Journal of
Sensor and Actuator Networks

MDPI

Article

Animal-Borne Adaptive Acoustic Monitoring

Devin Jean [1], Jesse Turner [2], Will Hedgecock [1], György Kalmár [3], George Wittemyer [2] and Ákos Lédeczi [1,*]

[1] Institute for Software Integrated Systems, Vanderbilt University, Nashville, TN 37235, USA; devin.c.jean@vanderbilt.edu (D.J.); ronald.w.hedgecock@vanderbilt.edu (W.H.)
[2] Department of Fish, Wildlife and Conservation Biology, Colorado State University, Fort Collins, CO 80523, USA; jesse.c.turner@colostate.edu (J.T.); g.wittemyer@colostate.edu (G.W.)
[3] Department of Technical Informatics, University of Szeged, 6720 Szeged, Hungary; kalmargy@inf.u-szeged.hu
* Correspondence: akos.ledeczi@vanderbilt.edu

Abstract

Animal-borne acoustic sensors provide valuable insights into wildlife behavior and environments but face significant power and storage constraints that limit deployment duration. We present a novel adaptive acoustic monitoring system designed for long-term, real-time observation of wildlife. Our approach combines low-power hardware, configurable firmware, and an unsupervised machine learning algorithm that intelligently filters acoustic data to prioritize novel or rare sounds while reducing redundant storage. The system employs a variational autoencoder to project audio features into a low-dimensional space, followed by adaptive clustering to identify events of interest. Simulation results demonstrate the system's ability to normalize the collection of acoustic events across varying abundance levels, with rare events retained at rates of 80–85% while frequent sounds are reduced to 3–10% retention. Initial field deployments on caribou, African elephants, and bighorn sheep show promising application across diverse species and ecological contexts. Power consumption analysis indicates the need for additional optimization to achieve multi-month deployments. This technology enables the creation of novel wilderness datasets while addressing the limitations of traditional static acoustic monitoring approaches, offering new possibilities for wildlife research, ecosystem monitoring, and conservation efforts.

Keywords: bioacoustics; wildlife monitoring; animal-borne sensors; low-power computing; unsupervised learning; acoustic classification; conservation technologies; variational autoencoders

Academic Editor: Lei Shu

Received: 2 April 2025
Revised: 16 May 2025
Accepted: 6 June 2025
Published: 24 June 2025

Citation: Jean, D.; Turner, J.; Hedgecock, W.; Kalmár, G.; Wittemyer, G.; Lédeczi, Á. Animal-Borne Adaptive Acoustic Monitoring. *J. Sens. Actuator Netw.* **2025**, *14*, 66. https://doi.org/10.3390/jsan14040066

1. Introduction

The application of acoustic monitoring in ecological sciences has grown exponentially in the last two decades [1]. Acoustic monitoring provides spatially and temporally specific information that has been leveraged for many questions, including detecting the presence or absence of animal species in an environment, evaluating animal behavior, and quantifying ecological phenology, as well as identifying ecological stressors and illegal activities [2]. A typical setup involves the deployment of recording devices at static locations (e.g., acoustic arrays) with large amounts of storage and ample power to enable continuous recording of an acoustic environment. Data processing is typically performed offline after the deployment has concluded; however, this approach suffers severe limitations due to its lack of scalability and real-time monitoring capabilities. Current uses often are limited to the coverage of relatively small geographic areas with a fixed number of sensors. Animal-borne sensors provide an alternative, distributed path for acoustic monitoring [3]. Current

use of GPS-based animal-borne trackers provides an exemplar of how such a data collection model can enhance ecological and conservation sciences [4]. Animal-borne trackers are ubiquitous globally, being one of the primary technologies used in wildlife research and ecosystem monitoring. Such units are increasingly being paired with other sensors (e.g., accelerometers) [5]. The inherent power limitation of wearable sensors, however, does not allow for long-term acoustic monitoring, with battery weight being the primary factor dictating total power availability. Recent acoustic animal-borne sensor implementations with power capacities of ∼4000–6000 mAh, for example, have been shown to last less than a month [6,7]. Since capturing and tranquilizing wild animals is a traumatic event for them, as well as being expensive and resource-intensive, multi-month- to year-long deployments are required. The aim of our project, then, is to devise an animal-borne adaptive acoustic monitoring system to enable long-term, real-time observation of the environment and behavior of wildlife.

A significant amount of recent research has been carried out with two independent aims, one of utilizing Artificial Intelligence (AI) to increase the accuracy of Environmental Sound Classification (ESC) systems, and one to address the need for ultra-low-power processing capabilities. These two research veins, however, have yet to converge into a single concerted push for robust acoustic sensing that is able to operate continuously over extended periods of time. On the one hand, this research has resulted in a number of open-source acoustic datasets becoming widely available, including Harvard's ESC-50 database [8] and NYU's UrbanSound8K [9] dataset, but on the other hand, the majority of recent AI-based innovations rely on a fixed number of predefined events to be classified, a known acoustic environment, a pre-trained neural network that is unable to dynamically adapt to its environment, and plentiful power and on-board storage.

Domain Problem

The goal of our work is to create a dynamic sensing infrastructure that is not strictly constrained to the detection of predefined, well-known events, but rather is able to dynamically learn what constitutes an "event of interest" worthy of storage for later analysis by a domain expert. In essence, the sensor must be intelligent enough to not just collect a set of predefined or heuristically-based statistics, but rather to learn which events are important enough to store and which should be ignored. For example, a wildlife monitoring network in a rainforest may not be trained to classify the sound of an automobile, but the network should be able to identify that such a sound is unusual or unexpected for the environment and store it for later evaluation, along with data from additional sensing modalities, such as GPS and accelerometers.

The system needs to be modular and highly configurable. Different applications and deployments may have different requirements and constraints. For example, one can easily attach a one-pound battery to an elephant, but not to a bat. Some deployment areas might have cell coverage or LoRaWAN, but in many regions, Iridium satellites may be the only option. Deployment duration and the acoustic environment may also be quite different. For example, a rainforest is very noisy compared to a savanna. The goal of the deployment may vary from long-term monitoring for research purposes to active protection from poachers. A robust system should be sufficiently configurable to address any of these deployment options within a single monitoring ecosystem.

2. Related Work

Animal-borne acoustic recorders, also known as acoustic biologgers, are increasingly being used in ecological studies [3,10]. Although this technology is used primarily in bioacoustics, the applications of biologging extend much further, enabling insights into

a variety of ecological fields including feeding behavior, movement patterns, physiology, and environmental sounds. Biologgers are particularly useful for monitoring vocalizations, which can encode a wealth of information about an animal's behavior, affective state [11], identity [12], and health [13]. In addition, feeding behavior can easily be detected through this technology, as sensors are typically incorporated into collars placed near the throat. This positioning allows researchers to quantify and classify food intake [14], as well as to assess related behaviors such as rumination [10] and drinking patterns [15]. Additionally, biologgers can capture sounds associated with movement [16] and physiology [17], allowing this technology to address questions about locomotion and time budgets. In addition, sounds associated with physiological processes can be captured, such as respiration, heart rate, and urination. Finally, biologgers record environmental sounds, allowing researchers to investigate the impact of environmental factors like anthropogenic noise [10] and weather parameters [6].

Beyond their use in ecological studies, biologgers have been used for decades to monitor livestock health and behavior, offering valuable insights into feeding patterns and the welfare of domesticated animals. The application of acoustics to measure livestock ingestive behavior has been well documented in species such as dairy cattle [18], sheep [19], and goats [20], showcasing the potential of this technology to study animal feeding habits. Variations in acoustic signals related to feeding can also serve as indicators of an individual's health [21] and reproductive state [22], with commercial PAM systems offering real-time estrus detection based on acoustic changes in rumination. The success of biologging in livestock management underscores its broader application in wildlife studies, where it holds promise for advancing our understanding of animal behavior [3].

Among the most popular devices in this field are the AudioMoth [23] and its compact variant, the MicroMoth [24]. The AudioMoth is a low-cost, open-source acoustic monitoring device capable of recording uncompressed audio across a wide frequency range, from audible to ultrasonic, onto a microSD card. Its versatility has made it a preferred choice for applications such as monitoring ultrasonic bat calls and capturing audible wildlife vocalizations. The MicroMoth retains the high-quality recording capabilities of the AudioMoth while boasting a significantly smaller and lighter design, measuring just 26 $\times$ 36 mm and weighing 5 g, excluding batteries, making it ideal for deployments where space and weight are critical factors, although providing continuous power for long-term deployments remains a significant issue. Other types of acoustic biologgers include acoustic transmitters which emit unique signals detected by underwater receivers to track aquatic animals [25]; acoustic accelerometer transmitters that combine telemetry with accelerometry to provide insights into animal activity levels [26]; Passive Integrated Transponder (PIT) tags [27] used in freshwater studies for individual animal identification; and Wildlife Computers tags [28], widely utilized in marine research for their versatility across species. Each of these devices offers unique features tailored to specific research needs, contributing significantly to the advancement of wildlife monitoring and conservation efforts.

Despite the availability of portable sensing devices and the utility of biologging, the high power demands of continuously recording audio pose a significant challenge for this technology. As biologging devices are animal-borne, lightweight instrumentation is a necessity, particularly for smaller species. This requirement, coupled with the logistical need for deployments that last many months to a year, create opposing targets for resource utilization which are difficult to simultaneously achieve when writing continuous audio. Furthermore, post-processing often involves analyzing only a small subset of recorded acoustic data, depending on the acoustic events of interest. The integration of embedded classification addresses this issue by only storing relevant acoustic events, significantly reducing power consumption and storage requirements, and enabling longer

deployments; however, it raises new questions about the ability of an embedded classifier to determine what qualifies as an event of interest when existing datasets are limited in their representation of natural wilderness sounds.

