

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

Rory Gibb^{1†}, Ella Browning^{1,2†}, Paul Glover-Kapfer³, and Kate E. Jones^{1,2*}

*Corresponding author: kate.e.jones@ucl.ac.uk

†Denotes joint first authorship.

¹Centre for Biodiversity and Environment Research, Department of Genetics, Evolution and Environment, University College London, London WC1E 6BT

²Institute of Zoology, Zoological Society of London, Regent's Park, London NW1 4RY

³WWF-UK, Living Planet Centre, Rufford House, Woking GU21 4LL

Article type: Reviews and Syntheses.

Running title: Acoustic ecological monitoring.

Word count: Abstract 200, Main Text 7482, Boxes 1464, Figures 476, Tables 43.

Number of figures 5, Number of tables 2, Number of text boxes 3.

Author contributions: KJ and PG-K conceived the study, and RG and EB conducted the literature review and wrote the first draft of the manuscript. All authors contributed to writing the manuscript.

Data accessibility: No data were analysed for this article.

Keywords: acoustic indices, bioacoustics, conservation technology, deep learning, ecoacoustics, ecological monitoring, machine learning, passive acoustic monitoring

Abstract

The recent boom in passive acoustic sensors has filled an important gap in the ecological survey toolbox, providing efficient, non-invasive and taxonomically-broad means to assess and monitor the responses of wildlife populations and communities to environmental change. However, their applications, especially at larger scales, have to date been limited by the complexity of extracting and analyzing ecologically-relevant data from large audio datasets. Here, we critically evaluate current and emerging opportunities, challenges and research priorities for passive acoustic monitoring (PAM) in ecological assessment and biodiversity monitoring. We show that both terrestrial and marine PAM applications are advancing rapidly, facilitated especially by emerging low-cost sensors, the application of machine learning innovations to automated wildlife call identification, and work towards developing acoustic indices of community diversity. However, the scope of PAM studies is often constrained by limited availability of reference sound libraries and audio processing tools, especially for the tropics, and lack of clarity around the limitations of many analytical methods. Future work to further increase possibilities for PAM globally should involve standardisation of survey and analysis protocols, collaborative development of publically-archived sound libraries and multi-year monitoring datasets, and development of a more robust theoretical and analytical framework for monitoring vocalising animal communities.

Introduction

There is a growing need for cost-effective and scalable survey techniques for comprehensive ecological assessments and long-term monitoring, in the face of worldwide declines in biodiversity and ecosystem services (Cardinale *et al.* 2012). As well as addressing fundamental ecological questions, data from survey and monitoring programmes have an essential role in evaluating the trends and drivers of species population change, informing the design of conservation strategies and assessing their efficacy, and addressing policy commitments to halt biodiversity loss (e.g. UN Convention on Biological Diversity) (Secretariat of the Convention on Biological Diversity 2014; Honrado *et al.* 2016).

Traditional survey methods, such as manual point counts, distance sampling and trapping, are limited by being resource intensive and/or invasive. However, rapid technological advances in recent decades have led to the uptake of multiple high-throughput sensing technologies including satellite remote sensing (Pettorelli *et al.* 2016), LIDAR (Hancock *et al.* 2017) and increasingly sophisticated camera traps for non-invasive animal monitoring (Steenweg *et al.* 2017).

Similarly, over the last decade passive acoustic sensors have become an important and widely-used component of the ecological survey toolbox. Many animals actively produce sound for communication, and in echolocating species, acoustic signals are also used for navigation and prey-search (Bradbury & Vehrencamp 1998). Vocalizing animals thus leak information into their surroundings regarding their presence as well as their behaviour, activity patterns and interactions with con- and heterospecifics (Kershenbaum *et al.* 2014; Tobias *et al.* 2014). Although acoustic survey methods are well-established in ecology, for example in point counts of birds, bats or amphibians, sounds are rarely recorded and are typically used for species identification by experienced surveyors in field settings (Gregory *et al.* 2004; Weir *et al.* 2009). Passive acoustic monitoring (PAM) instead involves using

acoustic sensors (recorders, ultrasound detectors, microphones and/or hydrophones) to record sound in the environment for later analysis (Blumstein *et al.* 2011) (Box 1). Ecologically-relevant data are subsequently derived from audio recordings by isolating sounds of interest (e.g. species calls) to produce a spatially- and temporally-explicit record of detections for input into statistical models (Bittle & Duncan 2013; Digby *et al.* 2013) and/or calculating environmental sound metrics (Sueur *et al.* 2014) (Figure 1).

To date, high costs and analytical difficulties have mainly confined the ecological applications of PAM to small-scale, short-term or methods-focused studies. However, as the technology improves, PAM appears increasingly well-suited to objectives-led ecological assessment and monitoring, which require that data collection and analysis protocols be standardisable, scalable and financially sustainable (Harrison *et al.* 2014). The falling cost of acoustic sensors is enabling large data collection initiatives involving both experts and volunteers (Jones *et al.* 2013; Newson *et al.* 2015; Hill *et al.* 2018), and standardised analyses of the resulting datasets are increasingly possible with emerging automated audio processing techniques. However, PAM still involves substantial logistical and technical difficulties, and it remains unclear how effectively current PAM approaches, which are usually applied at small scale, can translate to larger or long-term ecological survey and monitoring programmes. In this review, we therefore synthesise current research in order to highlight these emerging opportunities, evaluate the associated challenges and identify the main bottlenecks to upscaling PAM efforts. We first summarise the current ecological applications of PAM technologies, then sequentially discuss challenges and research priorities at each stage of a typical PAM pipeline (Figure 1), before closing with an overview of significant emerging trends for PAM in ecological research.

Ecological applications of passive acoustic monitoring

Sound recorders attached to microphones (terrestrial) or hydrophones (aquatic) have a rich history of use for surveying animals that are visually cryptic but acoustically active, in particular echolocating bats and cetaceans (Walters *et al.* 2013; Nowacek *et al.* 2016). The use of PAM in ecology has expanded in recent years with the emergence of purpose-designed passive acoustic sensors (hereafter referred to as '*acoustic sensors*') for wildlife and environmental monitoring. These are well-suited to ecological surveys, being non-invasive, autonomous, usually omnidirectional (surveying in a three-dimensional sphere around the sensor) and with a typically larger detection area and fewer taxonomic restrictions than camera traps, which without additional equipment are usually limited to detecting larger birds and mammals at close range (Rowcliffe *et al.* 2008; Lucas *et al.* 2015). As such, they can simultaneously survey entire vocalizing animal communities and their acoustic environments, including human-related sounds such as industrial noise and gunshots (Wrege *et al.* 2010, 2017).

Species detections derived from PAM are analogous to data from other survey methods, and thus have numerous general ecological applications (summarised in Table 1). To date these have included species occupancy and density estimation (Mellinger *et al.* 2011; Martin *et al.* 2013), mapping spatial distributions and behaviour in relation to environmental factors (Wrege *et al.* 2010; Campos-Cerqueira & Aide 2016), studies of phenology including migration movements (Risch *et al.* 2014), estimation of population trends (Barlow *et al.* 2015; Jaramillo-Legorreta *et al.* 2016), rapid biodiversity assessments (Sueur *et al.* 2008; Braulik *et al.* 2017b) and estimation of vocalising species inventories (MacSwiney G *et al.* 2008). In these cases PAM offers certain benefits over traditional surveys, for example the ability to conduct continuous monitoring over long periods for low manual effort and cost, and the increased likelihood of detecting rarer or less vocally-active species that might

otherwise be missed (Klingbeil & Willig 2015). The use of standardised analysis methods also avoids observer bias in species detection, meaning that audio surveys conducted by multiple surveyors and/or volunteers are less susceptible to the skill level biases that impact many forms of citizen science data (Celis-Murillo *et al.* 2009; Isaac *et al.* 2014). PAM data also have their own limitations in ecological contexts, including their unsuitability for species that are not acoustically active, the current inability to identify individual vocalising animals in most taxa (in contrast to visual recognition or mark-recapture), and the difficulties associated with audio analysis, which are discussed in depth below.

Importantly, PAM also offers numerous unique possibilities for ecological research and monitoring. In addition to surveys of visually cryptic or nocturnal species, these include studies of acoustic signalling among wild animal populations, such as vocalising behaviour, intraspecific variability in call repertoire, and the evolution of acoustic communities (Blumstein *et al.* 2011; Tobias *et al.* 2014; Prat *et al.* 2016; Linhart & Šálek 2017); species responses to the acoustic environment (Simpson *et al.* 2008; Nowacek *et al.* 2016); and landscape-scale monitoring of anthropogenic phenomena such as acoustic pollution, blast fishing and poaching, to assess their ecological impacts and inform interventions (Astaras *et al.* 2017; Braulik *et al.* 2017a) (Table 1). For instance, there is a rich literature on the effects of anthropogenic noise on population ecology and behaviour in cetaceans and increasingly birds (e.g. DeRuiter *et al.* 2013; Proppe *et al.* 2013; Marley *et al.* 2017). Sensor networks are also increasingly deployed to listen to ecosystems across large areas and long timescales, facilitating the study of vocalising animal communities across different habitats and biomes (Tucker *et al.* 2014; Nedelec *et al.* 2015) and the development of putative acoustic indices of biodiversity for ecological monitoring (Sueur *et al.* 2014) (Table 1).

Once archived, audio datasets can also provide valuable resources to answer ecological questions beyond their original study objectives. For example, bat monitoring data have been

reanalysed to study orthopteran populations in the UK and France (Penone *et al.* 2013; Newson *et al.* 2017) and predict impacts of urban development on bats to inform planning (Border *et al.* 2017). Gathering audio over years or decades can offer long-term insights into population ecology, behaviour and the effects of human disturbance and conservation initiatives which, especially for cryptic species such as forest elephants (Wrege *et al.* 2017), may be difficult to achieve by other means. Such large datasets may involve collating audio collected by volunteers (Jones *et al.* 2013) or from multiple sources (Van Parijs *et al.* 2015). Ultimately, archived PAM data also have the potential to contribute large quantities of occurrence and abundance data to global repositories (such as Global Biodiversity Information Facility) designed to facilitate biodiversity modelling and monitoring (Honrado *et al.* 2016).

Advances and challenges in passive acoustic monitoring methods

a) Passive acoustic sensor technology

The last decade has seen the emergence of numerous commercial terrestrial and aquatic acoustic sensors (Box 1) that, in contrast to early PAM studies that often repurposed field recorders (Riede 1993) or naval or seismological equipment (Sousa-Lima *et al.* 2013), are comparable to camera traps in durability and user-accessibility (Figure 1a). Improved battery life and data storage, on-board metadata collection and programmable recording schedules increasingly allow for extended autonomous deployments with flexible temporal sampling regimes (Klinck *et al.* 2012; Aide *et al.* 2013; Baumgartner *et al.* 2013; Suberg *et al.* 2014). However, hardware costs have often limited survey scalability, with ubiquitous, user-friendly sensor units such as Wildlife Acoustics Song Meters (SM4 and variants) often substantially more expensive than equivalent-spec camera traps. When synchronous multi-sensor surveys are unnecessary, one solution has been repeated redeployment of a small number of sensors

(Wrege *et al.* 2017). For example, the Norfolk Bat Survey loan out ultrasonic detectors to hundreds of volunteers, resulting in the collection of a high spatial resolution ultrasonic dataset of bat activity, covering a large area of the United Kingdom (Newson *et al.* 2015).

Looking forward, the arrival of low-cost and customisable sensor models promises to sharply reduce the cost of PAM surveys. Emerging open-source devices based on printed circuit boards and modular computers, including Solo (Whytock & Christie 2017) and AudioMoth (Hill *et al.* 2018), are significantly cheaper than commercial alternatives. For example, AudioMoth can be customised and manufactured in large numbers to reduce unit cost to around US\$30 (Hill *et al.* 2018), whereas per-unit costs of commercial sensors often exceed US\$1000. Initiatives like these will substantially lower the financial barriers to large-scale data collection and citizen scientist involvement. However, the use of less expensive components such as microelectromechanical systems (MEMS) microphones may involve trade-offs between sensor cost and data quality, for example if such components show inconsistent frequency response, lower signal-to-noise ratios, or are more vulnerable to environmental damage. A critical open question concerns the amount of data quality (signal-to-noise ratio, frequency resolution) that can be sacrificed without compromising the ability to derive sufficient ecological information (e.g. accurate species identification) from audio recordings (Figure 2). Addressing this question will require comparative analyses of data collected simultaneously using different sensor models (similar to Adams *et al.* 2012), and the answer may vary taxonomically since certain species are intrinsically harder to distinguish acoustically than others (see '*Automated sound identification*') (Kershenbaum *et al.* 2014).

(b) Survey protocols and data collection

Survey design and standardisation. An important challenge for the design of PAM programmes is understanding the comparability of audio data collected using different sensor models and sampling protocols, across different environment types (Figure 1b) (Browning *et al.* 2017). PAM sampling designs can include transect-based surveys (Jones *et al.* 2013; Yack *et al.* 2013), but more commonly involve static sensors (analogous to camera traps) deployed either standalone, in multi-sensor networks, or in linked microphone/hydrophone arrays to allow for two- or three-dimensional sound localisation (the latter reviewed in Blumstein *et al.* 2011). The advantages and disadvantages of different survey designs will depend on a study's taxonomic focus, environmental realm (terrestrial or marine; Box 2), spatial scale and objectives (Van Parijs *et al.* 2009), but often remain poorly understood. For example, research is needed to identify whether particular subsets of taxa are systematically over- or underrepresented by static or mobile survey methods, as shown for rare bat species (Braun de Torrez *et al.* 2017). Equally, while transects or sparsely-deployed static sensors may suffice for analysis of broad species occupancy or activity trends, modelling animal abundance or habitat use at finer spatial or temporal scales may often require the deployment of denser networks of calibrated static sensors, alongside estimates of additional parameters such as species detection distances (Lucas *et al.* 2015; Jaramillo-Legorreta *et al.* 2016).

