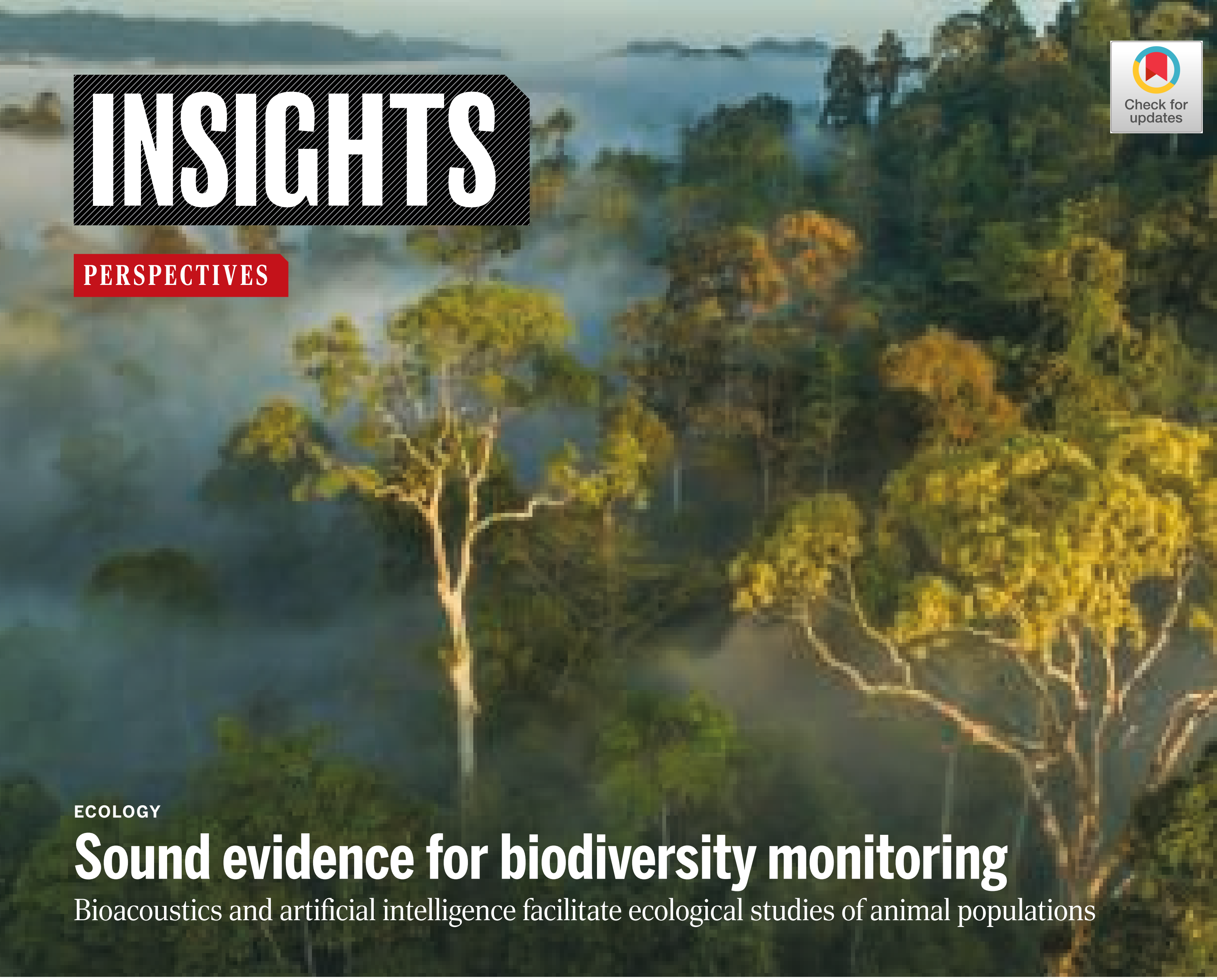
Sound evidence for biodiversity monitoring: bioacoustics and artificial intelligence facilitate ecological studies of animal populations