The rise of research into ESC has indeed produced a notable number of high-quality datasets for a variety of research tasks, ranging from urban deployments, like NYU's UrbanSound [9] and the Domestic Environment Sound Event Detection (DESED) [29] dataset, to bird classification tasks, such as BirdSong [30] and BirdVox [31], to detection of underwater marine calls, such as Discovery of Sound in the Sea (DOSITS) [32] and IOOS's SanctSound [33]. Characteristic-specific acoustic datasets have also been compiled, such as Harvard's ESC-50 [8], Google's AudioSet [34], and IEEE's AASP CASA Office dataset [35], containing highly impulsive, transient-specific signatures, as well as datasets which contain ambient background noises for contextual scene classification, such as the various TAU/TUT Urban Acoustic Scenes datasets [36] and NYU's UrbanSAS [37]. Notably missing from this list, however, are datasets capturing organically recorded wilderness sounds of animals in their natural habitats. While there have been efforts to capture this type of data, notably from the National Park Service's Sound Gallery project [38], this data is lacking important annotation information to make it usable for training generalizable machine learning models, and it still represents relatively small, predefined types of wilderness areas. For example, a dataset captured from Yellowstone National Park will not include wilderness sounds encountered by an animal as it migrates south for the winter. As such, one of the primary challenges in the initial deployment of an animal-borne embedded classifier is coming up with ways to define "events of interest" outside of the context of a fully supervised classification paradigm, which will be addressed in the remainder of this paper. Successful implementation and deployment of the proposed system will allow for the creation of novel wilderness datasets to allow researchers to explore supervised classification methodologies for animal-borne devices that are unavailable at this time.

Acoustic Event Detection and Processing

In this paper, we focus on a variational autoencoder (VAE) approach utilizing Mel-Frequency Cepstral Coefficients (MFCCs) for determining which sounds to record and which to discard. Our approach will be described in more detail in Section 3; here, we briefly review existing approaches in this domain.

Supervised classification methods, such as convolutional neural networks (CNNs) or recurrent neural networks (RNNs), are frequently used for sound-identification tasks. Unlike the unsupervised VAE-based approach, supervised methods require labeled datasets for training. CNNs effectively capture spatial (spectro-temporal) patterns within audio spectrograms, and RNNs—especially Long Short-Term Memory (LSTM) networks [39]—are adept at modeling temporal dependencies in sequential audio data. While supervised models typically achieve high accuracy, their dependency on extensive labeled datasets limits their applicability in scenarios with limited or costly data-annotation efforts.

Semi-supervised approaches utilize a small labeled dataset in combination with a larger unlabeled dataset, aiming to leverage both labeled and unlabeled data effectively. Weakly supervised methods, where the exact timing of events is unknown, have also been explored. Such methods can partially alleviate the labeling bottleneck associated with supervised methods and offer greater flexibility compared to fully supervised methods. However, these approaches still depend partially on manual annotations, unlike the fully unsupervised VAE-MFCC method.

Traditional clustering methods, such as k-means or Gaussian Mixture Models (GMMs), have been applied extensively in audio analysis for unsupervised categorization. These algorithms assume static feature distributions and simple separability of classes. More

recent eco-acoustic studies, however, demonstrate that density- and graph-aware clustering techniques, including DBSCAN, its hierarchical extension HDBSCAN [40,41], spectral clustering [42], and Dirichlet-process GMMs [43], cope better with the irregular, manifold-like distributions that characterize real-world soundscapes. Such non-parametric or adaptive algorithms not only remove the need to pre-specify the number of clusters, but also provide built-in outlier labels that can feed directly into a VAE's reconstruction-based novelty score as in Table 1.

Table 1. Comparison of clustering strategies suitable for edge acoustic sensors.

Category	Representative Algorithms	Strengths	Key Limitations
k-means/GMM	Lloyd *k*-means, EM-GMM	Extremely light to compute; single-pass updates	Must pre-set *k*; assumes convex clusters; sensitive to initialization
Density-based	DBSCAN, HDBSCAN	Detects arbitrary-shaped clusters; auto-labels outliers	Choice of ε is critical; scales poorly with dimension
Graph-based	Spectral Clustering	Captures manifold structure; separates overlapping sources	Requires offline $O(N^3)$ eigensolver; memory heavy for long logs
Non-parametric	Dirichlet-Process GMM	Grows cluster count online; fully probabilistic	Inference more complex; still Gaussian assumption

In terms of feature extraction methodologies, although MFCCs are widely adopted in audio analysis for their efficiency and biological plausibility, a growing body of work explores alternative feature front-ends that can be swapped into the same VAE framework with minimal architectural change. Harmonic-preserving transforms such as the Constant-Q Transform (CQT) [44] and ERB/Bark-scaled filterbanks retain fine pitch structures that aid in the separation of overlapping bird calls. Gammatone-based cepstra (GFCCs) are more robust to wind and rain, while scattering transforms [45] capture multi-scale modulations with only a modest computational overhead. Learnable front-ends such as LEAF [46] or the small SincNet variants can be quantized to $<$50 kB, and pre-trained embeddings like OpenL3 [47], VGGish [48], and PANNs [49] compress thousands of hours of supervision into compact vectors that cluster well, even when projected to 32 dimensions. Table 2 outlines some other commonly used acoustic feature extraction front-ends:

Table 2. Feature extraction front-ends for embedded acoustic monitoring.

Front-End	Typical Realization	Strengths	Key Limitations
MFCC (baseline)	13-Coefficient Mel Cepstra	Compact ($\sim$26 dB/frame); ubiquitous tooling	Loses fine harmonic detail; less wind-robust
CQT/ERB	Log-scaled Filterbanks	Preserves pitch; aids bird-call separation	Higher FFT cost; variable hop complicates buffering
GFCC	Gammatone Filterbank Cepstra	Robust to wind and rain; cochleagram-inspired	Slightly heavier than MFCC; fewer open libraries
Scattering Transform	Fixed Wavelet Cascade	Translation-invariant; good with few samples	2–3 $\times$ MFCC compute; filters not learnable
Learnable Front-ends	LEAF, Small SincNet	Data-driven; quantizes to <50 kB	Requires training; risk of habitat over-fitting
Pre-trained Embeddings	OpenL3, VGGish, PANNs	Leverage large-scale supervision; clusters well after projection	>1 MB unless compressed; potential domain mismatch

Novelty detection paradigms. While a VAE's reconstruction error already serves as a robust unsupervised anomaly score, alternative paradigms may further enhance detection accuracy on the edge. Probability-density models (e.g., normalizing flows), distance-based k-NN hashing, and domain (boundary) methods such as Deep SVDD [50] provide complementary operating points in the speed–memory–precision space. Generative-adversarial approaches (AnoGAN) [51] and distilled audio transformers under one million parameters [52] have recently shown promise on embedded hardware, although their training complexity remains higher than that of a single-stage VAE. Additional novelty detection paradigms applicable to low-power acoustic sensing are outlined in Table 3, although the unsupervised VAE-MFCC approach presented in this paper occupies a sweet spot among the various alternatives: it eliminates the need for labeled data and maintains computational simplicity and interpretability, yet it remains compatible with state-of-the-art clustering, feature extraction, and novelty-detection algorithms that can be swapped in as future research, deployment needs, or firmware updates dictate.

Table 3. Novelty-detection paradigms applicable to low-power recorders.

Paradigm	Representative Algorithms	Strengths	Key Limitations
Reconstruction-based	VAE Reconstruction Error	One-shot training; compact latent space; easy thresholding	May miss subtle anomalies with low reconstruction error
Density Estimation	GMM, Normalizing Flows	Calibrated likelihoods; Bayesian extensions	Sensitive to model mis-specification; memory grows with components
Distance-based	*k*-NN, LSH Hashing	Training-free; adapts instantly	$O(N)$ query time; metric choice critical
Boundary-based	One-Class SVM, Deep SVDD	Fixed memory footprint; robust to outliers post-training	Requires representative "normal" data; retraining costly
Generative Adversarial	AnoGAN, GANomaly	Captures fine-grained structure; flexible objective function	Training instability; flash footprint >5 MB
Energy/Score	Distilled Tiny-Transformers	State-of-the-art accuracy; transferable across habitats	Attention cost unless aggressively pruned

In a parallel research vein, recent advances in self-supervised learning have introduced highly effective approaches for audio representation that significantly reduce dependency on labeled data. Techniques such as contrastive learning frameworks, exemplified by COLA [53] and BYOL-A [54], leverage temporal proximity and data augmentation to generate robust, discriminative audio representations. Masked-modeling approaches inspired by BERT, including the Audio Spectrogram Transformer (AST) [55] and wav2vec 2.0 [56], predict masked segments of spectrograms or raw waveforms, demonstrating state-of-the-art performance and excellent generalization capabilities, especially beneficial in data-scarce domains such as wildlife acoustics. Moreover, foundation models like AudioMAE [57], which utilize large-scale pre-training followed by targeted fine-tuning, further bridge supervised and unsupervised paradigms, offering robust generalization across varied acoustic environments. While these advanced self-supervised methods may outperform traditional VAEs in terms of representational richness and adaptability, the VAE-MFCC approach presented in this paper still offers substantial advantages in computational simplicity, interpretability, ease of deployment, and effectiveness in resource-constrained settings.

3. Materials and Methods

3.1. System Requirements

Based on the challenges of long-term animal-borne acoustic monitoring discussed above, we have identified several key requirements for our adaptive acoustic monitoring system. These requirements address the fundamental constraints of power and storage limitations while ensuring the capture of ecologically significant acoustic data. Importantly, there is a tradeoff between weight, energy consumption, and memory capacity that varies depending on the species of animal chosen to carry the sensing device. Smaller animals may impose strict weight limits, necessitating lighter, more energy-efficient components and exclusive use of on-board memory, which can limit recording duration and data resolution. Conversely, larger species can accommodate more robust hardware with higher energy capacities and greater storage, facilitating longer and more detailed monitoring at the expense of increased device weight. As such, the target species in a given deployment will necessitate a balancing of these factors to optimize data collection without adversely impacting animal behavior or well-being.

Since the goal of this system is to provide a unified architecture for animal-borne acoustic monitoring, our design goals explicitly target the most resource-constraining (i.e., small) animal species, with the expectation that the inclusion of additional batteries will be the primary resource relaxation for deployments targeting larger animals. As such, the system has been designed to selectively record acoustic events of interest rather than to continuously capture audio data, although continuous and schedule-based recording are also supported. It should identify and record rare or unusual acoustic events that may be of scientific importance, while implementing intelligent clustering of similar sounds to avoid redundant storage. Only representative examples of common sound types should be stored, along with metadata and statistics about their occurrence patterns. This approach supports configurable event detection based on domain-specific criteria, allowing researchers to focus on the acoustic phenomena most relevant to their studies.

The system must also be highly configurable to accommodate different research objectives and deployment scenarios. Domain scientists should be able to specify target species or acoustic events of particular interest, as well as sounds to be excluded from recording, such as wind noise and sounds associated with the rubbing of the collar to the skin of the animal. The configuration should include required sampling rates based on the acoustic characteristics of target sounds, expected deployment duration and corresponding resource allocations, and available power budget and storage capacity. This flexibility ensures that the system can be tailored to specific research questions across diverse ecological contexts.

