retrieval and Synthesis for Audio Signals, 2020. http://marsyas.info.

[48] Marcelo T Lopes, Lucas L Gioppo, Thiago T Higushi, Celso AA Kaestner, Carlos N Silla Jr, and Alessandro L Koerich. Automatic bird species identification for large number of species. In Multimedia (ISM), 2011 IEEE International Symposium on, pages 117–122. IEEE, 2011.

[49] George Arthur Frederick Seber. The estimation of animal abundance. 1982.

[50] Byron K Williams, James D Nichols, and Michael J Conroy. Analysis and management of animal populations. Academic Press, 2002.

[51] Richard T Reynolds, J Michael Scott, and Ronald A Nussbaum. A variable circular-plot method for estimating bird numbers. Condor, pages 309–313, 1980.

[52] Deanna K Dawson and Murray G Efford. Bird population density estimated from acoustic signals. Journal of Applied Ecology, 46(6):1201–1209, 2009.

[53] Steven S Rosenstock, David R Anderson, Kenneth M Giesen, Tony Leukering, Michael F Carter, and F Thompson III. Landbird counting techniques: current practices and an alternative. The Auk, 119(1):46–53, 2002.

[54] Kenneth P Burnham, David R Anderson, and Jeffrey L Laake. Line transect estimation of bird population density using a fourier series. Stud. Avian Biol, 6:466–482, 1981.

[55] AR Hiby and AJ Ward. Analysis of cue-counting and blow rate estimation experiments carried out during the 1984/85 idcr minke whale assessment cruise. Report of the International Whaling Commission, 36:473–476, 1986.

[56] Stephen T Buckland and CM Handel. Point-transect surveys for songbirds: robust methodologies. The Auk, 123(2):345–357, 2006.

[57] J Andrew Royle. N-mixture models for estimating population size from spatially replicated counts. Biometrics, 60(1):108–115, 2004.

[58] Tiago A Marques, Len Thomas, Stephen W Martin, David K Mellinger, Jessica A Ward, David J Moretti, Danielle Harris, and Peter L Tyack. Estimating animal population density using passive acoustics. Biological Reviews, 88(2):287–309, 2013.

[59] Jérôme Sueur, Sandrine Pavoine, Olivier Hamerlynck, and Stéphanie Duvail. Rapid acoustic survey for biodiversity appraisal. PloS one, 3(12):e4065, 2008.

[60] Marion Depraetere, Sandrine Pavoine, Fréderic Jiguet, Amandine Gasc, Stéphanie Duvail, and Jérôme Sueur. Monitoring animal diversity using acoustic indices: implementation in a temperate woodland. Ecological Indicators, 13(1):46–54, 2012.

[61] Michael W Towsey. The calculation of acoustic indices to characterise acoustic recordings of the environment. 2012.

[62] Mangalam Sankupellay, Michael Towsey, Anthony Truskinger, and Paul Roe. Visual fingerprints of the acoustic environment: The use of acoustic indices to characterise natural habitats. In Big Data Visual Analytics (BDVA), 2015, pages 1–8. IEEE, 2015.

[63] Michael Towsey, Liang Zhang, Mark Cottman-Fields, Jason Wimmer, Jinglan Zhang, and Paul Roe. Visualization of long-duration acoustic recordings of the environment. Procedia Computer Science, 29:703–712, 2014.

[64] Antonio Celis-Murillo, Jill L Deppe, and Michael F Allen. Using soundscape recordings to estimate bird species abundance, richness, and composition. Journal of Field Ornithology, 80(1):64–78, 2009.

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

[66] Jason Wimmer, Michael Towsey, Paul Roe, and Ian Williamson. Sampling environmental acoustic recordings to determine bird species richness. Ecological Applications, 23(6):1419–1428, 2013.

[67] Anthony Truskinger, Mark Cottman-Fields, Philip Eichinski, Michael Towsey, and Paul Roe. Practical analysis of big acoustic sensor data for environmental monitoring. In Big Data and Cloud Computing (BdCloud), 2014 IEEE Fourth International Conference on, pages 91–98. IEEE, 2014.

[68] Volker B Deecke and Vincent M Janik. Automated categorization of bioacoustic signals: avoiding perceptual pitfalls. The Journal of the Acoustical Society of America, 119(1):645–653, 2006.