The detection distance of a sound from its emitter (e.g. calling animal) varies depending on its amplitude and frequency, the environmental medium (air, water or seawater) and features of the surrounding habitat, e.g. vegetation, topography, temperature and pressure (Box 1, Box 2) (Farcas *et al.* 2016). As a result, the effective sampling area around an acoustic sensor differs among species and call types, and varies over space and time due to habitat features and masking by non-target (including anthropogenic) sound (Figure 3). If unaccounted for, any resulting detection biases, for example towards animals that call at higher amplitudes and/or lower frequencies, may lead to biased abundance or species richness estimates.

Historically these challenges have often been overlooked in the PAM literature, but there are now increasing efforts to systematically quantify sources of bias and improve standardisation of marine and terrestrial PAM survey methods. These include development of guidelines for sensor calibration (Merchant *et al.* 2015), adoption of metadata standards (Roch *et al.* 2016), assessing the efficacy of different sampling designs (Van Parijs *et al.* 2009; Froidevaux *et al.* 2014; Braun de Torrez *et al.* 2017), quantifying differences in sensitivity between sensor models and over time due to environmental degradation (Adams *et al.* 2012; Merchant *et al.* 2015), and quantifying the effects of sensor proximity to habitat features (e.g. vegetation, water surface, topography) on sound detection (Darras *et al.* 2016; Farcas *et al.* 2016). As well as helping to improve field methods, these efforts will ultimately facilitate more robust analyses of large, multi-sensor acoustic datasets, which currently often assume species detectability to be constant over space and time (e.g. Newson *et al.* 2015; Davis *et al.* 2017).

Audio recording and data storage. Audio files have high data storage requirements compared to other monitoring data types, such as camera images. The conventional sampling rate for audible range and infrasound (44.1kHz) produces relatively manageable datasets (approximately 5MB per minute in 16bit mono), but recording full-spectrum ultrasound involves extremely high sampling rates (often >200kHz), resulting in much larger file sizes and associated costs of regular data retrieval and storage. There is thus a trade-off between data quality and storage capacity that is particularly acute for ultrasonic (e.g. bat and cetacean) monitoring programmes. Frequency division methods for ultrasound detection (see Box 1) have lower power and storage requirements which, provided sufficient information can be derived from the data, may suit extended or remote deployments (e.g. Jaramillo-Legorreta *et al.* 2016). However, these methods discard substantial amounts of frequency and amplitude information, which impacts the ability to discriminate different species and behaviours (Adams *et al.* 2012; Walters *et al.* 2013; Palmer *et al.* 2017). In future, tools for

isolating sounds from audio may become less sensitive to recording method, but in the meantime recording audio with the least possible information loss, i.e. full-spectrum in lossless data format (wav, flac or equivalent), is generally preferable, since this facilitates species call identification (Walters *et al.* 2012) and crucially also futureproofs the data by allowing for subsequent reanalysis with different tools or to address different questions.

Provided data are recorded and stored at sufficient quality (Figure 1c), the potential for future reanalyses is a key advantage of archived acoustic datasets. For example, a recent study combined multi-year hydrophone data from 11 separate sources to provide a detailed picture of the spatiotemporal distribution of the critically endangered *Eubalaena glacialis* (North Atlantic right whale) (Davis *et al.* 2017). Leveraging such opportunities from years or decades of monitoring data will require collaborative development of infrastructure to support the collation and public archiving of massive (multi-gigabyte to petabyte) audio datasets (e.g. <https://ngdc.noaa.gov/mgg/pad/>). This is an important opportunity and challenge for the PAM field going forward. Another possible solution to data capacity issues is to reduce the quantity of audio that is recorded and stored, for example by applying on-board thresholds that trigger recording only when possible sounds of interest are present, or on-board algorithms that detect and classify sounds in real-time (Baumgartner *et al.* 2013; Van Parijs *et al.* 2015; Hill *et al.* 2018). While discarding audio data is scientifically undesirable, some degree of on-board filtering may be necessary to prevent long-term datasets becoming unmanageably large, and combined with emerging wireless transmission functionality (Aide *et al.* 2013) could also facilitate real-time monitoring and reporting of wildlife or anthropogenic activity.

(c) Automated sound identification

Even relatively small PAM studies produce large volumes of audio data, which become logistically difficult to manage and analyse. Prior to further analysis, target sounds must first be located within audio recordings and classified to taxonomic and/or sound type categories (Aide *et al.* 2013; Salamon & Bello 2015), which requires pipelines to process sound files and metadata and annotate sounds of interest (e.g. calling animal species, location, precise date/time) (Figure 1d-e). Such workflows are analogous to camera trap data processing, in which photos are viewed and classified to species, usually by hand (Steenweg *et al.* 2017). Conducted manually, this process is time-consuming and subjective, and it is difficult to quantify biases related to analyst knowledge level (Brumm *et al.* 2017). These issues may be particularly problematic in conservation settings with limited resources and multiple analysts (Heinicke *et al.* 2015; Kalan *et al.* 2015). Automated systems for efficient post-hoc processing are therefore necessary prerequisites for scaling up PAM studies. Innovations in machine learning are increasingly applied to the recognition of bioacoustic signals, often from an applied ecology perspective (Mellinger *et al.* 2007; Walters *et al.* 2012; Aide *et al.* 2013; Heinicke *et al.* 2015). However, the complexity of acoustic environmental data also provides a useful real-world test case for new methods in computer science and signal processing. As a result, the involvement of the machine learning and computer vision communities in bioacoustic analysis is helping to drive methodological advances that will ultimately benefit ecologists (Marinexplore 2013; Stowell & Plumbley 2014; Goeau *et al.* 2016; Stowell *et al.* 2016).

An automated sound identification pipeline. A typical pipeline for automatically identifying target sounds within audio recordings (hereafter referred to as ‘*automated sound identification*’) involves several distinct stages (Figure 4). Most commonly, the audio recording is first processed to recover frequency information (using Fourier transform or equivalent; Box 1) and produce a time-frequency-amplitude representation (*spectrogram*)

(Figure 4a-b). The next stage is detection, i.e. locating relevant sounds within a recording (Figure 4c), using methods that range from threshold-based methods that detect energy peaks in specified frequency bands, to spectrogram cross-correlation against template sounds, to more complex statistical and machine learning techniques (Table 2). Detected calls can either be directly input into statistical models, or more commonly are then 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. Several standard feature types are well-established in audio processing, such as Mel-frequency cepstral coefficients in human voice recognition (Muda *et al.* 2010). However, bioacoustic analyses often instead involve extraction of hand-crafted features known to be useful for discriminating between species of interest, such as peak frequency, changes in fundamental frequency, call duration and peak amplitude (Riede *et al.* 2006; Baumgartner & Mussoline 2011; Potamitis *et al.* 2014) (Figure 4d). Each sound's combination of features is then classified using either supervised (previously trained on expert-labelled sound libraries) or unsupervised (based on the structure within the data) machine learning algorithms, which return the estimated likelihood that a sound belongs to its assigned category (summarised in Table 2) (e.g. Walters *et al.* 2012; Garland *et al.* 2015; Tolkova *et al.* 2017).

Designing and testing automated sound identification workflows early in a PAM project can substantially reduce later time costs of manual analysis or validation (Heinicke *et al.* 2015).

In ecology, automated tools are commonly developed to address study-specific objectives (e.g. Heinicke *et al.* 2015; Campos-Cerqueira & Aide 2016), which in an ideal world may be preferable, since machine learning models can be trained and tested on data that are representative of the actual survey dataset (although at the potential cost of reducing their generality for other datasets). However, in practice the training and testing of bespoke algorithms is time-consuming and often prohibitively complex for non-expert users. Both

proprietary (e.g. Raven Pro, Avisoft, Kaleidoscope) and open-source or freeware software (e.g. PAMGUARD, LFDACS, iBatsID, ARBIMON, Tadarida) (Gillespie *et al.* 2008; Baumgartner & Mussoline 2011; Walters *et al.* 2012; Aide *et al.* 2013; Bas *et al.* 2017) now offer an increasing range of sound identification tools for large taxonomic groups and geographical regions. Such tools have the advantage of accessibility for a wide user base, but may be trained on audio that differs substantially from the target data (e.g. different sensor model, less noisy recordings), which may impact their performance on other datasets. Equally, although proprietary software may be user-friendly, from a scientific perspective there are clear risks of relying on costly, closed-source tools whose underlying methods and limitations are often poorly reported.

Clear quantitative understanding of the performance of detection and classification algorithms is vital, because even thoroughly-tested bespoke systems are often error-prone on real-world audio recordings, which contain many non-target sounds from weather, water flow, anthropogenic sources and other animal species (Heinicke *et al.* 2015; Salamon & Bello 2015). During the detection process these may be falsely identified as sounds of interest, with simpler methods such as thresholding typically most sensitive, although prior application of noise reduction filters (to screen out background sound) can improve accuracy rates (reviewed in Stowell *et al.* 2016). However, even when detection precision is high (relatively few false positives), state-of-the-art methods often still struggle to distinguish faint or partially masked calls, leading to high false negative rates (low recall) (Digby *et al.* 2013; Goeau *et al.* 2016). Similarly, robust feature extraction is critical for accurate classification, but this process is sensitive to factors including caller distance from sensor, background noise and overlap between synchronous calls (Stowell & Plumbley 2014). Certain taxa are also more difficult to classify acoustically, for instance bird and cetacean vocal repertoires often contain large amounts of intraspecific variation, whereas bat echolocation call structures are

often relatively intra-specifically consistent due to their functional role in navigation (Walters *et al.* 2012; Kershenbaum *et al.* 2014; Garland *et al.* 2017). Species identification may be further complicated by ecological context, for example variations in vocal behaviour in response to environment or conspecifics, or the co-occurrence of species with very 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).

Processing workflows for PAM studies generally therefore involve a trade-off between time and accuracy: manual processing is often most accurate but can be subjective and slow, whereas fully automated processing is much faster but potentially error-prone (Digby *et al.* 2013). Extant tools are rarely accurate enough for fully automated analysis, meaning that most large-scale audio processing is semi-automated at best, involving regular manual cross-checking of outputs (e.g. Baumgartner *et al.* 2013; Kalan *et al.* 2015; Newson *et al.* 2015; Campos-Cerqueira & Aide 2016), with ambiguous classifications often resolved using expert opinion or rules of thumb (e.g. selecting the most likely species based on other calls in close temporal proximity). In future, incorporating newer machine learning techniques that take other surrounding calls into account (e.g. recursive neural networks) may facilitate the automation of this process. Looking forward, an achievable priority for the PAM field would be the community development and adoption of gold standard, publically archived bioacoustic sound libraries, to use as benchmarks for comparative testing of new and closed-source algorithms. Despite all these limitations, the growing availability of automated tools has sharply improved the efficiency of analysing large, complex acoustic datasets, broadening the possibilities for PAM studies. For example, it is possible to use automated tools to screen out detections or classifications that fall below a minimum confidence threshold, thereby substantially reducing the amount of data that must be manually examined. Crucially for ecological contexts, such an approach is modifiable depending on objectives,

for example monitoring endangered species may require minimising false negatives, at the cost of higher likelihood of false positives.

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 (Stowell & Plumbley 2014; Salamon & Bello 2015) and dynamic time warping based feature representations (Stathopoulos *et al.* 2017). Deep convolutional neural networks (CNNs) are particularly promising in this regard, since during model training they are able to learn discriminating spectro-temporal information directly from annotated spectrograms (bypassing a separate feature extraction stage) and are thus more robust to sound overlap and caller distance (Goeau *et al.* 2016) (Figure 4d). In recent tests, CNNs have markedly outperformed alternative methods on detection and/or classification of biotic and anthropogenic sounds in noisy urban sound recordings (Salamon & Bello 2016; Fairbrass *et al.* 2018) and animal calls in heterogeneous environmental audio datasets (Marinexplore 2013; Goeau *et al.* 2016; Mac Aodha *et al.* 2017). However, their ecological applications are currently constrained by the need for very large training datasets that are representative of the natural variability of real-world audio, i.e. recorded in various ambient sound situations and at multiple distances from vocalizing animals. In equivalent image classification tasks, a high number of examples per species (ideally at least 1000) is generally required for human-level performance (e.g. Russakovsky *et al.* 2015). Although such large datasets are easier to obtain for image classification (e.g. using online images; Krause *et al.* 2016), very few equivalent resources 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 recording quality. On a per-study basis this requirement can be partially addressed by data augmentation, i.e. adding background noise to training data to simulate

different distances and acoustic environments (Salamon & Bello 2016). Citizen science involvement is also possible, with online data labelling projects such as Bat Detective (www.batdetective.org) and Snapshot Serengeti (www.snapshotserengeti.org) involving volunteers in audio and photo annotation for CNN training data (Mac Aodha *et al.* 2017; Norouzzadeh *et al.* 2017). In addition, 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 requirement for large training datasets (e.g. Hariharan & Girshick 2016), although these have not yet been applied to ecological tasks. As deep learning methods become more widely used in ecology and training data gaps are addressed (as discussed below), it is likely that the uptake of CNNs and related innovations in PAM will continue increasing in the coming years.

Call libraries and training data. Perhaps the most fundamental current 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 collect, 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 invertebrates (Penone *et al.* 2013; Lehmann *et al.* 2014) and regions outside Europe and North America, despite an urgent need for tools to facilitate monitoring of biodiverse subtropical and tropical habitats (Zamora-Gutierrez *et al.* 2016). These gaps translate into equivalent biases in species classifier availability, and to our knowledge no widely-available automated tools exist for distinguishing intraspecific acoustic behaviours (e.g. social calls, search-phase and feeding echolocation calls in cetaceans and bats) (Figure 4e), although recent research into the

signalling function of bat vocalisations shows promise for the application of machine learning approaches to analysis of acoustic social behaviour (Prat *et al.* 2016).