To accommodate various research methodologies, the system should support multiple recording strategies. These include schedule-based recording for predictable, time-dependent monitoring; threshold-based triggering when sound levels exceed specified energy thresholds; adaptive triggering based on the detection of specific acoustic signatures; and hybrid approaches combining multiple strategies to optimize data collection. This versatility allows researchers to employ the most appropriate sampling method for their specific research questions. For example, if a researcher is primarily interested in sounds originating from the target animal wearing the device, then the noises of interest will typically be significantly louder than the background, suggesting that threshold-based sensing may be more appropriate than continuous, schedule-based, or signature-based sensing. In this case, only high-energy events would be presented to downstream decision-making algorithms, significantly reducing energy consumption and computational overhead. On the other hand, deployments targeting the capture of anthropogenic sounds would be more successful using schedule- or signature-based sensing.

Given the limited resources available in animal-borne systems, intelligent resource management is essential. The system must implement adaptive parameter adjustment based on remaining battery power and dynamic storage management that considers available SD card space. As resources diminish, progressive data compression or summarization may be employed, along with prioritization mechanisms for high-value acoustic events. This dynamic approach ensures that critical data collection continues throughout the deployment period, even as resources become increasingly constrained.

For contextual understanding of acoustic data, the system should integrate with complementary sensors. This includes synchronizing acoustic events with GPS location data, correlating audio recordings with accelerometer data to understand animal behavior, and supporting integration with other environmental sensors where applicable. This multi-sensor approach provides rich contextual information that enhances the interpretation of acoustic events.

These requirements form the foundation for our design approach, balancing the technical constraints of animal-borne sensing with the scientific objectives of long-term acoustic monitoring in wildlife research and conservation.

3.2. Sensor Node Hardware

Our proposed adaptive acoustic monitoring solution requires a lightweight yet powerful acoustic biologger, which is not currently available on the market. In response to the growing need for such a solution capable of supporting advanced Artificial Intelligence (AI) applications in biologging, we have developed a novel sensor board designed for animal-borne deployments. This board, with a small footprint of 18 × 23 mm on an 8-layer PCB, has been carefully designed to keep costs low—approximately $60 per board (excluding SD card and battery)—while integrating advanced features such as a high-performance microcontroller, a wide range of sensors, support for multiple types of microphones, and the ability to control an external actuator (e.g., VHF transmitter). The board operates from a 3.6 V battery and uses a multi-stage voltage regulation scheme to ensure efficient power utilization—the input voltage is converted to a 1.8 V core voltage with high efficiency and low noise characteristics. Additionally, a typical high-capacity C-cell battery weighs around 18 g. At only 2.4 g (∼13% of the battery weight), the board is light enough to be used even on small animals, with battery weight being the primary limiting factor.

A key advantage of our design is its processing capabilities and efficiency, powered by an Ambiq Apollo 4 Plus MCU [58]. This microcontroller significantly outperforms the MCU found in devices such as AudioMoth. For example, while AudioMoth is based on a low-power ARM Cortex-M0+ that typically runs at 48 MHz (62 µA/MHz) with about 256 KB of Flash memory and 32 KB of SRAM, the Apollo 4 Plus runs at a configurable rate between to 96 and 192 MHz (4 µA/MHz) with about 2 MB of MRAM and 2.75 MB of SRAM. In practical terms, this means that our board has approximately 4× the clock speed, 8× the non-volatile memory, and 85× the RAM while running at a lower power compared to AudioMoth. These improvements are critical for on-board AI processing, which requires significant compute power and memory to run deep learning models directly on the device.

In addition, the Apollo 4 Plus MCU features advanced digital signal processing (DSP) instructions and is optimized for ultra-low-power operation, even during active processing, making it exceptionally well suited for real-time audio analysis and other complex sensing tasks. A key feature is the integrated low-power audio-specific ADC, which includes an internal programmable gain array (PGA) capable of amplifying incoming audio signals over a range of 0–24 dB. This high-performance MCU is complemented by the overall design of the sensor board, which utilizes miniaturized components (such as BGA packages) and operates at a low voltage of 1.8 V to ensure extended battery life in field applications.

The board offers versatile sensor integration, supporting analog electret, analog MEMS, and digital PDM microphones, as well as on-board sensors such as the ADXL345 IMU and a magnetic sensor. An additional header enables seamless connectivity to external GPS tracking collars, extending the range of applicability to applications in wildlife monitoring and other field research scenarios.

The board and components have been field-tested across a variety of environmental conditions, ranging from high ambient temperatures ($\sim$100 °F) with high humidity during summer in the deep south of the United States to extremely low temperatures (~ -30 °F) and humidity during winter in Alaska. Aside from variations in battery capacity at these extreme temperatures, full board functionality (including SD card storage) was verified across this range of field conditions as in Figure 1.

Figure 1. Sensor node hardware. Top side (left board): Apollo 4 Plus MCU in the center, high-efficiency regulators along the bottom, external connectors (power, VHF, programmer, and microphone interfaces) on the left, plus an integration header at the top. Bottom side (right board): MicroSD card slot and low-power components in a low-noise area.

3.3. Sensing Firmware

In order to support the wide range of potential research goals that an animal-borne adaptive monitoring solution might be used for, we have developed highly configurable firmware to complement the low-power peripherals and sensing capabilities of the hardware described above. Notably, we strove to ensure that device configuration is intuitive and able to be carried out in the field by researchers of varying backgrounds and technological prowess. To that end, device configuration was implemented as a three-tiered hierarchical process, including (1) complete support for all configuration possibilities within the base firmware residing in flash memory on the hardware, (2) an editable runtime configuration file located on an SD card used for storing data in a given deployment, and (3) a graphical configuration dashboard for researchers to provision the devices in their deployment without needing to manually generate or manipulate text-based configuration files. The firmware was developed with the express intent of being able to support a wide range of configuration options using minimal power expenditure. Configuration options currently supported include:

- **Device Label:** Unique label for tagging all stored data to alleviate post-deployment processing and storage.
- **GPS Availability:** Toggleable setting to specify whether precise timing information is available via GPS or whether the on-board real-time clock (RTC) should be used.
- **Status LED Active Time:** Number of seconds for which status LEDs should be active on a deployed device. This is useful for ensuring that deployed devices are functioning properly without disturbing or affecting the animals during deployment.

- **Microphone Type and Gain:** Whether a digital or analog microphone is connected, as well as the target microphone gain in dB. This allows researchers to use the same hardware for deployments involving both very loud and very soft sounds of interest.
- **Magnetic Activation Settings:** Whether the device should be activated upon detection of a strong magnetic field, and for how long that field must be present to cause activation. If GPS is available, it may also be possible to schedule a deployment to start at a specific time instead of relying on manual magnetic activation.
- **Deployment Schedule:** Anticipated starting and ending date and time of the deployment, along with the deployed timezone, allowing researchers to schedule deployments more intuitively without considering UTC offsets.
- **RTC Time Alignment:** Whether the on-board RTC clock should be initialized to the scheduled deployment start time upon activation. If the device has GPS available, this should not typically be necessary; however, without GPS, this allows researchers to precisely align stored data timestamps with the actual start time of an experiment.
- **VHF Beacon Settings:** Date and time to optionally enable a VHF retrieval beacon upon termination of a deployment.
- **Audio Clip Settings:** Sampling rate and desired clip duration for each stored audio clip. There is also a toggleable setting allowing clips to exceed the target duration if continuing audio is detected, for example, during a prolonged animal vocalization.
- **Audio Recording Settings:** Whether audio clips should be generated continuously, intelligently, or according to a schedule, interval, or loudness threshold:

  Continuous: Records clips of the target duration continuously with no gaps.
  Schedule-Based: Records clips of the target duration continuously during an arbitrary schedule of explicit listening times.
  Interval-Based: Records a new audio clip every configurable **X** time units.
  Threshold-Based: Records a new clip of the target duration when a sound above a configurable loudness threshold (in dB) is detected.
  Intelligent: Records a new clip when the AI-based adaptive data collection algorithm indicates reception of an event of interest.
- **IMU Recording Settings:** How IMU data should be recorded, either (1) not at all, (2) initiated by and aligned with each new audio clip, or (3) initiated upon significant motion detection with a configurable acceleration threshold, as well as the sampling rate at which the IMU sensor should be polled.
- **Deployment Phases:** Whether all previous settings are valid throughout the entire deployment, or whether the deployment should be split into distinct phases, each with its own set of audio and IMU configuration settings. The ability to use phases to reconfigure almost all parameters of a deployment after it has already begun allows researchers to design experiments with more than a single target outcome using a single animal.

The firmware was implemented as baremetal C code to support the above configuration options, targeting the lowest possible power modes of the Ambiq Apollo 4 Plus microcontroller and peripherals. This includes utilizing Direct Memory Access (DMA) transfers of as much data as possible, shutting down memory banks and power domains when not in use, relying primarily on interrupts for waking the microcontroller, and ensuring that the microcontroller is operating in its deepest sleep mode for as much time as possible, waking only to service interrupts and DMA data-ready notifications.

All runtime configuration details are stored on each deployed device's SD card in a file formatted with text-based key-value pairs. To mitigate configuration errors and alleviate formatting details, an OS-agnostic, Python-based graphical user interface (GUI) has been developed to configure each device (See Figure 2). All settings previously described are

supported within the GUI, with special care taken to ensure that incompatible settings are unable to be selected, and out-of-range values cannot be specified. This GUI was successfully utilized to configure all deployments presented in the upcoming results section.

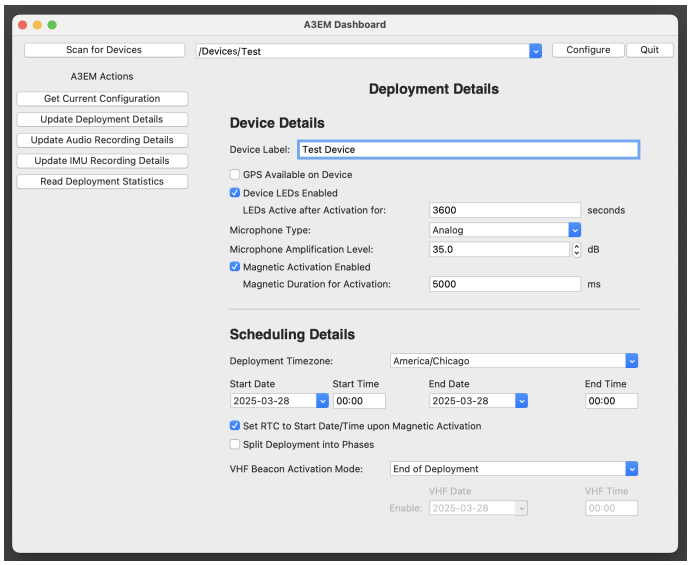

Figure 2. Deployment configuration dashboard.

