

## REVIEW

# Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring

Rory Gibb<sup>1\*</sup>  | Ella Browning<sup>1,2\*</sup>  | Paul Glover-Kapfer<sup>3,4</sup> | Kate E. Jones<sup>1,2</sup> 

<sup>1</sup>Department of Genetics, Evolution and Environment, Centre for Biodiversity and Environment Research, University College London, London, UK

<sup>2</sup>Institute of Zoology, Zoological Society of London, London, UK

<sup>3</sup>WWF-UK, Living Planet Centre, Woking, UK

<sup>4</sup>Flora & Fauna International, David Attenborough Building, Cambridge, UK

**Correspondence**

Kate E. Jones

Email: kate.e.jones@ucl.ac.uk

**Funding information**

Engineering and Physical Sciences Research Council, Grant/Award Number: EP/K503745/1; World Wildlife Fund; Natural Environment Research Council, Grant/Award Number: NE/P016677/1

Handling Editor: Luca Börger

**Abstract**

1. High-throughput environmental sensing technologies are increasingly central to global monitoring of the ecological impacts of human activities. In particular, the recent boom in passive acoustic sensors has provided efficient, noninvasive, and taxonomically broad means to study wildlife populations and communities, and monitor their responses to environmental change. However, until recently, technological costs and constraints have largely confined research in passive acoustic monitoring (PAM) to a handful of taxonomic groups (e.g., bats, cetaceans, birds), often in relatively small-scale, proof-of-concept studies.
2. The arrival of low-cost, open-source sensors is now rapidly expanding access to PAM technologies, making it vital to evaluate where these tools can contribute to broader efforts in ecology and biodiversity research. Here, we synthesise and critically assess the current emerging opportunities and challenges for PAM for ecological assessment and monitoring of both species populations and communities.
3. We show that terrestrial and marine PAM applications are advancing rapidly, facilitated by emerging sensor hardware, the application of machine learning innovations to automated wildlife call identification, and work towards developing acoustic biodiversity indicators. However, the broader scope of PAM research remains constrained by limited availability of reference sound libraries and open-source audio processing tools, especially for the tropics, and lack of clarity around the accuracy, transferability and limitations of many analytical methods.
4. In order to improve possibilities for PAM globally, we emphasise the need for collaborative work to develop standardised survey and analysis protocols, publicly archived sound libraries, multiyear audio datasets, and a more robust theoretical and analytical framework for monitoring vocalising animal communities.

**KEYWORDS**

acoustic indices, bioacoustics, biodiversity monitoring, deep learning, ecoacoustics, ecological monitoring, machine learning, passive acoustic monitoring

\*Denotes joint first authorship.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2018 The Authors. *Methods in Ecology and Evolution* published by John Wiley & Sons Ltd on behalf of British Ecological Society.

## 1 | INTRODUCTION

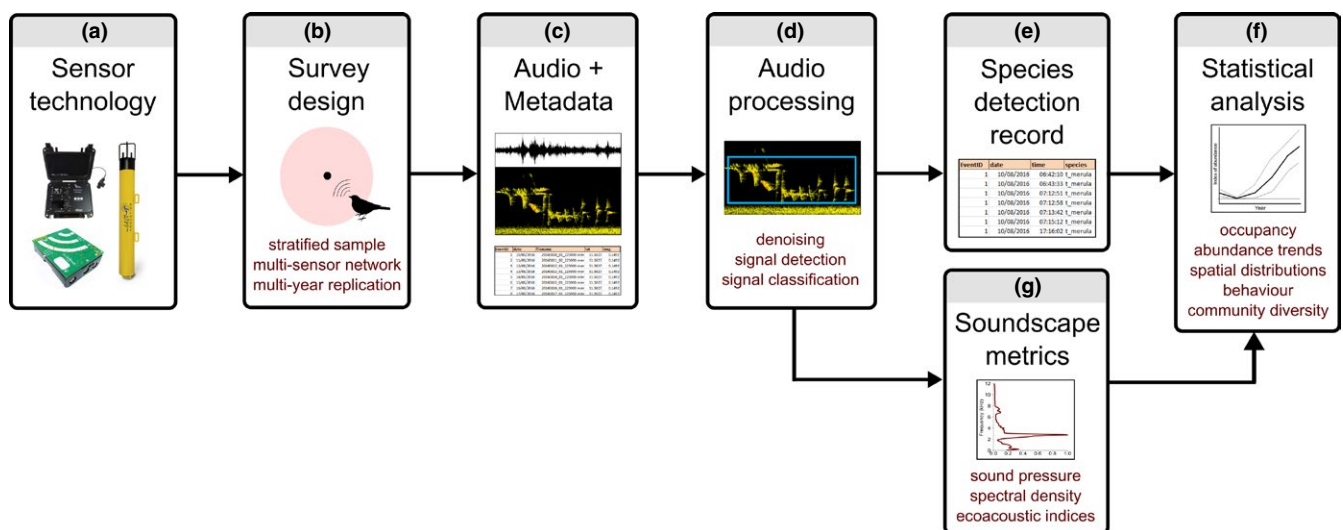
There is a growing need for cost-effective, scalable ecological monitoring techniques, in light of global declines in biodiversity (Cardinale et al., 2012). Alongside addressing fundamental ecological questions, survey and monitoring data are essential in evaluating trends and drivers of population change, informing conservation planning and efficacy assessment, and addressing biodiversity policy commitments (Honrado, Pereira, & Guisan, 2016). Traditional survey methods (e.g., manual counts, trapping) are limited by being resource intensive and invasive, but are now complemented by a suite of high-throughput sensing technologies including satellite sensing, LIDAR, and camera traps. Passive acoustic sensors have become an increasingly important component of this survey toolbox. Many animals emit acoustic signals that encode information about their presence and activities (Bradbury & Vehrencamp, 1998). Sound is also an important feature of the sensory environment, and anthropogenic acoustic phenomena are a critical yet understudied dimension of global change (e.g., Buxton et al., 2017).

Opportunities to acoustically survey wildlife and environments have historically been limited by technological costs and constraints, but this situation is fast improving. For example, the recently released AudioMoth low-cost sensor has seen broad uptake for study objectives ranging from population ecology to anthropogenic activity (Hill et al., 2018). Such initiatives now enable deployment of multisensor networks at scale, involving both experts and volunteers (Jones et al., 2013; Newson, Evans, & Gillings, 2015). Passive acoustic monitoring (PAM) is thus increasingly suited to objectives-driven survey and monitoring programmes, whose protocols must be standardisable, scalable, and financially sustainable (Honrado et al., 2016). However, the resulting massive audio datasets still present formidable logistical and analytical difficulties, and it remains unclear how effectively current PAM methodologies, which have mostly been developed in small-scale, taxonomically focused contexts (mostly bats

and cetaceans), can translate to the broader challenges of acoustic biodiversity monitoring. In this review, we synthesise current research to highlight emerging opportunities and critical knowledge gaps. We discuss current applications of PAM technologies, identify challenges and research priorities at each stage of the PAM pipeline (Figure 1), and lastly discuss significant emerging trends for PAM in ecological research.

## 2 | PASSIVE ACOUSTICS APPLICATIONS IN ECOLOGY

Many animals actively produce sound for communication, and echolocating species also emit sounds for navigation and prey search (Bradbury & Vehrencamp, 1998). Vocalising animals thus leak information into their surroundings regarding their presence, behaviour, and interactions in space and time (Kershenbaum et al., 2014). Long-established acoustic survey methods, for example, bird or amphibian point counts, typically involve experienced surveyors identifying species in the field (Gregory, Gibbons, & Donald, 2004). In contrast, PAM involves recording sound using passive acoustic sensors (recorders, ultrasound detectors, microphones and/or hydrophones; henceforth “acoustic sensors”) (Blumstein et al., 2011) and subsequently deriving relevant data from audio (e.g., species detections, environmental sound metrics) (Bittle & Duncan, 2013; Digby, Towsey, Bell, & Teal, 2013; Merchant et al., 2015) (Figure 1). Passive acoustics approaches have long been applied to studying visually cryptic animals such as cetaceans and echolocating bats (Nowacek, Christiansen, Bejder, Goldbogen, & Friedlaender, 2016; Walters et al., 2013), but in recent years their scope has expanded with the arrival of purpose-designed acoustic sensors. These are noninvasive, autonomous, usually omni-directional (sampling a three-dimensional sphere around the sensor), and offer the advantage of a larger detection area and fewer taxonomic restrictions than camera traps (which



**FIGURE 1** A typical passive acoustic monitoring workflow

are usually limited to detecting larger birds and mammals at close range) (Lucas, Moorcroft, Freeman, Rowcliffe, & Jones, 2015). As such, they can simultaneously survey entire vocalising animal communities and their acoustic environments (Wrege, Rowland, Keen, & Shiu, 2017).

Species detections derived from PAM are analogous to other forms of survey data, with applications ranging from species occupancy estimation to biodiversity assessment (detailed in Table 1). Their benefits over traditional surveys include continuous surveying for long periods with low manual effort, and the associated higher likelihood of detecting rarer or less vocally active species (Klingbeil & Willig, 2015). Standardised post hoc analysis also avoids the skill level biases in species identification that often impact citizen science data (Isaac, van Strien, August, de Zeeuw, & Roy, 2014). Conversely, current limitations of PAM data include their unsuitability for studying nonacoustic species, and the inability to identify individual calling animals for most taxa (in contrast to visual recognition or mark-recapture).

Beyond supporting established survey approaches, PAM also offers unique possibilities, including study of vocalising behaviour, intraspecific variability in call repertoire, and the evolution of acoustic communities (Blumstein et al., 2011; Linhart & Šálek, 2017; Prat, Taub, & Yovel, 2016; Tobias, Planqué, Cram, & Seddon, 2014); animal responses to the acoustic environment (Nowacek et al., 2016; Simpson, Meekan, Jeffs, Montgomery, & McCauley, 2008); and monitoring of anthropogenic phenomena such as sound pollution, blast fishing, and poaching (Astaras, Linder, Wrege, Orume, & Macdonald, 2017; Braulik et al., 2017) (Table 1). There is a rich literature on the effects of anthropogenic noise on cetacean and increasingly avian populations and behaviour (e.g., Pirotta, Merchant, Thompson, Barton, & Lusseau, 2015; Proppe, Sturdy, & St. Clair, 2013). Sensor networks can monitor ecosystems over large geographical and temporal scales, facilitating the characterisation of acoustic communities across habitats and biomes and the development of putative acoustic biodiversity indices (Nedelec et al., 2015; Sueur, Farina, Gasc, Pieretti, & Pavoine, 2014; Sueur, Pavoine, Hamerlynck, & Duvail, 2008) (Table 1). Researchers are also now starting to explore the opportunities afforded by archived audio datasets collected over years or decades, often by volunteers or multiple research groups (Jones et al., 2013; Van Parijs et al., 2015). For example, bat monitoring data have been repurposed to study orthoptera in the United Kingdom and France (Newson, Bas, Murray, & Gillings, 2017; Penone et al., 2013) and predict impacts of urban planning on bats (Border, Newson, White, & Gillings, 2017). Long-term datasets offer complex insights into population ecology, behaviour, and human impacts which, particularly for cryptic species, can otherwise be difficult to achieve (e.g., forest elephants; Wrege et al., 2017). Such archives could also contribute much-needed species data to global repositories for biodiversity modelling and monitoring (e.g., Global Biodiversity Information Facility).

