Trends in
Ecology & Evolution

Review

Individual identification in acoustic recordings

Elly Knight [1,*], Tessa Rhinehart [2,*], Devin R. de Zwaan [3,4], Matthew J. Weldy [5], Mark Cartwright [6], Scott H. Hawley [7], Jeffery L. Larkin [8,9], Damon Lesmeister [10], Erin Bayne [1], and Justin Kitzes [2]

Recent advances in bioacoustics combined with acoustic individual identification (AIID) could open frontiers for ecological and evolutionary research because traditional methods of identifying individuals are invasive, expensive, labor-intensive, and potentially biased. Despite overwhelming evidence that most taxa have individual acoustic signatures, the application of AIID remains challenging and uncommon. Furthermore, the methods most commonly used for AIID are not compatible with many potential AIID applications. Deep learning in adjacent disciplines suggests opportunities to advance AIID, but such progress is limited by training data. We suggest that broadscale implementation of AIID is achievable, but researchers should prioritize methods that maximize the potential applications of AIID, and develop case studies with easy taxa at smaller spatiotemporal scales before progressing to more difficult scenarios.

Highlights

Individual acoustic signatures, if present within animal species, can provide insights into evolution and behavior, and have great potential for differentiating individuals during monitoring and research.

Recent advances in bioacoustic technology combined with acoustic individual identification (AIID) have the potential to revolutionize the study of sound-producing animals; however, methods and application of AIID remain in their infancy.

Evidence of individual acoustic signatures across taxa and successes in adjacent acoustic disciplines suggest that opportunities exist for developing AIID.

Research and development into AIID should combine deep learning methods with the construction and sharing of labeled training datasets, and should focus on recording and classification methods that maximize the potential applications of AIID.

Broadscale implementation of AIID should be achievable in the near future and will allow biologists to answer important ecological and evolutionary questions with less bias and fewer negative population effects and resources than the current approaches.

The potential of AIID

Recent advances in **bioacoustics** (see Glossary) have revolutionized our ability to tackle fundamental ecological and evolutionary questions across large spatial and temporal scales [1]. Today, the deployment of **autonomous recording units (ARUs)** for **passive acoustic monitoring (PAM)** provides a scalable and increasingly popular method for monitoring animal populations at a fraction of the time and effort associated with traditional field observations [2,3]. Permanent archives of acoustic recordings from the field are being collected across the globe at staggering rates [4] and provide the opportunity to understand ecological phenomena at a global scale.

Most current bioacoustic applications focus on the identification of species within recordings. An open and critically important frontier in bioacoustics is **AIID**, which is the ability to differentiate individuals of the same species on the basis of recordings of the sounds they produce. Applications include the study of vocal individuality for its own sake, but also the assignment of identity, which is crucial to many monitoring techniques and research objectives such as estimating demographic rates. Many current applications in ecology and evolution that require individual identification rely on mark–recapture techniques such as bird banding, passive telemetry, or passive integrated transponder (PIT) tags as the gold standard [5,6]. Collecting such data can be labor-intensive [7], have adverse effects on individuals and populations [8], and can introduce bias via observer effects or behavioral responses to capture and tracking [9].

By contrast, AIID can potentially track individuals across space and time to estimate population size (e.g., [10]) or site fidelity (e.g., [11]), and has been shown to improve estimated survival rates relative to traditional mark–recapture methods (e.g., [12]). While initially suggested as a tool for ecological and evolutionary research and monitoring as an alternative to mark recapture techniques [13], AIID has since been implemented in complex applications, including censuses (e.g., [14]), studying dispersal (e.g., [15]), and estimating apparent survival (e.g., [12]). Indeed, the potential applications of AIID are numerous (Table 1). Despite these advances, some potential applications of AIID have yet to be achieved (Table 1), and the majority of existing applications

[1]Department of Biological Sciences, Alberta Biodiversity Monitoring Institute, University of Alberta, Edmonton, Alberta, T6G 2E6, Canada
[2]Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA, 15260, USA
[3]Department of Biology, Mount Allison University, Sackville, NB, E4L 1E4, Canada
[4]Acadia University, Wolfville, NS, B4P 2R6, Canada
[5]Department of Forest Ecosystems and Society, Oregon State University, Corvallis, OR, 97331-5704, USA
[6]Department of Informatics, New Jersey Institute of Technology, Newark, NJ, 07102, USA

have occurred within a select few systems (e.g., cetaceans), suggesting that there are many remaining challenges in AIID.

Given the potential of AIID, we developed a multidimensional framework to describe the relative difficulty of AIID (Figure 1). We used this framework to systematically review the literature (Box 1) and identify existing challenges to AIID. We suggest potential solutions to these challenges based on other acoustic signal processing disciplines and propose a pragmatic path forward for the development and application of AIID.

Strong evidence for acoustic signatures

AIID itself is not a new field of research. At small scales and in controlled settings, the field of bioacoustics has long been interested in whether individual **acoustic signatures** exist, with qualitative studies dating back to the late 1960s (e.g., [16,17]). Acoustic signatures may occur due to genetic differences [18], including structural differences in vocal tracts [19], as well as environmental variation [20] or cultural influence [21]. Nearly all published studies of acoustic signatures (96% of 598) have found evidence that individuals can be distinguished (Box 1), suggesting that the biological and quantitative capacity for AIID exists across taxa. Our review is consistent with prior work on mammals [22], which suggests that most sound-producing taxa have acoustic signatures that should allow AIID. However, because many studies deem AIID to be successful if its classification accuracy is higher than random chance, its classification accuracy can be low (e.g., <50%) and may be insufficient for the useful application of AIID and can potentially introduce bias into analyses [23]. The question therefore remains whether the application of AIID is possible for all species, or whether there are certain taxa that have insufficient interindividual variation in their acoustic signatures, excessive intraindividual variation in their acoustic signatures, and/or insurmountable species attributes (Box 1).

A multidimensional framework for AIID

We propose a conceptual framework to describe the difficulty of AIID based on three methodological components: the study's design, the spatiotemporal extent, and attributes of the taxa. The framework determines the difficulty of each potential ecological or evolutionary application of AIID (Table 1) based on the study's design and the spatiotemporal extent (Figure 1); that difficulty is then mediated by the attributes of the focal species (Box 1). The study design component has two categorical axes: the recording methods and classification methods required for the application. While many applications are feasible, with multiple combinations of recording and classification methods, there are often limitations to using a combination of methods that is easier to use (Table 1). The spatiotemporal component consists of two continuous axes that describe the spatial and temporal extent of a given application of AIID.

Improving the signal-to-noise ratio in passive recordings

Within the study design component, recording methods can be categorized into **targeted recording** and **passive recording** (Figure 1). Targeted approaches use handheld and often directional acoustic recorders (e.g., shotgun microphones), which are used to seek out and collect acoustic recordings. Targeted approaches are labor-intensive in the field and thus are often used for studies of acoustic behavior that occur over small temporal extents, including repertoire size and song structure (Table 1). Conversely, passive approaches use ARUs that are deployed in the environment to make recordings, often following a preset schedule. Relative to targeted recording, PAM can collect orders of magnitude more hours of recordings per unit of field effort and can also collect recordings of the entire soundscape, making it more suitable for applications at larger spatial and temporal scales like tracking migration and seasonal movements that require intensive sampling effort (Table 1).