Filling these data gaps is a priority for the PAM community, and will require increasing coordination around the collection and sharing of large, diverse and ground-truthed reference call libraries. In the meantime, online community databases such as MobySound (www.mobysound.org) and Watkins Marine Mammal Sound Database (www.whoi.edu/watkinssounds) for marine mammals, and Xeno-Canto for birds (www.xeno-canto.org) are growing sources for rich (albeit not necessarily standardised) training data (Mellinger & Clark 2006; Sayigh *et al.* 2016). Further research to improve the situation could include focusing on the development of noise-robust sound identification methods that perform well even with small and variable quality training datasets (e.g. Kaewtip *et al.* 2016), as well as generalised detection algorithms for entire taxonomic groups (e.g. Baumgartner & Mussoline 2011; Mac Aodha *et al.* 2017) that could subsequently be coupled to regional species classifiers. Another important focus will be lowering the technical barriers to ecologists developing and testing bespoke classifiers for their own datasets, for example via interactive machine learning software (Mac Aodha *et al.* 2014), or the functionality increasingly provided in bioacoustic analysis packages such as Kaleidoscope, ARBIMON and Tadarida (Aide *et al.* 2013; Bas *et al.* 2017).

(d) Inferring population information from acoustic data

Following processing, a typical output from a sound identification pipeline is 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 several sources of detection uncertainty that must be accounted for in survey design and statistical analysis. The first is the issue of imperfect detectability, as with other

ecological survey methods; the probability of successfully detecting a vocalizing animal varies depending on interactions between its distance from the sensor, vocalizing behaviour, call frequency and amplitude, and site-specific environmental factors (Kéry & Schmidt 2008; Darras *et al.* 2016). The second issue is that species vocalisations recorded in close spatial or temporal proximity are statistically non-independent, since they may have been emitted by the same individual (Lucas *et al.* 2015). Identification of individual animals from vocalisations alone is currently not possible for most species, and where this has been achieved (e.g. for some birds, cetaceans and wolves) usually requires extensive manual analysis (Kirschel *et al.* 2009; Root-Gutteridge *et al.* 2014; Petrusková *et al.* 2015; Clink *et al.* 2017). Furthermore, many soniferous animals produce sequences of several calls in quick succession (e.g. birdsong phrases, echolocation sequences) which must be merged into discrete detections prior to subsequent analysis (Newson *et al.* 2015; Jaramillo-Legorreta *et al.* 2016). The third major source of uncertainty relates to errors in automated sound identification, as discussed above (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.

Analytical techniques for working with PAM data (Figure 1f) must ideally account for these uncertainties. For example, although presence-absence data from acoustic sensors can be used to develop statistical species distribution models, the assumptions of uniform detectability in extant distribution models are likely to result in erroneous estimates (Lahoz-Monfort *et al.* 2014). Patch occupancy models are emerging as 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. Kalan *et al.* 2015; Campos-Cerqueira & Aide 2016). In addition, the arrival of less computationally expensive

Bayesian inference methods for fitting complex hierarchical regressions (e.g. Integrated Nested Laplace Approximation, Blangiardo *et al.* 2013) and occupancy models is increasingly enabling multiple sources of uncertainty to be incorporated into spatiotemporal model parameters (Isaac *et al.* 2014; Ruiz-Gutierrez *et al.* 2016). In future, such frameworks would ideally be extended to include, for example, the confidence associated with automated call detections and classifications.

One of the key applications of ecological survey data is the estimation of species abundance and population trends. Animal abundance estimation from passive acoustic count data is challenging, owing to the lack of a simple scaling relationship between call counts and animal density. The last decade has seen a rapidly growing literature on animal density estimation from acoustic data, in particular for cetaceans (reviewed in Marques *et al.* 2013). Spatially-explicit capture recapture models (across multi-sensor arrays and networks) (e.g. Marques *et al.* 2012; Stevenson *et al.* 2015) and other methods that adjust detected call density by the average calling rate of the target species (e.g. Thompson *et al.* 2010; Ward *et al.* 2012a) have been shown to provide accurate density estimates when validated against other non-acoustic methods. Another recently developed method applies a generalised random encounter model (gREM) (Lucas *et al.* 2015), a modification of an ideal gas model originally developed for camera trap data (Rowcliffe *et al.* 2008). To date, gREMs have only been tested on simulated data, but may prove useful tools given the growing use of REMs in camera trap studies (e.g. Zero *et al.* 2013). However, most of these analytical tools present practical difficulties for scaling up to large survey programmes with broad taxonomic coverage. Many have high data requirements and as such can be resource-intensive, often requiring the deployment and retrieval of multi-sensor networks and the estimation of species-specific input parameters, such as detection distances and average call rates (Lucas *et al.* 2015). In cetacean studies, call rates are often estimated from animals tagged with

acoustic loggers (Johnson & Tyack 2003), and although similar on-board acoustic tags for terrestrial animals are emerging (Lynch *et al.* 2013) these are often costly or too large to ethically deploy on many animals. As such, estimation of true abundance may often be best-suited to well-resourced projects with specifically defined, species-focused objectives, rather than broader-scope ecological monitoring.

Where more complex analytical methods are unfeasible, it is usually appropriate to develop informed indices of relative abundance. Detection counts within specified sampling periods are commonly used as proxies for relative density or activity rates, such as nightly bat detections (Newson *et al.* 2015) or temporally aggregated click rates in cetacean surveys (Jaramillo-Legorreta *et al.* 2016). While analysing nearly two decades of citizen science transect data from the UK National Bat Monitoring Programme, generalised additive models were used to model annual indices of relative per-survey detection rates, generating long-term population trends (Barlow *et al.* 2015). Such approaches often assume consistent detection distances and no variability in vocalizing behaviour 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). For example, a common problem with acoustic data is over-dispersion caused by individual animals vocalizing within audible range of a sensor for long periods, leading to artificially inflated detection rates. With careful survey design and replication, these issues may be less problematic for estimation of broad-scale activity or occupancy trends. However, more detailed data on species call rates, detection distances and environmental effects on sound transmission may often be required to estimate finer-scale spatial or temporal variations in animal density or behaviour.

(e) Acoustic ecological community assessment

Moving beyond a single-species focus and towards deriving community information (e.g. species diversity) from PAM data presents the considerable challenge of identifying multiple, or ideally all, vocalising species from sound recordings. However, for most taxa and geographical regions such an approach is currently either impossible due to a lack of reference call libraries, or extremely time-consuming due to the limited availability of robust automated sound identification tools. There is therefore a clear need for robust acoustic indicators of ecosystem condition and biodiversity (Figure 1g), especially to facilitate surveys of threatened and biodiverse regions for which we lack analytical tools but urgently require monitoring (e.g. tropical forests, coral reefs) (Harris *et al.* 2016). One solution may be to monitor populations of proposed indicator taxa such as orthoptera or bats as proxies for wider ecosystem health (Fischer *et al.* 1997; Jones *et al.* 2013), although their usefulness as ecological indicators is not well understood. In response to this gap, recent years have seen the development of soundscape-based analytical methods that aim to infer community ecological information from a habitat's global sound dynamics (Pijanowski *et al.* 2011; Sueur & Farina 2015) (Figure 5). Under the umbrella research theme of ecoacoustics, various summary indices have been designed to facilitate comparison of biotic sound between sites or 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 as such are essentially more complex extensions of conventional sound pressure and spectral density metrics used to measure acoustic habitats (Figure 5a(i)) (Merchant *et al.* 2015). Simpler indices involve comparison of power levels in broad frequency bands assumed to contain the majority of biotic (i.e. emitted by wildlife) or anthropogenic sound, often as putative indicators of habitat condition (e.g. Kasten *et al.* 2012) (Figure 5a(ii)). Acoustic entropy, complexity and diversity indices involve calculation of sound pressure ratios across both spectral and temporal dimensions (Sueur *et al.* 2008; Pieretti *et al.* 2011; Gasc *et al.* 2013) (Figure 5a(iii)). Such

metrics are derived from the theory that competition for acoustic space between sympatric signalling animals drives the evolution of signal divergence, and therefore that the spectro-temporal diversity of biotic sound in a habitat correlates to vocalizing species diversity (Sueur *et al.* 2008; Pijanowski *et al.* 2011). For example, the acoustic entropy and dissimilarity indices were developed as acoustic analogues of classical α - and β -diversity indices (Sueur *et al.* 2008). Another approach involves using indices for acoustic event detection, to rapidly screen datasets for relevant sounds (Farina *et al.* 2016).

Despite growing interest in these methods, their results to date have been mixed. Systematic tests in terrestrial environments occasionally find correlations between ecoacoustic indices and species richness and diversity metrics, suggesting that in some cases soundscape-based metrics can function as ecological indicators (Sueur *et al.* 2008; Depaetere *et al.* 2012; Gasc *et al.* 2013, 2015). However, many indices are highly sensitive to site-specific and temporal differences in vocalizing animal community composition and non-biotic sound levels (e.g. wind, rain, anthropogenic sound) (Lellouch *et al.* 2014; Gasc *et al.* 2015). 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. For example, systematic tests have shown that indices are very sensitive to the heterogeneous soundscapes typical of urban environments, limiting their suitability for monitoring in cities (Fairbrass *et al.* 2017). To date, most ecoacoustics studies have instead occurred in relatively undisturbed habitats such as forests, where anthropogenic sound may present fewer problems (e.g. Depaetere *et al.* 2012; Tucker *et al.* 2014). Similarly, it remains unclear whether extant ecoacoustic indices are suitable for the study of marine communities and soundscapes, due to the different acoustic properties of marine environments and increasing anthropogenic noise levels in the oceans globally (see Box 2).

Another more fundamental practical problem for biodiversity monitoring is that the theory of a link between community and biotic sound diversity that underpins many acoustic indices remains controversial (Tobias *et al.* 2014). At the landscape scale it is unclear if and how spectral and temporal biotic sound diversity (as derived from sound recordings) relates to either vocalizing 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; Sueur *et al.* 2014; Harris *et al.* 2016). Despite this lack of clarity, tools for calculation of acoustic indices are becoming increasingly accessible for non-expert users (e.g. in R packages *seewave* and *soundecology*). Their outputs must therefore be treated critically, especially for outcomes-focused monitoring, with acoustic index values at minimum ground-truthed on a per-study basis, either using expert-labelled audio subsets and/or other forms of ecological survey data (e.g. Sueur *et al.* 2008; Harris *et al.* 2016).

Soundscape-based community analyses and rapid biodiversity assessments nonetheless promise to be one of PAM's unique ecological applications, offering advantages over the visual and taxonomic limitations of camera traps, and the potential to provide ground-based biodiversity data to complement landscape data derived from satellite and aerial LIDAR sensing (Bush *et al.* 2017). In the absence of a broad spectrum of species classifiers and call libraries, leveraging these opportunities in the immediate future will likely require the use of acoustic indices or similar proxies. Their practical usefulness may be improved by ongoing testing of the reliability of indices for monitoring different environments (including marine habitats) and comparison of their values to ground-truthed community data. Another fruitful direction may be to gradually develop and test the performance of acoustic indicators, starting with labelled datasets from a few low-diversity, well-understood acoustic communities (e.g. with a handful of known species present), before moving towards analysis of progressively more complex communities. Other emerging ecoacoustics research attempts to address the

problem of biodiversity monitoring in different ways, for example by using CNNs to isolate biotic sound in noisy environments (Fairbrass *et al.* 2018), or applying unsupervised algorithms to learn regular acoustic patterns directly from survey data. Eldridge *et al.* (2016) used sparse coding methods to isolate periodic sound components within bird chorus recordings, which they suggest may correlate to particular sound types or species calls (Eldridge *et al.* 2016). Although early in development, such approaches might eventually enable the estimation of vocalizing species diversity without requiring extensive call libraries and species classifiers (although reference material would still be required to identify unsupervised classifications to species). Although it is still unclear whether this is a feasible outcome, if achieved this would represent a major step towards developing broadly-applicable acoustic ecological indicators.

Emerging and future ecological opportunities

Until recently, the use of PAM in outcomes-focused ecological assessment and monitoring has 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, the arrival of low-cost sensors is now pushing the main bottlenecks into the data processing and management stages. This parallels similar shifts in other data-rich sensing technologies (Pettorelli *et al.* 2016; Steenweg *et al.* 2017) and, as we have shown throughout this article, addressing these logistical and analytical barriers will increasingly require collaborative, community-led efforts. These include the standardisation of survey and data collection protocols, the public archiving of PAM datasets and reference call libraries, and the development of a more robust theoretical and empirical framework for acoustic community assessment. There is promising coordination around these challenges, such as through the recently-founded International Society of Ecoacoustics, and large, multi-institution marine monitoring initiatives such as the US Northeast Passive Acoustic Sensing

Network (NEPAN; Van Parijs *et al.* 2015). To conclude, we briefly outline some major emerging opportunities for PAM, as the field moves beyond proof-of-concept studies towards the applied use of acoustic data to inform environmental management, conservation, and monitoring of ecosystems under global change.

One key trend is the arrival of massive acoustic datasets collected and assembled by collaborative research networks and citizen science programmes (Table 1). In the future, these are likely to further increase in size and spatiotemporal resolution with the development of wireless data transmission from sensor networks (Aide *et al.* 2013) and more accurate automated data processing. Emerging applications of such datasets include the study of species distributions and 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); and large-scale monitoring of species of concern as bioindicators (Jones *et al.* 2013; Rossi *et al.* 2016), ecosystem services providers (e.g. pollinators), pests, invasive species or public health threats (e.g. mosquitoes; Mukundarajan *et al.* 2017). Equally, while key questions remain unanswered for now, the growth of terrestrial and marine ecoacoustics research is likely to ultimately provide both new tools for acoustic biodiversity monitoring, and more fundamental insights into the ecological and evolutionary processes that structure acoustic communities (Farina & James 2016).