3.4. Adaptive Data Collection

Existing approaches to audio data collection from animal-borne recording devices typically involve constant recording during the deployment, followed by retrieval of the device and offline processing using large and computationally expensive audio segmentation and classification models. While this approach is effective at capturing any audio events of interest, the requirement of continuous recording rapidly depletes limited power and storage resources on the device, ultimately shortening deployment time and causing valuable audio data to be lost once resources are exhausted. Instead, we propose an alternative adaptive filtering approach that leverages lightweight, low-power feature extraction and online clustering techniques to perform in situ relevance assessment of acoustic events. This enables selective data retention, significantly reducing the overhead associated with storing redundant or uninformative segments. This results in a more efficient system that not only extends deployment durations but also selectively preserves only the most valuable audio data. The underlying challenge is to find a balance between maximizing the retained informational content and prolonging operational lifetime, which is constrained by the device's limited energy and storage resources. However, with appropriately designed filtering mechanisms, it is possible to maximize information content, often quantified using entropy-based metrics in the literature, while simultaneously minimizing both power consumption and storage usage. This tradeoff is formalized in the following subsection. After presenting the abstract formulation of the optimization problem, we describe the proposed practical solution.

3.5. Theoretical Framework: Information–Resource Tradeoff

To formalize the design objective of our adaptive filtering system, we model the tradeoff between information retention, power consumption, and storage utilization on resource-constrained embedded platforms. Our goal is to maximize the total informative value of retained audio segments while minimizing the cost of power and storage consumption.

Let

- $I(f)$ denote the **total information content** retained under filtering policy f,
- $P(f)$ denote the **total power consumed** by the device under policy f,

- $S(f)$ denote the **total storage utilized** under policy f,
- $\lambda_P, \lambda_S \in \mathbb{R}^+$ be weight coefficients representing the relative cost of power and storage, respectively.

We define the following optimization objective:

$$\max_f \ \mathcal{J}(f) = I(f) - \lambda_P \cdot P(f) - \lambda_S \cdot S(f) \tag{1}$$

This objective balances the benefit of retaining high-value audio information with the energy and storage costs incurred during acquisition, decision making, and storage. Each audio segment x_t, observed at time t, is either retained or discarded:

$$r_t = f(x_t) \in \{0, 1\} \tag{2}$$

For simplicity, consider $t \in \{1, \ldots, T\}$ values, i.e., each clip is split into 1 s long segments. Then, let

- $i_t = I(x_t)$ be the estimated information content of segment x_t,
- $p_t = P_{\text{proc}} + r_t \cdot P_{\text{store}}$ be the power consumption for processing and (optionally) storing the segment,
- $s_t = r_t \cdot s$ be the associated storage cost, with s as the size of the audio clip in bytes.

Then, the components of the objective are defined as

$$I(f) = \sum_{t=1}^{T} r_t \cdot i_t \tag{3}$$

$$P(f) = T \cdot P_{\text{proc}} + P_{\text{store}} \cdot \sum_{t=1}^{T} r_t \tag{4}$$

$$S(f) = s \cdot \sum_{t=1}^{T} r_t \tag{5}$$

Substituting into the objective function

$$\mathcal{J}(f) = \sum_{t=1}^{T} [r_t \cdot (i_t - \lambda_P P_{\text{store}} - \lambda_S s)] - \lambda_P T \cdot P_{\text{proc}} \tag{6}$$

This expression reveals a natural decision rule: the filter f should retain segment x_t only if its estimated information value i_t exceeds the combined power and storage penalty:

$$i_t > \lambda_P P_{\text{store}} + \lambda_S s \tag{7}$$

It can also be observed that a constant term ($-\lambda_P T \cdot P_{\text{proc}}$) is always present in the objective function. This term represents the fixed cost of decision making incurred by the filtering algorithm itself. While the other penalty terms are determined by the hardware and deployment constraints, this term is influenced directly by the algorithmic design. This is where the adaptive filtering methodology exerts a significant influence on the overall optimization. A lightweight, low-power algorithm may reduce processing costs but typically suffers from limited precision, leading to suboptimal filtering decisions and increased downstream resource usage. Conversely, a more robust and accurate model may effectively reduce redundancy and preserve high-value information, but at the expense of significantly higher power consumption for decision-making. Thus, a careful tradeoff must be made between the complexity of the filtering model and its energy efficiency, as both

factors ultimately affect the system's operational lifetime and data utility. In Section 4, we evaluate and compare methods with different complexities.

The presented formulation offers an interpretable and tunable mechanism for controlling the resource efficiency of the system, allowing developers to adjust thresholds according to platform constraints and deployment objectives. The framework can be extended to include hard constraints that explicitly model the finite energy budget and storage capacity of the device. In real-world applications, the penalty parameters λ_P and λ_S can be estimated through various strategies: normalization-based methods (scaling all quantities to a common reference and using their ratios), constraint-based estimation (evaluating the marginal information loss under tighter power or storage limits), or empirical tuning based on observed system behavior.

In our implementation, audio events are stored on removable SD cards. Given that modern SD cards offer capacities up to 2 TB, well beyond the storage needs of typical bioacoustic logging tasks, the storage penalty is considered negligible. However, in scenarios where storage is constrained (e.g., on-board flash memory or wireless transmission), the storage cost λ_S may become a dominant factor in the optimization.

In practice, directly measuring the true information content $I(x_t)$ of an audio segment is challenging, particularly in unsupervised or unlabeled settings. As a practical proxy, we approximate information content by evaluating the diversity of retained acoustic events across distinct classes. Specifically, we design experiments where the input data is composed of multiple known sound classes, with one class significantly overrepresented in terms of sample count. Without filtering, such a distribution would bias retention toward the dominant class, reducing the effective information diversity. By applying our filtering approach, we aim to attenuate redundancy and retain a more balanced set of representative events. The effectiveness of this method is then assessed by analyzing the class distribution of the retained segments: an approximately uniform distribution across classes is indicative of successful suppression of redundant data and a better approximation of true information content. This evaluation strategy allows us to assess the filtering performance in a controlled yet realistic scenario.

3.6. Proposed Method

The goal of the proposed method is to optimize the tradeoff introduced in the theoretical formalism, maximizing information retention while minimizing power consumption and storage usage. In particular, we focus on reducing the processing power component P_{proc} by designing a lightweight yet effective filtering mechanism that balances computational cost with decision accuracy. To this end, we employ a fully unsupervised learning framework, which enables the system to operate without the need for labeled training data. This design choice allows the filtering agent to generalize across diverse and unpredictable acoustic environments, making it well suited for real-world deployments involving wildlife monitoring or urban soundscapes.

The machine learning agent used in our system, discussed in more detail in Section 4, is trained on a heterogeneous urban acoustic dataset compiled from public repositories, online audio hosting platforms, and recordings from prior field deployments. At a high level, the proposed method processes the incoming audio by first downsampling to 8 kHz and segmenting the stream into 1 s segments. From each segment, Mel-Frequency Cepstral Coefficients (MFCCs) [59] are extracted and fed into an encoder, which projects the input into a 16-dimensional latent space. These embeddings are subsequently evaluated by an online adaptive clustering algorithm, which identifies novel or information-rich segments in real time. Figure 3 illustrates the high-level architecture of the system.

The following two subsections describe the dimensionality reduction and online clustering components in greater detail.

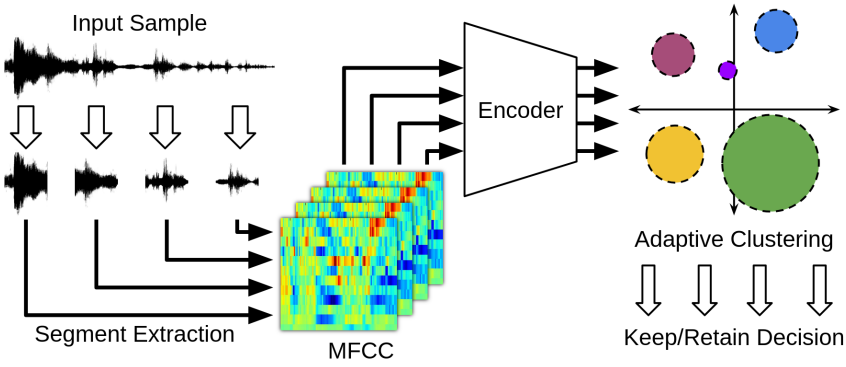

Figure 3. High-level architecture of the proposed filtering system: audio is downsampled, segmented, encoded into a latent space, and evaluated by an online clustering algorithm for real-time novelty detection.

3.6.1. Dimensionality Reduction

Effective dimensionality reduction is essential in this application to minimize computational overhead while preserving the salient structure of acoustic events for reliable novelty detection. Reducing the input data to a compact representation allows both efficient on-device inference and scalable clustering within constrained memory and energy budgets.

In the first step, the input audio signal obtained from the microcontroller's analog-to-digital converter is downsampled to 8 kHz to reduce the size of the spectrograms generated in subsequent stages, and the resulting signal is normalized. An iterative energy-based selection process is then applied to identify and extract high-energy segments, defined as peaks in the convolution with a triangular filter. At each iteration, a 1 s clip centered around the energy peak is extracted, and this region is zeroed out for subsequent energy calculations (though not excluded from future clip extraction). This energy-centric selection has two primary benefits: (i) it aligns potential events of interest temporally within each clip, thereby reducing variability in the encoded feature vectors due to time shifts; and (ii) it enables a sparser sampling of representative segments, thereby reducing both the inference time of the encoder and the memory footprint of the subsequent clustering stage.

Mel-Frequency Cepstral Coefficients (MFCCs) [59] are then extracted from each 1 s segment. These MFCC vectors are clamped and normalized to suppress the influence of high-energy outliers on the encoding process. The normalized MFCCs are passed through a compact encoder that maps them onto a 16-dimensional latent space. Specifically, we employ a variational autoencoder (VAE) [60], which enables unsupervised learning of low-dimensional representations without requiring labeled training data. The variational component improves clustering by enforcing local continuity in the latent space; Gaussian noise is injected during training to encourage embeddings of acoustically similar inputs to form tighter, more cohesive clusters.