[7]Chemistry and Physics Department, Belmont University, Nashville, TN, 37212, USA
[8]Department of Biology, Indiana University of Pennsylvania, Indiana, PA, 15705-1081, USA
[9]American Bird Conservancy, The Plains, VA, 20198, USA
[10]USDA Forest Service, Pacific Northwest Research Station, Corvallis Forestry Science Laboratory, Oregon State University, Corvallis, OR, 97330, USA

*Correspondence:
ecknight@ualberta.ca (E. Knight) and tessa.rhinehart@pitt.edu (T. Rhinehart).

A comparison of directional and omnidirectional recording methods suggests that AIID with passive recordings is an achievable goal [24], but there are three remaining obstacles related to the lower **signal-to-noise ratios** in passive recordings that make it more challenging than AIID with targeted recordings. First, a low signal-to-noise ratio can occur when the individual is far from the microphone, which occurs frequently during passive surveys due to the geometry of a circular survey area. Second and relatedly, sounds from other animals; geophonic sounds such as wind, rain, and stream noise; and anthropogenic sound can obscure the sounds of the target species. Third, overlapping sounds from different individuals of the same species also pose a challenge (i.e., the 'cocktail party' problem), especially when overlapping sounds are not considered to be noise but additional individuals to detect and identify [25–27]. Colonial species may be particularly challenging in this regard, although behavioral research has suggested that acoustic signatures exist in colonial environments because individuals use them to readily identify one another (e.g., penguins in [28]). By contrast, AIID with targeted recordings is easier than with passive recordings because targeted recordings are typically recorded at close range to a single focal individual, and as such, generally contain minimal sound masking and have a high signal-to-noise ratio [29]. Collectively, these features make it easier not only to apply a classification to targeted recordings rather than passive recordings, but also to train the classification models because clean individual labels are more readily obtained from targeted recordings (but see the section on training data later).

Opportunities to improve the signal-to-noise ratio in passive recordings include modifying the sampling designs, improving the hardware, and applying noise reduction and sound separation techniques. First, modifications to the study's design could improve the likelihood of recording the target sounds with a higher signal-to-noise ratio, such as scheduling passive recordings at times of day that maximize the signal-to-noise ratio (e.g., avoiding the dawn chorus) or thresholding recordings by sound level to only include individuals recorded at close range [30]. Second, continued improvements in ARUs that reduce recorder self-noise will facilitate AIID by providing clearer signals of acoustic signatures, particularly for quiet or distant sounds. Third, sound source separation and noise reduction techniques hold great promise for improving the signal-to-noise ratio of recordings. Approaches coupling multichannel ARUs with signal processing techniques (e.g., beamforming) have an improved signal-to-noise ratio and reduced masking (e.g., [31,32]). Applying preprocessing sound separation techniques to single-microphone recordings has also improved the performance of species classification [26]. Alternatively, a low signal-to-noise ratio due to attenuation in varying environmental conditions could be incorporated directly into the classification by modifying the training recordings (e.g., convolution from measured impulse responses [33]).

Adjacent disciplines provide techniques to improve classification

In addition to the recording method, the study design component includes the classification methods, which can be categorized into **closed-set classification** and **open-set classification** approaches (Figure 1). Closed-set classification involves using training and target datasets containing the same individuals, with a classification algorithm trained to directly classify sounds of individuals into known discrete classes, similar to traditional supervised learning problems such as species classification [34]. However, open-set classification involves a target dataset with some or all new individuals that the algorithm has not been trained with. Open-set classification is more akin to a mapping problem, where the classifier learns a high-dimensional feature space (the '**embedding**' space in deep learning) to be able to cluster similar acoustic cues from both known and new individuals. In the feature space, the distances between points reflect acoustic differences between sounds, and distances between clusters reflect differences between individuals [34,35]. Ideally, multiple vocalizations from a single individual will cluster

Glossary

Acoustic individual identification (AIID): the use of acoustic recordings to distinguish and identify individual animals.

Acoustic signature: unique and persistent individual-level characteristics of sounds that can be used to distinguish among individuals within a species.

Autonomous recording unit (ARU): a recording device deployed in the field designed to record audio autonomously, usually on a user-defined schedule.

Bioacoustics: the study of animal sounds, including the use of animal sounds to study other aspects of biology, including ecology and evolution.

Closed-set classification: AIID with a model that has been trained with the sounds of the same individuals it will be used to identify.

Embedding: a multidimensional numeric representation of a recorded sound that can be used to cluster sounds into similarity-based groups, such as individual animals.

Open-set classification: AIID of individuals that are new or unknown to the model.

Passive acoustic monitoring (PAM): the use of ARUs to survey and monitor the acoustic environment, often animals.

Passive recording: recordings made by an unattended recorder that indiscriminately record all sounds detectable by the recorder.

Signal-to-noise ratio: the ratio of the power of an acoustic signal of interest to the power of background sounds.

Targeted recording: recordings made by seeking out individual animals, often using a directional microphone.

Transfer learning: a machine learning technique where an existing model is used as a starting point for training a new model.

Table 1. Potential applications of AIID in ecology and evolution, the traditional approach used, and the frequency of that application in AIID research. Most research (89.1%) investigated AIID per se (i.e., without a particular application), and thus the frequencies are low

Potential application	Definition	Importance	Traditional approach	AIID approach	Frequency (%)
Density estimation (e.g., [11])	Number of individuals per unit of area	Status assessment, population management, and assessment of habitat quality	Point counts adjusted for imperfect detection Deriving abundance from the ARU's call rate Spot-mapping	Recording: targeted recording is likely to be impractical for large study areas Classification: counts obtained with open- or closed-set from recordings adjusted for imperfect detection	2.0
Census (e.g., [14])	Count of all individuals in a population	Status assessment and population management, particularly for endangered species	Direct observation and identification of all individuals via distinct markings Snapshot surveys	Recording: PAM arrays surveying continuous areas, targeted recording is likely to be inaccurate due to observer effects on species' detectability Classification: open-set only	2.0
Seasonal movement (e.g., [95])	Where and when an animal moves within its home range within a season	Understanding habitat requirements	Mark–recapture of tagged individuals Miniaturized satellite tracking technology Spot-mapping	Recording: targeted recording may be impractical due to the low probability of recapture across time for mobile species Classification: closed-set methods produce potentially inaccurate estimates if immigration rates are high	0.6
Demographic parameters (e.g., [12])	Population recruitment, immigration, emigration, survival	Quantifying population dynamics and influential factors	Mark–recapture of tagged individuals Miniaturized satellite tracking technology	Recording: targeted recording may be impractical due to the low probability of recapture across time for mobile species Classification: closed-set methods produce potentially inaccurate estimates if immigration rates are high	0.6
Migration tracking	Cyclical individual movements between locations across the annual cycle	Evaluating the factors affecting migratory populations, evolution of migration and speciation	Mark–recapture of tagged individualsMiniaturized satellite tracking technology	Recording: targeted recording is likely to be impossible due to the low probability of recapture Classification: closed-set methods are likely to be inaccurate due to repeat acoustic signatures	0.0
Interannual dispersal (e.g., [15])	Individual movements between years to different areas	Understanding population dynamics, including source–sink dynamics, invasion and colonization, gene flow, and species interactions	Mark–recapture of tagged individualsMiniaturized satellite tracking technology	Recording: targeted recording is likely to be impossible due to the low probability of recaptureClassification: closed-set methods are likely to be inaccurate due to repeat acoustic signatures	0.0
Communication behavior (e.g., [96])	Transfer of information between individual animals	Studying animal cognition, evolution, and sociology	Focal observations of marked or unmarked individuals	Recording: targeted or passive Classification: closed- or open-set to identify individual signalers	2.8
Individual state (e.g., [97])	A subset of communication behavior that involves the classification of the state of the signaler via acoustic characteristics	Inferring breeding status and understand population dynamics	Focal observations of marked or unmarked individuals Searching and monitoring nests	Recording: likely to be more effective with passive to detect changes in state over time Classification: closed- or open-set to identify individual signalers	2.2