Looking further forward, the arrival of networked sensors and on-board analysis pipelines also raises the possibility of using PAM-derived data for real-time environmental monitoring and adaptive management (Table 1). Real-time detections derived from large sensor networks could provide highly spatially- and temporally detailed data on wildlife and human activities

(e.g. London's Nature-Smart Cities bat monitoring project: <https://naturesmartcities.com>). Such data feeds could, for instance, be applied to adjust urban lighting regimes to reduce their impacts on bat activity, adaptively reroute shipping traffic to avoid threatened cetacean populations (Van Parijs *et al.* 2009; Davis *et al.* 2017), or report on hotspots of illegal logging or hunting (Astaras *et al.* 2017, Rainforest Connection <https://rfcx.org>). Beyond the likely institutional and political barriers, there are still substantial technical difficulties involved in developing the necessary infrastructure. For example, making accurate, fully automated call classification feasible for most species would require the development of rigorously-tested sound identification algorithms that are robust to the heterogeneity of real-world acoustic environments. Nonetheless, these possibilities represent promising potential futures for a technology that, in combination with data from other remote sensing technologies, is now helping to provide an increasingly sensitive global picture of the effects of human pressures on wildlife and ecosystems.

Disclosure statement

The authors declare no conflicts of interest.

Acknowledgements

This research was supported by funding from WWF UK, NERC (NE/P016677/1), and EPSRC (EP/K503745/1). The authors thank Alison Fairbrass, Oisín Mac Aodha, Stuart Newson, Alex Rogers, Andrew Hill, Peter Prince and Nick Tregenza for discussion and comments on previous versions of the manuscript, and our three anonymous reviewers for their comments and recommendations. We also gratefully thank the respondents of our 2016 WWF-UK online survey on best practices in PAM: Yves Bas, Geoff Billington, Zuzana Burivalova, Dena Clink, Johan De Ridder, Justin Halls, Simon Hoggett, Tom Hastings, David Jacoby, Ammie Kalan, Arik Kershenbaum, Simon Linke, Steve Lucas, Ricardo Machado,

Peter Owens, Christine Sutter, Paul Trethowan, Robbie Whytock and Peter Wrege, alongside
8 anonymous participants.

References

1. 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 Ecol. Evol.*,
3, 992–998

2. 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

3. Mac Aodha, O., Gibb, R., Barlow, K.E., Browning, E., Firman, M., Freeman, R., *et al.*
(2017). Bat Detective - Deep learning tools for bat acoustic signal detection. *bioRxiv Prepr.*
<http://dx.doi.org/10.1101/156869>

4. 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

- 683 5. Astaras, C., Linder, J.M., Wrege, P., Orume, R.D. & Macdonald, D.W. (2017). Passive
 684 acoustic monitoring as a law enforcement tool for Afrotropical rainforests. *Front. Ecol.*
 685 *Environ.*, 15, 233–234
- 686
- 687 6. Barlow, K.E., Briggs, P.A., Haysom, K.A., Hutson, A.M., Lechiara, N.L., Racey, P.A., *et*
 688 *al.* (2015). Citizen science reveals trends in bat populations: The National Bat Monitoring
 689 Programme in Great Britain. *Biol. Conserv.*, 182, 14–26
- 690
- 691 7. Bas, Y., Bas, D. & Julien, J.-F. (2017). Tadarida: A Toolbox for Animal Detection on
 692 Acoustic Recordings. *J. Open Res. Softw.*, 5
- 693
- 694 8. Baumgartner, M.F., Fratantoni, D.M., Hurst, T.P., Brown, M.W., Cole, T.V.N., Van Parijs,
 695 S.M., *et al.* (2013). Real-time reporting of baleen whale passive acoustic detections from
 696 ocean gliders. *J. Acoust. Soc. Am.*, 134, 1814–23
- 697
- 698 9. Baumgartner, M.F. & Mussoline, S.E. (2011). A generalized baleen whale call detection
 699 and classification system. *J. Acoust. Soc. Am.*, 129, 2889–902
- 700
- 701 10. Bittle, M. & Duncan, A. (2013). A review of current marine mammal detection and
 702 classification algorithms for use in automated passive acoustic monitoring. *Proc. Acoust. Sci.*
 703 *Technol. Amenity Nov 17-20 2013*, 1–8
- 704
- 705 11. Blangiardo, M., Cameletti, M., Baio, G. & Rue, H. (2013). Spatial and spatio-temporal

models with R-INLA. *Spat. Spatiotemporal. Epidemiol.*, 7, 33–49

12. Blumstein, D.T., Mennill, D.J., Clemins, P., Girod, L., Yao, K., Patricelli, G., *et al.* (2011). Acoustic monitoring in terrestrial environments using microphone arrays: applications, technological considerations and prospectus. *J. Appl. Ecol.*, 48, 758–767

13. 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. *Landsc. Urban Plan.*, 162, 44–55

14. Bradbury, J. & Vehrencamp, S. (1998). *Principles of Animal Communication*. Sinauer Associates Inc., U.S.

15. Braulik, G., Wittich, A., Macaulay, J., Kasuga, M., Gordon, J., Davenport, T.R.B., *et al.* (2017a). Acoustic monitoring to document the spatial distribution and hotspots of blast fishing in Tanzania. *Mar. Pollut. Bull.*, 125, 360–366

16. Braulik, G.T., Kasuga, M., Wittich, A., Kiszka, J.J., MacCaulay, J., Gillespie, D., *et al.* (2017b). Cetacean rapid assessment: An approach to fill knowledge gaps and target conservation across large data deficient areas. *Aquat. Conserv. Mar. Freshw. Ecosyst.*

17. Braun de Torrez, 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

- 729 sampling for monitoring bats. *PeerJ*, 5, e3940
- 730
- 731 18. Browning, E., Gibb, R., Glover-Kapfer, P. & Jones, K.E. (2017). Passive acoustic
- 732 monitoring in ecology and conservation. *WWF Conserv. Technol. Ser.*, 1, 1–75
- 733
- 734 19. Brumm, H., Zollinger, S.A., Niemelä, P.T. & Sprau, P. (2017). Measurement artefacts
- 735 lead to false positives in the study of birdsong in noise. *Methods Ecol. Evol.*, (online early
- 736 view)
- 737
- 738 20. Bush, A., Sollmann, R., Wilting, A., Bohmann, K., Cole, B., Balzter, H., *et al.* (2017).
- 739 Connecting Earth observation to high-throughput biodiversity data. *Nat. Ecol. Evol.*, 1, 176
- 740
- 741 21. Campos-Cerqueira, M. & Aide, T.M. (2016). Improving distribution data of threatened
- 742 species by combining acoustic monitoring and occupancy modelling. *Methods Ecol. Evol.*, 7,
- 743 1340–1348
- 744
- 745 22. Cardinale, B.J., Duffy, J.E., Gonzalez, A., Hooper, D.U., Perrings, C., Venail, P., *et al.*
- 746 (2012). Biodiversity loss and its impact on humanity. *Nature*, 486, 59–67
- 747
- 748 23. Celis-Murillo, A., Deppe, J.L. & Allen, M.F. (2009). Using soundscape recordings to
- 749 estimate bird species abundance, richness, and composition. *J. F. Ornithol.*, 80, 64–78
- 750

- 751 24. Clink, D.J., Bernard, H., Crofoot, M.C. & Marshall, A.J. (2017). Investigating Individual
752 Vocal Signatures and Small-Scale Patterns of Geographic Variation in Female Bornean
753 Gibbon (*Hylobates muelleri*) Great Calls. *Int. J. Primatol.*, 38, 656–671
754
- 755 25. Darras, K., Pütz, P., Fahrurrozi, Rembold, K. & Tschardtke, T. (2016). Measuring sound
756 detection spaces for acoustic animal sampling and monitoring. *Biol. Conserv.*, 201, 29–37
757
- 758 26. Davis, G.E., Baumgartner, M.F., Bonnell, J.M., Bell, J., Berchok, C., Bort Thornton, J., *et*
759 *al.* (2017). Long-term passive acoustic recordings track the changing distribution of North
760 Atlantic right whales (*Eubalaena glacialis*) from 2004 to 2014. *Sci. Rep.*, 7, 1–12
761
- 762 27. Depraetere, M., Pavoine, S., Jiguet, F., Gasc, A., Duvail, S. & Sueur, J. (2012).
763 Monitoring animal diversity using acoustic indices: Implementation in a temperate woodland.
764 *Ecol. Indic.*, 13, 46–54
765
- 766 28. DeRuiter, S.L., Southall, B.L., Calambokidis, J., Zimmer, W.M.X., Sadykova, D.,
767 Falcone, E.A., *et al.* (2013). First direct measurements of behavioural responses by Cuvier’s
768 beaked whales to mid-frequency active sonar. *Biol. Lett.*, 9, 20130223
769
- 770 29. Digby, A., Towsey, M., Bell, B.D. & Teal, P.D. (2013). A practical comparison of
771 manual and autonomous methods for acoustic monitoring. *Methods Ecol. Evol.*, 4, 675–683
772
- 773 30. Efford, M.G., Dawson, D.K. & Borchers, D.L. (2009). Population density estimated from

locations of individuals on a passive detector array. *Ecology*, 90, 2676–2682

31. 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. *Peer J*, 4, e2108

32. Erbe, C., Verma, A., McCauley, R., Gavrilov, A. & Parnum, I. (2015). The marine
soundscape of the Perth Canyon. *Prog. Oceanogr.*, 137, 38–51

33. Fairbrass, A., Firman, M., Williams, C., Brostow, G.J., Titheridge, H. & Jones, K.E.
(2018). CityNet - Deep Learning Tools for Urban Ecoacoustic Assessment. *bioRxiv Prepr.*
<http://dx.doi.org/10.1101/248708>

34. Fairbrass, A., Rennett, P., Williams, C., Titheridge, H. & Jones, K.E. (2017). Biases of
acoustic indices measuring biodiversity in urban areas. *Ecol. Indic.*, 83, 169–177

35. Farcas, A., Thompson, P.M. & Merchant, N.D. (2016). Underwater noise modelling for
environmental impact assessment. *Environ. Impact Assess. Rev.*, 57, 114–122

36. Farina, A. & James, P. (2016). The acoustic communities: Definition, description and
ecological role. *BioSystems*, 147, 11–20

37. Farina, A., Pieretti, N., Salutari, P., Tognari, E. & Lombardi, A. (2016). The Application of the Acoustic Complexity Indices (ACI) to Ecoacoustic Event Detection and Identification (EEDI) Modeling. *Biosemiotics*, 9, 227–246
38. Fischer, F.P., Schulz, U., Schubert, H., Knapp, P. & Schmoger, M. (1997). Quantitative Assessment of Grassland Quality: Acoustic Determination of Population Sizes of Orthopteran Indicator Species. *Ecol. Appl.*, 7, 909
39. Froidevaux, J.S.P., Zellweger, F., Bollmann, K. & Obrist, M.K. (2014). Optimizing passive acoustic sampling of bats in forests. *Ecol. Evol.*, 4, 4690–4700
40. Garland, E.C., Castellote, M. & Berchok, C.L. (2015a). Beluga whale (*Delphinapterus leucas*) vocalizations and call classification from the eastern Beaufort Sea population. *J. Acoust. Soc. Am.*, 137, 3054–3067
41. Garland, E.C., Castellote, M. & Berchok, C.L. (2015b). Beluga whale (*Delphinapterus leucas*) vocalizations and call classification from the eastern Beaufort Sea population. *J. Acoust. Soc. Am.*, 137, 3054–3067
42. Garland, E.C., Rendell, L., Lilley, M.S., Poole, M.M., Allen, J. & Noad, M.J. (2017). The devil is in the detail: Quantifying vocal variation in a complex, multi-levelled, and rapidly evolving display. *J. Acoust. Soc. Am.*, 142, 460–472

43. 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. *Biol. Conserv.*, 191, 306–312
44. Gasc, A., Sueur, J., Jiguet, F., Devictor, V., Grandcolas, P., Burrow, C., *et al.* (2013). Assessing biodiversity with sound: Do acoustic diversity indices reflect phylogenetic and functional diversities of bird communities? *Ecol. Indic.*, 25, 279–287
45. Gillespie, D., Gordon, J., McHugh, R., McLaren, D., Mellinger, D.K., Redmond, P., *et al.* (2008). PAMGUARD: Semiautomated, open source software for real-time acoustic detection and localisation of cetaceans. In: *Proceedings of the Institute of Acoustics*. pp. 54–62
46. Goeau, H., Glotin, H., Vellinga, W.-P., Planque, R. & Joly, A. (2016). LifeCLEF Bird Identification Task 2016: The arrival of deep learning. *CLEF2016 Work. Notes*, 440–449
47. Gregory, R.D., Gibbons, D.W. & Donald, P.F. (2004). *Bird census and survey techniques*. Oxford University Press, Oxford, UK
48. Hancock, S., Anderson, K., Disney, M. & Gaston, K.J. (2017). Measurement of fine-spatial-resolution 3D vegetation structure with airborne waveform lidar: Calibration and validation with voxelised terrestrial lidar. *Remote Sens. Environ.*, 188, 37–50
49. Hariharan, B. & Girshick, R. (2016). Low-shot Visual Recognition by Shrinking and

- 842 Hallucinating Features. *arXiv Prepr.* <http://arxiv.org/abs/1606.02819>
- 843
- 844 50. Harris, S.A., Shears, N.T. & Radford, C.A. (2016). Ecoacoustic indices as proxies for
- 845 biodiversity on temperate reefs. *Methods Ecol. Evol.*, 7, 713–724
- 846
- 847 51. Harrison, P.J., Buckland, S.T., Yuan, Y., Elston, D.A., Brewer, M.J., Johnston, A., *et al.*
- 848 (2014). Assessing trends in biodiversity over space and time using the example of British
- 849 breeding birds. *J. Appl. Ecol.*, 51, 1650–1660
- 850
- 851 52. Heinicke, S., Kalan, A.K., Wagner, O.J.J., Mundry, R., Lukashevich, H. & Kühl, H.S.
- 852 (2015). Assessing the performance of a semi-automated acoustic monitoring system for
- 853 primates. *Methods Ecol. Evol.*, 6, 753–763
- 854
- 855 53. Hill, A.P., Prince, P., Piña Covarrubias, E., Doncaster, C.P., Snaddon, J.L. & Rogers, A.
- 856 (2018). AudioMoth: Evaluation of a smart open acoustic device for monitoring biodiversity
- 857 and the environment. *Methods Ecol. Evol.*
- 858
- 859 54. Honrado, J.P., Pereira, H.M. & Guisan, A. (2016). Fostering integration between
- 860 biodiversity monitoring and modelling. *J. Appl. Ecol.*, 53, 1299–1304
- 861
- 862 55. Isaac, N.J.B., van Strien, A.J., August, T.A., de Zeeuw, M.P. & Roy, D.B. (2014).
- 863 Statistics for citizen science: Extracting signals of change from noisy ecological data.
- 864 *Methods Ecol. Evol.*, 5, 1052–1060