### 3 | PASSIVE ACOUSTIC SENSOR TECHNOLOGIES AND SURVEY APPROACHES

#### 3.1 | Passive acoustic sensor hardware

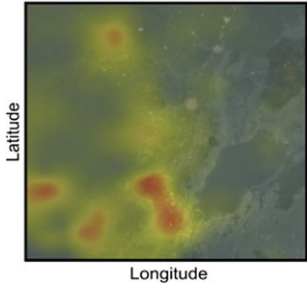
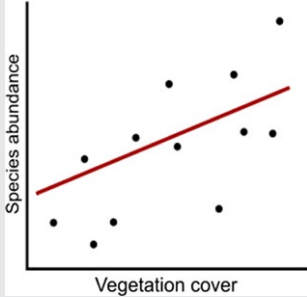
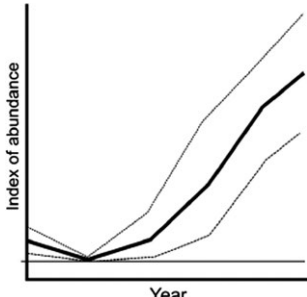
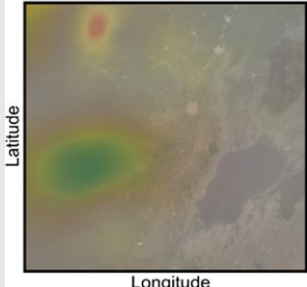
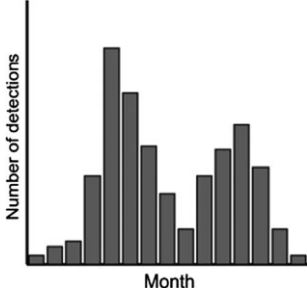
In contrast to early PAM studies that repurposed field recorders (Riede, 1993) or naval or seismological equipment (Sousa-Lima, Fernandes, Norris, & Oswald, 2013), commercial acoustic sensors are now comparable to camera traps in durability and user-accessibility (Figure 1a). Improved battery life and storage, on-board metadata collection and programmable schedules allow for extended autonomous deployments with flexible sampling regimes (Aide et al., 2013; Baumgartner et al., 2013). However, hardware costs have limited scalability, with ubiquitous models such as Wildlife Acoustics Song Meters often substantially more expensive than equivalent-spec camera traps. When synchronous multisensor surveys are unnecessary, one common solution is repeated redeployment of a handful of sensors, for example, the Norfolk Bat Survey loan out ultrasonic detectors to hundreds of volunteers (Newson et al., 2015).

Looking forward, emerging open-source, microcomputer-based sensors are significantly cheaper than commercial alternatives (Sethi, Ewers, Jones, Orme, & Picinali, 2018; Whytock & Christie, 2017). For instance, the AudioMoth can be mass-produced to reduce unit cost to around US\$30 (Hill et al., 2018), thereby drastically lowering the initial financial barriers to large multisensor surveys, although maintenance costs (e.g., regular replacement of batteries and SD cards) may substantially increase in larger projects. Furthermore, in some cases the use of inexpensive components (e.g., microelectromechanical systems (MEMS) microphones) might involve trade-offs between sensor cost and data quality, for example, if these show inconsistent frequency response, lower signal-to-noise ratios, or are vulnerable to environmental damage. A critical open question concerns how much data quality can be sacrificed without compromising the ability to derive sufficient information from audio (i.e., accurate species identification) (e.g., Figure 2). Addressing this question requires comparative analyses of data collected simultaneously with different sensor models (Adams, Jantzen, Hamilton, & Fenton, 2012), and the answer may vary taxonomically since certain species are intrinsically harder to distinguish acoustically than others (see below in “Automated sound identification”) (Kershenbaum et al., 2014).

#### 3.2 | Survey design and data standardisation

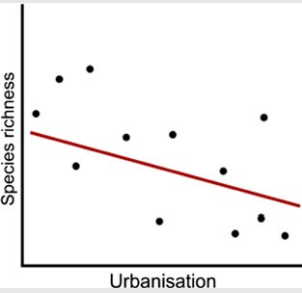
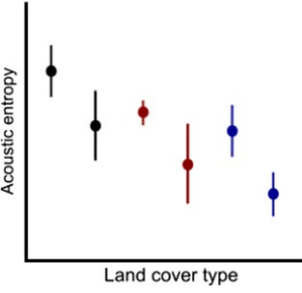
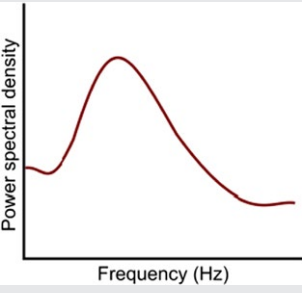
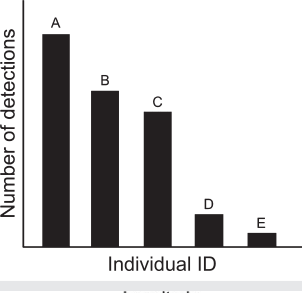
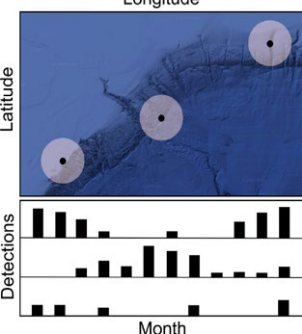
Understanding the comparability of audio data collected using different sensor models and sampling protocols, across different environments, is an ongoing challenge (Figure 1b) (Browning, Gibb, Glover-Kapfer, & Jones, 2017). As well as transect surveys (Jones et al., 2013), PAM studies now commonly deploy static sensors (analogous to camera traps) either standalone, in multisensor networks, or in linked arrays to allow for sound localisation (reviewed in Blumstein et al., 2011). The most appropriate combination of sensor

**TABLE 1** Ecological applications of passive acoustic monitoring

Analysis	Data type	Example result	Example applications	Key challenges
Occupancy	Presence/absence, single species		Species inventories (MacSwiney G 2008). Spatial trends in species occupancy (e.g., endangered or data-deficient species) and relationship with environmental covariates (Campos-Cerqueira & Aide, 2016; Kalan et al., 2015)	Minimising error rates in species call ID
Abundance/density estimation	Spatially and temporally explicit detection counts, single species		Estimating density and abundance of monitored species, and relationship to environmental covariates (Lucas et al., 2015; Marques et al., 2013)	Minimising error rates in species call ID. Individuals cannot be identified, so abundance estimates must account for nonindependence of detected calls
Temporal abundance trends	Detection counts, single species (per replicate survey)		Monitoring endangered or indicator species (Jones et al., 2013). Estimating abundance trends from multiyear monitoring data (e.g., generalised additive models, Barlow et al. 2015)	Minimising error rates in species call ID. Difficult to estimate the true relationship between detection rate and animal abundance from acoustic data only
Spatial/temporal behaviour trends	Detection counts of different behaviours, single species		Modelling relationship between behaviour, habitat covariates and/or the acoustic environment (e.g., anthropogenic noise) (Wrege et al., 2017)	Minimising error rates in species and behavioural call ID. Poor availability of automated tools for differentiating acoustic behaviours
Phenology/activity patterns (species)	Temporally explicit detection record, single species		Monitoring circadian and seasonal trends in behaviour, e.g., migration timing (Petrušková et al., 2016)	Minimising error rates in species call ID

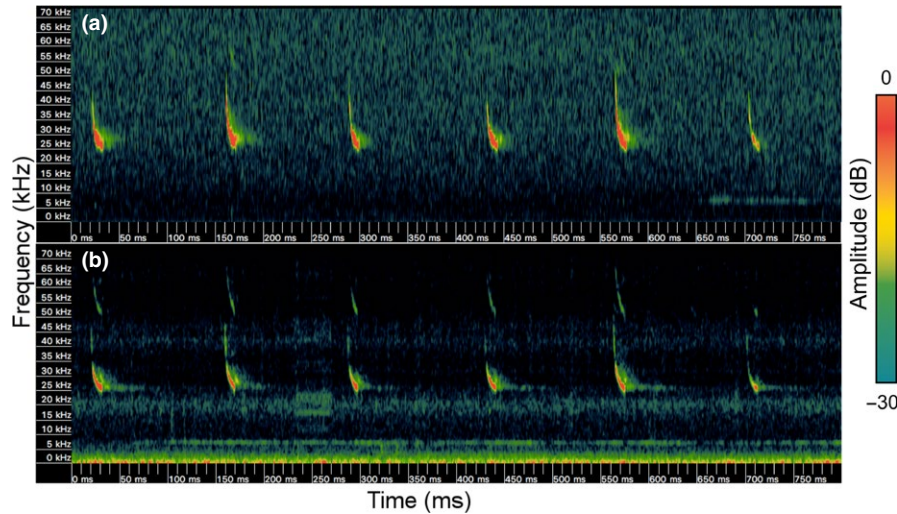
(Continues)

**TABLE 1** (Continued)

Analysis	Data type	Example result	Example applications	Key challenges
Species richness	Presence/absence, multiple species		Relationships between species richness and habitat covariates	Automated call ID tools and reference call libraries are currently unavailable for most taxa and regions
Acoustic community diversity	Acoustic indices (e.g., complexity, entropy, diversity, NDSI)		Measuring spatiotemporal trends in acoustic indices as proxies for community diversity, e.g., relationship between indices and habitat, or community vocalising phenology (Nedelec et al., 2015; Sueur et al., 2014)	Relationships between index values and community diversity poorly understood. Indices are strongly sensitive to variation in nonbiotic sound (e.g., from anthropogenic sources)
Environmental sound	Metrics of sound pressure and spectral density; also acoustic indices (e.g., NDSI)		Measuring the acoustic environment (e.g., anthropogenic sound) and relationships with wildlife abundance and behaviour (Merchant et al., 2015; Pirota et al., 2015)	More complex metrics (i.e., acoustic indices) may be sensitive to variation in different sound types (e.g., weather)
Intraspecific individual identification	Detection counts, identified to individual by differences in call structures or repertoire		Study of individual call repertoires, social behaviour, or facilitating density estimation, e.g., in birds and cetaceans (King et al., 2013; Petrusková et al., 2016)	Currently not possible for most species, due to limited reference data and/or poor knowledge of individual variation in calls/repertoire
Large or real-time sensor network	Detection counts, collated or transmitted from multiple sensors		Seasonal and spatial distributions of species or behaviour (Davis et al., 2017). Real-time monitoring of species occurrence or anthropogenic activity (Astaras et al., 2017)	Costs of data storage and transmission infrastructure. For real-time monitoring, patchy availability of automated analysis tools and data transmission capacity

type and survey design will depend on a study's taxonomic focus, environmental realm (terrestrial or marine), spatial scale, and objectives (Figure 1; Table 1) (Van Parijs et al., 2009), but the advantages

and disadvantages of different acoustic survey designs are often poorly understood. For example, it is mostly unknown whether certain subsets of species are systematically over- or underrepresented



**FIGURE 2** Example comparison of recordings between sensor models. Spectrograms show simultaneous ultrasonic recordings from two co-deployed sensors: a commercial device with electret microphone (Batlogger M, Elekon AG) (a), and a low-cost model with MEMS microphone (AudioMoth) (b). Spectrograms show time (x-axis), frequency (y-axis), and amplitude on a linear colour scale (amplitude normalised with peak at 0 dB, values below  $-30$  dB shown in black for visualisation). Bat echolocation calls are bright patches between 20 and 60 kHz. The comparison highlights differences in frequency sensitivity, with higher frequencies more consistently resolved in (a), and larger amounts of low- to midrange background noise in (b)

by different survey techniques, as recently shown for rare bat species when using mobile transects (Braun de Torrez, Wallrichs, Ober, & McCleery, 2017). Equally, while transects or sparsely deployed static sensors may suffice for occupancy or activity estimation, modelling abundance, activity or space-use at finer scales may require denser networks of calibrated static sensors, often combined with additional parameters such as species call detection distances (Jaramillo-Legorreta et al., 2016; Lucas et al., 2015).