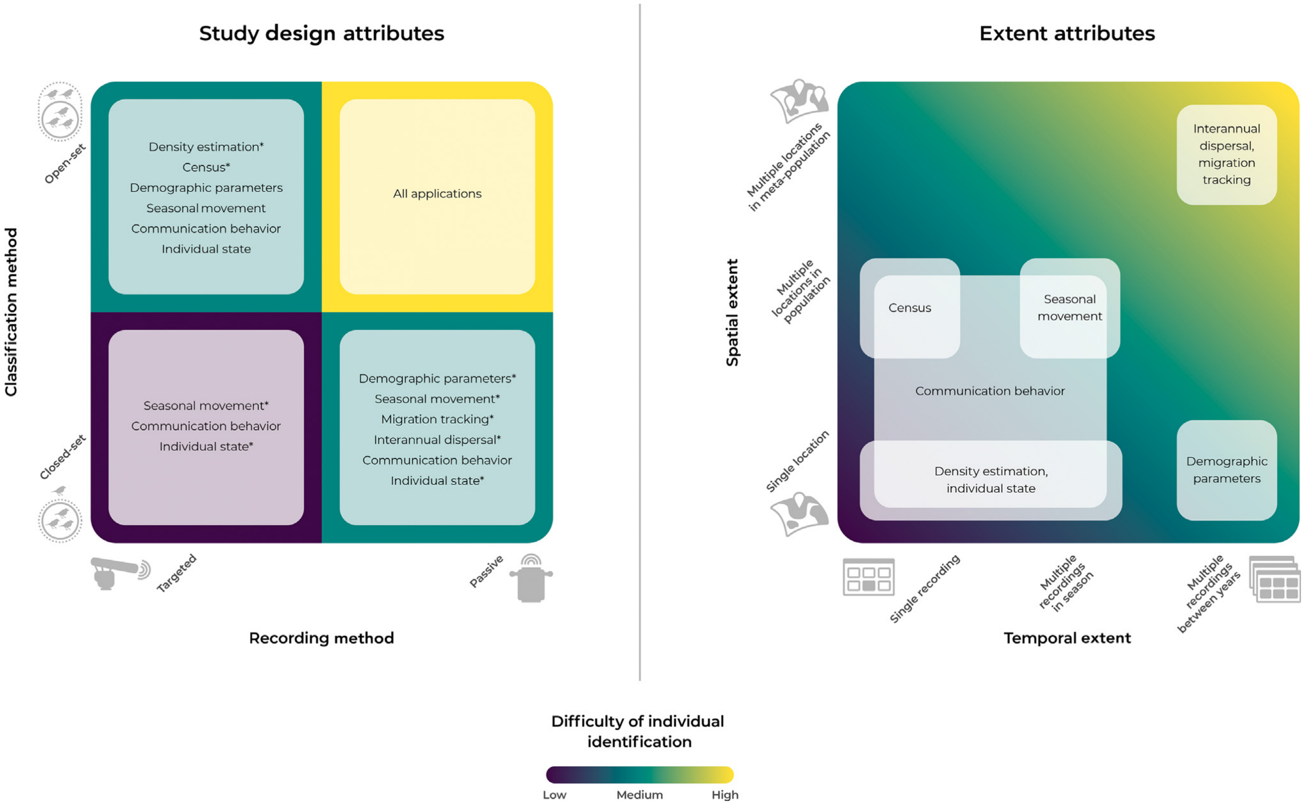

Figure 1. Relative difficulty of the applications of acoustic individual identification (AIID) based on the study design used to collect acoustic recordings and classify individuals, and the spatiotemporal extent of the application across which AIID is required. The asterisk (*) indicates combinations of the study design's attributes that are feasible but potentially suboptimal or limited for that application, depending on the species' attributes and the spatiotemporal scale (see Table 1 for details).

together, even for novel calls or individuals, whereas vocalizations from different individuals will form separate clusters. Open-set classification is substantially more useful for applications of AIID to ecological questions, particularly for PAM recordings, because ecologists rarely have access to fully censused wild populations to train closed-set classification models (Table 1).

Strong parallels to the AIID problem exist in adjacent acoustic disciplines such as music information retrieval and human speaker recognition, as well as in nonacoustic tasks such as facial recognition and reidentification of wildlife from camera trap imagery. In particular, human speaker recognition [36] has many direct parallels to AIID, often incorporating passive recording and high signal masking [37]. Extensive speaker recognition research in recent years has produced state-of-the-art closed-set and open-set models that have achieved high accuracy rates on datasets with thousands of individuals [38,39], which suggests that broadscale implementation of AIID is achievable. Given these parallels, we encourage bioacoustic researchers to collaborate, disseminate knowledge, and share datasets with experts in adjacent acoustic and computing science disciplines.

Given that open-set classification provides more opportunities for the application of AIID (Table 1, Figure 1, and Box 1), researchers should look toward deep learning advancements in open-set classification from disciplines such as speaker recognition and verification [40,41]. In particular, Box 1. Methods and use of acoustic individual identification

We reviewed 598 studies that tested for individual acoustic signatures, 96.0% of which found that individuals could be distinguished, representing a broad range of species, from well-studied taxa such as bats and songbirds to less frequently studied fish and invertebrates (see supplemental information online). Many studies also tested for differences across demographic stages and consistently found individual acoustic differences in age and sex categories. Of the 22 studies that did not find evidence of acoustic signatures, 12 tested for the acoustic signatures of avian taxa (primarily nonpasserines), four of aquatic mammals, three of bats, and three of terrestrial mammals.

Of the 598 acoustic signature papers, 358 also used classification for AIID per se. Among these AIID papers, most used targeted recording (93.3%). The majority of targeted recording designs involved following known individuals and recording them with directional handheld recorders that produce high signal-to-noise ratios. Other approaches included recording captive or captured individuals in recording chambers, placing ARUs in locations likely to target specific individuals, or attaching passive recorders to individual animals. True untargeted passive recording approaches only comprised 6.7% of studies.

Similarly, most studies used the easier closed-set classification approach (93.3%). Those closed-set classifiers included multivariate methods (79.8%), supervised machine learning (6.8%), probabilistic assessments based on variations among the measured sound features (3.7%), deep learning (4.6%), cluster analysis (2.0%), template matching (1.7%), or manual classification by experts (1.1%). Discriminant function analysis (DFA) was used in 79.0% of instances of closed-set classification, and was often preceded by a potential for individual encoding (PIC) analysis. The probability that DFA was used in a closed-set classification peaked in 2002 [95% confidence interval (CI): 0.86–0.95] and has declined since (95% CI: 0.46–0.77); however, DFA remains the dominant classification approach for AIID (supplemental information). By contrast, open-set classification was used in only 6.7% of AIID studies. Of these open-set studies, many (17.1%) were conducted using manual classification by experts, in which an individual investigator was responsible for examining the sounds and assigning them to individuals. Multivariate (20.0%), supervised machine learning (22.9%), and unsupervised clustering (17.1%) approaches were also commonly used. Although it is a different approach to AIID that does not rely on acoustic signature, we considered spatial clustering of localized vocalizations obtained from time-synchronized ARU arrays as an example of open-set classification for territorial songbirds or short-duration recordings of other species (5.7%). Despite the potential for deep learning to be used in open-set classification, we found only one study in our review that used this approach, although there is at least one additional existing example [50].