865

866 56. Jaramillo-Legorreta, A., Cardenas-Hinojosa, G., Nieto-Garcia, E., Rojas-Bracho, L.,
 867 Hoef, J. Ver, Moore, J., *et al.* (2016). Passive acoustic monitoring of the decline of Mexico's
 868 critically endangered vaquita. *Conserv. Biol.*, 31, 183–191

869

870 57. Johnson, M.P. & Tyack, P.L. (2003). A digital acoustic recording tag for measuring the
 871 response of wild marine mammals to sound. *IEEE J. Ocean. Eng.*, 28, 3–12

872

873 58. Jones, K.E., Russ, J.A., Basgta, A.-T., Bilhari, Z., Catto, C., Csösz, I., *et al.* (2013).
 874 Indicator Bats Program: A System for the Global Acoustic Monitoring of Bats. In:
 875 *Biodiversity Monitoring and Conservation: Bridging the Gap between Global Commitment*
 876 *and Local Action* (eds. Collen, B., Pettorelli, N., Baillie, J.E.M. & Durant, S.M.). Wiley-
 877 Blackwell, Oxford, UK, pp. 213–247

878

879 59. Kaewtip, K., Alwan, A., O'Reilly, C. & Taylor, C.E. (2016). A robust automatic birdsong
 880 phrase classification: A template-based approach. *J. Acoust. Soc. Am.*, 140, 3691–3701

881

882 60. Kalan, A.K., Mundry, R., Wagner, O.J.J., Heinicke, S., Boesch, C. & Köhl, H.S. (2015).
 883 Towards the automated detection and occupancy estimation of primates using passive
 884 acoustic monitoring. *Ecol. Indic.*, 54, 217–226

885

886 61. Kasten, E.P., Gage, S.H., Fox, J. & Joo, W. (2012). The remote environmental assessment
 887 laboratory's acoustic library: An archive for studying soundscape ecology. *Ecol. Inform.*, 12,

888 50–67

889

890 62. Kershenbaum, A., Blumstein, D.T., Roch, M.A., Akçay, Ç., Backus, G., Bee, M.A., *et al.*
 891 (2014). Acoustic sequences in non-human animals: a tutorial review and prospectus. *Biol.*
 892 *Rev.*, 91, 13–52

893

894 63. Kéry, M. & Schmidt, B. (2008). Imperfect detection and its consequences for monitoring
 895 for conservation. *Community Ecol.*, 9, 207–216

896

897 64. King, S.L., Sayigh, L.S., Wells, R.S., Fellner, W., Janik, V.M., Janik, V., *et al.* (2013).
 898 Vocal copying of individually distinctive signature whistles in bottlenose dolphins. *Proc.*
 899 *Biol. Sci.*, 280, 20130053

900

901 65. Kirschel, A.N.G., Earl, D.A., Yao, Y., Escobar, I.A., Vilches, E., Vallejo, E.E., *et al.*
 902 (2009). Using Songs To Identify Individual Mexican Antthrush *Formicarius Moniliger*:
 903 Comparison of Four Classification Methods. *Bioacoustics*, 19, 1–20

904

905 66. Klinck, H., Mellinger, D.K., Klinck, K., Bogue, N.M., Luby, J.C., Jump, W.A., *et al.*
 906 (2012). Near-real-time acoustic monitoring of beaked whales and other cetaceans using a
 907 SeagliderTM. *PLoS One*, 7, e36128

908

909 67. Klingbeil, B.T. & Willig, M.R. (2015). Bird biodiversity assessments in temperate forest:
 910 the value of point count versus acoustic monitoring protocols. *PeerJ*, 3, e973

911

912 68. Krause, J., Sapp, B., Howard, A., Zhou, H., Toshev, A., Duerig, T., *et al.* (2016). The913 unreasonable effectiveness of noisy data for fine-grained recognition. *arXiv Prepr.*914 <http://arxiv.org/abs/1511.06789>

915

916 69. Lahoz-Monfort, J.J., Guillera-Arroita, G. & Wintle, B.A. (2014). Imperfect detection

917 impacts the performance of species distribution models. *Glob. Ecol. Biogeogr.*, 23, 504–515

918

919 70. Lehmann, G.U.C., Frommolt, K.H., Lehmann, A.W. & Riede, K. (2014). Baseline data

920 for automated acoustic monitoring of Orthoptera in a Mediterranean landscape, the Hymettos,

921 Greece. *J. Insect Conserv.*, 18, 909–925

922

923 71. Lellouch, L., Pavoine, S., Jiguet, F., Glotin, H. & Sueur, J. (2014). Monitoring temporal

924 change of bird communities with dissimilarity acoustic indices. *Methods Ecol. Evol.*, 5, 495–

925 505

926

927 72. Linhart, P. & Šálek, M. (2017). The assessment of biases in the acoustic discrimination of

928 individuals. *PLoS One*, 12, e0177206

929

930 73. Lucas, T.C.D., Moorcroft, E.A., Freeman, R., Rowcliffe, J.M. & Jones, K.E. (2015). A

931 generalised random encounter model for estimating animal density with remote sensor data.

932 *Methods Ecol. Evol.*, 6, 500–509

933

74. Lynch, E., Angeloni, L., Fristrup, K., Joyce, D. & Wittemyer, G. (2013). The use of on-animal acoustical recording devices for studying animal behavior. *Ecol. Evol.*, 3, 2030–2037

75. MacSwiney G, M.C., Clarke, F.M. & Racey, P.A. (2008). What you see is not what you get: the role of ultrasonic detectors in increasing inventory completeness in Neotropical bat assemblages. *J. Appl. Ecol.*, 45, 1364–1371

76. Marinexplore. (2013). *The Marinexplore and Cornell University Whale Detection Challenge: competition summary*. Kaggle. Available at: <https://www.kaggle.com/c/whale-detection-challenge/discussion/4472>. Last accessed 20 January 2018

77. Marley, S.A., Salgado Kent, C.P., Erbe, C. & Parnum, I.M. (2017). Effects of vessel traffic and underwater noise on the movement, behaviour and vocalisations of bottlenose dolphins in an urbanised estuary. *Sci. Rep.*, 7, 1–14

78. Marques, T.A., Thomas, L., Martin, S.W., Mellinger, D.K., Jarvis, S., Morrissey, R.P., *et al.* (2012). Spatially explicit capture-recapture methods to estimate minke whale density from data collected at bottom-mounted hydrophones. *J. Ornithol.*, 152, 445–455

79. Marques, T.A., Thomas, L., Martin, S.W., Mellinger, D.K., Ward, J.A., Moretti, D.J., *et al.* (2013). Estimating animal population density using passive acoustics. *Biol. Rev.*, 88, 287–309

- 957 80. Martin, S.W., Marques, T.A., Thomas, L., Morrissey, R.P., Jarvis, S., DiMarzio, N., *et al.*
 958 (2013). Estimating minke whale (*Balaenoptera acutorostrata*) boing sound density using
 959 passive acoustic sensors. *Mar. Mammal Sci.*, 29, 142–158
- 960
- 961 81. McWilliam, J.N. & Hawkins, A.D. (2013). A comparison of inshore marine soundscapes.
 962 *J. Exp. Mar. Bio. Ecol.*, 446, 166–176
- 963
- 964 82. McWilliam, J.N., McCauley, R.D., Erbe, C. & Parsons, M.J.G. (2017). Patterns of
 965 biophonic periodicity on coral reefs in the Great Barrier Reef. *Sci. Rep.*, 7, 17459
- 966
- 967 83. Mellinger, D.K. & Clark, C.W. (2006). MobySound: A reference archive for studying
 968 automatic recognition of marine mammal sounds. *Appl. Acoust.*, 67, 1226–1242
- 969
- 970 84. Mellinger, D.K., Nieuwkirk, S.L., Klinck, K., Klinck, H., Dziak, R.P., Clapham, P.J., *et al.*
 971 (2011). Confirmation of right whales near a nineteenth-century whaling ground east of
 972 southern Greenland. *Biol. Lett.*, 7, 411–3
- 973
- 974 85. Mellinger, D.K., Stafford, K.M., Moore, S.E., Dziak, R.P. & Matsumoto, H. (2007). An
 975 overview of fixed passive acoustic observation methods for cetaceans. *Oceanography*, 20,
 976 236–45
- 977
- 978 86. Merchant, N.D., Fristrup, K.M., Johnson, M.P., Tyack, P.L., Witt, M.J., Blondel, P., *et al.*
 979 (2015). Measuring acoustic habitats. *Methods Ecol. Evol.*, 6, 257–265

87. Muda, L., Begam, M. & Elamvazuthi, I. (2010). Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques
88. 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
89. Nedelec, S.L., Simpson, S.D., Holderied, M., Radford, A.N., Lecellier, G., Radford, C., *et al.* (2015). Soundscapes and living communities in coral reefs: temporal and spatial variation. *Mar. Ecol. Prog. Ser.*, 524, 125–135
90. 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 Ecol. Evol.*, 8, 1051–1062
91. 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. *Biol. Conserv.*, 191, 38–49
92. 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 Prepr. <http://arxiv.org/abs/1703.05830>*, 1–12

- 1003 93. Nowacek, D.P., Christiansen, F., Bejder, L., Goldbogen, J.A. & Friedlaender, A.S.
 1004 (2016). Studying cetacean behaviour: new technological approaches and conservation
 1005 applications. *Anim. Behav.*, 120, 235–244
- 1006
- 1007 94. Palmer, K.J., Brookes, K. & Rendell, L. (2017). Categorizing click trains to increase
 1008 taxonomic precision in echolocation click loggers. *J. Acoust. Soc. Am.*, 142, 863–877
- 1009
- 1010 95. Van Parijs, S., Clark, C., Sousa-Lima, R., Parks, S., Rankin, S., Risch, D., *et al.* (2009).
 1011 Management and research applications of real-time and archival passive acoustic sensors
 1012 over varying temporal and spatial scales. *Mar. Ecol. Prog. Ser.*, 395, 21–36
- 1013
- 1014 96. Van Parijs, S.M., Baumgartner, M., Cholewiak, D., Davis, G., Gedamke, J., Gerlach, D.,
 1015 *et al.* (2015). NEPAN: A U.S. Northeast Passive Acoustic Sensing Network for Monitoring,
 1016 Reducing Threats and the Conservation of Marine Animals Passive Acoustic Monitoring as a
 1017 Tool for Research and Management in Marine Ecosystems. *Mar. Technol. Soc. J.*, 49, 70–86
- 1018
- 1019 97. Penone, C., Le Viol, I., Pellissier, V., Julien, J.F., Bas, Y. & Kerbiriou, C. (2013). Use of
 1020 large-scale acoustic monitoring to assess anthropogenic pressures on orthoptera communities.
 1021 *Conserv. Biol.*, 27, 979–987
- 1022
- 1023 98. Petrusková, T., Pišvejcová, I., Kinštová, A., Brinke, T. & Petrusek, A. (2015). Repertoire-
 1024 based individual acoustic monitoring of a migratory passerine bird with complex song as an
 1025 efficient tool for tracking territorial dynamics and annual return rates. *Methods Ecol. Evol.*, 7,

- 1026 274–284
- 1027
- 1028 99. Pettorelli, N., Owen, H.J.F. & Duncan, C. (2016). How do we want satellite remote
 1029 sensing to support biodiversity conservation globally? *Methods Ecol. Evol.*, 7, 656–665
- 1030
- 1031 100. Pieretti, N., Farina, A. & Morri, D. (2011). A new methodology to infer the singing
 1032 activity of an avian community: the Acoustic Complexity Index (ACI). *Ecol. Indic.*, 11, 868–
 1033 873
- 1034
- 1035 101. Pijanowski, B.C., Farina, A., Gage, S.H., Dumyahn, S.L. & Krause, B.L. (2011). What
 1036 is soundscape ecology? An introduction and overview of an emerging new science. *Landsc.*
 1037 *Ecol.*, 26, 1213–1232
- 1038
- 1039 102. Pirotta, E., Merchant, N.D., Thompson, P.M., Barton, T.R. & Lusseau, D. (2015).
 1040 Quantifying the effect of boat disturbance on bottlenose dolphin foraging activity. *Biol.*
 1041 *Conserv.*, 181, 82–89
- 1042
- 1043 103. Potamitis, I., Ntalampiras, S., Jahn, O. & Riede, K. (2014). Automatic bird sound
 1044 detection in long real-field recordings: Applications and tools. *Appl. Acoust.*, 80, 1–9
- 1045
- 1046 104. Prat, Y., Taub, M. & Yovel, Y. (2016). Everyday bat vocalizations contain information
 1047 about emitter, addressee, context, and behavior. *Sci. Rep.*, 6, 39419

1048

1049 105. Proppe, D.S., Sturdy, C.B. & St. Clair, C.C. (2013). Anthropogenic noise decreases
 1050 urban songbird diversity and may contribute to homogenization. *Glob. Chang. Biol.*, 19,
 1051 1075–1084

1052

1053 106. Riede, K. (1993). Monitoring Biodiversity : Analysis of Amazonian Rainforest Sounds.
 1054 *Ambio*, 22, 546–548

1055

1056 107. Riede, K., Nischk, F., Dietrich, C., Thiel, C. & Schwenker, F. (2006). Automated
 1057 annotation of Orthoptera songs: first results from analysing the DORSA sound repository. *J.*
 1058 *Orthoptera Res.*, 15, 105–113

1059

1060 108. Risch, D., Castellote, M., Clark, C.W., Davis, G.E., Dugan, P.J., Hodge, L.E., *et al.*
 1061 (2014). Seasonal migrations of North Atlantic minke whales: novel insights from large-scale
 1062 passive acoustic monitoring networks. *Mov. Ecol.*, 2, 24