Given the severe memory and computational constraints of the embedded platform, the encoder follows a convolutional architecture composed of five 3×3 convolutional layers, each followed by three 1×1 convolutional layers to implement pointwise non-linear mappings. This is followed by three fully connected layers that reduce a 32-dimensional flattened intermediate representation to the final 16-dimensional output, comprising both the mean and (log-scale) standard deviation vectors of the latent distribution. All non-terminal layers employ the Leaky ReLU activation function for computational efficiency. The trained VAE model initially occupies 28 KB and is subsequently quantized using 8-bit signed integer arithmetic to reduce memory usage and enable real-time deployment on embedded hardware. This quantization results in a final model size of 24 KB, making it

feasible for execution on low-power microcontroller platforms with limited computational and memory resources.

In addition to the VAE-based encoder, we implemented two classical dimensionality reduction/feature extraction baselines for comparative evaluation. These include simple time-domain methods such as extracting root mean square (RMS) energy and zero-crossing (ZC) rate, as well as a frequency-domain extractor based on spectral flux. While the VAE models are computationally more demanding, they are expected to offer more robust and consistent filtering performance across diverse acoustic environments. Conversely, the classical feature-based methods represent the opposite end of the complexity–accuracy spectrum, providing lightweight alternatives that trade off precision for minimal resource consumption. These complementary approaches allow us to explore the efficiency frontier defined by the optimization of P_{proc} under the constraints discussed in the theoretical formalism.

3.6.2. Novelty Detection: Online Clustering

The low-dimensional latent space produced by the encoder serves as the input domain for our proposed online clustering algorithm. This step enables real-time, unsupervised novelty detection by tracking representative acoustic patterns in a compact, memory-efficient manner. Operating in the reduced 16-dimensional feature space of the VAE embeddings not only minimizes computational burden but also facilitates adaptive filtering in resource-constrained environments.

The clustering filter maintains a fixed-size array of N D-dimensional cluster centers, denoted as $c_i \in \mathbb{R}^D$, each associated with a non-negative weight $w_i \in \mathbb{R}_0^+$. At initialization, all centers are set to zero, i.e., $c_i = \vec{0}$ and $w_i = 0$, representing an untrained filter state that accepts any input as novel. For each new feature vector $p \in \mathbb{R}^D$ arriving from the encoder, the algorithm identifies the set of "nearby" cluster centers as

$$A = \{\, i : \|c_i - p\| < t\sqrt{w_i} \,\}$$

where $t \in \mathbb{R}^+$ is a configurable threshold and $\|\cdot\|$ denotes the Euclidean (i.e., ℓ^2) norm.

If $A = \varnothing$, the feature vector p is treated as novel and inserted into the filter as a new cluster center with an initial weight of 1, replacing the oldest cluster center to maintain a fixed array size. If $A \neq \varnothing$, the feature vector is considered redundant and merged with the identified cluster centers. The updated cluster center is computed as a weighted average

$$\bar{c} = \frac{p + \sum_{i \in A} c_i w_i}{1 + \sum_{i \in A} w_i},$$

with an updated weight

$$\overline{w} = \min\left(W, 1 + \sum_{i \in A} w_i\right),$$

where $W \in \mathbb{R}^+$ denotes the maximum allowable weight to prevent indefinite growth.

The merged cluster center $\bar{c}$ replaces the previous clusters in A, and any resulting vacant entries in the cluster array are reinitialized to zero. Each latent vector (thus a short segment of a longer clip) is labeled either "retain" (if novel) or "discard" (if redundant). A final retention decision is made at the audio clip level: the clip is stored if the proportion of "retain" responses among its constituent 1 s segments exceeds a user-defined voting threshold $v \in [0, 1]$.

This clustering approach allows the system to dynamically adapt to evolving acoustic environments, efficiently preserving only informative and novel audio segments within a bounded memory and power budget.

4. Results

4.1. Simulation

To test the performance of the adaptive filtering approach introduced in Section 3.6, we have created an audio simulation environment which predicts the behavior of this system over long-term deployments. This simulation is constructed by iteratively presenting the adaptive filter with 5 s audio clips which have been manually labeled with various audio event classes of potential interest. As a fully unsupervised system, these audio class labels are exclusively used to measure the filtering response when presented with non-uniform input distributions. The datasets used in this simulation include the urban acoustic dataset mentioned in Section 3.6, as well as clips from the EDANSA ecoacoustic dataset [61].

Even excluding the hyper-parameters associated with training the VAE-based machine learning model, the adaptive filtering algorithm introduced in Section 3.6 has several critical hyper-parameters that require tuning: the number of cluster centers N, the maximum cluster weight W, the baseline proximity threshold t, and the voting threshold v. There are also a few other parameters elsewhere in the inference pipeline; for instance, the size of the triangular filter used for the initial energy-based segmentation procedure, as well as the number of 1 s segments to extract per input clip. Because of this, our first step was to perform a parallel grid search to determine reasonable values for each parameter.

For each set of candidate parameters, a 24 h simulation was performed containing three audio event classes: frog, aircraft, and rooster sounds. To measure the adaptive clustering algorithm's ability to filter the input into "interesting" (i.e., infrequent) sounds, the frequency of rooster sounds was set to 8 times the frequency of each of the other two audio classes. After performing this grid search, a heuristic was used to identify the "optimal" set of parameters for the goal of regularizing the imbalance of audio event occurrences.

To test the generalizability of these settings, we then used these optimal parameters to perform similar tests using an $8\times$ abundance of each of six audio event classes (the three classes from the grid search tuning process and another three unseen classes) over 7-day simulated deployments. For comparison, two alternative, heuristical methods were used: one based on spectral flux, and another based on RMS energy and the zero-crossing rate. The spectral flux approach generates a spectrogram for the entire 5 s clip and integrates out time to arrive at a final 120-dimensional time-invariant frequency spectrum. This spectrum is then used directly as the feature vector for clustering, as the calculation of the ℓ^2 norm naturally produces spectral flux values between two embeddings. For the other approach, we simply use the mean RMS energy of the 5 s clip in conjunction with the mean zero-crossing rate, resulting in a two-dimensional embedding; naturally, this approach is significantly faster to compute than either the VAE-based or spectral flux embeddings. The results of these tests using our proposed VAE-based system and these two heuristical approaches are presented in Figure 4 and compared numerically in Table 4 and visually in Figure 5. Notably, all included methods make use of the adaptive clustering filter algorithm introduced in Section 3.6.2, with the only difference being the method of producing feature vectors for the clustering procedure.

Table 4. Numeric filtering results measured by the KL-divergence from an ideal uniform distribution.

Test	Unfiltered	VAE	Q-VAE	Spectral Flux	RMS + ZC
(a)	0.7367	0.0446	0.0166	0.7106	0.3289
(b)	0.7301	0.0491	0.0930	0.0675	0.0689
(c)	0.7179	0.2099	0.1628	0.6165	0.3974
(d)	0.7306	0.1352	0.1080	0.1965	0.1014
(e)	0.7184	0.0456	0.1432	0.0988	0.2004
(f)	0.7263	0.0573	0.0866	0.1124	0.2041

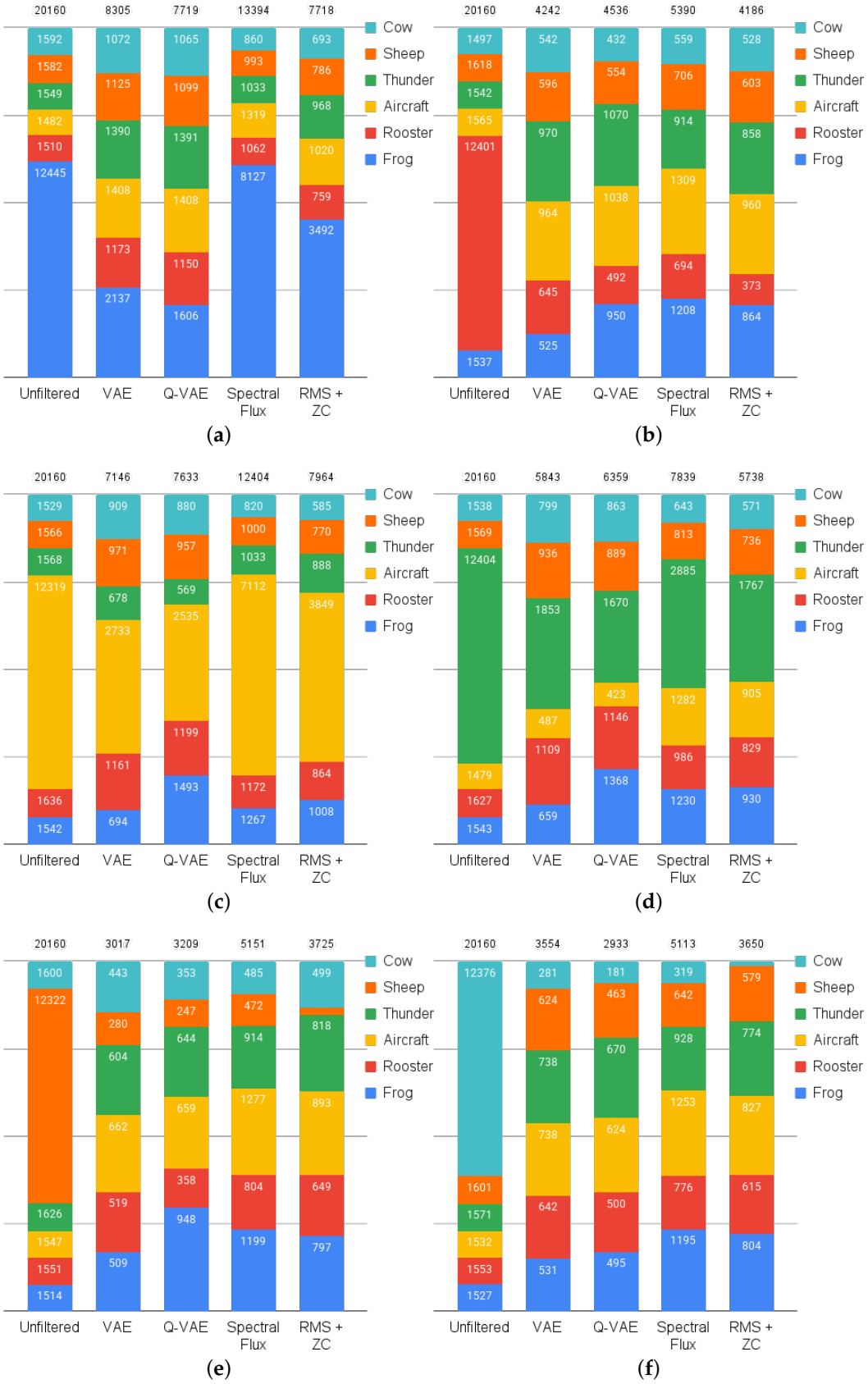

Figure 4. Filtering performance on simulated 7-day continuous deployments. Each plot has a different 8× abundance of the 6 included audio classes, and presents the output distribution for each of four algorithms: our proposed VAE-based system, its 8-bit quantized form (Q-VAE), a spectral flux filter, and an RMS energy and zero-crossing filter.