The mean number of individuals included to train and test the AIID models was 20.3 individuals (SD = 23.4), with a minimum of 2 and a maximum of 263. Moreover, there was a strong taxonomic bias in published studies, with birds (38.8%), terrestrial mammals (36.9%), bats (10.3%), and aquatic mammals (9.8%) disproportionately represented. We found only four AIID studies on amphibians, six on fish, one on invertebrates, and two on reptiles.

The minority (10.9%) of studies used AIID for ecological or evolutionary applications (Table 1). Of the 38 studies using AIID as a tool, most studied the population size or estimations of density (36.8%), particularly when using open-set classification. Studies of communication behavior were also common (26.3%), including evaluating the decision rules of territorial neighbors, the reliability of alarm calls and kin recognition over time, or parent–offspring identification in breeding colonies. The probability that a study used AIID for an application was higher for open-set than closed-set classification, and significantly higher for passive recording than targeted recording (see supplemental information online).

embedding shows promise for improving the accuracy of open-set classification. Pretrained embedding spaces from speaker recognition models [42] or bioacoustic species classifiers [43] may provide a good jumping-off point for AIID models; embedding spaces from pretrained human voice recognition models have been successfully used to distinguish individual acoustic signatures, both with **transfer learning** techniques [44] and without them [45]. Applying transfer learning to existing embedding spaces may be particularly successful if combined with metrics for measuring individual signatures that are less sensitive to species-specific variation in the acoustic signature [46], because the features required to discriminate individuals (e.g. tone, harmonic emphasis) may be different from those that discriminate species (e.g. syllabic content), although there is considerable overlap.

Open-set classification for AIID can also be improved by emerging techniques within bioacoustics and adjacent disciplines. Speaker and facial recognition often use specialized objective functions such as triplet loss, which can encourage the distances in embedding spaces to better represent between-individual differences (e.g., [37,47]). The use of autoencoders could also encourage trained models to represent the key features that distinguish individuals (e.g., [48]). The development of methods to cluster the embeddings themselves into individuals also remains an active area of research, with affinity propagation clustering showing particular promise [49]. Jointly training the embedding space with a clustering algorithm also shows great promise for open-set classification [50].

Regardless of the classification method, variation in the acoustic signature of a single individual is a major challenge for both AIID applications and other disciplines such as speaker recognition. Intraindividual variation can be due to cultural changes or aging [51,52], variation in body size [53], differences in the signature associated with behavior [54], or adaptive changes for coping with signal degradation due to environmental conditions [55]. In speech recognition, intraindividual variation in an acoustic signature is described as 'intrinsic mismatch,' including mismatches in the language being spoken, emotion, vocal effort, or physiological changes [56]. In a recent human speaker classification challenge, performance dropped substantially when models built on one age of speakers were applied to another age [38]. However, behavioral studies suggest that long-term acoustic kin recognition does exist in several species [57–59], suggesting that overcoming intrinsic mismatches in acoustic signatures is possible. One approach to contend with this mismatch is to model aging as a generative process to predict future acoustic signatures [60,61]; similar modeling could be applied to behavioral states. Nonmechanistic approaches to improve the performance of the speaker verification include specialized transfer learning methods called domain adaptation [38,62–64], which can enable the model to bridge the gap between mismatched signatures from the same individual. Transfer learning methods may be helpful not only for contending with intraindividual variation, but also for improving the models' ability to generalize across datasets (e.g., targeted to passive recording datasets).

AIID is limited by training data

One of the major challenges in AIID is the lack of large datasets labeled with individual identities, which are needed to train the types of powerful but data-hungry deep learning models that are successful in other disciplines [34]. The mean number of individuals used to train AIID models was 20.3 (Box 2), which is several orders of magnitude less than the datasets used to train the existing state-of-the-art human speaker models [38,39]. Furthermore, the strong taxonomic bias in previous AIID studies suggests that the existing labeled datasets are also likely to be restricted by taxa, as in other areas of ecology and evolution [65–67].

There are several potential approaches to building large, labeled bioacoustic datasets for AIID. For some species, the annotators may be able to perceive the differences between individuals aurally [68], potentially with improved accuracy if a spectrogram is used for interpretation (e.g., [69,70]). Additionally, captive populations represent an excellent opportunity for collecting training data, particularly for rare or endangered species. For wild animals, existing long-term marked populations could be sampled with PAM to build training datasets, particularly if the individuals are fitted with high-precision movement-tracking loggers to link sound recordings to the identity of the emitter (e.g., [71,72]). Targeted deployment of ARUs within individual specific core areas of territorial species could also be conducted, even for unmarked populations, to provide recordings of individuals known by their territory's identity. Spatial clustering of individual sounds localized from time-synchronized ARU arrays is also an excellent opportunity to build labeled datasets from unmarked populations, at least for shorter temporal extents [73]. For some species, multisensor monitoring could simultaneously collect paired datasets such as acoustic and imagery data for additional information that could improve classification. Although clean individual labels are more readily obtained from targeted recordings, we encourage the use of both targeted and passive recording approaches for building AIID training datasets because including soundscape data in training classifiers can improve the domain shifts that occur when applying deep learning models trained on targeted data to passive acoustic data [74].

Other methods can be used to maximize the utility of the existing data, but are not without challenges. For example, data augmentation can effectively multiply the size of a dataset by modifying the original recordings [75,76], but these modifications must be consistent with the variations within and among individuals [34]. Methods of data synthesis can also be used to generate large amounts of data with known labels, but these data must share the features that contribute to the acoustic signatures in the target taxa, which can vary widely by species, even for closely related taxa. Model training approaches similar to transfer learning from other models, such as BirdNet (e.g., [43,44]), few-shot learning (e.g., [77,78]), and self-supervised learning (e.g., [79,80]) would also reduce the need for large training datasets.

Regardless of the approach chosen to build and maximize the training datasets, we call for open access to labeled datasets to move the field of AIID forward. Researchers with existing marked

Box 2. Species' attributes affect the difficulty of acoustic individual identification Five categories of the attributes of the focal taxa or species affect the difficulty of AIID (see Table I and Figure I)

Table I. Species' attributes that affect the difficulty of AIID.

Attribute	Challenges for AIID	Study design considerations
Acoustic signature	Acoustic signatures that change across time will be difficult to classify Genetic or learned similarities among individuals may cause misclassification	AIID at larger temporal extents may be challenging if the acoustic signature changes across time Modeling change or updating classifiers will help with temporal change in the acoustic signatures
Species' density	High-density study systems contain more individuals for classification Overlapping sounds in high-density areas may be challenging to distinguish	In low-density study systems, passive recorders must be placed in areas where individuals are likely to occur The spatial coverage of the chosen recording unit should be considered Open-set classification may be required
Detectability of sounds	Low cue rates and low cue amplitude decrease the probability of an individual's cue being clearly detected and accurately classified High signal masking decreases the probability that a cue is accurately classified	More acoustic recorders and longer recording durations increase the chance of capturing individuals when cue rates are low if passive recording is used Open-set classification approaches may be preferred because gathering complete training datasets for closed-set classification is difficult when cue rates are low Masking of the signal can be particularly problematic for species with low-frequency acoustic cues that overlap with anthropogenic and/or recording unit noise
Movement during recording	Rotation of the head, vocalization during movement, or frequent switching between locations of sound production during singing causes variable degradation of the acoustic signature	Densely placed recorders or even localization arrays may be required if passive recording is used
Movement between recordings	Movement between recordings (e.g., through dispersal, large home range) increases the probability of encountering new individuals	Large spatial and temporal extents may be required if acoustic mark–recapture is desired Open-set classification and passive recording may be required for high movement rates