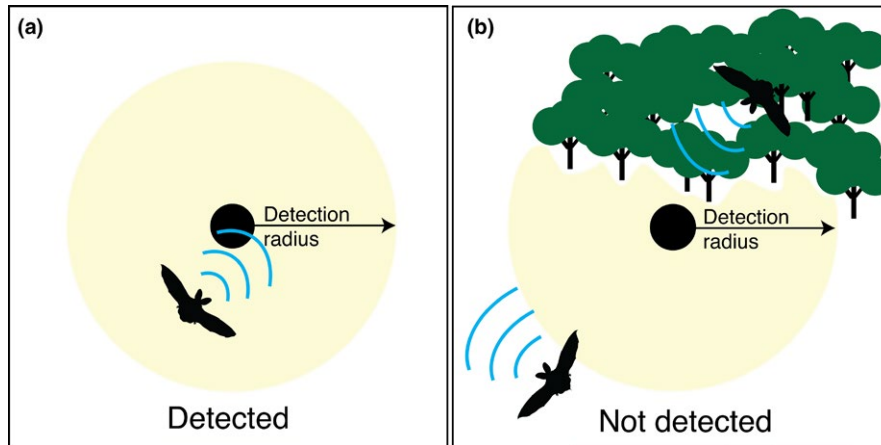
Sound waves attenuate as they travel through the environment, until at a certain distance from the caller they are no longer detectable above ambient background noise. This distance varies depending on the sound's amplitude and frequency (higher frequencies attenuate more rapidly), the environmental medium (sound velocity in seawater is over four times greater than in air), the caller's position relative to the sensor (e.g., differences in depth underwater) and environmental features such as vegetation, topography, bathymetry, temperature, and pressure (Farcas, Thompson, & Merchant, 2016) (Supporting Information Appendix S1). Sounds can also be masked by nontarget sound, from anthropogenic sources as well as other vocalising animals. The effective sampling area around an acoustic sensor therefore varies among species and call types, and across space and time (Figure 3). If unaccounted for, any resulting detection biases (e.g., towards animals that call at higher amplitudes and/or lower frequencies) may cause biased population or diversity estimates.

Although previously often overlooked in the PAM literature, there are now increasing efforts to systematically quantify sources of bias and improve survey standardisation. These include sensor calibration guidelines (Merchant et al., 2015), metadata standards (Roch et al., 2016), assessing the efficacy of sampling designs (Braun de Torrez et al., 2017; Froidevaux, Zellweger, Bollmann, & Obrist, 2014; Van Parijs et al., 2009), quantifying sensitivity

differences between sensor models and over time due to environmental degradation (Adams et al., 2012; Merchant et al., 2015), and quantifying effects of sensor proximity to habitat features (e.g., vegetation, water surface, topography) on sound detection (Darras, Pütz, Fahrurrozi, Rembold, & Tscharnke, 2016; Farcas et al., 2016). Ultimately, these efforts should facilitate more robust, data-driven approaches to analysing large, multisensor acoustic datasets, which currently tend to assume constant species detectability over space and time (e.g., Davis et al., 2017; Newson et al., 2015).

### 3.3 | Trade-offs in audio recording and data storage

During digital sound recording, incoming sound waves are transduced into an electrical signal that is recorded at a specified sampling rate (in Hz) and bit-depth (number of bits per sample). These parameters determine a recording's frequency (pitch) and amplitude (volume) resolution, with much higher sampling rates required to resolve ultrasonic frequencies (those above human hearing range;  $>20,000$  Hz) compared to audible range frequencies (20–20,000 Hz) (Supporting Information Appendix S1). The conventional sampling rate for audible sound (44.1 kHz) produces relatively manageable file sizes (c. 5 MB per minute in 16bit mono), but recording full-spectrum ultrasound in bat and cetacean surveys (sampling rates often  $>200$  kHz) produces very large files, resulting in a trade-off between data quality and storage capacity. Some ultrasound detectors use less data-intensive recording methods based on frequency division, which divide the incoming signal frequency by a specified factor; their lower storage requirements may suit extended or remote deployments, provided sufficient information can be derived from the data (e.g., Jaramillo-Legorreta et al., 2016). However, resulting losses of frequency and amplitude information can impact the discrimination of species and



**FIGURE 3** The effects of distance from sensor, call amplitude, and habitat clutter on vocalising animal detectability. Sound emitted within the detection space (yellow radius) of a sensor (black circle) are successfully recorded (a), whereas sounds outside this radius are missed (b). Habitat clutter causes acoustic interference, particularly for higher-frequency sounds (e.g., ultrasonic echolocation), and may decrease detection probability (b). Figure modified with permission from Browning et al., (2017)

behaviours (Adams et al., 2012; Walters et al., 2013). In future, these analytical tools may become less sensitive to recording method, but currently ensuring minimal information loss during recording and storage (Supporting Information Appendix S1) both facilitates species identification (Walters et al., 2012) and futureproofs the data by allowing for later reanalysis with improved tools.

Crucially, recording and storing audio at sufficient quality (Figure 1c), alongside detailed metadata on surveys, sensor type and recording parameters, also provides opportunities to address additional questions. For example, a recent study collated multi-year hydrophone data to estimate the distribution of the critically endangered North Atlantic right whale *Eubalaena glacialis* (Davis et al., 2017). Leveraging decades of PAM survey data will require collaborative development and maintenance of web infrastructure for the collation and public archiving of massive (multi-gigabyte to petabyte) environmental audio datasets (e.g., <https://ngdc.noaa.gov/mgg/pad/>). Another possible solution to data capacity issues could be to reduce the amount of audio that is stored, for example, by applying on-board thresholds or algorithms that only trigger recording when potential sounds of interest are present (Baumgartner et al., 2013; Hill et al., 2018). Discarding audio data is scientifically undesirable, but some degree of prior filtering can prevent datasets becoming unmanageably large, and combined with wireless data transmission (Aide et al., 2013) could facilitate real-time ecological monitoring and reporting.

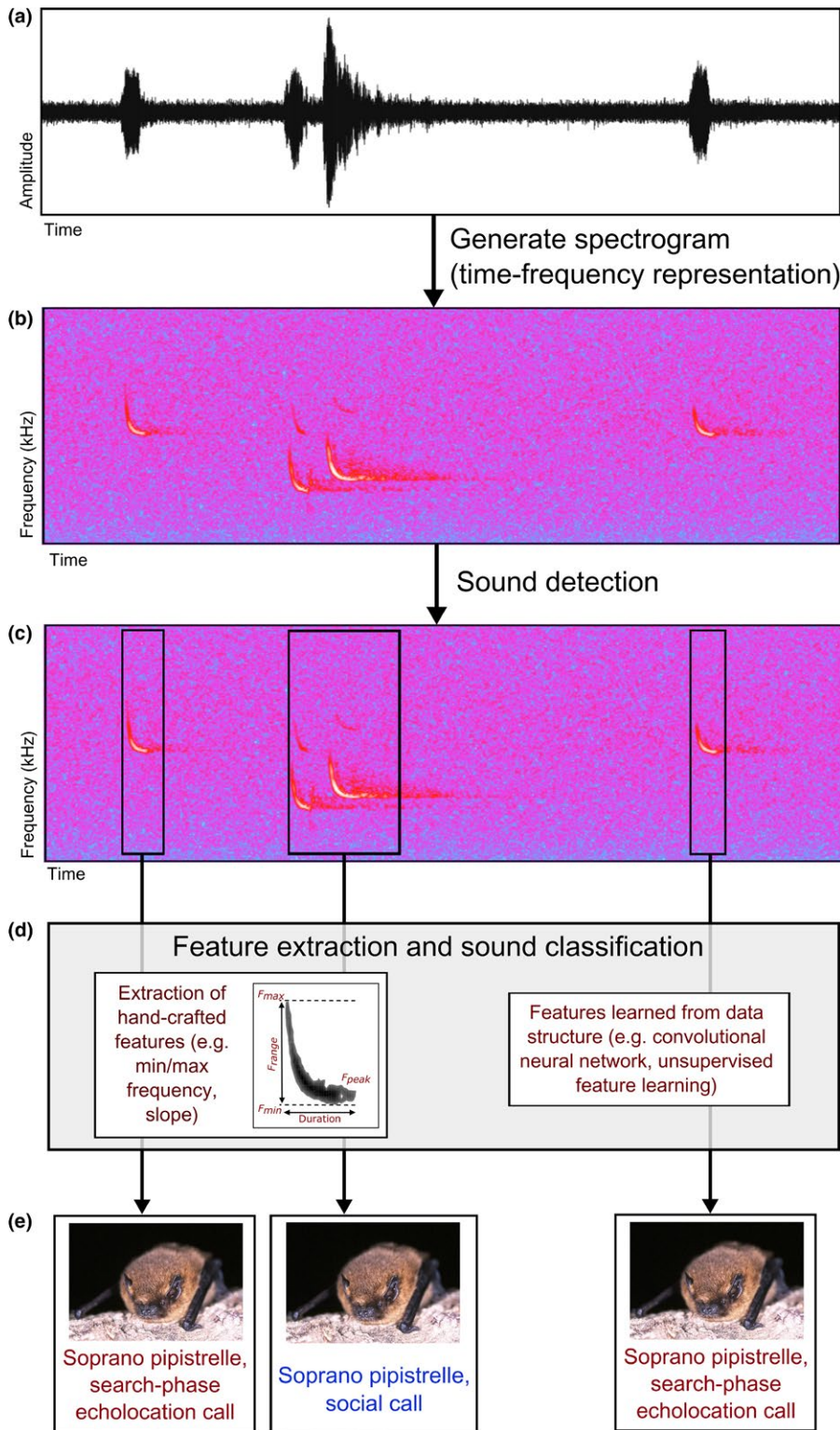
## 4 | DETECTING AND CLASSIFYING ACOUSTIC SIGNALS WITHIN AUDIO DATASETS

For studies focusing on specific species or taxonomic groups, target sounds must be identified from recordings (Aide et al., 2013; Salamon & Bello, 2015), which requires pipelines to process sound files and metadata and output useful annotations (e.g.,

calling animal species, location, precise date/time) (Figure 1d,e). Conducted manually, this process is time-consuming and subjective, and it is difficult to quantify biases related to analyst knowledge level, which may be particularly problematic in resource-limited conservation settings (Heinicke et al., 2015; Kalan et al., 2015). Efficient automated systems are therefore prerequisites for scaling up PAM studies, with innovations in machine learning increasingly applied to bioacoustic signal recognition (Aide et al., 2013; Bittle & Duncan, 2013; Heinicke et al., 2015; Walters et al., 2012). The complexity of environmental audio offers a useful real-world test for new methods, and the involvement of the machine learning and computer vision communities in PAM is driving analytical advances that benefit ecologists (Goeau, Glotin, Vellinga, Planque, & Joly, 2016; Marinexplore, 2013; Stowell & Plumbley, 2014; Stowell, Wood, Stylianou, & Glotin, 2016).

### 4.1 | Developing a pipeline for automated sound identification

A pipeline for automatically identifying target sounds within audio recordings (hereafter referred to as “*automated sound identification*” or “*auto-ID*”) involves several stages (Figure 4). Audio waveforms are commonly preprocessed to recover frequency information and produce a time-frequency-amplitude representation (*spectrogram*) (Figure 4a,b), usually via Fourier analysis or similar techniques (Supporting Information Appendix S1). Relevant sounds must first be *detected*, that is, located in time within the recording (a task sometimes alternatively termed “segmentation”) (Figure 4c), using methods ranging in complexity from simple thresholding to complex statistical models (Table 2). Next, detected sounds are typically *classified* to a relevant category (e.g., species, call type) (Figure 4d,e) based on a combination of spectro-temporal features extracted from the sound. These features may be generic (e.g., Mel-frequency cepstral coefficients) (Muda, Begam, & Elamvazuthi, 2010), but are often hand-crafted



**FIGURE 4** An automated sound detection and classification pipeline. Frequency information is recovered from the sound waveform (a), generating a time-frequency representation (spectrogram, with amplitude shown as colour intensity) (b). Sounds of interest are detected (c), features are extracted (d), then calls are classified to a category (species and call type, here either bat echolocation or social calls) (e). Figure modified with permission from (Browning et al., 2017).

Photo © Hugh Clark/www.bats.org.uk, reproduced with permission

to facilitate species discrimination (e.g., peak frequency, call duration, peak amplitude) (Baumgartner & Mussoline, 2011; Walters et al., 2012) (Figure 4d). Sounds are classified using either supervised (previously trained on expert-labelled sound libraries) or unsupervised (based on the structure within the data) algorithms, which return the estimated likelihood that a sound belongs to its assigned category (Table 2).