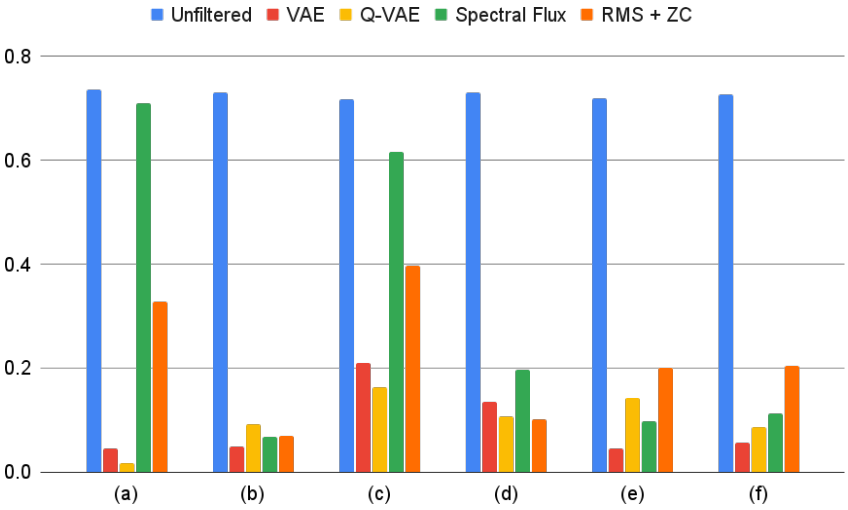

Figure 5. Numeric filtering results measured by the KL-divergence from an ideal uniform distribution.

From these results, we see that in every case, the VAE-based systems, both unquantized (VAE) and quantized (Q-VAE), were able to significantly reduce the imbalance of input audio event classes. Indeed, in many cases (e.g., Figures 4 and 5a,b,e,f)) we even observe a nearly equal balance of all six audio classes in the filtered output. This good filtering performance for the quantized model confirms the viability and utility of our proposed system for use in on-edge adaptive novelty-based audio filtering in long-running deployments. Looking further, we also see relatively good performance from the two heuristical methods (i.e., Spectral Flux and RMS + ZC); however, these approaches did not fare well across the board, and saw significant filtering degradation for specific audio classes (e.g., Figures 4 and 5a,c). However, the good performance of these heuristical methods in the majority of tests indicates that these comparatively cheaper algorithms could be used as drop-in replacements for the more robust VAE-based approach, improving overall power consumption at the expense of decreased filtering performance in some pathological audio classes.

4.2. Power Consumption

As one of the primary goals of this research is to enable months- to year-long animal-borne deployments, much effort was put into ensuring that every aspect of the system consumes as little power as possible. Sections 3.2–3.4 describe the design choices made in the development of each of the hardware, firmware, and data collection algorithms with this goal in mind. To validate our design, we ran a series of power consumption experiments to assess the impact of various configuration choices and runtime modes on the longevity of potential deployments, with results summarized in Table 5:

Table 5. Power consumption results using various runtime configurations.

Runtime Configuration	Avg. Power @ 3.6 V
Fully Idle with No Mic	0.14 mA
Audio Sampling @ 48 kHz with Analog Mic	0.76 mA
Audio Sampling @ 48 kHz + SD Card Storage	3.40 mA
Adaptive Data Collection + Audio + SD Card	3.81 mA (max, no duty-cycling)

From these results, it is apparent that the largest consumer of power is actually writing to the SD card, followed by continuous use of the adaptive data collection algorithm, although each second of audio requires only ∼180 ms to complete one invocation of this algorithm, causing its overall impact on the power budget to be quite small. Continuous

storage of audio data without any intelligent decision making consumes ∼3.4 mA at 3.6 V. Assuming an ideal device battery capacity of 8500 mAh, this would allow our devices to store continuous data for approximately 100 days on a single charge. However, the goal of this system is not to naively store all data, but rather to intelligently decide which data is relevant, interesting, and important to store, given the configured deployment parameters. As such, a future research task will be to develop an extremely low-power filtering algorithm based on known acoustic properties like spectral flux to preprocess incoming audio and decide whether a clip contains potentially interesting information that should be further processed by the adaptive data collection algorithm. This will allow the microcontroller to remain in deep-sleep mode for significantly longer periods of time. Consequently, the worst-case average power consumption should remain closer to the 1–3 mA range, with the adaptive data collection algorithm only being employed when there is relevant work to conduct and the SD card only being accessed upon positive detection of an event of interest. Of course, deployment-specific needs and configuration choices will greatly impact the level of duty-cycling achievable.

4.3. Deployments

Prototypes have been deployed on caribou (Rangifer tarandus), African elephants (Loxodonta africana), and bighorn sheep (Ovis canadensis) as part of various ecological projects (See Figure 6). Note that these initial deployments used a simple schedule-based recording approach and not the VAE-based adaptive processing. Additionally, these deployments were initiated before substantial power-saving improvements were made to the firmware, netting a decrease of roughly 2.5× in required power from the levels in these deployments. As such, deployment longevity for continuously recording audio can be expected to be roughly two and a half times longer using the current state of the research. The primary purpose of these experiments was to test the operation and durability of the hardware and gather acoustic data to enable offline testing of the algorithms.

Figure 6. Collaring a bighorn sheep (**left**) and a deployed unit on an elephant in Kenya (**right**).

Audiologgers were used on caribou as part of a project investigating the impact of insect harassment. Caribou distribution is known to be influenced by insect abundance [62], which is often modeled using environmental proxies. However, insect presence can also be detected acoustically, enabling audiologgers to provide estimates of insect distributions

on a finer scale. Two audiologgers were deployed on captive individuals at the Large Animal Research Station (LARS) in Fairbanks, Alaska. Both systems were housed in a custom 3D-printed enclosure and potted with a re-enterable potting compound for data retrieval. They were powered by a 4200 mAh battery pack consisting of two Tadiran TL-2100/S cells, and the audio was sampled continuously. One system lasted 33 days and gathered 44.68 GB of audio, while the other lasted 23 days and gathered 59.98 GB of audio. Frigid temperatures during the deployment (as low as -30 °F) significantly reduced power efficiency and therefore system longevity.

In Samburu National Reserve, Kenya, a prototype was deployed on a female African elephant. While audiologgers used with captive elephants have significantly advanced our understanding of elephant communication [63,64], this marks the first application of this technology on wild individuals. By combining GPS data with vocalization detections, we aim to examine spatiotemporal variations in elephant vocalizations, advancing our understanding of behavior and supporting conservation efforts. This system was housed in a modified aluminum housing and integrated onto a GPS collar with a custom drop-off mechanism so the data could be recovered prior to retrieval of the entire collar. It was powered with a 6300 mAh battery pack consisting of three Tadiran TL-2100/S cells. Unfortunately, the detachment mechanism on this device failed, and the sensor is still on the animal at the time of this writing, so we do not have any information on system longevity or data quality. Five more systems are scheduled to be deployed on additional African elephants in the coming months.

Finally, a prototype was also deployed on a female bighorn sheep in Fort Collins, Colorado, at the Colorado Parks and Wildlife (CPW) foothills facility. Pneumonia is a significant disease for bighorn sheep in the mountain west [65], and symptoms such as coughing can be detected acoustically. Detection of respiratory symptoms has been successfully used to diagnose illnesses in other species [66,67], indicating that audiologgers could play a crucial role in monitoring disease outbreaks in bighorn sheep herds. Similar to the caribou deployments, this logger was contained in a 3D-printed housing and powered with a 4200 mAh battery pack consisting of two Tadiran TL-2100/S cells. This deployment lasted 38 days and gathered 50.81 GB of audio. Another audiologger is currently deployed on an additional ewe with a chronic cough in order to determine the efficacy of this approach.

5. Discussion

The simulation results presented in Figures 4 and 5 demonstrate the potential of our adaptive acoustic monitoring approach for animal-borne applications. The adaptive filtering algorithm successfully reduced input imbalances across different audio event classes, with particularly strong performance in distinguishing between common and rare acoustic events. This is crucial for wildlife monitoring applications where researchers are often interested in detecting infrequent but ecologically significant sounds amidst common background noise.

Our findings highlight the system's ability to maintain effectiveness even in long-running, continuous operation. This suggests that the adaptive clustering algorithm is successfully learning and adapting to the acoustic environment over time, rather than simply reaching saturation points for specific sound classes. Notably, the quantized version of the VAE model (Q-VAE) yielded similarly good filtering performance compared to the unquantized model. This finding is particularly encouraging, as it demonstrates that this adaptive filtering approach can be effectively implemented on resource-constrained embedded devices without significant performance degradation, further validating our approach for real-world deployment.

In terms of power consumption and device longevity, all consumption values listed in Table 5 were measured under the burden of continuous environmental sensing. A much more likely scenario includes deployments that are able to use either threshold-based or schedule-based sensing to greatly extend the battery life of a device. For example, ambient environmental noise levels tend to be much lower at night than during the day. A deployment could be configured such that the devices remain in deep-sleep mode during nighttime hours and are only awoken upon detection of an acoustic event with significant amplitude above a configurable threshold. Our sensor hardware is able to wake from deep-sleep upon a hardware trigger from an analog comparator connected directly to the microphone. As such, we could continue sensing overnight at the more advantageous power consumption level of $\sim$0.76 mA, more than doubling the battery life of the deployment. There are any number of other schedule-based policies that could be employed in this manner, depending on the needs and requirements of the target deployment.