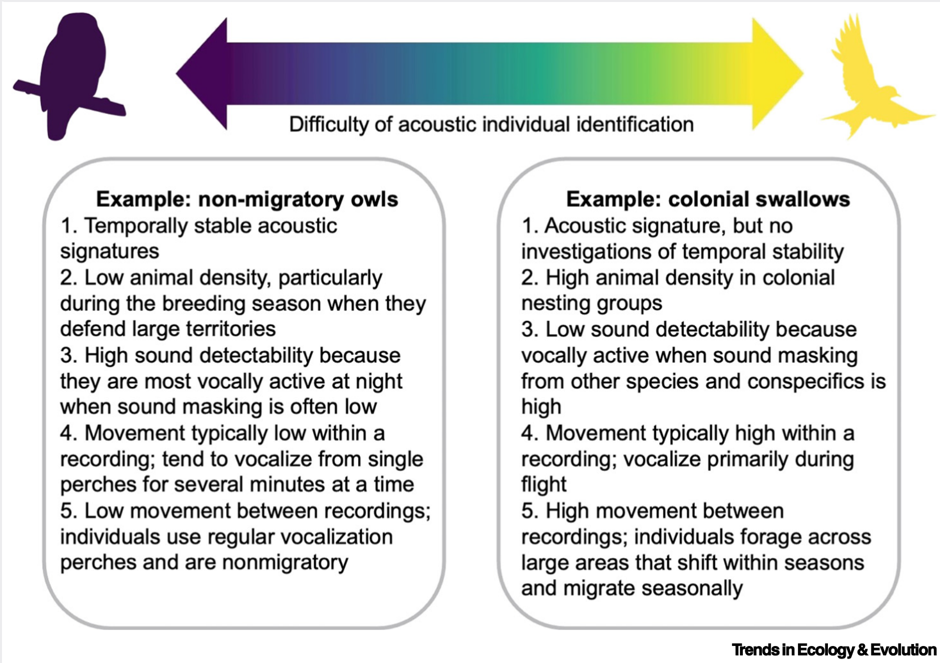

Figure I. Examples of easy and hard taxa for successful acoustic individual identification.

populations have a unique opportunity to contribute in this space by collaborating with acoustic researchers to build labeled datasets of known individuals. Data-sharing is becoming increasingly common and has spurred advances in both ecology (reviewed by [81]) and machine learning (e.g., [82]. We propose that a flexible data catalog of open-access datasets is the best approach to sharing individually labeled acoustic datasets because data catalogs operate under the findable, accessible, interoperable, and reusable (FAIR) principles of data-sharing while also allowing for flexibility in where the actual dataset is stored [83]. Furthermore, a data catalog would eliminate the need for a large central data repository, which can be limiting because acoustic data files are large [2,3]. We strongly encourage the current and future owners of individually labeled datasets to contribute them to the Environmental Data Initiative (EDI) repository (https://edirepository.org/) [84]. The EDI is an open data repository suited for publishing and archiving environmental and ecological data from around the globe. EDI uses ezEML, a web-based tool for creating and managing data documentation using the Ecological Metadata Language. EDI's data repository, including its highly rated curation service, and ezEML are freely available for small- to medium-volume data and support offline data archives over 100 GB. All data owners should include the keyword 'AIID' in their metadata submission so that datasets relevant to AIID can be easily queried and so there is potential to create a custom AIID data catalog. We also recommend that the data owners use our framework to include metadata for the recording method ('targeted' vs. 'passive') and to indicate the spatial ('location', 'population', and 'metapopulation') and temporal extent ('recording', 'season', 'year') across which the individuals are consistently labeled. The datasets should be fully labeled, and the metadata should list whether the annotations were compiled by segmenting the recording into windows of equal duration (e.g., 3 seconds, 'window-level annotation') and listing all individuals within each window, or by annotating each call of each individual with the start and end time of the vocalization ('call-level annotation'). An example of an individually labeled dataset with metadata following our framework is available [85]. We particularly encourage owners of individually labeled PAM datasets to contribute them to EDI to facilitate the development of AIID for passive recordings. We also encourage the sharing and development of repeated recordings of individuals across seasons and years from marked and captive populations (e.g., [34]) for understanding and modeling drift in acoustic signatures due to aging or cultural changes.

Improving and diversifying the application of AIID

Applying AIID for ecological and evolutionary studies remains in its infancy, as evidenced by the low proportion of studies that have used it for an ecological or evolutionary applications (Table 1) as opposed to proof-of-concept studies that AIID can be achieved (89.1%; Box 1).

The difficulty of a given application of AIID is strongly driven by the spatiotemporal extent of that application (Figure 1). The temporal extent influences complexity as the duration of the recording increases and also as the study's duration increases to include AIID across multiple recordings collected at different times within a season, or even across seasons and years. Spatial extents that cross multiple populations can be particularly challenging, as the probability of encountering the same individual at multiple survey locations is determined by factors such as natal dispersal.

Increasing difficulty within the spatiotemporal plan is caused by three mechanisms. First, the spatiotemporal extent is generally correlated with the population size, which decreases the chance of