Although methods are fast improving, poor or variable accuracy of auto-ID tools remains a major issue. In particular, the detection stage presents formidable difficulties (Stowell et al., 2016). In real-world PAM audio this process frequently involves distinguishing large numbers of spectrally and temporally overlapping calls, emitted by multiple vocalising species in acoustically heterogeneous settings (e.g., birds in the dawn chorus, swarming bats), which is an



**TABLE 2** Signal detection and classification techniques commonly used in bioacoustic analysis

Method	Application	Summary	Advantages	Disadvantages	Example references
Thresholding	Detection	Detection occurs when energy within specified frequency band(s) exceeds a specified threshold	Computationally inexpensive; does not require large training datasets	Often sensitive to nontarget background noise and signal overlap	Digby et al., 2013
Spectrogram cross-correlation	Detection, classification	Detection occurs when correlation coefficient against a template spectrogram exceeds a specified value (e.g., 0.9)	Computationally inexpensive; does not require large training datasets	Relies on sufficiently representative template data	Aide et al., 2013
Hidden Markov models	Detection, classification	Probabilistically infers whether a signal of interest is present, based on an underlying multistate model	Incorporates temporal detail on signal/sequence	Complex for nonexperts to develop. Requires sufficient training/reference data	Zilli, Parson, Merrett, & Rogers, 2014
Supervised learning with prior feature extraction (e.g., support vector machines, random forest)	Classification	Supervised algorithms classify unknown signals based on their similarity to previously learned training data (expert-verified call libraries)	Can be trained on large and varied reference datasets	Poor availability of verified call libraries for many taxa. Feature extraction methods are often noise-sensitive	Bittle & Duncan, 2013; Walters et al., 2012
Unsupervised learning (e.g., clustering algorithms)	Classification	Groups signals based on the similarity of their features, using unsupervised (clustering) algorithms	Does not require training data, as clustering is based on variation within the survey dataset	Does not leverage prior knowledge, and clusters must subsequently be identified to a useful category (e.g., species)	Pirotta et al., 2015
Supervised learning without prior feature extraction (e.g., convolutional neural networks)	Detection, classification	Signals detected and classified based on similarity to a learned training dataset	Distinguishing features learned directly from spectrogram data, so bypasses noise-sensitive feature extraction stage	Sensitive to overfitting to training data, so requires very large training datasets to account for within-class variability and variable background sound	Goeau et al., 2016; Mac Aodha et al., 2018

extremely challenging task for most extant algorithms. Nontarget environmental, biotic, and anthropogenic sounds can mask target sounds or generate false positives (Heinicke et al., 2015; Salamon & Bello, 2015; Stowell, Stylianou, Wood, Pamuła, & Glotin, 2018), although there is evidence that prior noise reduction filtering can improve accuracy (reviewed in Stowell et al., 2016). Even when detection precision is high (few false positives), state-of-the-art methods regularly fail to distinguish faint, transient or partially masked calls, leading to high false-negative rates (low recall) (Digby et al., 2013; Goeau et al., 2016). At the classification stage, robust feature extraction is crucial to classification accuracy, but is similarly sensitive to factors including caller distance, background noise, and temporal overlap between calls (Stowell & Plumbley, 2014). Species classification may be intrinsically more difficult for taxa with highly variable vocal repertoires (e.g., birds, cetaceans) relative to those with more intraspecifically consistent call structures (e.g., bat echolocation calls) (Kershenbaum et al., 2014; Walters et al., 2012). Classification

may be further complicated by ecological context, for example, differences in vocalising behaviour in response to environment or conspecifics, or the co-occurrence of species with similar call structures, such as bats of the genus *Myotis* or sympatric right and humpback whales (Van Parijs et al., 2009; Walters et al., 2012).

The often substantially poorer performance of detection and classification algorithms on target audio recorded in novel contexts (e.g., difficult sensor models or more background noise than the training data), is a critical emerging problem as data collection capacities continue to grow (Stowell et al., 2018). In ecology, auto-ID tools are commonly developed for study-specific objectives and trained on data representative of the actual survey dataset, thereby avoiding this issue of transferability (e.g., Campos-Cerqueira & Aide, 2016; Heinicke et al., 2015). However, algorithm development is time-consuming and prohibitively complex for nonexpert users. Both proprietary (e.g., Raven Pro, Avisoft, Kaleidoscope, ARBIMON) and

open-source software or freeware (e.g., PAMGUARD, LFDSC, iBatsID, Tadarida) (Aide et al., 2013; Bas, Bas, & Julien, 2017; Baumgartner & Mussoline, 2011; Gillespie et al., 2009; Walters et al., 2012) offer a growing range of inbuilt auto-ID tools for large taxonomic groups and geographical regions. Although user-friendly, their transferability to novel datasets remains unclear, and there are clear risks of relying on costly, closed-source tools whose underlying methods are poorly reported. Looking forward, an achievable priority is the community development and adoption of gold standard, publicly archived bioacoustic sound libraries, to use as benchmarks for comparative testing of new and closed-source algorithms.

PAM workflows therefore involve a time-accuracy trade-off: manual processing is often most accurate but can be subjective and slow, whereas fully automated processing is much faster but error-prone (Digby et al., 2013). Currently, large PAM analyses are usually semi-automated at best, involving regular manual cross-checking (Campos-Cerqueira & Aide, 2016; Kalan et al., 2015) and resolving ambiguous classifications using expert opinion or rules of thumb (e.g., selecting the most likely species based on other calls in close temporal proximity). Newer machine learning techniques that account for other surrounding calls (e.g., recursive neural networks) could facilitate the automation of this process. Current auto-ID systems are nonetheless improving processing efficiency, for example, by filtering out detections below a minimum probability threshold (adjustable depending on study objectives) to reduce the volume of data for manual inspection.

## 4.2 | Emerging innovations in sound identification

Looking forward, several emerging methods are substantially improving detection and classification accuracies by learning representations from spectrogram data, such as unsupervised feature extraction (Salamon & Bello, 2015; Stowell & Plumbley, 2014) and dynamic time warping based feature representations (Stathopoulos, Zamora-Gutierrez, Jones, & Girolami, 2017). Deep convolutional neural networks (CNNs) are particularly promising, since these can learn discriminating spectro-temporal information directly from annotated spectrograms (bypassing a separate feature extraction stage), improving their robustness to sound overlap and caller distance (Goeau et al., 2016) (Figure 4d). In recent tests, CNNs have markedly outperformed alternative methods on detection and classification of biotic and anthropogenic sounds in urban recordings (Fairbrass et al., 2018; Salamon & Bello, 2016) and animal calls in noisy monitoring datasets (Goeau et al., 2016; Mac Aodha et al., 2018; Marinexplore, 2013). Their performance in more complex tasks that involve distinguishing multiple overlapping vocalisations (e.g., songs in the dawn chorus) has not yet been tested, although their success in similarly challenging computer vision and individual human voice recognition tasks is a promising sign (e.g., Lukic, Vogt, Dürr, & Stadelmann, 2016). However, currently such applications in ecology are constrained by CNN sensitivity to overfitting to training data, and the consequent requirement

for very large training datasets that represent natural variability in species call repertoires, background sound, and caller distance (Krause et al., 2016; Russakovsky et al., 2015). Although more accessible for image or voice classification (e.g., using online images or audio) (Krause et al., 2016), very few such datasets exist for environmental sound, since the practical difficulty of reference data collection means that verified wildlife call libraries, when available, are typically small in size and lack variability in call type, recording quality, and acoustic environment. Some studies have partially addressed this issue by augmenting training data with background noise to simulate different distances and acoustic environments (Salamon & Bello, 2016). Online data labelling projects such as Bat Detective ([www.batdetective.org](http://www.batdetective.org)) and Snapshot Serengeti ([www.snapshotserengeti.org](http://www.snapshotserengeti.org)) have also involved citizen scientists in annotation of CNN training data (Mac Aodha et al., 2018; Norouzzadeh et al., 2017).

Further research to improve the situation could include the development of noise-robust auto-ID methods that perform well even with small and variable quality training datasets (e.g., Kaewtip, Alwan, O'Reilly, & Taylor, 2016), and generalised detection algorithms for entire taxonomic groups (Baumgartner & Mussoline, 2011; Mac Aodha et al., 2018) that could subsequently be coupled to regional species classifiers. Additionally, emerging low-shot and zero-shot visual learning approaches aim to learn classification models from very few examples of a class of interest, reducing the need for large training datasets (e.g., Hariharan & Girshick, 2016). More broadly, the limited understanding of the transferability of extant auto-ID systems emphasises that, irrespective of the underlying algorithms, a critical focus must be on lowering the technical barriers to ecologists developing and testing bespoke tools, for example, via interactive machine learning software (Mac Aodha et al., 2014). Such functionality is beginning to emerge in bioacoustic analysis packages (e.g., Kaleidoscope, ARBIMON, Tadarida) (Aide et al., 2013; Bas et al., 2017).

## 4.3 | Sound libraries and training data: identifying and filling the gaps

Perhaps the most fundamental knowledge gap for PAM is the limited availability of comprehensive, expert-verified species call databases for reference and training data. Much remains unknown about the intra- and interspecific call diversity of even well-studied taxa (Kershenbaum et al., 2014), and ground-truthed call databases are difficult and laborious to assemble, requiring the collection of high-quality audio recordings of animals identified to species either visually or through capture (e.g., Zamora-Gutierrez et al., 2016). Where such verified datasets exist they are biased towards vertebrates (particularly cetaceans, bats, and birds), with especially scarce resources for anurans and invertebrates (Lehmann, Frommolt, Lehmann, & Riede, 2014; Penone et al., 2013) and regions outside Europe and North America, despite the urgent need for tools to facilitate monitoring of subtropical and tropical habitats (Zamora-Gutierrez et al., 2016). These gaps translate into equivalent biases in classifier availability, and to our knowledge no widely available tools

exist for distinguishing intraspecific acoustic behaviours (e.g., social from echolocation calls in cetaceans and bats) (Figure 4e), although machine learning methods have successfully been applied to analysis of bat acoustic social behaviour (Prat et al., 2016).

Filling these data gaps is a priority for the entire PAM community, which would strongly benefit from collaborative efforts to collect verified call data for neglected taxa and regions (e.g., tropical terrestrial biomes). Additionally, the establishment of centralised sound libraries with consensus data and metadata standards (e.g., date/time of recording, geographic location, recording parameters, sensor position) (Roch et al., 2016), would improve the accessibility and comparability of reference sound libraries. Online databases such as MobySound ([www.mobysound.org](http://www.mobysound.org)) and Watkins Marine Mammal Sound Database ([www.who.edu/watkinsounds](http://www.who.edu/watkinsounds)) for marine mammals, and Xeno-Canto for birds ([www.xeno-canto.org](http://www.xeno-canto.org)) highlight the benefits of adopting open-data approaches in this area, offering rich (albeit not necessarily standardised) training data (Mellinger & Clark, 2006; Sayigh et al., 2016).

## 5 | ACOUSTIC ECOLOGICAL INFERENCE FROM POPULATIONS TO COMMUNITIES

### 5.1 | Inferring population information from acoustic data

Following processing, a typical sound identification pipeline outputs a spatially and temporally explicit record of species call detections (Figure 1e). Population inference from PAM-derived species occurrence or count data presents its own difficulties, since acoustic surveys involve multiple sources of detection uncertainty. The first is imperfect detectability: the probability of successfully detecting a vocalising animal depends on its distance from the sensor, vocalising behaviour, call parameters, and site-specific environmental factors (Darras et al., 2016; Kéry & Schmidt, 2008). The second issue is that species vocalisations recorded in close spatial or temporal proximity are statistically nonindependent since they may come from the same individual (Lucas et al., 2015); for example, detection rates may be artificially inflated by individual animals vocalising close to a sensor for long periods. However, acoustic identification of individuals is currently not possible for most taxa, and where possible (e.g., for some birds, primates, cetaceans, and wolves) usually requires extensive manual analysis (e.g., Clink, Bernard, Crofoot, & Marshall, 2017; Petrusková, Pišvejcová, Kinštová, Brinke, & Petrušek, 2016; Root-Gutteridge et al., 2014). Furthermore, many vocalising animals produce multical sequences (e.g., birdsong phrases, echolocation passes) which must be merged into discrete detections (Jaramillo-Legorreta et al., 2016; Newson et al., 2015). The third major source of uncertainty relates to errors in automated sound identification (Figure 4) (Digby et al., 2013). Predicted detections and classifications below a suitable confidence threshold can be removed prior to modelling, however, site-specific differences in false-positive and -negative rates (e.g., due to environmental noise) may still impact model estimates.