[69] Chia-Feng Juang and Tai-Mou Chen. Birdsong recognition using prediction-based recurrent neural fuzzy networks. Neurocomputing, 71(1):121–130, 2007.

[70] Alex L McIlraith and Howard C Card. Birdsong recognition using backpropagation and multivariate statistics. IEEE Transactions on Signal Processing, 45(11):2740–2748, 1997.

[71] Julian D Olden, Joshua J Lawler, and N LeRoy Poff. Machine learning methods without tears: a primer for ecologists. The Quarterly review of biology, 83(2):171–193, 2008.

[72] Botond Fazeka, Alexander Schindler, Thomas Lidy, and Andreas Rauber. A multi-modal deep neural network approach to bird-song identification. arXiv preprint arXiv:1811.04448, 2018.

[73] Mangalam Sankupellay and Dmitry Konovalov. Bird call recognition using deep convolutional neural network, resnet-50. In Proceedings of ACOUSTICS, volume 7, 2018.

[74] Xianjun Xia, Roberto Togneri, Ferdous Sohel, Yuanjun Zhao, and Defeng Huang. A survey: Neural network-based deep learning for acoustic event detection. Circuits, Systems, and Signal Processing, pages 1–21, 2019.

[75] Chang-Hsing Lee, Chin-Chuan Han, and Ching-Chien Chuang. Automatic classification of bird species from their sounds using two-dimensional cepstral coefficients. IEEE Transactions on Audio, Speech, and Language Processing, 16(8):1541–1550, 2008.

[76] Gonzalo Vaca-Castaño and Domingo Rodriguez. Using syllabic mel cepstrum features and k-nearest neighbors to identify anurans and birds species. In Signal Processing Systems (SIPS), 2010 IEEE Workshop on, pages 466–471. IEEE, 2010.

[77] Erika Vilches, Ivan A Escobar, Edgar E Vallejo, and Charles E Taylor. Data mining applied to acoustic bird species recognition. In Pattern Recognition, 2006. ICPR 2006. 18th International Conference on, volume 3, pages 400–403. IEEE, 2006.

[78] Carlos N Silla and Celso AA Kaestner. Hierarchical classification of bird species using their audio recorded songs. In Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on, pages 1895–1900. IEEE, 2013.

[79] Miguel A Acevedo, Carlos J Corrada-Bravo, Héctor Corrada-Bravo, Luis J Villanueva-Rivera, and T Mitchell Aide. Automated classification of bird and amphibian calls using machine learning: A comparison of methods. Ecological Informatics, 4(4):206–214, 2009.

[80] Dan Stowell and Mark D Plumbley. Automatic large-scale classification of bird sounds is strongly improved by unsupervised feature learning. PeerJ, 2:e488, 2014.

[81] Seppo Fagerlund. Bird species recognition using support vector machines. EURASIP Journal on Applied Signal Processing, 2007(1):64–64, 2007.

[82] Christian Kampichler, Ralf Wieland, Sophie Calmé, Holger Weissenberger, and Stefan Arriaga-Weiss. Classification in conservation biology: a comparison of five machine-learning methods. Ecological Informatics, 5(6):441–450, 2010.

[83] Jinkui Cheng, Bengui Xie, Congtian Lin, and Liqiang Ji. A comparative study in birds: call-type-independent species and individual recognition using four machine-learning methods and two acoustic features. Bioacoustics, 21(2):157–171, 2012.

[84] NANCY VAUGHAN, GARETH JONES, and STEPHEN HARRIS. Identification of british bat species by multivariate analysis of echolocation call parameters. Bioacoustics, 7(3):189–207, 1997.

[85] Michael J O'Farrell, Bruce W Miller, and William L Gannon. Qualitative identification of free-flying bats using the anabat detector. Journal of Mammalogy, 80(1):11–23, 1999.

[86] Donald R Griffin. Listening in the dark: the acoustic orientation of bats and men. 1958.

[87] Martin K Obrist, Ruedi Boesch, and Peter F Flückiger. 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 mamm, 68(4):307–322, 2004.