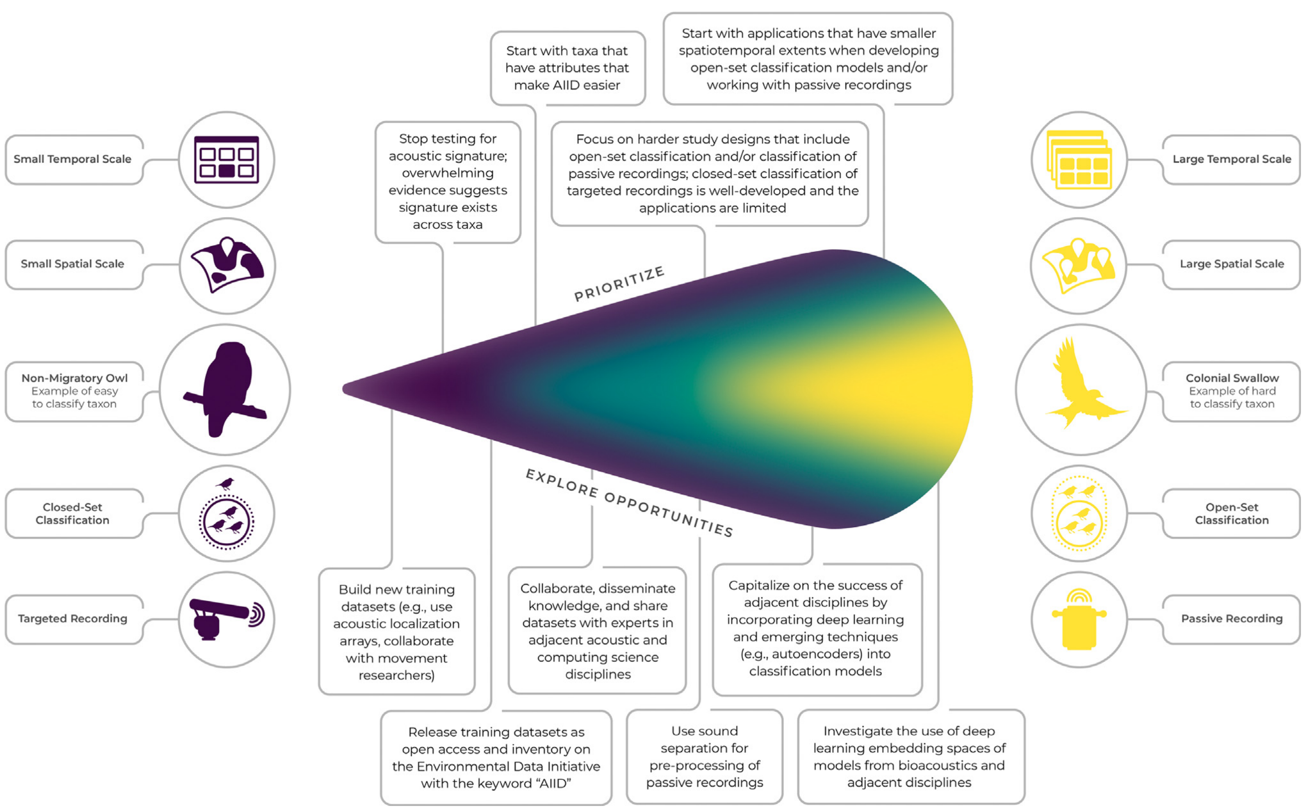

Figure 2. Priorities and opportunities for development and diversification of acoustic individual identification (AIID) in ecology and evolution.

correct classification and classification performance [10,86,87]. Second, the spatiotemporal extent is also correlated with the probability of individual movements in space and turnover in time, which increases the complexity of the open-set problem. Third, the spatiotemporal extent is correlated with variation in acoustic cues, both across space via geographic differences in vocalization or repertoire [21,88,89], and also across time as individual sounds change across and even within seasons [52,57,90]. The relative effect of these three factors also interacts with the attributes of the focal species, such as the movement rate and stability of acoustic signatures (Box 1).

Due to these processes, the suitable recording and classification methods for a given AIID application are also determined by the spatiotemporal extent. At restricted scales, ecological datasets might have the characteristics of a closed-set problem, but as the spatial or temporal scales are expanded, stochastic demographic processes and increased spatial coverage will result in violation of the closed-set problem's assumptions. Similarly, targeted recording might be practical at small spatial and temporal scales, but due to the labor intensity of most targeted recording approaches, it will quickly become untenable across large spatial scales or for applications that require long recording periods. We therefore suggest the dominant approach to the design of AIID studies, namely closed-set classification of targeted recordings, is not realistic for many ecological and conservation applications (Figure 1 and Table 1). Indeed, we found that the probability of a study using open-set classification or passive recording was higher for papers that applied AIID to an ecological or evolutionary question (Box 1).

Successful implementation of AIID for a diverse set of ecological and evolutionary applications will therefore require advances in both recording and classification methods, as discussed earlier, as well as an understanding of how AIID affects the outcomes of those applications. Statistical approaches to using mark–recapture techniques should be adapted to incorporate classification error [91–93]. Ecologists should also use sensitivity analyses to understand how that classification error affects both the conclusions of studies and the downstream management decisions, and compare these results with those using traditional approaches of individual identification.

Concluding remarks

AIID has been demonstrated to be both possible and useful, but scaling these demonstrations across the range of potential ecological and evolutionary applications (Table 1 and Figure 1) remains challenging, partially due to the dominance of closed-set classification and targeted recordings that limit the spatiotemporal extent of potential applications. Adjacent acoustic disciplines such as human speaker recognition suggest that broad-scale implementation of AIID is a solvable problem, and there are already several successful examples within the literature (e.g., [11,14,94]). Our framework outlines the potential of AIID, and we suggest that a successful road forward for the development and diversification of AIID in ecology and evolution should use that framework to prioritize easier taxa and applications when exploring opportunities for AIID that use open-set classification and/or passive recording methods (Figure 2). There are, however, questions as to the generalizability of successful AIID across more difficult taxa and applications and the impact of classification error on downstream analyses and outcomes (see Outstanding questions). Regardless, we are confident that broad-scale implementation of AIID will be achievable in the near future for many applications and will allow biologists to answer new and important ecological and evolutionary questions with less bias and fewer negative population effects and resources than current approaches.

Data availability

The results of the systematic literature review are available at https://zenodo.org/records/10626982.

Outstanding questions

Do signals themselves limit the accuracy of AIID, or can higher accuracy be achieved with better or more training data and more advanced methods?

Can intraindividual variations in acoustic signatures be overcome with advances in classification methods?

Is AIID achievable for all taxa, or are some ecological traits insurmountable barriers to the successful classification of individuals?

Can pretrained AIID models transfer to new species or datasets with no or limited training data?

Is it possible to build AIID classifiers that are insensitive to species?

Are estimates obtained using traditional approaches similar to those from AIID?

What are sufficiently low error rates of the identification of individuals for ecological and evolutionary applications?

How does the remaining classification error affect the outcome of ecological and evolutionary applications of AIID?

Can classification error be propagated through statistical analyses for applications of AIID?

Acknowledgments

This project was initiated at a cross-disciplinary acoustics conference, AudioXD 1.0, held at the University of Pittsburgh in August 2022. We thank all the attendees of that conference and the AIID working group that contributed ideas and suggestions during the initial brainstorming sessions. Thank you to the Editor-in-Chief and the three anonymous reviewers whose thoughtful comments and suggestions greatly improved the clarity, structure, and messaging of this manuscript. Thank you to Sonder Creative for producing the publication-quality figures for the manuscript, and to Mark Servilla and Colin Smith of the EDI for support in developing the FAIR data sharing recommendations.

Declaration of interests

The authors declare no competing interests.

Supplementary information

Supplemental information associated with this article can be found online at https://doi.org/10.1016/j.tree.2024.05.007.