Statistical analyses (Figure 1f) must account for these uncertainties. For example, patch occupancy models are useful tools for

spatially explicit distribution modelling with PAM-derived data, since these incorporate detection probability parameters that can be estimated from repeat surveys (e.g., Campos-Cerqueira & Aide, 2016; Kalan et al., 2015). Also, the emergence of more accessible and less computationally expensive Bayesian inference methods for complex hierarchical and occupancy models is increasingly enabling multiple sources of uncertainty to be incorporated into spatiotemporal models (e.g., Isaac et al., 2014; Ruiz-Gutierrez, Hooten, & Campbell Grant, 2016). Such frameworks can be extended to include, for example, the confidence associated with automated call detections and classifications (Banner et al., 2018).

A core application of ecological survey data is abundance and population trend estimation. Abundance estimation from PAM count data is difficult due to the lack of a simple relationship between call counts and animal density; the last decade has seen a growing toolbox of methods to address this issue (reviewed in Marques et al., 2013). Spatially explicit capture-recapture models (across multisensor arrays and networks) (Stevenson et al., 2015) and other methods that adjust detected call density by the average calling rate of the target species (Thompson, Schwager, & Payne, 2010; Ward et al., 2012) have been shown to provide accurate density estimates when validated against nonacoustic methods. Another recent study developed a generalised extension of a random encounter model (REM) originally designed for camera trap data (Lucas et al., 2015). However, these methods are often data-intensive, requiring the deployment and retrieval of multisensor networks and the estimation of species-specific parameters such as detection distances and average call rates (Lucas et al., 2015). In cetacean studies, call rates are often estimated by tagging animals with acoustic loggers (Johnson & Tyack, 2003), but in terrestrial realms these remain too large to ethically deploy on many species. Estimation of true abundance may be best suited to well-resourced projects with clear, species-focused objectives, rather than broader scope ecological monitoring.

Informed indices of abundance may suffice where these more complex analytical methods are unfeasible. Detection counts within specified sampling periods are often used as proxies for relative density or activity, such as nightly bat detections (Newson et al., 2015) or temporally aggregated click rates in cetacean surveys (Jaramillo-Legorreta et al., 2016). Such approaches generally assume consistent detection between individuals and over time, even though the relationship between detection rates and relative abundance may vary widely between species and habitats (Marques et al., 2013). However, with careful survey design and replication, these issues may be less problematic for estimation of broad-scale activity or occupancy trends.

### 5.2 | Acoustic ecological community and biodiversity assessment

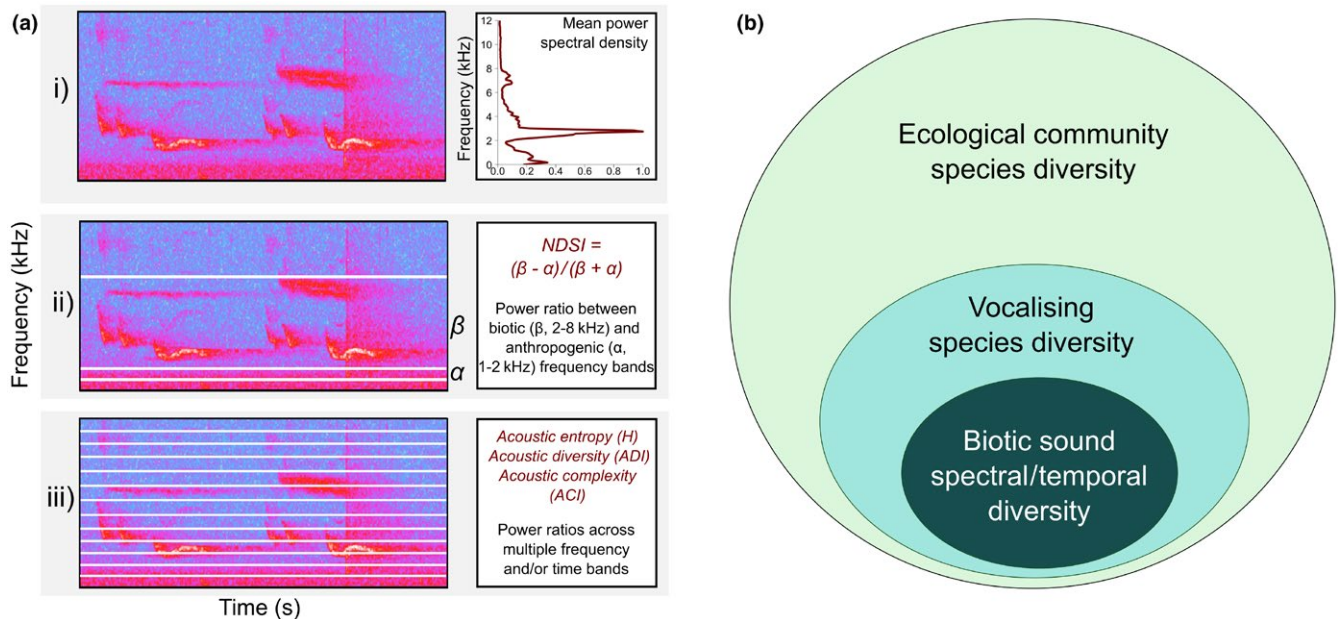
Moving beyond a species focus and towards deriving community information (e.g., species diversity) from PAM data presents the challenge of classifying calls from multiple, or ideally all, vocalising species. For most taxa and geographical regions this is currently

either impossible or extremely time-consuming due to the lack of reference data and auto-ID tools, which emphasises the need for acoustic biodiversity indicators (Figure 1g) to facilitate surveys of data-deficient (often highly biodiverse) regions (Harris, Shears, & Radford, 2016). Monitoring proposed indicator taxa such as bats or orthoptera offers one potential solution (Fischer, Schulz, Schubert, Knapp, & Schmoger, 1997; Jones et al., 2013) but their usefulness as ecological indicators is not clearly established. Recent years have therefore seen the development of soundscape-based methods that seek to infer community information from a habitat's global sound dynamics (Pijanowski, Farina, Gage, Dumyahn, & Krause, 2011) (Figure 5). Under the theme of ecoacoustics, various summary indices have been designed to facilitate comparison of biotic sound between sites and over time (reviewed in Sueur et al., 2014). Most involve calculation of power ratios between multiple frequency and/or time bins across a recording, and thus are essentially more complex extensions of conventional sound pressure and spectral density metrics (Kasten, Gage, Fox, & Joo, 2012; Merchant et al., 2015; Pieretti, Farina, & Morri, 2011; Sueur et al., 2008) (Figure 5). Acoustic indices are derived from the theory that competition for acoustic space between sympatric signalling animals drives the evolution of signal divergence (acoustic niche partitioning), and therefore that the spectro-temporal diversity of biotic sound in a habitat correlates with vocalising species diversity (Pijanowski et al., 2011; Sueur et al., 2008). For example, acoustic entropy and dissimilarity indices are designed as acoustic analogues of classical  $\alpha$ - and  $\beta$ -diversity indices (Sueur et al., 2008).

Despite growing interest in these methods, their results to date have been mixed. Systematic tests in both terrestrial and marine

environments occasionally find correlations between acoustic indices and species diversity, suggesting that soundscape-based metrics can sometimes function as ecological indicators (Gasc, Pavoine, Lellouch, Grandcolas, & Sueur, 2015; Gasc et al., 2013; Harris et al., 2016; Sueur et al., 2008). However, many indices are highly sensitive to site-specific and temporal differences in vocalising animal community composition and nontarget sound levels (e.g., weather, anthropogenic sound, other vocalising species) (Gasc et al., 2015; Lellouch, Pavoine, Jiguet, Glotin, & Sueur, 2014; Staaterman et al., 2017). It is therefore difficult to directly compare acoustic index values between sites and surveys, which limits the reliability of indices in PAM studies that span multiple localities, dates and habitat types. Most ecoacoustics studies to date have focused on relatively undisturbed habitats such as forests, where anthropogenic sound may present fewer problems; in contrast, systematic tests suggest that indices are highly sensitive to heterogeneous urban soundscapes, limiting their suitability for monitoring in cities (Fairbrass, Rennett, Williams, Titheridge, & Jones, 2017). Similarly, there is growing interest in marine soundscape analysis, for instance, in studies of reef phenology (McWilliam, McCauley, Erbe, & Parsons, 2017), the use of acoustic cues by fish (Simpson et al., 2008), mapping biotic sound across oceanic habitats (Nedelec et al., 2015), and development of biodiversity indicators (Sueur et al., 2014). However, these efforts are complicated by the acoustic connectedness of underwater habitats, with long-range sounds and anthropogenic noise potentially swamping local variations in biotic sound (Harris et al., 2016; McWilliam & Hawkins, 2013).

More fundamentally, the theorised link between community and biotic sound diversity remains controversial. The acoustic



**FIGURE 5** Indices of biotic and environmental sound. Conventional metrics such as power spectral density (ai) can measure the acoustic environment. Ecoacoustic indices range from simple power ratios across broad frequency bands (e.g., Normalised Difference Soundscape Index; aii) to finer-band spectral/temporal diversity and entropy (aiii). Their practical applications are limited by poor understanding of the relationships between the diversity of recorded biotic sound, the diversity of vocalising species, and wider community diversity (b)

niche partitioning hypothesis that underpins acoustic indices has rarely been empirically tested, and the sensory, environmental and evolutionary processes that structure vocalising animal communities are poorly understood (Tobias et al., 2014). It remains unclear if and how landscape-scale biotic sound diversity relates to either vocalising species diversity or wider community diversity, and how this relationship varies taxonomically, geographically, and between terrestrial and marine realms (Figure 5b) (Gasc et al., 2013; Harris et al., 2016; Sueur et al., 2014). Despite this lack of clarity, tools for calculation of acoustic indices are increasingly accessible in bioacoustic software packages; similar to auto-ID softwares their outputs should be treated critically, with index values at a minimum ground-truthed against either expert-labelled audio subsets and/or other forms of survey data (e.g., Harris et al., 2016; Sueur et al., 2008). If these practical and theoretical problems can be resolved, acoustic community analyses promise to be one of PAM's unique ecological applications, with potential to offer rich local biodiversity information to complement landscape data from satellite and aerial LIDAR sensing (Bush et al., 2017). For now, leveraging these opportunities will likely require the use of acoustic indices or similar proxies. Ongoing work to improve these prospects could include systematic evaluation of the performance of indices across taxa and habitats (including tests in well-characterised, low-diversity communities), alongside fundamental research into the structure and evolution of acoustic communities (Farina & James, 2016).

Looking forward, newer machine learning methods may offer alternative means to tackle the problem of soundscape monitoring. For instance, a recent study used CNNs to separate and quantify biotic and anthropogenic sound in urban audio, thereby explicitly bypassing the issue of background noise sensitivity (although their transferability to different cities or environments remains unknown) (Fairbrass et al., 2018). Another promising avenue involves unsupervised learning of acoustic patterns directly from survey data. For example, Eldridge, Casey, Moscoso, and Peck (2016) used sparse coding to isolate periodic sound components within bird chorus recordings, which they suggest may correlate with particular sound types or species calls. Although embryonic, such approaches might eventually facilitate estimation of vocalising species diversity without requiring comprehensive auto-ID tools (although reference material would be required to link unsupervised classifications to species). It is still unclear whether this could be feasible, but if so it would represent a major step towards broadly applicable acoustic ecological indicators.