[88] Alice C Hughes, Chutamas Satasook, Paul JJ Bates, Pipat Soisook, Tuanjit Sritongchuay, Gareth Jones, and Sara Bumrungsri. Using echolocation calls to identify thai bat species: Vespertilionidae, emballonuridae, nycteridae and megadermatidae. Acta Chiropterologica, 13(2):447–455, 2011.

[89] Damiano G Preatoni, Mosè Nodari, Roberta Chirichella, Guido Tosi, Luc A Wauters, and Adriano Martinoli. Identifying bats from time-expanded recordings of search calls: comparing classification methods. Journal of Wildlife Management, 69(4):1601–1614, 2005.

[90] N Jennings, Stuart Parsons, and MJO Pocock. Human vs. machine: identification of bat species from their echolocation calls by humans and by artificial neural networks. Canadian Journal of Zoology, 86(5):371–377, 2008.

[91] Robert D Redgwell, Joseph M Szewczak, Gareth Jones, and Stuart Parsons. Classification of echolocation calls from 14 species of bat by support vector machines and ensembles of neural networks. Algorithms, 2(3):907–924, 2009.

[92] Charlotte L Walters, Robin Freeman, Alanna Collen, Christian Dietz, M Brock Fenton, Gareth Jones, Martin K Obrist, Sébastien J Puechmaille, Thomas Sattler, Björn M Siemers, et al. A continental-scale tool for acoustic identification of european bats. Journal of Applied Ecology, 49(5):1064–1074, 2012.

[93] Sophia Yin and Brenda McCowan. Barking in domestic dogs: context specificity and individual identification. Animal Behaviour, 68(2):343–355, 2004.

[94] Patrick J Clemins, Michael T Johnson, Kirsten M Leong, and Anne Savage. Automatic classification and speaker identification of african elephant (loxodonta africana) vocalizations. The Journal of the Acoustical Society of America, 117(2):956–963, 2005.

[95] ED Chesmore and C Nellenbach. Acoustic methods for the automated detection and identification of insects. In III International Symposium on Sensors in Horticulture 562, pages 223–231, 1997.

[96] Ilyas Potamitis, Todor Ganchev, and Nikos Fakotakis. Automatic acoustic identification of crickets and cicadas. In Signal Processing and Its Applications, 2007. ISSPA 2007. 9th International Symposium on, pages 1–4. IEEE, 2007.

[97] Julie N Oswald, Shannon Rankin, and Jay Barlow. The effect of recording and analysis bandwidth on acoustic identification of delphinid species. The Journal of the Acoustical Society of America, 116(5):3178–3185, 2004.

[98] Julie N Oswald, Shannon Rankin, Jay Barlow, and Marc O Lammers. A tool for real-time acoustic species identification of delphinid whistles. The Journal of the Acoustical Society of America, 122(1):587–595, 2007.

[99] Marie A Roch, Melissa S Soldevilla, Jessica C Burtenshaw, E Elizabeth Henderson, and John A Hildebrand. Gaussian mixture model classification of odontocetes in the southern california bight and the gulf of california. The Journal of the Acoustical Society of America, 121(3):1737–1748, 2007.

[100] Marcelo Teider Lopes, Carlos Nascimento Silla Junior, Alessandro Lameiras Koerich, and Celso Antonio Alves Kaestner. Feature set comparison for automatic bird species identification. In Systems, Man, and Cybernetics (SMC), 2011 IEEE International Conference on, pages 965–970. IEEE, 2011.

[101] Marcos Gridi-Papp. SoundRuler, 2011. http://soundruler.sourceforge.net.

[102] Mark A Bee. Equipment review. Bioacoustics, 14(2):171–178, 2004.

[103] Jerome Sueur. seewave: Sound Analysis and Synthesis, 2019. https://cran.r-project.org/web/packages/seewave.

[104] Avisoft Bioacoustics. Avisoft, 2020. http://www.avisoft.com.

[105] Cornell Lab of Ornithology. Raven Pro, Lite, and Exhibit, 2020.

[106] Wildlife Acoustics, Inc. Kaleidoscope Pro Analysis Software, 2020. https://www.wildlifeacoustics.com/products/kaleidoscope-pro.

[107] Jason Wimmer, Michael Towsey, Birgit Planitz, Paul Roe, and Ian Williamson. Scaling acoustic data analysis through collaboration and

automation. In e-Science (e-Science), 2010 IEEE Sixth International Conference on, pages 308–315. IEEE, 2010.