References

1. Ross, S.R.P.-J. *et al.* (2023) Passive acoustic monitoring provides a fresh perspective on fundamental ecological questions. *Funct. Ecol.* 37, 959–975
2. Shonfield, J. and Bayne, E.M. (2017) Autonomous recording units in avian ecological research: current use and future applications. *Avian Conserv. Ecol.* 12, 14
3. Sugai, L.S.M. *et al.* (2019) Terrestrial passive acoustic monitoring: review and perspectives. *BioScience* 69, 15–25
4. Sugai, L.S.M. and Llusia, D. (2019) Bioacoustic time capsules: using acoustic monitoring to document biodiversity. *Ecol. Indic.* 99, 149–152
5. Pollock, K.H. (1976) Building models of capture–recapture experiments. *J. R. Stat. Soc. Ser. Stat.* 25, 253–259
6. Otis, D.L. *et al.* (1978) Statistical inference from capture data on closed animal populations. *Wildl. Monogr.* 62, 3–135
7. McClintock, B.T. and White, G.C. (2009) A less field-intensive robust design for estimating demographic parameters with mark–resight data. *Ecology* 90, 313–320
8. Sikes, R.S. and the Animal Care and Use Committee of the American Society of Mammalogists (2016) 2016 Guidelines of the American Society of Mammalogists for the use of wild mammals in research and education. *J. Mammal.* 97, 663–688
9. Camacho, C. *et al.* (2017) Lifelong effects of trapping experience lead to age-biased sampling: lessons from a wild bird population. *Anim. Behav.* 130, 133–139
10. Marin-Cudraz, T. *et al.* (2019) Acoustic monitoring of rock ptarmigan: a multi-year comparison with point-count protocol. *Ecol. Indic.* 101, 710–719
11. Bailey, H. *et al.* (2021) Identifying and predicting occurrence and abundance of a vocal animal species based on individually specific calls. *Ecosphere* 12, e03685
12. Vögeli, M. *et al.* (2008) Who are we sampling? Apparent survival differs between methods in a secretive species. *Oikos* 117, 1816–1823
13. Terry, A.M.R. and McGregor, P.K. (2002) Census and monitoring based on individually identifiable vocalizations: the role of neural networks. *Anim. Conserv.* 5, 103–111
14. Hensel, A.L.J. *et al.* (2022) Passive acoustic monitoring provides predictable and reliable underestimates of population size and longevity in wild savannah sparrows. *Ornithol. Appl.* 124, 1–11
15. Odom, K.J. *et al.* (2013) Distinctiveness in the territorial calls of great horned owls within and among years. *J. Raptor Res.* 47, 21–30
16. Hutchinson, R.E. *et al.* (1968) The basis for individual recognition by voice in the Sandwich tern (*Sterna sandvicensis*). *Behaviour* 32, 150–157
17. Watson, M. (1969) Significance of antiphonal song in the eastern whipbird, *Psophodes olivaceus*. *Behaviour* 35, 157–178
18. Hyacinthe, C. *et al.* (2023) Acoustic signatures in Mexican cavefish populations inhabiting different caves. *PLoS One* 18, e0289574
19. Suthers, R.A. (1994) Variable asymmetry and resonance in the avian vocal tract: a structural basis for individually distinct vocalizations. *J. Comp. Physiol. A.* 175, 457–466
20. Smith-Vidaurre, G. *et al.* (2021) Individual vocal signatures show reduced complexity following invasion. *Anim. Behav.* 179, 15–39
21. Aplin, L.M. (2019) Culture and cultural evolution in birds: a review of the evidence. *Anim. Behav.* 147, 179–187
22. Linhart, P. *et al.* (2022) The potential for acoustic individual identification in mammals. *Mamm. Biol.* 102, 667–683
23. Morrison, T.A. *et al.* (2011) Estimating survival in photographic capture–recapture studies: overcoming misidentification error. *Methods Ecol. Evol.* 2, 454–463
24. Ehnes, M. and Foote, J.R. (2015) Comparison of autonomous and manual recording methods for discrimination of individually distinctive ovenbird songs. *Bioacoustics* 24, 111–121
25. Park, T.J. *et al.* (2022) A review of speaker diarization: recent advances with deep learning. *Comput. Speech Lang.* 72, 101317
26. Denton, T. *et al.* (2022) Improving bird classification with unsupervised sound separation. In *Proceedings of the ICASSP 2022 – 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 636–640, IEEE, Singapore
27. Bermant, P.C. (2021) BioCPPNet: automatic bioacoustic source separation with deep neural networks. *Sci. Rep.* 11, 23502
28. Aubin, T. (2004) Penguins and their noisy world. *An. Acad. Bras. Ciênc.* 76, 279–283
29. Darras, K.F.A. *et al.* (2020) High microphone signal-to-noise ratio enhances acoustic sampling of wildlife. *PeerJ* 8, e9955
30. Hedley, R.W. *et al.* (2021) Distance truncation via sound level for bioacoustic surveys in patchy habitat. *Bioacoustics* 30, 303–323
31. Cai, S. *et al.* (2013) New wireless acoustic array node for localization, beamforming and source separation for bio-complexity bird data collection and study. In *Proceedings of the 2013 IEEE China Summit and International Conference on Signal and Information Processing*, pp. 210–214, IEEE, Beijing, China
32. Suzuki, R. *et al.* (2017) HARKBird: exploring acoustic interactions in bird communities using a microphone array. *J. Robot. Mechatron.* 29, 213–223
33. Somervuo, P. *et al.* (2023) Effects of landscape and distance in automatic audio based bird species identification. *J. Acoust. Soc. Am.* 154, 245–254
34. Stowell, D. *et al.* (2019) Automatic acoustic identification of individuals in multiple species: improving identification across recording conditions. *J. R. Soc. Interface* 16, 20180940
35. Ptacek, L. *et al.* (2016) Automatic recognition of bird individuals on an open set using as-is recordings. *Bioacoustics* 25, 55–73
36. Bai, Z. and Zhang, X.-L. (2021) Speaker recognition based on deep learning: an overview. *Neural Netw.* 140, 65–99
37. Jati, A. *et al.* (2019) Multi-task discriminative training of hybrid DNN-TVM model for speaker verification with noisy and far-field speech. In *Proceedings of Interspeech 2019*, pp. 2463–2467, *Graz, Austria*
38. Huh, J. *et al.* (2023) VoxSRC 2022: The Fourth VoxCeleb speaker recognition challenge. *arXiv*, Published online March 6 2023. https://doi.org/10.48550/arXiv.2302.10248
39. Yakovlev, I. *et al.* (2023) VoxTube: a multilingual speaker recognition dataset. In *INTERSPEECH 2023*, pp. 2238–2242. Published online August 2023 https://doi.org/10.21437/Interspeech.2023