## 6 | EMERGING AND FUTURE OPPORTUNITIES FOR PASSIVE ACOUSTICS

Finally, we outline some major emerging opportunities, as PAM moves beyond proof-of-concept studies towards applications in management and conservation. Until recently, outcomes-driven acoustic monitoring projects have mostly occurred where PAM is

either the only feasible approach, or provides clear advantages over other methods despite higher costs (i.e., bat and cetacean surveys, and field bioacoustics studies). However, low-cost sensors have pushed the bottlenecks into the analysis and management stages, and as we have emphasised, addressing these logistical and analytical barriers now increasingly requires collaborative, community-led efforts. Marine research remains a source of key innovations, including auto-ID software development (Baumgartner & Mussoline, 2011; Gillespie et al., 2009), acoustic sensor tags (Johnson & Tyack, 2003), density estimation methods (Marques et al., 2013), real-time reporting (Baumgartner et al., 2013; <http://dcs.whoi.edu/>), and collation of multisource datasets (Davis et al., 2017). Increased integration between marine and terrestrial PAM communities would be beneficial to jointly addressing pressing challenges, such as standardisation of survey protocols, establishment of publically archived audio datasets and sound libraries, development of an improved theoretical and analytical framework for measuring vocalising animal communities, and research around operationalising PAM data for conservation. There is already promising coordination, for example, via the International Society of Ecoacoustics, and multi-institution initiatives such as the US Northeast Passive Acoustic Sensing Network (NEPAN; Van Parijs et al., 2015).

Currently, we are seeing the arrival of massive acoustic datasets collected across research networks and citizen science programmes (Table 1). As auto-ID tools and wireless data transmission improve, the increasing scope of these datasets could facilitate, for example, the tracking of range shifts under climate change (Davis et al., 2017), long-term studies of population ecology and habitat use (Wrege et al., 2017), year-on-year tracking of population trends (Jaramillo-Legorreta et al., 2016), conservation planning and efficacy assessment (Astaras et al., 2017; Border et al., 2017), behaviour and phenology studies in taxa beyond birds and cetaceans (Nedelec et al., 2015), as well as monitoring of species of concern as ecosystem services providers (e.g., pollinators), pests, invasive species or public health threats (Mukundarajan, Hol, Castillo, Newby, & Prakash, 2017).

Looking further forward, emerging networked sensors and on-board analysis pipelines raise the possibility of using PAM-derived data for real-time monitoring and adaptive management (Table 1). Detections derived from sensor networks can provide highly spatially and temporally detailed data on wildlife activity (e.g., London's Nature-Smart Cities bat monitoring network: <https://naturesmartcities.com>). Real-time data feeds could, for instance, be applied to adjust urban lighting regimes to reduce impacts on bat activity, mitigate human-wildlife conflict, adaptively reroute shipping traffic to avoid threatened cetacean populations (Davis et al., 2017; Van Parijs et al., 2009), or report on illegal logging or hunting (Astaras et al., 2017, Rainforest Connection <https://rfcx.org>). Beyond the institutional and political barriers, developing such an infrastructure would still face substantial technical difficulties, especially since the ultimate goal of developing comprehensive suites of robust auto-ID tools is likely many years or even decades away. Nonetheless, these possibilities represent exciting futures for a technology that, alongside other

sensing technologies, is providing increasingly sensitive insights into the effects of human pressures on wildlife and ecosystems.

## ACKNOWLEDGEMENTS

This research was supported by WWF-UK, NERC (NE/P016677/1), and EPSRC (EP/K503745/1). The authors thank A. Fairbrass, O. Mac Aodha, S. Newson, A. Rogers, A. Hill, P. Prince, N. Tregenza, D. Blumstein, L. Borger, and four anonymous reviewers for discussion and comments on previous versions of the manuscript, and gratefully thank the respondents of our 2016 WWF-UK online survey on best practices in PAM (Browning et al., 2017).

## DISCLOSURE

The authors declare no conflicts of interest.

## AUTHORS' CONTRIBUTIONS

All authors conceived the study and were involved in the development and writing of the manuscript. R.G. and E.B. conducted the literature review and user survey, and planned and wrote the initial manuscript.

## DATA ACCESSIBILITY

Our manuscript does not contain any data.

## ORCID

Rory Gibb  <http://orcid.org/0000-0002-0965-1649>

Ella Browning  <http://orcid.org/0000-0002-7959-9292>

Kate E. Jones  <http://orcid.org/0000-0001-5231-3293>

## REFERENCES

- Adams, A. M., Jantzen, M. K., Hamilton, R. M., & Fenton, M. B. (2012). Do you hear what I hear? Implications of detector selection for acoustic monitoring of bats. *Methods in Ecology and Evolution*, 3, 992–998. <https://doi.org/10.1111/j.2041-210X.2012.00244.x>
- Aide, T. M., Corrada-Bravo, C., Campos-Cerqueira, M., Milan, C., Vega, G., & Alvarez, R. (2013). Real-time bioacoustics monitoring and automated species identification. *PeerJ*, 1, e103. <https://doi.org/10.7717/peerj.103>
- Astaras, C., Linder, J. M., Wrege, P., Orume, R. D., & Macdonald, D. W. (2017). Passive acoustic monitoring as a law enforcement tool for Afrotropical rainforests. *Frontiers in Ecology and the Environment*, 15, 233–234. <https://doi.org/10.1002/fee.1495>
- Banner, K. M., Irvine, K. M., Rodhouse, T. J., Wright, W. J., Rodriguez, R. M., & Litt, A. R. (2018). Improving geographically extensive acoustic survey designs for modeling species occurrence with imperfect detection and misidentification. *Ecology and Evolution*, 8, 6144–6156. <https://doi.org/10.1002/ece3.4162>
- Bas, Y., Bas, D., & Julien, J.-F. (2017). Tadarida: A toolbox for animal detection on acoustic recordings. *Journal of Open Research Software*, 5, 6. <https://doi.org/10.5334/jors.154>
- Baumgartner, M. F., Fratantoni, D. M., Hurst, T. P., Brown, M. W., Cole, T. V. N., Van Parijs, S. M., & Johnson, M. (2013). Real-time reporting of baleen whale passive acoustic detections from ocean gliders. *The Journal of the Acoustical Society of America*, 134, 1814–1823. <https://doi.org/10.1121/1.4816406>
- Baumgartner, M. F., & Mussoline, S. E. (2011). A generalized baleen whale call detection and classification system. *The Journal of the Acoustical Society of America*, 129, 2889–2902. <https://doi.org/10.1121/1.3562166>
- Bittle, M., & Duncan, A. (2013). A review of current marine mammal detection and classification algorithms for use in automated passive acoustic monitoring. *Proceedings of Acoustics: Science, Technology and Amenity*, 2013, 1–8.
- Blumstein, D. T., Mennill, D. J., Clemins, P., Girod, L., Yao, K., Patricelli, G., & Kirschel, A. N. G. (2011). Acoustic monitoring in terrestrial environments using microphone arrays: Applications, technological considerations and prospectus. *Journal of Applied Ecology*, 48, 758–767. <https://doi.org/10.1111/j.1365-2664.2011.01993.x>
- Border, J. A., Newson, S. E., White, D. C. J., & Gillings, S. (2017). Predicting the likely impact of urbanisation on bat populations using citizen science data, a case study for Norfolk, UK. *Landscape and Urban Planning*, 162, 44–55. <https://doi.org/10.1016/j.landurbplan.2017.02.005>
- Bradbury, J., & Vehrencamp, S. (1998). *Principles of animal communication*. Sunderland, MA: Sinauer Associates Inc.
- Braulik, G., Wittich, A., Macaulay, J., Kasuga, M., Gordon, J., Davenport, T. R. B., & Gillespie, D. (2017). Acoustic monitoring to document the spatial distribution and hotspots of blast fishing in Tanzania. *Marine Pollution Bulletin*, 125, 360–366. <https://doi.org/10.1016/j.marpolbul.2017.09.036>
- Braun de Torre, E. C., Wallrichs, M. A., Ober, H. K., & McCleery, R. A. (2017). Mobile acoustic transects miss rare bat species: Implications of survey method and spatio-temporal sampling for monitoring bats. *PeerJ*, 5, e3940. <https://doi.org/10.7717/peerj.3940>
- Browning, E., Gibb, R., Glover-Kapfer, P., & Jones, K. E. (2017). Passive acoustic monitoring in ecology and conservation. *WWF Conservation Technology Series*, 1(2), 1–75. Retrieved from <https://www.wwf.org.uk/conservationtechnology/documents/Acousticmonitoring-WWF-guidelines.pdf>
- Bush, A., Sollmann, R., Wilting, A., Bohmann, K., Cole, B., Balzter, H., & Yu, D. W. (2017). Connecting Earth observation to high-throughput biodiversity data. *Nature Ecology & Evolution*, 1, 0176. <https://doi.org/10.1038/s41559-017-0176>
- Buxton, R. T., McKenna, M., Mennill, D. J., Frstrup, K. M., Crooks, K., Angeloni, L., & Wittemyer, G. (2017). Noise pollution is pervasive in U.S. protected areas. *Science*, 356, 531–533. <https://doi.org/10.1126/science.aah4783>
- Campos-Cerqueira, M., & Aide, T. M. (2016). Improving distribution data of threatened species by combining acoustic monitoring and occupancy modelling. *Methods in Ecology and Evolution*, 7, 1340–1348. <https://doi.org/10.1111/2041-210X.12599>
- Cardinale, B. J., Duffy, J. E., Gonzalez, A., Hooper, D. U., Perrings, C., Venail, P., & Naeem, S. (2012). Biodiversity loss and its impact on humanity. *Nature*, 486, 59–67. <https://doi.org/10.1038/nature11373>
- Clink, D. J., Bernard, H., Crofoot, M. C., & Marshall, A. J. (2017). Investigating individual vocal signatures and small-scale patterns of geographic variation in female Bornean gibbon (*Hylobates muelleri*) great calls. *International Journal of Primatology*, 38, 656–671. <https://doi.org/10.1007/s10764-017-9972-y>
- Darras, K., Pütz, P., Fahrurrozi, Rembold, K., & Tschartke, T. (2016). Measuring sound detection spaces for acoustic animal sampling and monitoring. *Biological Conservation*, 201, 29–37. <https://doi.org/10.1016/j.biocon.2016.06.021>
- Davis, G. E., Baumgartner, M. F., Bonnell, J. M., Bell, J., Berchok, C., Bort Thornton, J., & Van Parijs, S. M. (2017). Long-term passive acoustic recordings track the changing distribution of North Atlantic right