Jeppe H. Rasmussen, Dan Stowell, Elodie F. Briefer

DOI:
https://doi.org/10.1126/science.adh2716



INSIGHTS

PERSPECTIVES

ECOLOGY

Sound evidence for biodiversity monitoring

Bioacoustics and artificial intelligence facilitate ecological studies of animal populations

By **Jeppe H. Rasmussen[1,2], Dan Stowell[3,4], Elodie F. Briefer[1]**

With 28% of all assessed plant and animal species now at risk of extinction, new noninvasive and efficient tools to monitor populations are urgently needed. For animal populations, studying their vocalizations through automated monitoring and machine learning offers one such solution (*1*). Machine-learning techniques have the potential to handle huge amounts of data and uncover sound patterns, allowing for faster, cheaper, and better ecological studies based on acoustics. However, challenges remain in using this technology to monitor animal populations.

Sound is among the most efficient communication channels because of its saliency, flexibility, and potential for long-range transmission. As a result, acoustic communication is widespread in the animal kingdom (*2*). The study of production, dispersion, and reception of animal sounds—animal bioacoustics (henceforth bioacoustics)—began more than a century ago. This discipline consists of recording animal vocalizations, performing detailed analyses of sound characteristics, and running playback experiments to investigate discrimination and perception. Current bioacoustics techniques enable scientists to identify an ever-increasing range of information contained in animal vocalizations, some of which are relevant for population monitoring, such as the emitter's population, sex, individual identity, and health.

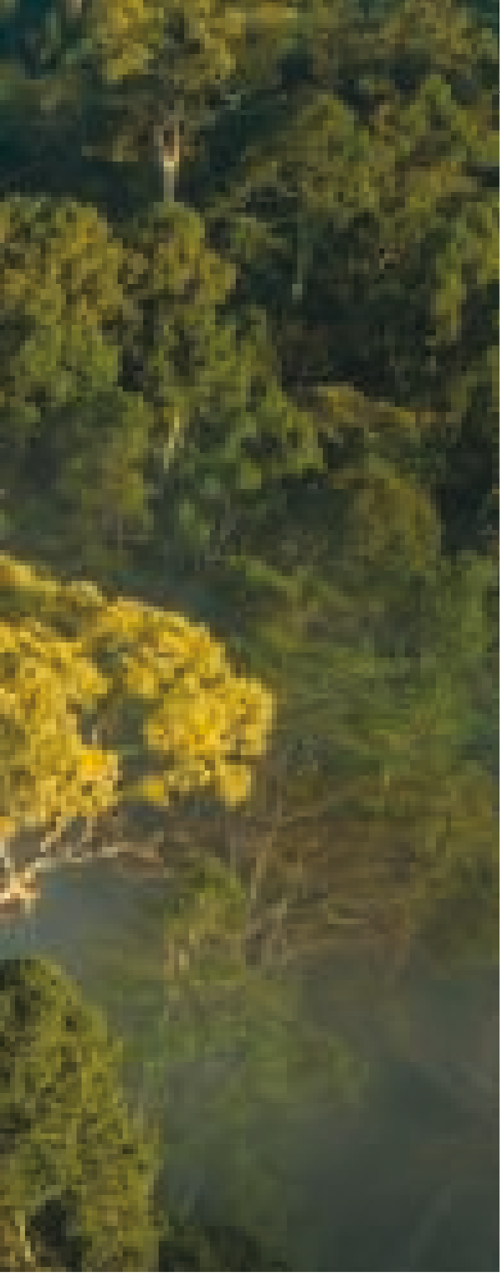
Population monitoring relies on an estimate of the presence or absence of species (occupancy) in particular locations, to evaluate changes in population size and species distribution over time. Despite some limitations, animal vocalizations are a good proxy for this because they can be collected noninvasively and provide an estimate of the occupancy of a species based on the presence or absence of animal vocalizations in a sound clip. Bioacoustics is especially useful in visually inaccessible habitats, such as the ocean, or in dense vegetation, where visual monitoring (such as with camera traps) is less efficient. Modern, self-contained acoustic recording units also have the capability to be left in place for months or even years before retrieval. This allows for continuous monitoring of animals and ecosystems at ecologically relevant spatiotemporal scales, which is a substantial advantage because deployment and retrieval can be extremely laborious and costly.