1063

1064 109. Roch, M.A., Batchelor, H., Baumann-Pickering, S., Berchok, C.L., Cholewiak, D.,
 1065 Fujioka, E., *et al.* (2016). Management of acoustic metadata for bioacoustics. *Ecol. Inform.*,
 1066 31, 122–136

1067

1068 110. Root-Gutteridge, H., Bencsik, M., Chebli, M., Gentle, L.K., Terrell-Nield, C., Bourit,
 1069 A., *et al.* (2014). Identifying individual wild Eastern grey wolves (*Canis lupus lycaon*) using
 1070 fundamental frequency and amplitude of howls. *Bioacoustics*, 23, 55–66

1071

1072 111. Rossi, T., Connell, S.D. & Nagelkerken, I. (2016). Silent oceans: ocean acidification
 1073 impoverishes natural soundscapes by altering sound production of the world's noisiest marine
 1074 invertebrate. *Proc. R. Soc. B Biol. Sci.*, 283, 20153046

1075

1076 112. Rowcliffe, J.M., Field, J., Turvey, S.T. & Carbone, C. (2008). Estimating animal density
 1077 using camera traps without the need for individual recognition. *J. Appl. Ecol.*, 45, 1228–1236

1078

1079 113. Ruiz-Gutierrez, V., Hooten, M.B. & Campbell Grant, E.H. (2016). Uncertainty in
 1080 biological monitoring: a framework for data collection and analysis to account for multiple
 1081 sources of sampling bias. *Methods Ecol. Evol.*, 7, 900–909

1082

1083 114. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., *et al.* (2015).
 1084 ImageNet Large Scale Visual Recognition Challenge. *Int. J. Comput. Vis.*, 115, 211–252

1085

1086 115. Salamon, J. & Bello, J.P. (2015). Unsupervised feature learning for urban sound
 1087 classification. *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.*, 171–175

1088

1089 116. Salamon, J. & Bello, J.P. (2016). Deep Convolutional Neural Networks and Data
 1090 Augmentation for Environmental Sound Classification. *arXiv Prepr.*
 1091 <http://arxiv.org/abs/1604.07160>

1092

1093 117. Sanders, C.E. & Mennill, D.J. (2014). Acoustic monitoring of nocturnally migrating

- 1094 birds accurately assesses the timing and magnitude of migration through the Great Lakes.
- 1095 *Condor*, 116, 371–383
- 1096
- 1097 118. Sayigh, L., Daher, M.A., Allen, J., Gordon, H., Joyce, K., Stuhlmann, C., *et al.* (2016).
- 1098 The Watkins Marine Mammal Sound Database: An online, freely accessible resource. p.
- 1099 40013
- 1100
- 1101 119. Secretariat of the Convention on Biological Diversity. (2014). *Global Biodiversity*
- 1102 *Outlook 4*. Montreal
- 1103
- 1104 120. Simpson, S.D., Meekan, M.G., Jeffs, A., Montgomery, J.C. & McCauley, R.D. (2008).
- 1105 Settlement-stage coral reef fish prefer the higher-frequency invertebrate-generated audible
- 1106 component of reef noise. *Anim. Behav.*, 75, 1861–1868
- 1107
- 1108 121. Skowronski, M.D. & Harris, J.G. (2006). Acoustic detection and classification of
- 1109 microchiroptera using machine learning: Lessons learned from automatic speech recognition.
- 1110 *J. Acoust. Soc. Am.*, 119, 1817–1833
- 1111
- 1112 122. Sousa-Lima, R.S., Fernandes, D.P., Norris, T.F. & Oswald, J.N. (2013). A review and
- 1113 inventory of fixed autonomous recorders for passive acoustic monitoring of marine
- 1114 mammals. *Aquat. Mamm.*, 39, 23–53
- 1115
- 1116 123. Staaterman, E., Ogburn, M., Altieri, A., Brandl, S., Whippo, R., Seemann, J., *et al.*

- 1117 (2017). Bioacoustic measurements complement visual biodiversity surveys: preliminary
 1118 evidence from four shallow marine habitats. *Mar. Ecol. Prog. Ser.*, 575, 207–215
- 1119
- 1120 124. Stathopoulos, V., Zamora-Gutierrez, V., Jones, K.E. & Girolami, M. (2017). Bat
 1121 echolocation call identification for biodiversity monitoring: a probabilistic approach. *J. R.*
 1122 *Stat. Soc. Ser. C Appl. Stat.*, (online early view)
- 1123
- 1124 125. Steenweg, R., Hebblewhite, M., Kays, R., Ahumada, J., Fisher, J.T., Burton, C., *et al.*
 1125 (2017). Scaling up camera traps: monitoring the planet’s biodiversity with networks of
 1126 remote sensors. *Front. Ecol. Environ.*, 15, 26–34
- 1127
- 1128 126. Stevenson, B.C., Borchers, D.L., Altwegg, R., Swift, R.J., Gillespie, D.M. & Measey,
 1129 G.J. (2015). A general framework for animal density estimation from acoustic detections
 1130 across a fixed microphone array. *Methods Ecol. Evol.*, 6, 38–48
- 1131
- 1132 127. Stowell, D. & Plumbley, M.D. (2014). Automatic large-scale classification of bird
 1133 sounds is strongly improved by unsupervised feature learning. *PeerJ*, 2, e488
- 1134
- 1135 128. Stowell, D., Wood, M., Stylianou, Y. & Glotin, H. (2016). Bird detection in audio: a
 1136 survey and a challenge. In: *Proceedings of the 2016 IEEE International Workshop on*
 1137 *Machine Learning for Signal Processing*
- 1138
- 1139 129. Suberg, L., Wynn, R.B., Kooij, J. van der, Fernand, L., Fielding, S., Guihen, D., *et al.*

- 1140 (2014). Assessing the potential of autonomous submarine gliders for ecosystem monitoring
 1141 across multiple trophic levels (plankton to cetaceans) and pollutants in shallow shelf seas.
 1142 *Methods Oceanogr.*, 10, 70–89
- 1143
- 1144 130. Sueur, J. & Farina, A. (2015). Ecoacoustics: the Ecological Investigation and
 1145 Interpretation of Environmental Sound. *Biosemitotics*, 8, 493–502
- 1146
- 1147 131. Sueur, J., Farina, A., Gasc, A., Pieretti, N. & Pavoine, S. (2014). Acoustic indices for
 1148 biodiversity assessment and landscape investigation. *Acta Acust. United With Acust.*, 100,
 1149 772–781
- 1150
- 1151 132. Sueur, J., Pavoine, S., Hamerlynck, O. & Duvail, S. (2008). Rapid acoustic survey for
 1152 biodiversity appraisal. *PLoS One*, 3, e4065
- 1153
- 1154 133. Thompson, M.E., Schwager, S.J. & Payne, K.B. (2010). Heard but not seen: An acoustic
 1155 survey of the African forest elephant population at Kakum Conservation Area, Ghana. *Afr. J.*
 1156 *Ecol.*, 48, 224–231
- 1157
- 1158 134. Tobias, J.A., Planqué, R., Cram, D.L. & Seddon, N. (2014). Species interactions and the
 1159 structure of complex communication networks. *Proc. Natl. Acad. Sci. U. S. A.*, 111, 1020–
 1160 1025
- 1161
- 1162 135. Tolkova, I., Bauer, L., Wilby, A., Kastner, R. & Seger, K. (2017). Automatic

- 1163 classification of humpback whale social calls. *J. Acoust. Soc. Am.*, 141, 3605
- 1164
- 1165 136. Tucker, D., Gage, S.H., Williamson, I. & Fuller, S. (2014). Linking ecological condition
- 1166 and the soundscape in fragmented Australian forests. *Landsc. Ecol.*, 29, 745–758
- 1167
- 1168 137. Walters, C.L., Collen, A., Lucas, T.C.D., Mroz, K., Sayer, C.A. & Jones, K.E. (2013).
- 1169 Challenges of using bioacoustics to globally monitor bats. In: *Bat Evolution, Ecology and*
- 1170 *Conservation* (eds. Adams, R.A. & Pedersen, S.C.). Springer New York
- 1171
- 1172 138. Walters, C.L., Freeman, R., Collen, A., Dietz, C., Brock Fenton, M., Jones, G., *et al.*
- 1173 (2012). A continental-scale tool for acoustic identification of European bats. *J. Appl. Ecol.*,
- 1174 49, 1064–1074
- 1175
- 1176 139. Ward, J.A., Thomas, L., Jarvis, S., DiMarzio, N., Moretti, D., Marques, T.A., *et al.*
- 1177 (2012). Passive acoustic density estimation of sperm whales in the Tongue of the Ocean,
- 1178 Bahamas. *Mar. Mammal Sci.*, 28, E444–E455
- 1179
- 1180 140. Weir, L., Fiske, I.J. & Royle, A.J. (2009). Trends in Anuran Occupancy from
- 1181 Northeastern States of the North American Amphibian Monitoring Program. *Herpetol.*
- 1182 *Conserv. Biol.*, 4, 389–402
- 1183
- 1184 141. Whytock, R.C. & Christie, J. (2017). Solo: an open source, customisable and
- 1185 inexpensive audio recorder for bioacoustic research. *Methods Ecol. Evol.*, 8, 308–312

1186

1187 142. Wrege, P.H., Rowland, E.D., Keen, S. & Shiu, Y. (2017). Acoustic monitoring for
 1188 conservation in tropical forests: examples from forest elephants. *Methods Ecol. Evol.*, (online
 1189 early view)

1190

1191 143. Wrege, P.H., Rowland, E.D., Thompson, B.G. & Batruch, N. (2010). Use of acoustic
 1192 tools to reveal otherwise cryptic responses of forest elephants to oil exploration. *Conserv.*
 1193 *Biol.*, 24, 1578–1585

1194

1195 144. Yack, T.M., Barlow, J., Calambokidis, J., Southall, B. & Coates, S. (2013). Passive
 1196 acoustic monitoring using a towed hydrophone array results in identification of a previously
 1197 unknown beaked whale habitat. *J. Acoust. Soc. Am.*, 134, 2589–2595

1198

1199 145. Zamora-Gutierrez, V., Lopez-Gonzalez, C., Macswiney Gonzalez, M.C., Fenton, B.,
 1200 Jones, G., Kalko, E.K. V, *et al.* (2016). Acoustic identification of Mexican bats based on
 1201 taxonomic and ecological constraints on call design. *Methods Ecol. Evol.*, 7, 1082–1091

1202

1203 146. Zero, V.H., Sundaresan, S.R., O'Brien, T.G. & Kinnaird, M.F. (2013). Monitoring an
 1204 Endangered savannah ungulate, Grevy's zebra *Equus grevyi*: choosing a method for
 1205 estimating population densities. *Oryx*, 47, 410–419

1206

1207 147. Zilli, D., Parson, O., Merrett, G. V. & Rogers, A. (2014). A Hidden Markov Model-
 1208 Based Acoustic Cicada Detector for Crowdsourced Smartphone Biodiversity Monitoring. *J.*

1209 *Artif. Intell. Res.*, 51, 805–827

1210

1211 148. Zimmer, W.M.X. (2011). *Passive acoustic monitoring of cetaceans*. Cambridge

1212 University Press, Cambridge

1213

Figures and tables

Figure 1: Acoustic ecological monitoring. The stages of a typical passive acoustic monitoring (PAM) workflow involve objectives-driven data collection over multiple years and sites, audio processing to generate a spatially- and temporally-explicit species detection record (often aggregated into count data), calculation of global sound pressure metrics or ecoacoustic indices, and statistical analysis to yield ecologically-relevant trends.

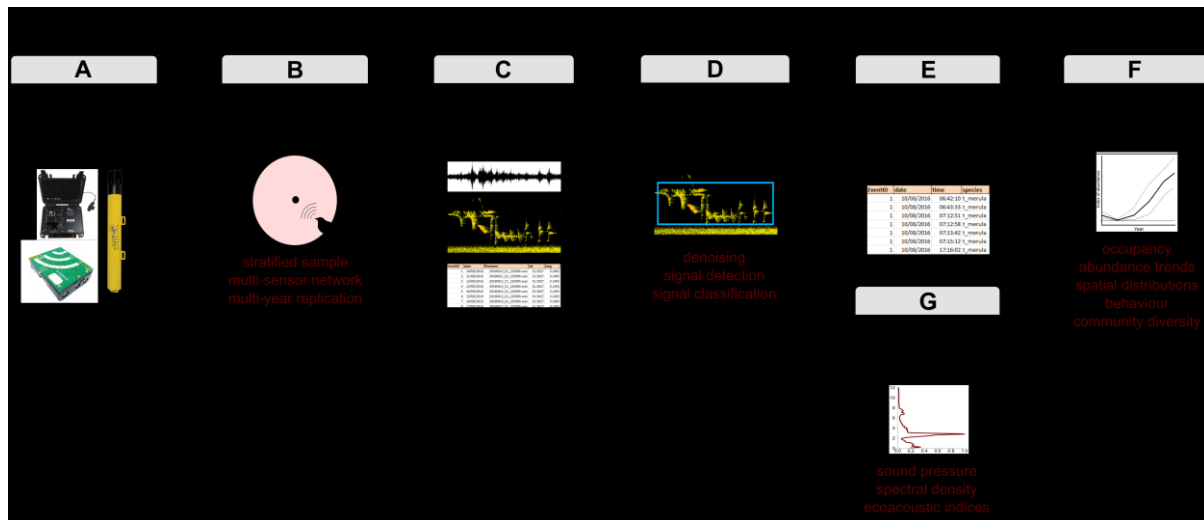


Figure 2: Comparison of recording quality across different sensor models. Spectrograms of two full-spectrum ultrasound bat call recordings made simultaneously on two co-deployed sensors: a commercial device with electret microphone (Batlogger M, Elekon AG Switzerland) (A), and a low-cost model based on a microprocessor board and MEMS microphone (AudioMoth) (B). Spectrograms show time on the x-axis, frequency on the y-axis, and amplitude on a linear colour scale (amplitude normalised with peak at 0dB, values below -30dB shown in black for visualisation). Echolocation calls are visible as bright patches between 20 and 60kHz. Overall performance is similar, but the comparison highlights differences in frequency sensitivity that could impact species identification, with higher frequencies more consistently resolved in (A). The low-cost sensor recording also contains larger amounts of background sound, which show up as heterogeneous bands in the low-to-mid frequency range.