- whales (*Eubalaena glacialis*) from 2004 to 2014. *Scientific Reports*, 7, 1–12. <https://doi.org/10.1038/s41598-017-13359-3>
- Digby, A., Towsey, M., Bell, B. D., & Teal, P. D. (2013). A practical comparison of manual and autonomous methods for acoustic monitoring. *Methods in Ecology and Evolution*, 4, 675–683. <https://doi.org/10.1111/2041-210X.12060>
- Eldridge, A., Casey, M., Moscoso, P., & Peck, M. (2016). A new method for ecoacoustics? Towards the extraction and evaluation of ecologically-meaningful soundscape components using sparse coding methods. *PeerJ*, 4, e2108. <https://doi.org/10.7717/peerj.2108>
- Fairbrass, A., Firman, M., Williams, C., Brostow, G. J., Titheridge, H., & Jones, K. E. (2018). CityNet - Deep Learning Tools for Urban Ecoacoustic Assessment. *BioRxiv Preprint*: <http://dx.doi.org/10.1101/248708>.
- Fairbrass, A., Rennett, P., Williams, C., Titheridge, H., & Jones, K. E. (2017). Biases of acoustic indices measuring biodiversity in urban areas. *Ecological Indicators*, 83, 169–177. <https://doi.org/10.6084/M9.FIGSHARE.C.3361488.V1>
- Farcas, A., Thompson, P. M., & Merchant, N. D. (2016). Underwater noise modeling for environmental impact assessment. *Environmental Impact Assessment Review*, 57, 114–122. <https://doi.org/10.1016/j.eiar.2015.11.012>
- Farina, A., & James, P. (2016). The acoustic communities: Definition, description and ecological role. *BioSystems*, 147, 11–20. <https://doi.org/10.1016/j.biosystems.2016.05.011>
- Fischer, F. P., Schulz, U., Schubert, H., Knapp, P., & Schmoeger, M. (1997). Quantitative assessment of grassland quality: Acoustic determination of population sizes of orthopteran indicator species. *Ecological Applications*, 7, 909. <https://doi.org/10.2307/2269442>
- Froidevaux, J. S. P., Zellweger, F., Bollmann, K., & Obrist, M. K. (2014). Optimizing passive acoustic sampling of bats in forests. *Ecology and Evolution*, 4, 4690–4700. <https://doi.org/10.1002/ece3.1296>
- Gasc, A., Pavoine, S., Lellouch, L., Grandcolas, P., & Sueur, J. (2015). Acoustic indices for biodiversity assessments: Analyses of bias based on simulated bird assemblages and recommendations for field surveys. *Biological Conservation*, 191, 306–312. <https://doi.org/10.1016/j.biocon.2015.06.018>
- Gasc, A., Sueur, J., Jiguet, F., Devictor, V., Grandcolas, P., Burrow, C., & Pavoine, S. (2013). Assessing biodiversity with sound: Do acoustic diversity indices reflect phylogenetic and functional diversities of bird communities? *Ecological Indicators*, 25, 279–287. <https://doi.org/10.1016/j.ecolind.2012.10.009>
- Gillespie, D., Gordon, J., McHugh, R., McLaren, D., Mellinger, D. K., Redmond, P., & Thode, A. (2009). PAMGUARD: Semiautomated, open source software for real-time acoustic detection and localization of cetaceans. *Journal of the Acoustical Society of America*, 125, 2547. <https://doi.org/10.1121/1.4808713>
- Goeau, H., Glotin, H., Vellinga, W.-P., Planque, R., & Joly, A. (2016). LifeCLEF Bird Identification Task 2016: The arrival of deep learning. CLEF2016 Working Notes, 440–449. Retrieved from <http://ceur-ws.org/Vol-1609/>
- Gregory, R. D., Gibbons, D. W., & Donald, P. F. (2004). *Bird census and survey techniques*. Oxford, UK: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780198520863.001.0001>
- Hariharan, B., & Girshick, R. (2016). Low-shot Visual Recognition by Shrinking and Hallucinating Features. *ArXiv Preprint*. Retrieved from <http://arxiv.org/abs/1606.02819>
- Harris, S. A., Shears, N. T., & Radford, C. A. (2016). Ecoacoustic indices as proxies for biodiversity on temperate reefs. *Methods in Ecology and Evolution*, 7, 713–724. <https://doi.org/10.1111/2041-210X.12527>
- Heinicke, S., Kalan, A. K., Wagner, O. J. J., Mundry, R., Lukashevich, H., & Kühl, H. S. (2015). Assessing the performance of a semi-automated acoustic monitoring system for primates. *Methods in Ecology and Evolution*, 6, 753–763. <https://doi.org/10.1111/2041-210X.12384>
- Hill, A. P., Prince, P., Piña Covarrubias, E., Doncaster, C. P., Snaddon, J. L., & Rogers, A. (2018). AudioMoth: Evaluation of a smart open acoustic device for monitoring biodiversity and the environment. *Methods in Ecology and Evolution*, 9, 1199–1211. <https://doi.org/10.1111/2041-210X.12955>
- Honrado, J. P., Pereira, H. M., & Guisan, A. (2016). Fostering integration between biodiversity monitoring and modeling. *Journal of Applied Ecology*, 53, 1299–1304. <https://doi.org/10.1111/1365-2664.12777>
- Isaac, N. J. B., van Strien, A. J., August, T. A., de Zeeuw, M. P., & Roy, D. B. (2014). Statistics for citizen science: Extracting signals of change from noisy ecological data. *Methods in Ecology and Evolution*, 5, 1052–1060. <https://doi.org/10.1111/2041-210X.12254>
- Jaramillo-Legorreta, A., Cardenas-Hinojosa, G., Nieto-Garcia, E., Rojas-Bracho, L., Ver Hoef, J., Moore, J., & Taylor, B. (2016). Passive acoustic monitoring of the decline of Mexico's critically endangered vaquita. *Conservation Biology*, 31, 183–191. <https://doi.org/10.1111/cobi.12789>
- Johnson, M. P., & Tyack, P. L. (2003). A digital acoustic recording tag for measuring the response of wild marine mammals to sound. *IEEE Journal of Oceanic Engineering*, 28, 3–12. <https://doi.org/10.1109/JOE.2002.808212>
- Jones, K. E., Russ, J. A., Basgta, A.-T., Bilhari, Z., Catto, C., Csósz, I., ... Zavarzin, O. (2013). Indicator bats program: A system for the global acoustic monitoring of bats. In B. Collen, N. Pettorelli, J. E. M. Baillie & S. M. Durant (Eds.), *Biodiversity monitoring and conservation: Bridging the gap between global commitment and local action* (pp. 213–247). Oxford, UK: Wiley-Blackwell. <https://doi.org/10.1002/9781118490747>
- Kaewtip, K., Alwan, A., O'Reilly, C., & Taylor, C. E. (2016). A robust automatic birdsong phrase classification: A template-based approach. *The Journal of the Acoustical Society of America*, 140, 3691–3701. <https://doi.org/10.1121/1.4966592>
- Kalan, A. K., Mundry, R., Wagner, O. J. J., Heinicke, S., Boesch, C., & Kühl, H. S. (2015). Towards the automated detection and occupancy estimation of primates using passive acoustic monitoring. *Ecological Indicators*, 54, 217–226. <https://doi.org/10.1016/j.ecolind.2015.02.023>
- Kasten, E. P., Gage, S. H., Fox, J., & Joo, W. (2012). The remote environmental assessment laboratory's acoustic library: An archive for studying soundscape ecology. *Ecological Informatics*, 12, 50–67. <https://doi.org/10.1016/j.ecoinf.2012.08.001>
- Kershenbaum, A., Blumstein, D. T., Roch, M. A., Akçay, Ç., Backus, G., Bee, M. A., & Zamora-Gutierrez, V. (2014). Acoustic sequences in non-human animals: A tutorial review and prospectus. *Biological Reviews*, 91, 13–52. <https://doi.org/10.1111/brv.12160>
- Kéry, M., & Schmidt, B. (2008). Imperfect detection and its consequences for monitoring for conservation. *Community Ecology*, 9, 207–216. <https://doi.org/10.1556/ComEc.9.2008.2.10>
- King, S. L., Sayigh, L. S., Wells, R. S., Fellner, W., Janik, V. M., Janik, V., & Prinage, S. (2013). Vocal copying of individually distinctive signature whistles in bottlenose dolphins. *Proceedings of the Royal Society B*, 280, 20130053. <https://doi.org/10.1098/rspb.2013.0053>
- Klingbeil, B. T., & Willig, M. R. (2015). Bird biodiversity assessments in temperate forest: The value of point count versus acoustic monitoring protocols. *PeerJ*, 3, e973. <https://doi.org/10.7717/peerj.973>
- Krause, J., Sapp, B., Howard, A., Zhou, H., Toshev, A., Duerig, T., & ... , L. (2016). The unreasonable effectiveness of noisy data for fine-grained recognition. *ArXiv Preprint*. doi:arXiv:1511.06789v3
- Lehmann, G. U. C., Frommolt, K. H., Lehmann, A. W., & Riede, K. (2014). Baseline data for automated acoustic monitoring of Orthoptera in a Mediterranean landscape, the Hymettos, Greece. *Journal of Insect Conservation*, 18, 909–925. <https://doi.org/10.1007/s10841-014-9700-2>
- Lellouch, L., Pavoine, S., Jiguet, F., Glotin, H., & Sueur, J. (2014). Monitoring temporal change of bird communities with dissimilarity acoustic indices. *Methods in Ecology and Evolution*, 5, 495–505. <https://doi.org/10.1111/2041-210X.12178>