A common limitation of the use of bioacoustics for population monitoring is

PHOTO: CHIEN LEE/MINDEN PICTURES

that the parameter extraction and statistical analyses necessary to process field recordings require considerable manual processing. Machine learning offers an ideal solution to many of these issues, through automated processing and learning from data. Current artificial intelligence (AI) technology allows sufficiently reliable sound detection, although precise counting or tracking of individuals is harder. Thus, low-cost, unattended acoustic detection is a valuable addition to manual surveys for, among others, anurans (such as frogs), bats, birds, marine mammals, and many insects.

[1]Behavioural Ecology Group, Section for Ecology and Evolution, Department of Biology, University of Copenhagen, Copenhagen, Denmark. [2]Center for Coastal Research and Center for Artificial Intelligence Research, University of Agder, Kristiansand, Norway. [3]Tilburg University, Tilburg, Netherlands. [4]Naturalis Biodiversity Center, Leiden, Netherlands. Email: elodie.briefer@bio.ku.dk; d.stowell@tilburguniversity.edu



In this tropical forest in Sabah, Malysia, 34 out of the intended 39 bird and anuran species were detected by using AI and bioacoustics.

Traditional statistical machine learning has been applied to bioacoustics for many years. However, it was only around 2016 that the application of deep learning to big datasets recorded from many species in the wild enabled larger-scale inferences and insights to be generated (*3*). There are now large projects that record animal sounds (for example, the Australian Acoustic Observatory, SAFE Acoustics Borneo, and Sound of Norway) and a common "recipe" for acoustic deep learning and convolutional neural networks (CNNs) (*4*), plus off-the-shelf classifiers that work well for species classification in birds and bats (such as BirdNET and BTO Acoustic Pipeline).

Adapting an AI algorithm to perform a specific task such as recognizing the species, sex, or health of the emitter of a certain vocalization involves a few essential steps: The network architecture of the AI must be chosen, it must be trained on relevant data, and classification performance must be tested on unseen data. All of this is now possible to accomplish without writing a single line of code. AI software based on graphical user interface is now available, simplifying the use of AI and allowing researchers to use it properly with only limited training.

Most studies that use AI for population monitoring transform sound into "spectrograms" before classification; these two-dimensional time-frequency representations are "image-like" and well handled by CNNs. Three main recognition tasks, characterized by increasing levels of complexity but also versatility, exist for these images: whole-image classification, object detection, and segmentation (*5*). Whole-image classification labels the entire image using a standard CNN. This label can be whatever is being classified, such as species, sex, health, or individual identity. Whole-image classification is easy to work with and is the most used in bioacoustics and AI studies so far. The next step, object detection, uses networks such as region-based CNNs or You Only Look Once to identify potential regions of interest before applying the standard CNN architecture for classification. This yields a label for each region of interest, which is advantageous if more than one type of relevant sound is present per spectrogram. The third approach is segmentation, using networks such as fully convolutional networks or U-Nets. This technique classifies every pixel in the image, allowing for more precise analysis of the fine-scale structure of animal calls, but also entails more work in creating the training dataset as compared with the two other recognition tasks. Whereas whole-image classification needs a label for each image and object detection needs a label for each region of interest, segmentation needs a label for each pixel in the image.

Applying AI directly on raw audio files demands specific variants of neural networks. One such variant, recurrent neural networks (RNNs), contains a feedback loop that allows it to reason on the basis of previous data to classify upcoming data, hence lending itself for classifying time series. Yet despite the established performance of various RNNs, such as bidirectional RNN or long short-term memory (*6*), many researchers are now transitioning to Transformer networks (*7*) known from ChatGPT. These networks work solely on attention mechanisms and do not use any convolutions or pooling operations to extract features from the input.

Bioacoustics combined with AI has already been implemented in many ecological studies. For example, off-the-shelf classifiers, such as BirdNET (*8*), are readily available, making them easier to implement in new studies that focus on bird sounds. Establishing which bird species are present in a hard-to-access ecosystem can be slow and laborious work; bioacoustics and AI have already been applied to this kind of task. For example, 34 out of the intended 39 species of birds and anurans were detected in a tropical forest in Sabah, Malaysia, by using CNN (*9*). Urban areas also pose a challenge for traditional bioacoustics approaches because of a high degree of noise masking animal vocalizations. In this situation, AI tools can isolate and classify the biotic sounds present in the soundscape (*10*).