40. Kishan, K.C. *et al.* (2022) OpenFEAT: improving speaker identification by open-set few-shot embedding adaptation with transformer. In *Proceedings of ICASSP 2022*, pp. 7062–7066, Singapore
41. Peri, R. *et al.* (2023) VoxWatch: an open-set speaker recognition benchmark on VoxCeleb. *arXiv*, Published online June 30 2023. https://doi.org/10.48550/arXiv.2307.00169
42. Spillmann, B. *et al.* (2017) Who shall I say is calling? Validation of a caller recognition procedure in Bornean flanged male orangutan (*Pongo pygmaeus wurmbii*) long calls. *Bioacoustics* 26, 109–120
43. Ghani, B. *et al.* (2023) Feature embeddings from large-scale acoustic bird classifiers enable few-shot transfer learning. *Sci. Rep.* 13, 22876
44. Dufourq, E. *et al.* (2022) Passive acoustic monitoring of animal populations with transfer learning. *Ecol. Inform.* 70, 101688
45. Sarkar, E. and Magimai-Doss, M. (2023) Can self-supervised neural representations pre-trained on human speech distinguish animal callers? In *Proceedings of INTERSPEECH 2023*, pp. 1189–1193, *Dublin, Ireland*
46. Linhart, P. *et al.* (2019) Measuring individual identity information in animal signals: overview and performance of available identity metrics. *Methods Ecol. Evol.* 10, 1558–1570
47. Schroff, F. *et al.* (2015) FaceNet: a unified embedding for face recognition and clustering. In *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA*, pp. 815–823
48. Xie, J. *et al.* (2020) High accuracy individual identification model of crested ibis (*Nipponia nippon*) based on autoencoder with self-attention. *IEEE Access* 8, 41062–41070
49. Clink, D.J. and Klinck, H. (2021) Unsupervised acoustic classification of individual gibbon females and the implications for passive acoustic monitoring. *Methods Ecol. Evol.* 12, 328–341
50. Bedoya, C.L. and Molles, L.E. (2021) Acoustic censusing and individual identification of birds in the wild. *bioRxiv*, Published online October 31 2021. https://doi.org/10.1101/2021.10.29.466450
51. Deng, Z. *et al.* (2019) Within-season decline in call consistency of individual male common cuckoos (*Cuculus canorus*). *J. Ornithol.* 160, 317–327
52. Raymond, S. *et al.* (2020) Vocal instability over time in individual male European nightjars, *Caprimulgus europaeus*: recommendations for acoustic monitoring and surveys. *Bioacoustics* 29, 280–295
53. Barbraud, C. *et al.* (2000) Variation in call properties of the snow petrel, *Pagodroma nivea*, in relation to sex and body size. *Aust. J. Zool.* 48, 421–430
54. Linhart, P. *et al.* (2013) Being angry, singing fast? Signalling of aggressive motivation by syllable rate in a songbird with slow song. *Behav. Process.* 100, 139–145
55. Mouterde, S.C. *et al.* (2014) Learning to cope with degraded sounds: female zebra finches can improve their expertise at discriminating between male voices at long distance. *J. Exp. Biol.* 217, 3169–3177
56. Hansen, J.H.L. and Hasan, T. (2015) Speaker recognition by machines and humans: a tutorial review. *IEEE Signal Process. Mag.* 32, 74–99
57. Boeckle, M. and Bugnyar, T. (2012) Long-term memory for affiliates in ravens. *Curr. Biol.* 22, 801–806
58. Insley, S.J. (2000) Long-term vocal recognition in the northern fur seal. *Nature* 406, 404–405
59. Briefer, E.F. *et al.* (2012) Mother goats do not forget their kids' calls. *Proc. R. Soc. B Biol. Sci.* 279, 3749–3755
60. Das, B. *et al.* (2013) Aging speech recognition with speaker adaptation techniques: study on medium vocabulary continuous Bengali speech. *Pattern Recogn. Lett.* 34, 335–343
61. Werner, L. *et al.* (2019) Automated speech recognition systems and older adults: a literature review and synthesis. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* 63, 42–46
62. Wang, Q. *et al.* (2023) Generalized domain adaptation framework for parametric back-end in speaker recognition. *IEEE Trans. Inf. Forensics Secur.* 18, 3936–3947
63. Zhang, J.-T. *et al.* (2023) Adversarial training based on meta-learning in unseen domains for speaker verification. In *Communications in Computer and Information Science* (1765) (Zhenhua, L. *et al.*, eds), pp. 124–131, Springer
64. Bahmaninezhad, F. *et al.* (2021) An investigation of domain adaptation in speaker embedding space for speaker recognition. *Speech Comm.* 129, 7–16
65. Bonnet, X. *et al.* (2002) Taxonomic chauvinism. *Trends Ecol. Evol.* 17, 1–3
66. Troudet, J. *et al.* (2017) Taxonomic bias in biodiversity data and societal preferences. *Sci. Rep.* 7, 9132
67. Pichler, M. and Hartig, F. (2023) Machine learning and deep learning—a review for ecologists. *Methods Ecol. Evol.* 14, 994–1016
68. Armstrong, J.T. (1965) Breeding home range in the nighthawk and other birds: its evolutionary and ecological significance. *Ecology* 46, 619–629
69. Petrusková, T. *et al.* (2016) Repertoire-based individual acoustic monitoring of a migratory passerine bird with complex song as an efficient tool for tracking territorial dynamics and annual return rates. *Methods Ecol. Evol.* 7, 274–284
70. Sayigh, L.S. *et al.* (2007) Facts about signature whistles of bottlenose dolphins, *Tursiops truncatus*. *Anim. Behav.* 74, 1631–1642
71. Siemers, B.M. and Kerth, G. (2006) Do echolocation calls of wild colony-living Bechstein's bats (*Myotis bechsteinii*) provide individual-specific signatures? *Behav. Ecol. Sociobiol.* 59, 443–454
72. Reid, D.S. *et al.* (2022) Breeding status shapes territoriality and vocalization patterns in spotted owls. *J. Avian Biol.* 2022, e02952
73. Rhinehart, T.A. *et al.* (2020) Acoustic localization of terrestrial wildlife: current practices and future opportunities. *Ecol. Evol.* 10, 6794–6818
74. Lauha, P. *et al.* (2022) Domain-specific neural networks improve automated bird sound recognition already with small amounts of local data. *Methods Ecol. Evol.* 13, 2799–2810
75. Takahashi, N. *et al.* (2016) Deep convolutional neural networks and data augmentation for acoustic event detection. In *Proceedings of Proc. Interspeech 2016*, pp. 2982–2986
76. Salamon, J. and Bello, J.P. (2017) Deep convolutional neural networks and data augmentation for environmental sound classification. *IEEE Signal Process. Lett.* 24, 279–283
77. Stowell, D. and Plumbley, M.D. (2014) Automatic large-scale classification of bird sounds is strongly improved by unsupervised feature learning. *PeerJ* 2, e488
78. Nolasco, I. *et al.* (2022) Few-shot bioacoustic event detection at the DCASE 2022 challenge. *arXiv*, Published online July 14 2022. https://doi.org/10.48550/arXiv.2207.07911
79. de Sa, V. (1993) Learning classification with unlabeled data. In *Advances in Neural Information Processing Systems* (6) (Cowan, J. *et al.*, eds), pp. 112–119, MIT Press
80. Doersch, C. *et al.* (2015) Unsupervised visual representation learning by context prediction. In *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 1422–1430, Santiago, Chile
81. Michener, W.K. (2015) Ecological data sharing. *Ecol. Inform.* 29, 33–44
82. Prior, F. *et al.* (2020) Open access image repositories: high-quality data to enable machine learning research. *Clin. Radiol.* 75, 7–12
83. Wilkinson, M.D. *et al.* (2016) The FAIR guiding principles for scientific data management and stewardship. *Sci. Data* 3, 160018
84. Gries, C. *et al.* (2023) The Environmental Data Initiative: connecting the past to the future through data reuse. *Ecol. Evol.* 13, e9592
85. Bayne, E.M. and Knight, E.C. (2024) Manual annotation of individual ovenbirds in acoustic recordings. *Environ. Data Initiative*, Published online 7 May, 2024. https://doi.org/10.6073/pasta/fea50e96eb343b579b0b69f317b70cda
86. Linhart, P. and Šálek, M. (2017) The assessment of biases in the acoustic discrimination of individuals. *PLoS One* 12, e0177206
87. Budka, M. *et al.* (2015) Is it possible to acoustically identify individuals within a population? *J. Ornithol.* 156, 481–488
88. Podos, J. and Warren, P.S. (2007) The evolution of geographic variation in birdsong. In *Advances in the Study of Behavior* (37), pp. 403–458
89. Wilkins, M.R. *et al.* (2013) Evolutionary divergence in acoustic signals: causes and consequences. *Trends Ecol. Evol.* 28, 156–166

90. Stowell, D. (2022) Computational bioacoustics with deep learning: a review and roadmap. *PeerJ* 10, e13152
91. Ferguson, P.F.B. *et al.* (2015) Occupancy models for data with false positive and false negative errors and heterogeneity across sites and surveys. *Methods Ecol. Evol.* 6, 1395–1406
92. Royle, J.A. and Link, W.A. (2006) Generalized site occupancy models allowing for false positive and false negative errors. *Ecology* 87, 835–841
93. Doser, J.W. *et al.* (2021) Integrating automated acoustic vocalization data and point count surveys for estimation of bird abundance. *Methods Ecol. Evol.* 12, 1040–1049
94. Merchan, F. *et al.* (2019) Detection and identification of manatee individual vocalizations in Panamanian wetlands using spectrogram clustering. *J. Acoust. Soc. Am.* 146, 1745–1757
95. Mikkelsen, G. *et al.* (2013) Can individually characteristic calls be used to identify long-distance movements of corncrakes *Crex crex*? *J. Ornthol.* 154, 751–760
96. Kidawa, D. *et al.* (2023) Mine or my neighbours' offspring: an experimental study on parental discrimination of offspring in a colonial seabird, the little auk *Alle alle*. *Sci. Rep.* 13, 15088
97. Probert, R. *et al.* (2021) Vocal correlates of arousal in bottlenose dolphins (*Tursiops* spp.) in human care. *PLoS One* 16, e0250913