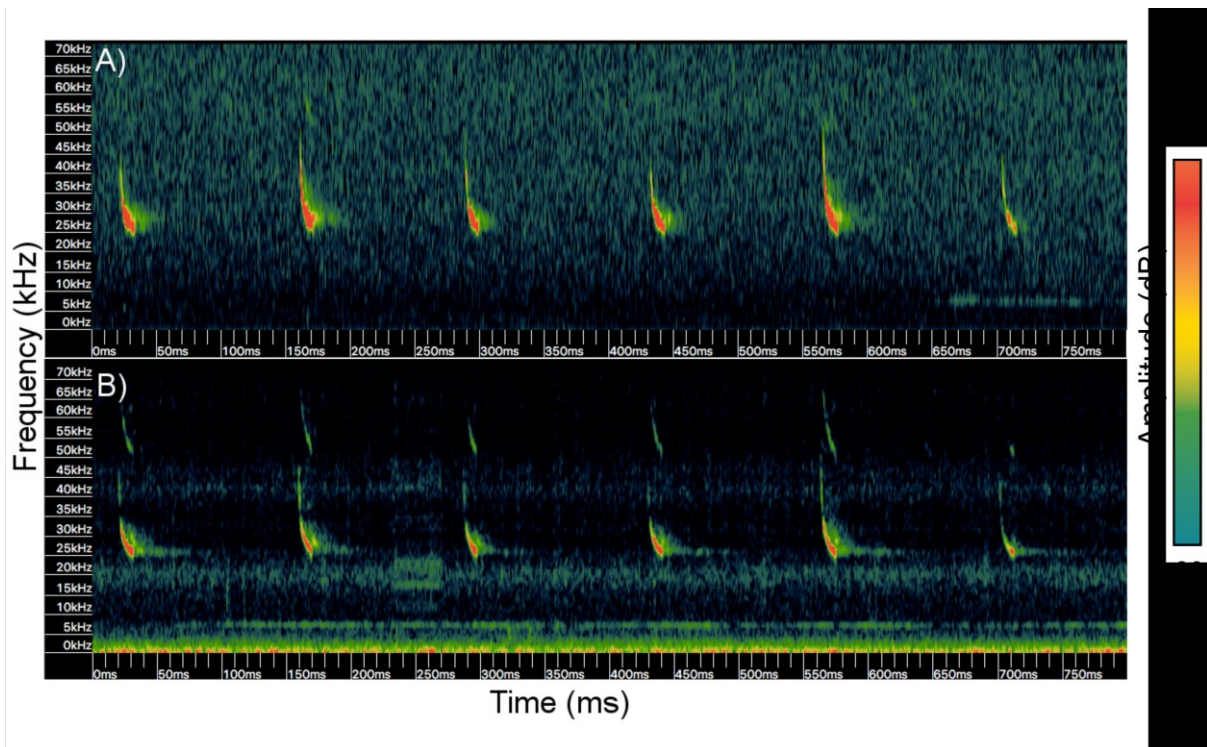


Figure 3: The effects of distance from sensor, call amplitude and habitat clutter on vocalizing animal detectability. Sensor is shown as a black circle with detection radius shaded in pink. Sound emitted within a microphone/hydrophone's detection space will be successfully recorded (A), whereas animals calling outside this radius will not be recorded (B). Habitat clutter causes acoustic interference, particularly for higher-frequency sounds (e.g. ultrasonic echolocation), and may decrease the likelihood of detection (B). Figure reproduced with permission from (Browning *et al.* 2017).

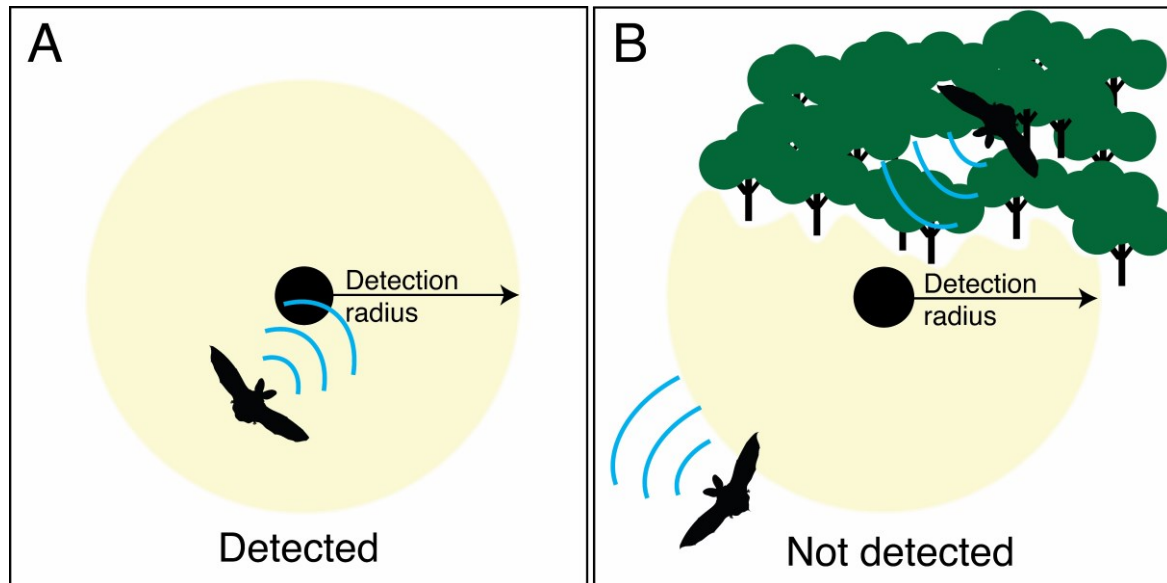


Figure 4: Automated detection and classification of animal calls. A generalised example of automated sound identification, in this case for ultrasonic bat calls. Signal processing recovers frequency information from time-amplitude sound wave recordings (A) to generate a time-frequency representation (spectrogram, with amplitude shown as colour intensity) (B). Each call of interest is located in the recording (C) and features are extracted (either hand-crafted, generic or learned from the data using unsupervised techniques) (D), then calls are classified to a category (species and call type, in this case either search-phase echolocation or social calls) (E). Figure modified with permission from (Browning *et al.* 2017). Photo © Hugh Clark/www.bats.org.uk, reproduced with permission.

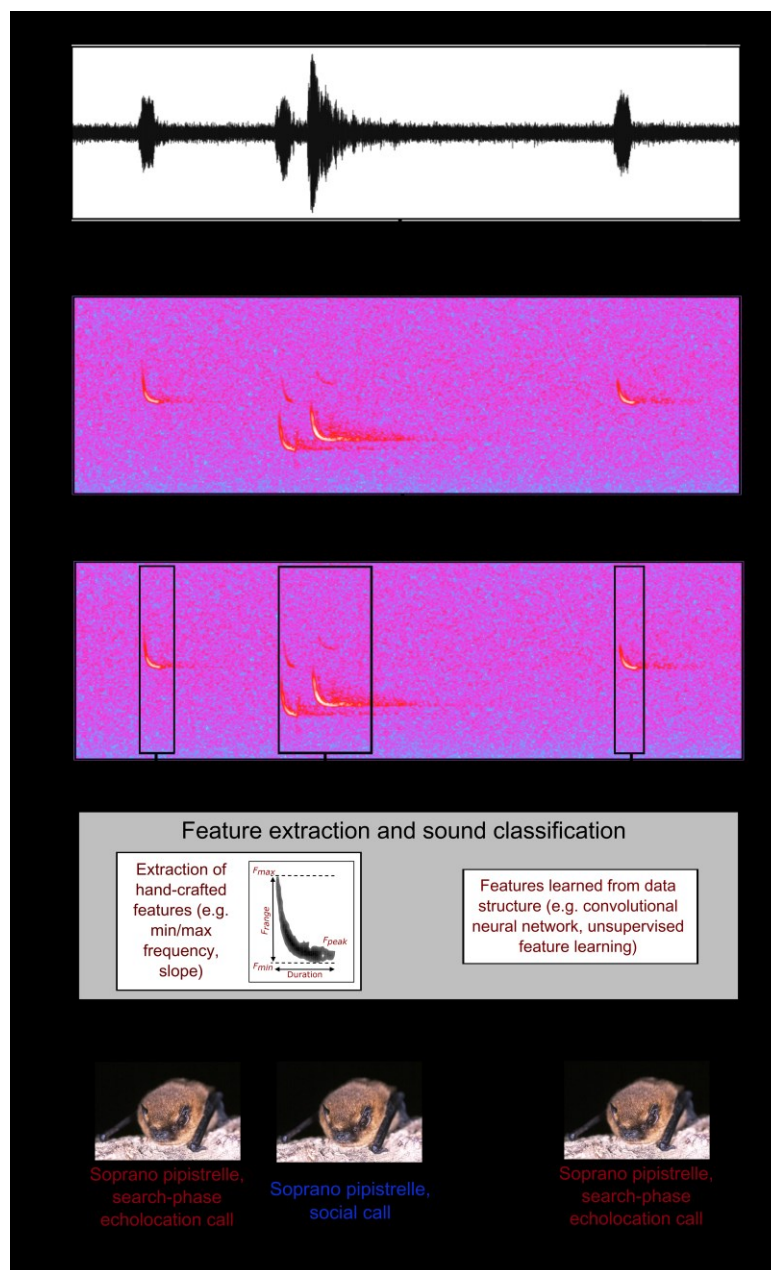
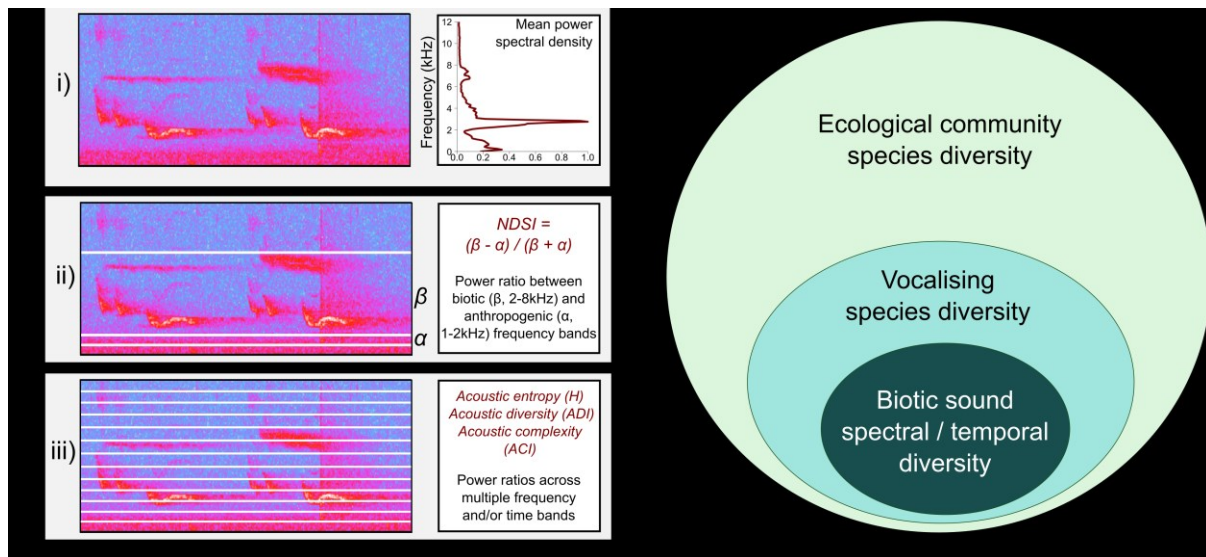
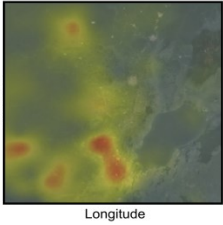
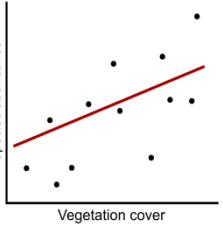
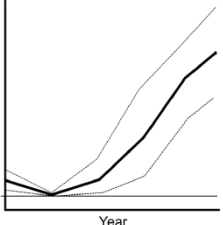
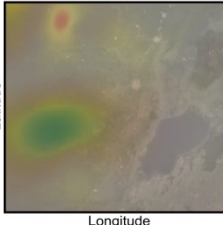
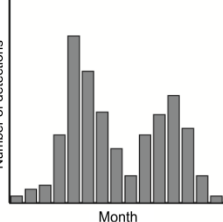


Figure 5: Indices of biotic and environmental sound. Conventional sound pressure metrics such as power spectral density (Ai) can measure the acoustic environment (e.g. anthropogenic sound). Ecoacoustic indices calculate pressure across spectrally- and temporally-partitioned sound recordings, from simple metrics using broad frequency bands (e.g. Normalised Difference Soundscape Index; Aii) to finer-band spectral and temporal diversity and entropy (Aiii). Their suitability for ecological community assessment is limited by poor understanding of the relationships between the diversity of recorded biotic sound, the diversity of vocalizing species, and broader community diversity (B).



1274 **Table 1: Ecological applications of PAM data.**

Analysis	Data type	Example result	Example applications	Key challenges
Occupancy	Presence/absence, single species		Estimating spatial trends in species occupancy (e.g. endangered or data-deficient species) and relationship to environment covariates (Kalan <i>et al.</i> 2015; Campos-Cerqueira & Aide 2016)	Minimizing error rates in species call ID.
Abundance	Spatially and temporally-explicit detection counts, single species		Estimating density and abundance of monitored species, and relationship to environment covariates (Marques <i>et al.</i> 2013; Lucas <i>et al.</i> 2015)	Minimizing error rates in species call ID. Individuals cannot be identified, so abundance estimates must account for non-independence of detected calls.
Indices of abundance	Detection counts, single species (per replicate survey)		Monitoring endangered or indicator species (Jones <i>et al.</i> 2013). Estimating abundance trends from multi-year monitoring data (e.g. generalised additive models, Barlow <i>et al.</i> 2015).	Minimizing 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 <i>et al.</i> 2017).	Minimizing 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 (Sanders & Mennill 2014)	Minimizing error rates in species call ID.