In marine ecology, bioacoustics is an invaluable source of information (*11*). The oceans are an environment where directly observing animals is difficult, but species-specific sound works as a good proxy for both species or individual presence and even behavior because specific behaviors are often associated with identifiable types of sound. For example, migration routes, behavior, and relative population density of whales have long been studied by deploying self-contained, noninvasive recording devices in the oceans, called passive acoustic monitoring (PAM) (*12*). Traditional algorithms, such as matched filtering and spectrogram correlation, have been effective in detecting highly stereotyped songs emitted by male baleen whales. However, because of the greater inter- and intra-emitter variability in social calls produced by both sexes, automatic detection by using traditional algorithms has proven challenging. Recent studies that used deep-learning techniques have demonstrated promising results for automatic detection of social calls (*13*). Another advantage of AI versus manual analysis is the potential for near-instantaneous results. This allows, for example, real-time warnings of a large baleen whale presence in shipping channels, which could help avoid collisions.

Using sound for biodiversity monitoring presents some limitations, such as the differential use of sound production by specific individuals in many species (such as juveniles versus adults, or females versus males), or restriction of sound production to a specific season (such as breeding season). Many of these limitations are, however, also encountered when using visual surveys because both behavior and appearance can change between seasons and according to an individual's age and sex, resulting in biases. These biases can differ between methods, depending on the focus or situation, and should be considered when selecting the approach. For example, vocalization monitoring may be more accurate for detecting males than females, whereas visual surveys may be more accurate for detecting adults than juveniles.

There are a few technical bottlenecks for widespread application of bioacoustics AI to population monitoring. The lack of expert labeled data, particularly for rare acoustic phenomena, and for less-documented areas such as marine sound or the Global South limit its use. How to overcome this obstacle? Crowdsourcing is one answer. There is a variety of resources that provide training data for AI. Among those are publicly available sources such as Kaggle and OpenML, which cover a broad range of species and ecosystems. Another good example is DCLDE 2015, a large dataset that consists of expert-labeled underwater soundscapes recorded off the coast of California, which is optimal for creating AI detectors for the whale species present in those waters. Unfortunately, the task of using AI for biodiversity monitoring has so far been geographically segregated into two distinct regions: Sound recording and hands-on conservation work are often conducted in the Global South, whereas AI coding is primarily performed at institutions located in the Global North. Cloud computing, allowing for access to powerful hardware from standard laptops, combined with the current trend of making AI more accessible through graphical user interfaces has the potential to unite the two regions. This would allow biodiversity-monitoring projects that use AI to be completed on location, ensuring a better inclusion of local communities and stakeholders, hence boosting the likelihood of achieving conservation policy and management success as a result.

Recent advances in AI techniques may also help to mitigate the lack of labeled data; data augmentation, a procedure of adding augmented versions of existing data to the training set, increases the effective size of the data while simultaneously enforcing certain generalization properties of the network. The current shift toward extensive use of pretrained networks also helps; this technique, called transfer learning, enables pretrained networks to solve new tasks using only a minimum amount of new training data. The availability and performance of these pretrained networks are increasing.

The combination of animal bioacoustics and AI also has broader applications than assessment of biodiversity, such as monitoring animal welfare on the basis of emotional information contained in animal calls (*14*) or aiding neuroethology by discovering patterns and groupings in vocalizations (*15*). However, vocalizations on their own cannot reveal much without detailed observations of the behavior of the animal vocalizing and the context of production. Thus, research should be informed by ethological observations when possible, and most importantly, the labeled datasets should be shared with the scientific community. These newly developed tools will enable efficient and powerful monitoring of changes in the environment, population size, and species distribution, which is crucial for conservation. ■