Another important takeaway from Table 5 is the relatively low power consumption required by continuous use of the adaptive clustering algorithm, especially when compared to the power requirements of writing to the SD card. The mere act of powering on the SD card and writing one second of audio consumes almost 2.64 mA. This is due to the extremely high current draw required for SD card communications, which unfortunately cannot be lowered other than by writing less frequently. In contrast, classifying one second of audio only takes about 180 ms and requires roughly 0.41 mA of average current per second. In fact, the active-mode detection–classification–recording loop is temporally dominated by this classification task. Detection, when used as a prefiltering step for the classifier, is hardware-based using the analog comparator peripheral of the Apollo 4 microcontroller, and as such, is instantaneous. SD card storage speed depends on the size of the audio chunk being stored, which itself is determined by the sampling rate chosen by the researcher; however, under average conditions of storing 16 kHz single-channel audio samples once per second, writing to the SD card takes around 40 ms. Since the classification algorithm in its most optimized form takes around 180 ms per invocation, the total loop time per second of audio is $\sim$220 ms. It should be noted that although the classification step introduces a small latency between the real-time onset of an event and the determination of whether to store the corresponding audio, our Apollo 4 microcontroller is able to store a minimum of 1 full second of 96 kHz audio in SRAM, with much longer buffers being storable at lower, more realistic sampling rates. As such, historical audio data of several seconds is maintained in typical deployment scenarios, and as such, this latency does not cause loss of data.

Nonetheless, power consumption still remains a critical challenge for long-term deployments that may be expected to reach up to a year. While our hardware design and firmware implementation emphasize efficiency, our current power measurements indicate that continuous operation of the adaptive data collection algorithm and storage to an SD card would significantly limit deployment duration. The development of a preliminary low-power filtering stage based on spectral flux or similar acoustic properties will be essential to trigger the more power-intensive adaptive algorithm only when potentially relevant audio is detected. Additionally, storage of clips to the SD card should be minimized as much as possible, as this is the largest contributor to power consumption. One mitigating technique that we plan to explore is using leftover SRAM memory on the microcontroller to increase the size of the temporary audio storage buffers. By decreasing the frequency that the SD card must be woken up, we gain substantial power savings. As an example, decreasing the frequency of SD card writes from 4 Hz to 1 Hz netted almost 4.4 mA in power savings during experimentation, although diminishing returns are expected as the power-up cost is amortized over longer and longer periods.

The initial field deployments on caribou, African elephants, and bighorn sheep represent important first steps in validating this technology in real-world conditions. However, these deployments are ongoing, and comprehensive analysis of their results is forthcoming. Additionally, research has progressed significantly since the first of these deployments, including creation of the quantized Q-VAE model and power-friendly improvements to the firmware, such as lowering the SD card writing frequency, as discussed above. As such, these initial deployments represent the absolute worst-case technical performance we can hope to achieve with these devices in their current state, providing us rather with a means of assessing their practical usability for ecological research and conservation applications.

Our development process to date has addressed many of the requirements outlined in Section 3.1, particularly in the areas of modular hardware design, configurable firmware, and intelligent data filtering. However, several key requirements remain to be fully implemented:

- The adaptive resource management features that adjust parameters based on remaining battery power and storage capacity are still under development.
- Integration with complementary sensors (GPS, accelerometer) has been implemented at the hardware level, but the firmware for synchronized multi-sensor data collection requires further refinement.
- The noise reduction techniques needed for robust performance in varied acoustic environments require additional development and field testing.

Future work will focus on addressing these requirements while analyzing data from ongoing field deployments. We will refine the adaptive filtering algorithm based on real-world acoustic data from different species and environments, develop more sophisticated noise reduction techniques, and optimize power management for extended deployment durations. Additionally, we plan to create an open database of wilderness sounds collected through these deployments to address the current gaps in acoustic datasets for ecological research. Successful completion of these remaining tasks will enable truly long-term, adaptive acoustic monitoring of wildlife, providing unprecedented insights into animal behavior, communication, and environmental interactions that have been previously impossible to continuously observe in natural settings.

Author Contributions: All the authors participated in the manuscript preparation as follows. Conceptualization, Á.L. and G.W.; methodology, G.K., D.J. and W.H.; hardware, G.K. and J.T.; software, D.J. and W.H.; validation, D.J. and J.T.; data curation, J.T., D.J. and W.H.; writing—original draft preparation, all; writing—review and editing, all. All authors have read and agreed to the published version of the manuscript.

Funding: This material is based upon work supported by the National Science Foundation under Grant No. 2312391 and 2312392.

Institutional Review Board Statement: The animal study protocol was approved by the Institutional Animal Care and Use Committee (IACUC) of Colorado State University (IACUC No. 5204 Animal-borne Adaptive Acoustic Environmental Monitoring; approved 15 December 2023).

Data Availability Statement: The data contains in this article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Hoefer, S.; McKnight, D.T.; Allen-Ankins, S.; Nordberg, E.J.; Schwarzkopf, L. Passive acoustic monitoring in terrestrial vertebrates: A review. *Bioacoustics* **2023**, *32*, 506–531.
2. Sugai, L.S.M.; Silva, T.S.F.; Ribeiro, J.W., Jr.; Llusia, D. Terrestrial Passive Acoustic Monitoring: Review and Perspectives. *BioScience* **2019**, *69*, 15–25. [CrossRef]

3. Greif, S.; Yovel, Y. Using on-board sound recordings to infer behaviour of free-moving wild animals. *J. Exp. Biol.* **2019**, *222*, jeb184689. [CrossRef] [PubMed]
4. Hebblewhite, M.; Haydon, D.T. Distinguishing technology from biology: A critical review of the use of GPS telemetry data in ecology. *Philos. Trans. R. Soc. B Biol. Sci.* **2010**, *365*, 2303–2312.
5. Rafiq, K.; Pitcher, B.J.; Cornelsen, K.; Hansen, K.W.; King, A.J.; Appleby, R.G.; Abrahms, B.; Jordan, N.R. Animal-borne technologies in wildlife research and conservation. In *Conservation Technology*; Oxford University Press: Oxford, UK, 2021.
6. Wijers, M.; Trethowan, P.; Markham, A.; Du Preez, B.; Chamaillé-Jammes, S.; Loveridge, A.; Macdonald, D. Listening to Lions: Animal-Borne Acoustic Sensors Improve Bio-logger Calibration and Behaviour Classification Performance. *Front. Ecol. Evol.* **2018**, *6*, 171. [CrossRef]
7. Northrup, J.; Avrin, A.; Anderson, C.; Brown, E.; Wittemyer, G. On-animal acoustic monitoring provides insight to ungulate foraging behavior. *J. Mammal.* **2019**, *100*, 1479–1489. [CrossRef]
8. Piczak, K.J. ESC: Dataset for Environmental Sound Classification. In Proceedings of the 23rd Annual ACM Conference on Multimedia, Brisbane, QLD, Australia, 26–30 October 2015; ACM Press: New York, NY, USA, 2015; pp. 1015–1018. [CrossRef]
9. Salamon, J.; Jacoby, C.; Bello, J.P. A Dataset and Taxonomy for Urban Sound Research. In Proceedings of the 22nd ACM International Conference on Multimedia (ACM-MM'14), Orlando, FL, USA, 3–7 November 2014; pp. 1041–1044.
10. Lynch, E.; Angeloni, L.; Fristrup, K.; Joyce, D.; Wittemyer, G. The use of on-animal acoustical recording devices for studying animal behavior. *Ecol. Evol.* **2013**, *3*, 2030–2037. [CrossRef]
11. Briefer, E.F.; Sypherd, C.C.R.; Linhart, P.; Leliveld, L.M.C.; Padilla de la Torre, M.; Read, E.R.; Guerin, C.; Deiss, V.; Monestier, C.; Rasmussen, J.H.; et al. Classification of pig calls produced from birth to slaughter according to their emotional valence and context of production. *Sci. Rep.* **2022**, *12*, 3409.
12. Lehmann, K.D.S.; Jensen, F.H.; Gersick, A.S.; Strandburg-Peshkin, A.; Holekamp, K.E. Long-distance vocalizations of spotted hyenas contain individual, but not group, signatures. *Proc. R. Soc. B Biol. Sci.* **2022**, *289*, 20220548. [CrossRef]
13. Ngo, H.Q.T.; Nguyen, T.P.; Nguyen, H. Research on a Low-Cost, Open-Source, and Remote Monitoring Data Collector to Predict Livestock's Habits Based on Location and Auditory Information: A Case Study from Vietnam. *Agriculture* **2020**, *10*, 180.
14. Chelotti, J.O.; Vanrell, S.R.; Galli, J.R.; Giovanini, L.L.; Rufiner, H.L. A pattern recognition approach for detecting and classifying jaw movements in grazing cattle. *Comput. Electron. Agric.* **2018**, *145*, 83–91. [CrossRef]
15. Yan, X.; Zhang, H.; Li, D.; Wu, D.; Zhou, S.; Sun, M.; Hu, H.; Liu, X.; Mou, S.; He, S.; et al. Acoustic recordings provide detailed information regarding the behavior of cryptic wildlife to support conservation translocations. *Sci. Rep.* **2019**, *9*, 5172. [CrossRef] [PubMed]
16. Thiebault, A.; Huetz, C.; Pistorius, P.; Aubin, T.; Charrier, I. Animal-borne acoustic data alone can provide high accuracy classification of activity budgets. *Anim. Biotelemetry* **2021**, *9*, 28.
17. Couchoux, C.; Aubert, M.; Garant, D.; Réale, D. Spying on small wildlife sounds using affordable collar-mounted miniature microphones: An innovative method to record individual daylong vocalisations in chipmunks. *Sci. Rep.* **2015**, *5*, 10118. [CrossRef]
18. Green, A.C.; Johnston, I.N.; Clark, C.E.F. Invited review: The evolution of cattle bioacoustics and application for advanced dairy systems. *Animal* **2018**, *12*, 1250–1259. [CrossRef]
19. Galli, J.R.; Cangiano, C.A.; Milone, D.H.; Laca, E.A. Acoustic monitoring of short-term ingestive behavior and intake in grazing sheep. *Livest. Sci.* **2011**, *140*, 32–41. [CrossRef]
20. Navon, S.; Mizrach, A.; Hetzroni, A.; Ungar, E.D. Automatic recognition of jaw movements in free-ranging cattle, goats and sheep, using acoustic monitoring. *Biosyst. Eng.* **2013**, *114*, 474–483. [CrossRef]
21. Herlin, A.; Brunberg, E.; Hultgren, J.; Högberg, N.; Rydberg, A.; Skarin, A. Animal Welfare Implications of Digital Tools for Monitoring and Management of Cattle and Sheep on Pasture. *Animals* **2021**, *11*, 829. [CrossRef]
22. Wang, J.; Chen, H.; Wang, J.; Zhao, K.; Li, X.; Liu, B.; Zhou, Y. Identification of oestrus cows based on vocalisation characteristics and machine learning technique using a dual-channel-equipped acoustic tag. *Animal* **2023**, *17*, 100811. [CrossRef]
23. Hill, A.P.; Prince, P.; Piña Covarrubias, E.; Doncaster, C.P.; Snaddon, J.L.; Rogers, A. AudioMoth: Evaluation of a smart open acoustic device for monitoring biodiversity and the environment. *Methods Ecol. Evol.* **2018**, *9*, 1199–1211. [CrossRef]
24. Open Acoustic Devices. AudioMoth: MicroMoth. Available online: https://www.openacousticdevices.info/audiomoth (accessed on 5 January 2025).
25. Lennox, R.J.; Aarestrup, K.; Alós, J.; Arlinghaus, R.; Aspillaga, E.; Bertram, M.G.; Birnie-Gauvin, K.; Brodin, T.; Cooke, S.J.; Dahlmo, L.S.; et al. Positioning aquatic animals with acoustic transmitters. *Methods Ecol. Evol.* **2023**, *14*, 2514–2530. [CrossRef]
26. Lennox, R.J.; Eldøy, S.H.; Dahlmo, L.S.; Matley, J.K.; Vollset, K.W. Acoustic accelerometer transmitters and their growing relevance to aquatic science. *J. Mov. Ecol.* **2023**, *11*. [CrossRef] [PubMed]
27. Roussel, J.M.; Haro, A.; Cunjak, R.A. Field test of a new method for tracking small fishes in shallow rivers using passive integrated transponder technology. *Can. J. Fish. Aquat. Sci.* **2000**, *57*, 1326–1339. [CrossRef]
28. Wildlife Computers. Wildlife Telemetry Solutions. Available online: https://wildlifecomputers.com/ (accessed on 5 January 2025).