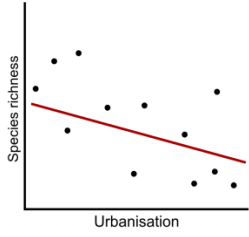
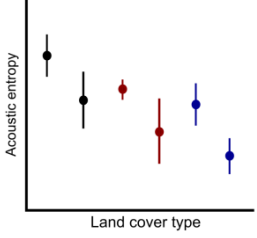
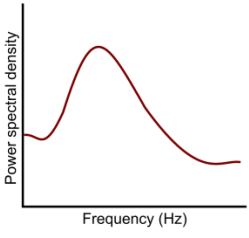
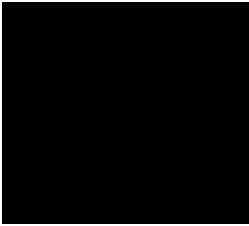
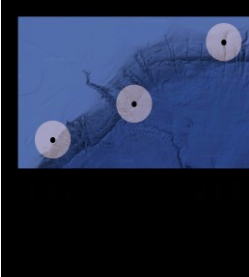
Species richness	Presence/absence, multiple species		Species inventories (MacSwiney G 2008). Relationships between species richness and habitat covariates.	Automated call ID tools and reference call libraries are currently unavailable for most taxa and regions.
Acoustic community diversity	Acoustic indices (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 (Sueur <i>et al.</i> 2014; Nedelec <i>et al.</i> 2015).	Relationships between index values and community diversity poorly understood. Indices are strongly sensitive to variation in non-biotic 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 <i>et al.</i> 2015; Marley <i>et al.</i> 2017).	More complex metrics (i.e. acoustic indices) may be sensitive to variation in different sound types (e.g. weather).
Individual animal 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, cetaceans, bats (King <i>et al.</i> 2013; Petrusková <i>et al.</i> 2015; Prat <i>et al.</i> 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 <i>et al.</i> 2017). Real-time monitoring of species occurrence or anthropogenic activity (Astaras <i>et al.</i> 2017).	High data storage costs. For real-time monitoring, patchy availability of automated analysis tools and data transmission capacity.

Table 2: Signal detection and classification techniques commonly used in bioacoustic analysis. The table summarises widely-used current methods for the detection and classification of wildlife and environmental sounds, including key advantages and disadvantages and example references.

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 non-target background noise and signal overlap.	(Digby <i>et al.</i> 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 <i>et al.</i> 2013)
Hidden Markov models	Detection, classification	Probabilistically infers whether a signal of interest is present, based on an underlying multi-state model.	Incorporates temporal detail on signal/sequence.	Complex for non-experts to develop. Requires sufficient training/reference data.	(Zilli <i>et al.</i> 2014; Tolkova <i>et al.</i> 2017)
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.	(Walters <i>et al.</i> 2012; Bittle & Duncan 2013; Garland <i>et al.</i> 2015b)
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).	(Skowronski & Harris 2006; Pirotta <i>et al.</i> 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.	Requires very large training datasets to account for within-class variability.	(Goeau <i>et al.</i> 2016; Mac Aodha <i>et al.</i> 2017)

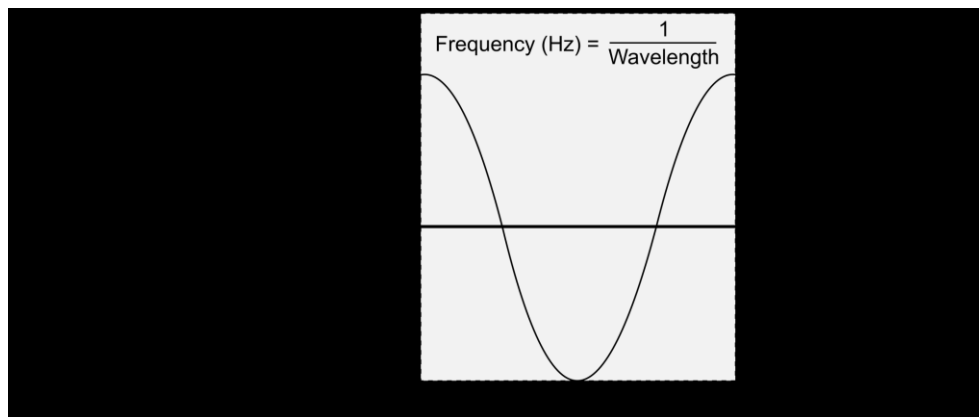
Box 1: The principles of sound recording

Sound is generated when the vibrations of a sound-producing object (such as an animal's vocalizing organs) alternatively compress and rarefy a medium, producing waves of pressure that propagate outward from the source (Bradbury & Vehrencamp 1998). A sound wave's *amplitude* is proportional to its energy content, is perceived as volume and is conventionally measured in decibels (dB). The *frequency* is the number of wave cycles per unit time (measured in Hz or kHz), is perceived as pitch, and is the inverse of *wavelength* (the length of one complete wave cycle) (see Box 1 inset figure). As sound travels it attenuates: its amplitude decreases as energy is lost to the environment, until at a certain distance the sound is no longer detectable. Attenuation, and therefore detection distance, is affected by properties of both wave and medium. Lower frequencies attenuate more slowly than higher frequencies, and sound waves travel over four times faster in water than air due to the medium's greater density (Bradbury & Vehrencamp 1998). In a PAM context, the likelihood of detecting a vocalizing animal from a given distance is therefore affected by its calling amplitude and frequency, whether it is aquatic or terrestrial, and environmental factors affecting sound transmission such as temperature, pressure and habitat structure (Darras *et al.* 2016).

In both animal auditory systems and electronic sound recording, incoming sound waves are transduced into an electrical signal whose amplitude is proportional to the amplitude of the sound wave. In animal auditory systems this is the function of tympanic and/or basilar (cochlear) membranes (Bradbury & Vehrencamp 1998). In microphones, current is induced by the displacement of a diaphragm, while hydrophones (underwater microphones) use piezoelectric transducers that produce a current when mechanical force is applied. Different transducers, whether electronic or biological, are optimally sensitive to different frequency

1306 ranges. The human cochlea optimally detects frequencies between 20 and 20,000Hz (*audible*
1307 *range*); frequencies above this range are termed *ultrasonic* (e.g. bat echolocation), and below
1308 it *infrasonic* (e.g. elephant rumbles). During digital recording, the signal's amplitude is
1309 sampled at a specified bit-depth (number of bits of information per sample) and sampling rate
1310 (number of samples per second; kHz), enabling its digital reconstruction in the time-
1311 amplitude domain. Bit-depth affects a recording's amplitude resolution and therefore
1312 dynamic range, while the sampling rate affects frequency resolution. Frequency information
1313 is recovered through signal processing, usually via Fourier transform, which for any given
1314 time period calculates the power spectrum of a signal's frequency components (although
1315 other methods, such as wavelet transforms, are also sometimes used in bioacoustic signal
1316 processing). Calculated in a sliding window this generates a spectrogram, a time-frequency
1317 representation enabling sounds to be visually identified. Importantly for wildlife monitoring,
1318 to fully resolve frequency information the sampling rate must be double the highest
1319 frequency of interest (the Nyquist frequency). As a result, sampling rate is conventionally
1320 44.1kHz for audible range, but much higher (up to 400kHz) for full-spectrum ultrasound (e.g.
1321 bat echolocation), requiring much more processing power and storage memory to record.
1322 Some ultrasound detectors therefore use less data-intensive recording methods based on
1323 frequency division, which divide the incoming signal frequency by a specified factor, but in
1324 the process often discard much of the signal's temporal and amplitude information.

1329 *Box 1 inset figure:* An example of a sinusoidal sound wave, showing amplitude, frequency
1330 and wavelength (figure modified with permission from (Browning *et al.* 2017)).



1331

1332

1333

1334

1335

1336

1337

1338

1339

1340

1341

1342

1343

Box 2: Surveying species and soundscapes in the marine realm

Marine passive acoustic monitoring, especially in cetacean studies, continues to be a source of key innovations including the development of automated call identification software (Gillespie *et al.* 2008; Baumgartner & Mussoline 2011) and on-animal acoustic sensor tags (Johnson & Tyack 2003; Nowacek *et al.* 2016), deployment of large sensor networks (Van Parijs *et al.* 2015), real-time species monitoring and reporting (Baumgartner *et al.* 2013; <http://dcs.whoi.edu/>), and collation of large, multi-study datasets (Davis *et al.* 2017).

However, the acoustic properties of the underwater environment mean that marine PAM involves particular challenges regarding ecological inference from audio data. Sound velocity in seawater is over four times that of air, due to the medium's higher density, and varies with depth due to changes in temperature, salinity and pressure (for simplicity it is commonly assumed that sound velocity increases linearly with depth; see Zimmer 2011). Furthermore, variations in sound absorption and scattering by the seafloor and sea surface can significantly affect a sound's propagation distance (Farcas *et al.* 2016). As a result, marine habitats are noisy and acoustically complex. Inference of population information from marine PAM data must account for variations in sensor detection range due to sensor depth, the depth of vocalising animals, ambient sound levels, the surrounding bathymetry, and seasonal differences in temperature and weather (Farcas *et al.* 2016).

There is increasing PAM research into marine soundscapes and acoustic communities, for instance around reef community phenology (McWilliam *et al.* 2017), the use of acoustic environmental cues by fish (Simpson *et al.* 2008), snapping shrimp as ecological indicators (Rossi *et al.* 2016), and patterns of biotic and environmental sound across oceanic habitats (Erbe *et al.* 2015; Nedelec *et al.* 2015). With accelerating degradation of ocean ecosystems, there is also interest in developing ecoacoustic indices for monitoring marine biodiversity (following terrestrial work; see Sueur *et al.* 2014). However, such efforts are complicated by

the high acoustic interconnectedness of underwater habitats, with long-range sounds potentially swamping local-scale variations in biotic sound (McWilliam & Hawkins 2013). Some reef studies have found correlations between certain indices and biodiversity metrics (Harris *et al.* 2016), but in other habitats and indices such relationships were either not present or confounded by non-target sound (Harris *et al.* 2016; Staaterman *et al.* 2017). Indeed, ongoing increases in oceanic anthropogenic sound levels highlight that marine acoustic indicators must be particularly robust to background noise (McWilliam & Hawkins 2013; Harris *et al.* 2016). Even beyond the broader conceptual issues around ecoacoustic indices (see '*Acoustic ecological community assessment*'), it therefore remains unclear whether index-based approaches can be effective in the marine realm, highlighting the need for further research in this area. There would be clear benefits to increased collaboration among marine and terrestrial PAM researchers towards jointly addressing pressing questions and challenges, including the development of noise-robust analytical tools, how to effectively measure vocalising animal communities, and how to operationalise PAM data for conservation goals.

1391 **Box 3: Glossary of key terms**

1392 *Acoustic community* – an assemblage of vocalising animals.

1393 *Acoustic index* – a mathematical function that summarises spectral and/or temporal features
 1394 of a sound recording, often as putative proxies for diversity of biotic sound and/or vocalising
 1395 species (reviewed in Sueur *et al.* 2014).

1396 *Audible range* – refers to sounds within the frequency range of human hearing (20 –
 1397 20,000Hz).

1398 *Automated sound identification* – the process of automatically detecting (locating) sounds of
 1399 interest within a recording and classifying them to an ecologically relevant category (e.g.
 1400 species, call behaviour).

1401 *Bit depth* – in digital sound recording, the number of bits of information contained in each
 1402 sample from an incoming sound wave (commonly 16-bit in PAM sensors). Recording at
 1403 higher bit depths improves the amplitude resolution of a sound recording.

1404 *Ecoacoustics* – an emerging research area focused on deriving ecological community
 1405 information (e.g. estimates or proxies of species diversity) from environmental sound
 1406 recordings.

1407 *Features (in sound classification)* – parameters that describe a sound's spectral and temporal
 1408 characteristics, which are then used to assign sounds to a category via comparison to other
 1409 sounds in either the same dataset (unsupervised classification) or a known reference dataset
 1410 (supervised classification).

1411 *Frequency (of a sound wave)* – number of wave cycles per unit time, measured in hertz (Hz;
 1412 cycles per second).

1413 *Machine learning* – a family of data analysis techniques that employ algorithms to learn
 1414 patterns from, and then make predictions on, data. These can be unsupervised (based solely
 1415 on the patterns within the data, e.g. clustering algorithms) or supervised (previously trained
 1416 on labelled datasets of known objects, e.g. species calls). In PAM, such methods are typically
 1417 used to detect (locate) and then classify sounds of interest in audio recordings.

1418 *Passive acoustic monitoring (PAM)* – the recording and analysis of environmental sound
 1419 using acoustic and/or ultrasonic sensors, in order to derive ecological information about
 1420 species or vocalising animal communities.

1421 *Passive acoustic sensor (acoustic sensor)* – any combination of sound recorder, detector,
 1422 microphone and/or hydrophone used to detect and record sound in the environment.

1423 *Sampling rate* – in digital sound recording, the number of digital samples per second (Hz)
 1424 recorded from an incoming acoustic signal. Higher sampling rates improve a recording's
 1425 frequency resolution, at the cost of increased energy and storage requirements. For recording
 1426 ultrasound, very high sampling rates (often >200kHz) are required to adequately resolve
 1427 frequency information.

1428 *Spectral* – relating to the frequency of a sound.

1429 *Ultrasound (ultrasonic)* – refers to sounds above the frequency range of human hearing
 1430 (exceeding 20,000kHz), e.g. bat and cetacean echolocation calls.

1431