REFERENCES AND NOTES

1. C. Rutz *et al.*, *Science* **381**, 152 (2023).
2. Z. Chen, J. J. Wiens, *Nat. Commun.* **11**, 369 (2020).
3. H. Goëau *et al.*, *Working notes of CLEF: Conference and Labs of the Evaluation Forum*, Évora, Portugal, 5–8 September 2016, pp. 440–449.
4. D. Stowell, *PeerJ* **10**, e13152 (2022).
5. M. Goodwin *et al.*, *ICES J. Mar. Sci.* **79**, 319 (2022).
6. S. Madhusudhana *et al.*, *J. R. Soc. Interface* **18**, 20210297 (2021).
7. A. Vaswani *et al.*, *Adv. Neural Inf. Process. Syst.* **30** (2017).
8. S. Kahl, C. M. Wood, M. Eibl, H. Klinck, *Ecol. Inform.* **61**, 101236 (2021).
9. S. S. Sethi *et al.*, *Oikos* 2022 **3**, e08525 (2022).
10. A. J. Fairbrass *et al.*, *Methods Ecol. Evol.* **10**, 186 (2019).
11. C. M. Duarte *et al.*, *Science* **371**, eaba4658 (2021).
12. S. M. Wiggins, in *Frontiers of Engineering: Reports on Leading-Edge Engineering from the 2009 Symposium* (National Academies Press, 2010), pp. 13–22.
13. J. H. Rasmussen, A. Širović, *J. Acoust. Soc. Am.* **149**, 3635 (2021).
14. E. F. Briefer *et al.*, *Sci. Rep.* **12**, 3409 (2022).
15. T. Sainburg, T. Q. Gentner, *Front. Behav. Neurosci.* **15**, 811737 (2021).

ACKNOWLEDGMENTS

The authors are supported by the Carlsberg Foundation (CF22-1389 to J.H.R. and CF20-0538 to E.F.B.). Thanks to anonymous reviewers, M. Goodwin, G. Chrupała, J. Sills, R. Oomen, E. van Massenhoeve, and R. Mandel for useful suggestions.

10.1126/science.adh2716

CANCER

Unmasking immune suppression

Inhibition of a mutated metabolic enzyme puts the sting back in antitumor immunity

By **Jason R. Pitarresi**[1], *and* **Katherine A. Fitzgerald**[2]

Isocitrate dehydrogenase 1 (*IDH1*) is a commonly mutated metabolic gene in cancer. Mutant IDH1 produces the oncometabolite (*R*)-2-hydroxyglutarate (R-2HG), which inhibits histone demethylases, metabolic enzymes, and DNA repair pathways, culminating in an altered epigenetic state in tumor cells (*1*–*6*). Mutant IDH1 tumor cells also secrete R-2HG, and this cell-extrinsic activity leads to the exclusion of T cells from the tumor microenvironment (TME) and impairs $CD8^+$ T cell–mediated tumor cell killing (*5*, *6*). On page 159 of this issue, Wu *et al.* (*7*) report additional tumor cell–intrinsic activities of R-2HG. They reveal that R-2HG leads to epigenetic silencing of the gene encoding the intracellular innate immune sensor, cyclic GMP-AMP synthase (cGAS), as well as endogenous retrovirus genes (ERVs). Inhibition of mutant IDH1 restored cGAS activation by a subset of ERVs, which induced innate immune responses and antitumor immunity. These findings provide insight into the tumor cell–intrinsic epigenetic mechanisms that drive immune evasion in mutant IDH1 tumors.

IDH1 gain-of-function mutations occur in a variety of cancer types, most notably in ~27% of patients with intrahepatic cholangiocarcinoma (ICC) and >80% of patients with low-grade glioma (LGG). Production of the R-2HG oncometabolite competitively inhibits enzymes that are dependent on the important cofactor α-ketoglutarate, such as histone demethylases and DNA demethylating enzymes [ten-eleven translocation (TET) methylcytosine dioxygenases] that alter the epigenetic landscape and promote

[1]Division of Hematology-Oncology, Department of Medicine, University of Massachusetts Chan Medical School, Worcester, MA, USA. [2]Division of Innate Immunity, Department of Medicine, University of Massachusetts Chan Medical School, Worcester, MA, USA. Email: kate.fitzgerald@umassmed.edu