29. Turpault, N.; Serizel, R.; Shah, A.P.; Salamon, J. Sound event detection in domestic environments with weakly labeled data and soundscape synthesis. In Proceedings of the Workshop on Detection and Classification of Acoustic Scenes and Events, New York City, NY, USA, 25–26 October 2019.
30. Briggs, F.; Fern, X.; Raich, R. BirdSong: H. J. Andrews (HJA) Experimental Forest Recordings. 2012. Available online: https://paperswithcode.com/dataset/birdsong (accessed on 27 March 2025).
31. Lostanlen, V.; Salamon, J.; Farnsworth, A.; Kelling, S.; Bello, J.P. BirdVox-full-night: A dataset and benchmark for avian flight call detection. In Proceedings of the IEEE ICASSP, Calgary, AB, Canada, 15–20 April 2018. [CrossRef]
32. Vigness-Raposa, K.J.; Scowcroft, G.; Miller, J.H.; Ketten, D. Discovery of Sound in the Sea: An Online Resource. In *The Effects of Noise on Aquatic Life*; Popper, A.N., Hawkins, A., Eds.; Springer: New York, NY, USA, 2012; pp. 135–138. [CrossRef]
33. Integrated Ocean Observing System (IOOS), NOAA, and the U.S. Navy. Sanctuary Soundscape Monitoring Project (SanctSound). Available online: https://sanctsound.ioos.us/ (accessed on 27 March 2025).
34. Gemmeke, J.F.; Ellis, D.P.W.; Freedman, D.; Jansen, A.; Lawrence, W.; Moore, R.C.; Plakal, M.; Ritter, M. Audio Set: An ontology and human-labeled dataset for audio events. In Proceedings of the IEEE ICASSP 2017, New Orleans, LA, USA, 5–9 March 2017.
35. Institute of Electrical and Electronics Engineers (IEEE). AASP CASA Challenge Office Event Detection Datasets. 2013. Available online: https://dcase.community/challenge2013/download (accessed on 27 March 2025).
36. Various. DCASE Open Datasets Listing. 2015–2025. Available online: https://dcase-repo.github.io/dcase_datalist/ (accessed on 27 March 2025).
37. Fuentes, M.; Steers, B.; Zinemanas, P.; Rocamora, M.; Bondi, L.; Wilkins, J.; Shi, Q.; Hou, Y.; Das, S.; Serra, X.; et al. Urban sound & sight: Dataset and benchmark for audio-visual urban scene understanding. In Proceedings of the ICASSP 2022—2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Singapore, 23–27 May 2022; pp. 141–145.
38. National Park Service. Sound Gallery. 2022. Available online: https://nps.gov/subjects/sound/gallery.html (accessed on 27 March 2025).
39. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]
40. Ester, M.; Kriegel, H.P.; Sander, J.; Xu, X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In Proceedings of the KDD, Portland, OR, USA, 2–4 August 1996; pp. 226–231.
41. Campello, R.J.; Moulavi, D.; Sander, J. Hierarchical Density Estimates for Data Clustering, Visualization, and Outlier Detection. *IEEE Trans. Knowl. Data Eng.* **2015**, *27*, 1–14. [CrossRef]
42. von Luxburg, U. A Tutorial on Spectral Clustering. *Stat. Comput.* **2007**, *17*, 395–416. [CrossRef]
43. Gershman, S.; Blei, D. A Tutorial on Bayesian Nonparametric Models. *J. Math. Psychol.* **2012**, *56*, 1–12. [CrossRef]
44. Schörkhuber, C.; Klapuri, A. Constant-Q Transform Toolbox for Music Processing. In Proceedings of the ICMC, New York, NY, USA, 1–5 June 2010.
45. Andén, J.; Mallat, S. Deep Scattering Spectrum. *IEEE Trans. Signal Process.* **2014**, *62*, 4114–4128. [CrossRef]
46. Zeghidour, N.; Tachette, O.; Rosenkranz, F.; Dupoux, E. LEAF: A Learnable Audio Front-End for Speech and Audio Classification. In Proceedings of the ICLR, Virtual Event, 3–7 May 2021.
47. Cramer, J.; Wu, H.H.; Salamon, J.; Bello, J.P. Look, Listen and Learn More: Design Choices for Deep Audio Embeddings. In Proceedings of the IEEE International Conference on on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 3852–3856.
48. Hershey, S.; Chaudhuri, S.; Ellis, D.P.; Gemmeke, J.F.; Jansen, A.; Moore, R.C.; Plakal, M.; Platt, D.; Saurous, R.A.; Seybold, B.; et al. CNN architectures for large-scale audio classification. In Proceedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA, 5–9 March 2017; pp. 131–135.
49. Kong, Q.; Cao, Y.; Iqbal, T.;Wang, Y.; Wang, W.; Plumbley, M.D. PANNs: Large-scale Pretrained Audio Neural Networks for Audio Pattern Recognition. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2020**, *28*, 2880–2894. [CrossRef]
50. Ruff, L.; Vandermeulen, R.A.; Görnitz, N.; Deecke, L.; Siddiqui, S.A.; Binder, A.; Müller, E.; Kloft, M. Deep One-Class Classification. In Proceedings of the ICML, Stockholm, Sweden, 10–15 July 2018; pp. 4393–4402.
51. Schlegl, T.; Seeböck, P.; Waldstein, S.M.; Schmidt-Erfurth, U.; Langs, G. Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery. In Proceedings of the IPMI, Boone, NC, USA, 25–30 June 2017; pp. 146–157.
52. Chen, J.; Sun, M.; Cui, B. Efficient Audio Transformers Under One Million Parameters for Edge Devices. In Proceedings of the EMBC, Sydney, Australia, 24–27 July 2023.
53. Saeed, A.; Grangier, D.; Zeghidour, N. Contrastive learning of general-purpose audio representations. In Proceedings of the ICASSP 2021—2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada, 6–11 June 2021; pp. 3875–3879.
54. Niizumi, D.; Takeuchi, D.; Ohishi, Y.; Harada, N.; Kashino, K. BYOL for Audio: Self-Supervised Learning for General-Purpose Audio Representation. In Proceedings of the 2022 International Joint Conference on Neural Networks (IJCNN), Padova, Italy, 18–23 July 2022; pp. 1–8.
55. Gong, Y.; Chung, Y.A.; Glass, J. Ast: Audio spectrogram transformer. *arXiv* **2021**, arXiv:2104.01778.

56. Baevski, A.; Zhou, Y.; Mohamed, A.; Auli, M. wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 12449–12460.
57. Huang, P.Y.; Xu, H.; Li, J.; Baevski, A.; Auli, M.; Galuba, W.; Metze, F.; Feichtenhofer, C. Masked autoencoders that listen. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 28708–28720.
58. Team, A.E. Apollo4 Plus. Available online: https://ambiq.com/apollo4-plus/ (accessed on 31 March 2025).
59. Davis, S.; Mermelstein, P. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE Trans. Acoust. Speech, Signal Process.* **1980**, *28*, 357–366. [CrossRef]
60. Kingma, D.P.; Welling, M. Auto-Encoding Variational Bayes. *arXiv* **2022**, arXiv:1312.6114.
61. Çoban, E.B.; Perra, M.; Pir, D.; Mandel, M.I. EDANSA-2019: The Ecoacoustic Dataset from Arctic North Slope Alaska. In Proceedings of the 7th Detection and Classification of Acoustic Scenes and Events 2022 Workshop (DCASE2022), Nancy, France, 3–4 November 2022.
62. Johnson, H.E.; Lenart, E.A.; Gustine, D.D.; Adams, L.G.; Barboza, P.S. Survival and reproduction in Arctic caribou are associated with summer forage and insect harassment. *Front. Ecol. Evol.* **2022**, *10*, 899585.
63. Soltis, J.; Leong, K.; Savage, A. African elephant vocal communication I: Antiphonal calling behaviour among affiliated females. *Anim. Behav.* **2005**, *70*, 579–587. [CrossRef]
64. Leong, K.M.; Ortolani, A.; Burks, K.D.; Mellen, J.D.; Savage, A. Quantifying acoustic and temporal characteristics of vocalizations for a group of captive african elephants *Loxodonta africana*. *Bioacoustics* **2003**, *13*, 213–231. [CrossRef]
65. Cassirer, E.F.; Manlove, K.R.; Almberg, E.S.; Kamath, P.L.; Cox, M.; Wolff, P.; Roug, A.; Shannon, J.; Robinson, R.; Harris, R.B.; et al. Pneumonia in bighorn sheep: Risk and resilience. *J. Wildl. Manag.* **2018**, *82*, 32–45.
66. Devi, I.; Dudi, K.; Singh, Y.; Lathwal, S.S. Bioacoustics features as a tool for early diagnosis of pneumonia in riverine buffalo (*Bubalus bubalis*) calves. *Buffalo Bull.* **2021**, *40*, 399–407.
67. Ferrari, S.; Silva, M.; Sala, V.; Berckmans, D.; Guarino, M. Bioacoustics: A tool for diagnosis of respiratory pathologies in pig farms. *J. Agric. Eng.* **2009**, *40*, 7–10. [CrossRef]