[108] T Mitchell Aide, Carlos Corrada-Bravo, Marconi Campos-Cerqueira, Carlos Milan, Giovany Vega, and Rafael Alvarez. Real-time bioacoustics monitoring and automated species identification. PeerJ, 1:e103, 2013.

[109] Jinglan Zhang, Kai Huang, Mark Cottman-Fields, Anthony Truskinger, Paul Roe, Shufei Duan, Xueyan Dong, Michael Towsey, and Jason Wimmer. Managing and analysing big audio data for environmental monitoring. In Computational Science and Engineering (CSE), 2013 IEEE 16th International Conference on, pages 997–1004. IEEE, 2013.

[110] Anthony Truskinger, Mark Cottman-Fields, Daniel Johnson, and Paul Roe. Rapid scanning of spectrograms for efficient identification of bioacoustic events in big data. In eScience (eScience), 2013 IEEE 9th International Conference on, pages 270–277. IEEE, 2013.

[111] Peter Dugan, John Zollweg, Herve Glotin, Marian Popescu, Denise Risch, Yann LeCun, and Christopher Clark. High performance computer acoustic data accelerator (hpc-ada): A new system for exploring marine mammal acoustics for big data applications. Proc. ICML Unsupervised learning for Bioacoustics, 2014.

[112] Angie K Reyes and Jorge E Camargo. Visualization of audio records for automatic bird species identification. In Signal Processing, Images and Computer Vision (STSIVA), 2015 20th Symposium on, pages 1–6. IEEE, 2015.

[113] Omar Y Al-Jarrah, Paul D Yoo, Sami Muhaidat, George K Karagiannidis, and Kamal Taha. Efficient machine learning for big data: A review. Big Data Research, 2(3):87–93, 2015.

[114] Steve Kelling, Wesley M Hochachka, Daniel Fink, Mirek Riedewald, Rich Caruana, Grant Ballard, and Giles Hooker. Data-intensive science: a new paradigm for biodiversity studies. BioScience, 59(7):613–620, 2009.

[115] Eric P Kasten, Philip K McKinley, and Stuart H Gage. Ensemble extraction for classification and detection of bird species. Ecological Informatics, 5(3):153–166, 2010.

[116] J Peter, Holger Klinck, John A Zollweg, Christopher W Clark, et al. Data mining sound archives: A new scalable algorithm for parallel-distributing processing. In Data Mining Workshop (ICDMW), 2015 IEEE International Conference on, pages 768–772. IEEE, 2015.

[117] Dan Stowell, Mike Wood, Yannis Stylianou, and Hervé Glotin. Bird detection in audio: a survey and a challenge. In Machine Learning for Signal Processing (MLSP), 2016 IEEE 26th International Workshop on, pages 1–6. IEEE, 2016.

RAMA RAO KVSN has received his graduation and Masters from Andhra University, India. His experience spans across academic and industry domains. He is currently involved in research on bioacoustics at the University of Tasmania, Australia. His research interests include bioacoustics, machine learning and stream processing.

JAMES MONTGOMERY received a BInfTech (Hons) degree from Bond University, Australia in 2000, and was awarded a PhD in computer science from Bond in 2005. He has held postdoctoral appointments at Swinburne University of Technology and the Australian National University. He is currently a Senior Lecturer with ICT at the University of Tasmania, Australia. His research interests span evolutionary computation, bioacoustics, machine learning, and web services.

SAURABH GARG is a Senior Lecturer with the University of Tasmania, Australia. He is one of the few PhD students who completed in less than three years at the University of Melbourne. He has authored over 40 papers in highly cited journals and conferences. His research interests include resource management, scheduling, utility and grid computing, Cloud computing, green computing, wireless networks, and ad hoc networks.

MICHAEL CHARLESTON Michael Charleston is an Associate Professor at the University of Tasmania. He has held academic posts at Massey University, University of Texas at Austin TX, Glasgow University, Oxford University (where he was awarded a Royal Society University Research Fellowship), and University of Sydney. He has authored and coauthored over 100 journal papers in international peer-reviewed journals. His research interests include bioinformatics methods, phylogenetics and cophylogenetic epidemiology of infectious disease, biological modeling, ecology, graph theory and software development.