- Linhart, P., & Šálek, M. (2017). The assessment of biases in the acoustic discrimination of individuals. *PLoS ONE*, *12*, e0177206. <https://doi.org/10.1371/journal.pone.0177206>
- Lucas, T. C. D., Moorcroft, E. A., Freeman, R., Rowcliffe, J. M., & Jones, K. E. (2015). A generalised random encounter model for estimating animal density with remote sensor data. *Methods in Ecology and Evolution*, *6*, 500–509. <https://doi.org/10.1111/2041-210X.12346>
- Lukic, Y., Vogt, C., Dürr, O., & Stadelmann, T. (2016). Speaker identification and clustering using convolutional neural networks. In F. Palmieri, A. Uncini, K. Diamantaras, & J. Larsen (Eds.), *2016 IEEE 26th international workshop on machine learning for signal processing* (pp. 13–16). <https://doi.org/10.1109/MLSP.2016.7738816>
- Mac Aodha, O., Gibb, R., Barlow, K. E., Browning, E., Firman, M., Freeman, R., & Jones, K. E. (2018). Bat Detective—Deep learning tools for bat acoustic signal detection. *PLoS Computational Biology*, *14*, 1–19. <https://doi.org/10.1371/journal.pcbi.1005995>
- Mac Aodha, O., Stathopoulos, V., Terry, M., Jones, K. E., Brostow, G. J., & Girolami, M. (2014). Putting the scientist in the loop – Accelerating scientific progress with interactive machine learning. In *Proceedings of the 2014 22nd international conference on pattern recognition* (pp. 9–17). Stockholm, Sweden: ICPR. <https://doi.org/10.1109/ICPR.2014.12>
- Marinexplore. (2013). The Marinexplore and Cornell University Whale Detection Challenge: competition summary. Retrieved from <https://www.kaggle.com/c/whale-detection-challenge/discussion/4472>
- Marques, T. A., Thomas, L., Martin, S. W., Mellinger, D. K., Ward, J. A., Moretti, D. J., & Tyack, P. L. (2013). Estimating animal population density using passive acoustics. *Biological Reviews*, *88*, 287–309. <https://doi.org/10.1111/brv.12001>
- McWilliam, J. N., & Hawkins, A. D. (2013). A comparison of inshore marine soundscapes. *Journal of Experimental Marine Biology and Ecology*, *446*, 166–176. <https://doi.org/10.1016/j.jembe.2013.05.012>
- McWilliam, J. N., McCauley, R. D., Erbe, C., & Parsons, M. J. G. (2017). Patterns of biophonic periodicity on coral reefs in the Great Barrier Reef. *Scientific Reports*, *7*, 17459. <https://doi.org/10.1038/s41598-017-15838-z>
- Mellinger, D. K., & Clark, C. W. (2006). MobySound: A reference archive for studying automatic recognition of marine mammal sounds. *Applied Acoustics*, *67*, 1226–1242. <https://doi.org/10.1016/j.apacoust.2006.06.002>
- Merchant, N. D., Fristrup, K. M., Johnson, M. P., Tyack, P. L., Witt, M. J., Blondel, P., & Parks, S. E. (2015). Measuring acoustic habitats. *Methods in Ecology and Evolution*, *6*, 257–265. <https://doi.org/10.1111/2041-210X.12330>
- Muda, L., Begam, M., & Elamvazuthi, I. (2010). Voice recognition algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) techniques. *Journal of Computing*, *2*, 138–143.
- Mukundarajan, H., Hol, F. J. H., Castillo, E. A., Newby, C., & Prakash, M. (2017). Using mobile phones as acoustic sensors for high-throughput mosquito surveillance. *eLife*, *6*, 1–26. <https://doi.org/10.7554/eLife.27854>
- Nedelec, S. L., Simpson, S. D., Holderied, M., Radford, A. N., Lecellier, G., Radford, C., & Lecchini, D. (2015). Soundscapes and living communities in coral reefs: Temporal and spatial variation. *Marine Ecology Progress Series*, *524*, 125–135. <https://doi.org/10.3354/meps11175>
- Newson, S. E., Bas, Y., Murray, A., & Gillings, S. (2017). Potential for coupling the monitoring of bush-crickets with established large-scale acoustic monitoring of bats. *Methods in Ecology and Evolution*, *8*, 1051–1062. <https://doi.org/10.1111/2041-210X.12720>
- Newson, S. E., Evans, H. E., & Gillings, S. (2015). A novel citizen science approach for large-scale standardised monitoring of bat activity and distribution, evaluated in eastern England. *Biological Conservation*, *191*, 38–49. <https://doi.org/10.1016/j.biocon.2015.06.009>
- Norouzzadeh, M. S., Nguyen, A., Kosmala, M., Swanson, A., Packer, C., & Clune, J. (2017). Automatically identifying wild animals in camera trap images with deep learning. *ArXiv Preprint*, <http://arxiv.org/abs/1703.05830>, 1–12.
- Nowacek, D. P., Christiansen, F., Bejder, L., Goldbogen, J. A., & Friedlaender, A. S. (2016). Studying cetacean behaviour: New technological approaches and conservation applications. *Animal Behaviour*, *120*, 235–244. <https://doi.org/10.1016/j.anbehav.2016.07.019>
- Penone, C., Le Viol, I., Pellissier, V., Julien, J. F., Bas, Y., & Kerbiriou, C. (2013). Use of large-scale acoustic monitoring to assess anthropogenic pressures on orthoptera communities. *Conservation Biology*, *27*, 979–987. <https://doi.org/10.1111/cobi.12083>
- Petrusková, T., Pišvejcová, I., Kinštová, A., Brinke, T., & Petrusek, A. (2016). Repertoire-based individual acoustic monitoring of a migratory passerine bird with complex song as an efficient tool for tracking territorial dynamics and annual return rates. *Methods in Ecology and Evolution*, *7*, 274–284. <https://doi.org/10.1111/2041-210X.12496>
- Pieretti, N., Farina, A., & Morri, D. (2011). A new methodology to infer the singing activity of an avian community: The Acoustic Complexity Index (ACI). *Ecological Indicators*, *11*, 868–873. <https://doi.org/10.1016/j.ecolind.2010.11.005>
- Pijanowski, B. C., Farina, A., Gage, S. H., Dumyahn, S. L., & Krause, B. L. (2011). What is soundscape ecology? An introduction and overview of an emerging new science. *Landscape Ecology*, *26*, 1213–1232. <https://doi.org/10.1007/s10980-011-9600-8>
- Pirota, E., Merchant, N. D., Thompson, P. M., Barton, T. R., & Lusseau, D. (2015). Quantifying the effect of boat disturbance on bottlenose dolphin foraging activity. *Biological Conservation*, *181*, 82–89. <https://doi.org/10.1016/j.biocon.2014.11.003>
- Prat, Y., Taub, M., & Yovel, Y. (2016). Everyday bat vocalizations contain information about emitter, addressee, context, and behavior. *Scientific Reports*, *6*, 39419. <https://doi.org/10.1038/srep39419>
- Proppe, D. S., Sturdy, C. B., & St. Clair, C. C. (2013). Anthropogenic noise decreases urban songbird diversity and may contribute to homogenization. *Global Change Biology*, *19*, 1075–1084. <https://doi.org/10.1111/gcb.12098>
- Riede, K. (1993). Monitoring biodiversity: Analysis of Amazonian rainforest sounds. *Ambio*, *22*, 546–548.
- Roch, M. A., Batchelor, H., Baumann-Pickering, S., Berchok, C. L., Cholewiak, D., Fujioka, E., & Soldevilla, M. S. (2016). Management of acoustic metadata for bioacoustics. *Ecological Informatics*, *31*, 122–136. <https://doi.org/10.1016/j.ecoinf.2015.12.002>
- Root-Gutteridge, H., Bencsik, M., Chebli, M., Gentle, L. K., Terrell-Nield, C., Bourit, A., & Yarnell, R. W. (2014). Identifying individual wild Eastern grey wolves (*Canis lupus lycaon*) using fundamental frequency and amplitude of howls. *Bioacoustics*, *23*, 55–66. <https://doi.org/10.1080/09524622.2013.817317>
- Ruiz-Gutierrez, V., Hooten, M. B., & Campbell Grant, E. H. (2016). Uncertainty in biological monitoring: A framework for data collection and analysis to account for multiple sources of sampling bias. *Methods in Ecology and Evolution*, *7*, 900–909. <https://doi.org/10.1111/2041-210X.12542>
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., & Fei-Fei, L. (2015). ImageNet large scale visual recognition challenge. *International Journal of Computer Vision*, *115*, 211–252. <https://doi.org/10.1007/s11263-015-0816-y>
- Salamon, J., & Bello, J. P. (2015). Unsupervised feature learning for urban sound classification. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 171–175. <https://doi.org/10.1017/cbo9781107415324.004>
- Salamon, J., & Bello, J. P. (2016). Deep Convolutional Neural Networks and Data Augmentation for Environmental Sound Classification. *ArXiv Preprint*, <http://arxiv.org/abs/1604.07160>, (August). <https://doi.org/10.21437/interspeech.2016805>
- Sayigh, L., Daher, M. A., Allen, J., Gordon, H., Joyce, K., Stuhlmann, C., & Tyack, P. (2016). The Watkins Marine Mammal Sound Database:



- An online, freely accessible resource (p. 040013). <https://doi.org/10.1121/2.0000358>
- Sethi, S. S., Ewers, R. M., Jones, N. S., Orme, D., & Picinali, L. (2018). Robust, real-time and autonomous monitoring of ecosystems with an open, low-cost, networked device. *Methods in Ecology & Evolution*. <https://doi.org/10.1111/2041-210X.13089>
- Simpson, S. D., Meekan, M. G., Jeffs, A., Montgomery, J. C., & McCauley, R. D. (2008). Settlement-stage coral reef fish prefer the higher-frequency invertebrate-generated audible component of reef noise. *Animal Behaviour*, 75, 1861–1868. <https://doi.org/10.1016/j.anbehav.2007.11.004>
- Sousa-Lima, R. S., Fernandes, D. P., Norris, T. F., & Oswald, J. N. (2013). A review and inventory of fixed autonomous recorders for passive acoustic monitoring of marine mammals. *Aquatic Mammals*, 39, 23–53. <https://doi.org/10.1578/am.39.1.2013>
- Staaterman, E., Ogburn, M., Altieri, A., Brandl, S., Whippo, R., Seemann, J., & Duffy, J. (2017). Bioacoustic measurements complement visual biodiversity surveys: Preliminary evidence from four shallow marine habitats. *Marine Ecology Progress Series*, 575, 207–215. <https://doi.org/10.3354/meps12188>
- Stathopoulos, V., Zamora-Gutierrez, V., Jones, K. E., & Girolami, M. (2017). Bat echolocation call identification for biodiversity monitoring: A probabilistic approach. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 67, 165–183. <https://doi.org/10.1111/rssc.12217>
- Stevenson, B. C., Borchers, D. L., Altwegg, R., Swift, R. J., Gillespie, D. M., & Measey, G. J. (2015). A general framework for animal density estimation from acoustic detections across a fixed microphone array. *Methods in Ecology and Evolution*, 6, 38–48. <https://doi.org/10.1111/2041-210X.12291>
- Stowell, D., & Plumbley, M. D. (2014). Automatic large-scale classification of bird sounds is strongly improved by unsupervised feature learning. *PeerJ*, 2, e488. <https://doi.org/10.7717/peerj.488>
- Stowell, D., Stylianou, Y., Wood, M., Pamuła, H., & Glotin, H. (2018). Automatic acoustic detection of birds through deep learning: the first Bird Audio Detection challenge. *Methods in Ecology & Evolution*. <https://doi.org/10.1111/2041-210X.13103>
- Stowell, D., Wood, M., Stylianou, Y., & Glotin, H. (2016). Bird detection in audio: A survey and a challenge. In *Proceedings of the 2016 IEEE International Workshop on Machine Learning for Signal Processing*. Retrieved from <http://arxiv.org/abs/1608.03417>
- Sueur, J., Farina, A., Gasc, A., Pieretti, N., & Pavoine, S. (2014). Acoustic indices for biodiversity assessment and landscape investigation. *Acta Acustica United With Acustica*, 100, 772–781. <https://doi.org/10.3813/AAA.918757>
- Sueur, J., Pavoine, S., Hamerlynck, O., & Duvail, S. (2008). Rapid acoustic survey for biodiversity appraisal. *PLoS ONE*, 3, e4065. <https://doi.org/10.1371/journal.pone.0004065>
- Thompson, M. E., Schwager, S. J., & Payne, K. B. (2010). Heard but not seen: An acoustic survey of the African forest elephant population at Kakum Conservation Area, Ghana. *African Journal of Ecology*, 48, 224–231. <https://doi.org/10.1111/j.1365-2028.2009.01106.x>
- Tobias, J. A., Planqué, R., Cram, D. L., & Seddon, N. (2014). Species interactions and the structure of complex communication networks. *Proceedings of the National Academy of Sciences of the United States of America*, 111, 1020–1025. <https://doi.org/10.1073/pnas.1314337111>
- Van Parijs, S. M., Baumgartner, M., Cholewiak, D., Davis, G., Gedamke, J., Gerlach, D., & ... , M. (2015). NEPAN: A U.S. Northeast passive acoustic sensing network for monitoring, reducing threats and the conservation of marine animals passive acoustic monitoring as a tool for research and management in marine ecosystems. *Marine Technology Society Journal*, 49, 70–86. <https://doi.org/10.4031/MTSJ.49.2.16>
- Van Parijs, S. M., Clark, C. W., Sousa-Lima, R. S., Parks, S. E., Rankin, S., Risch, D., & Van Opzeeland, I. (2009). Management and research applications of real-time and archival passive acoustic sensors over varying temporal and spatial scales. *Marine Ecology Progress Series*, 395, 21–36. <https://doi.org/10.3354/meps08123>
- Walters, C. L., Collen, A., Lucas, T. C. D., Mroz, K., Sayer, C. A., & Jones, K. E. (2013). Challenges of using bioacoustics to globally monitor bats. In R. A. Adams & S. C. Pedersen (Eds.), *Bat evolution, ecology and conservation*. New York, NY: Springer New York. <https://doi.org/10.1007/978-3-319-02904-7>
- Walters, C. L., Freeman, R., Collen, A., Dietz, C., Brock Fenton, M., Jones, G., & Jones, K. E. (2012). A continental-scale tool for acoustic identification of European bats. *Journal of Applied Ecology*, 49, 1064–1074. <https://doi.org/10.1111/j.1365-2664.2012.02182.x>
- Ward, J. A., Thomas, L., Jarvis, S., DiMarzio, N., Moretti, D., Marques, T. A., & Tyack, P. (2012). Passive acoustic density estimation of sperm whales in the Tongue of the Ocean, Bahamas. *Marine Mammal Science*, 28, E444–E455. <https://doi.org/10.1111/j.1748-7692.2011.00560.x>
- Whytock, R. C., & Christie, J. (2017). Solo: An open source, customisable and inexpensive audio recorder for bioacoustic research. *Methods in Ecology and Evolution*, 8, 308–312. <https://doi.org/10.1111/2041-210X.12678>
- Wrege, P. H., Rowland, E. D., Keen, S., & Shiu, Y. (2017). Acoustic monitoring for conservation in tropical forests: Examples from forest elephants. *Methods in Ecology and Evolution*, 8, 1292–1301. <https://doi.org/10.1111/2041-210X.12730>
- Zamora-Gutierrez, V., Lopez-Gonzalez, C., Macswiney Gonzalez, M. C., Fenton, B., Jones, G., Kalko, E. K. V., & Jones, K. E. (2016). Acoustic identification of Mexican bats based on taxonomic and ecological constraints on call design. *Methods in Ecology and Evolution*, 7, 1082–1091. <https://doi.org/10.1111/2041-210X.12556>
- Zilli, D., Parson, O., Merrett, G. V., & Rogers, A. (2014). A hidden Markov model-based acoustic cicada detector for crowdsourced smartphone biodiversity monitoring. *Journal of Artificial Intelligence Research*, 51, 805–827. <https://doi.org/10.1613/jair.4434>

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**How to cite this article:** Gibb R, Browning E, Glover-Kapfer P, Jones KE. Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. *Methods Ecol Evol*. 2018;00:1–17. <https://doi.org/10.1111/2041-210X.13101>