Improving the Accuracy and Efficiency of Bird Song Analysis
with Machine Learning based Event Detection

By
Anju Mathew

A THESIS

submitted to

Oregon State University

Honors College

in partial fulfillment of
the requirements for the
degree of

Honors Baccalaureate of Science in Computer Science
(Honors Scholar)

Presented May 23, 2023
Commencement June 2023

AN ABSTRACT OF THE THESIS OF

Anju Mathew for the degree of Honors Baccalaureate of Science in Computer Science presented on May 23, 2023. Title:
Improving the Accuracy and Efficiency of Bird Song Analysis with Machine Learning based Event Detection

Abstract approved:

---

Patrick Donnelly

Ornithology is an exciting field with novel research emerging everyday. Researchers in bioacoustics often spend hours within the wilderness recording bird calls to analyze later in their lab. The burden of sifting through hours of audio recordings from the field continues to remain a time-consuming and manual task, despite the inherent value of such research. In this work, we investigate the ability of a mobile application to provide an accelerated method for ornithologists to upload acoustic field recordings for analysis. Driven by recent advancements in technology, there is a growing need for innovative solutions that can streamline this research process. Our proposed solution integrates machine learning algorithms with bioacoustic signal processing with the goal of accelerating the process of analyzing hours of field recordings. In this thesis, we focus on the integration of future machine learning algorithms into an event detection pipeline. This larger project involves three separate parts: (1) the user interface and back end architecture developed by a small team of students for the Oregon State CS 46x capstone series, (2) an event detection algorithm completed as a synergistic extension of the capstone, and (3) a machine learning system developed as part of a graduate thesis. Together, these projects integrate to enhance the accuracy and efficiency of bioacoustic event detection by leveraging machine intelligence with deep learning. Furthermore, we evaluate the performance of these approaches using real-world datasets and discuss our approach in the context of traditional event detection techniques. This research project contributes valuable insights into the feasibility and effectiveness of integrating machine learning algorithms into a cloud-based acoustic event detection pipeline and highlights the potential benefits of this integration for various applications to expedite time-consuming bioacoustic research processes.

Key Words: audio processing, bioacoustics, bird song, event detection, database

Corresponding e-mail address: mathewa@oregonstate.edu

# Improving the Accuracy and Efficiency of Bird Song Analysis with Machine Learning based Event Detection

By
Anju Mathew

A THESIS

submitted to

Oregon State University

Honors College

in partial fulfillment of
the requirements for the
degree of

Honors Baccalaureate of Science in Computer Science
(Honors Scholar)

Presented May 23, 2023
Commencement June 2023

Honors Baccalaureate of Science in Computer Science project of Anju Mathew presented on May 23, 2023.

APPROVED:

---

Patrick Donnelly, Mentor, representing School of Electrical Engineering and Computer Science

---

Hannah Scott, Committee Member, representing School of Electrical Engineering and Computer Science

---

Matthew Orr, Committee Member, representing Department of Integrative Biology

---

Toni Doolen, Dean, Oregon State University Honors College



---

Anju Mathew, Author

# Contents

# 1 Introduction

Ornithology is the scientific study of birds. The behavior of animals has always fascinated humans and for countless millennia we have tried to understand the mannerisms they use to move, speak, and communicate. Historically, we have been interested in animal husbandry for our survival, but more recently there has been growing interest regarding animal to animal relationships.

There are over ten thousand different extant species of birds remaining today. These species come in a vast array of sizes, shapes, and colors, ranging from the tiny bee hummingbird (*Mellisuga helenae*) which weighs less than a penny, to the ostrich (*Struthio camelus*) which stands up to nine feet tall. Birds inhabit virtually every environment on Earth, diving the depths of the oceans and soaring among the highest mountain peaks. Across the world, birds are known for their unique abilities, such as flight, song, and complex mating rituals. [6].

| | |
|---|---|
| 1 | Struthionidae, Ostriches |
| 2 | Rheidae, Rheas |
| 3 | Casuariidae, Cassowaries |
| 2 | Dromiceidae, Emus |
| 3 | Apterygidae, Kiwis |
| 32 | Tinamidae, Tinamous |
| 17 | Spheniscidae, Penguins |
| 3 | Gaviidae, Loons |
| 20 | Colymbidae, Grebes |
| 14 | Diomedeidae, Albatrosses |
| 56 | Procellariidae, Shearwaters, Fulmars |
| 18 | Hydrobatidae, Storm Petrels |
| 5 | Pelecanoididae, Diving Petrels |
| 3 | Phaëthontidae, Tropic-birds |
| 6 | Pelecanidae, Pelicans |
| 9 | Sulidae, Boobies, Gannets |
| 30 | Phalacrocoracidae, Cormorants |
| 1 | Anhingidae, Snake-birds |
| 5 | Fregatidae, Frigate-birds |
| 59 | Ardeidae, Herons, Bitterns |
| 1 | Balaenicipitidae, Whale-headed Storks |
| 1 | Scopidae, Hammerheads |
| 16 | Ciconiidae, Storks, Jabirus |
| 28 | Threskiornithidae, Ibises, Spoonbills |
| 6 | Phoenicopteridae, Flamingos |
| 3 | Anhimidae, Screamers |
| 145 | Anatidae, Ducks, Geese, Swans |
| 6 | Cathartidae, New World Vultures |
| 1 | Sagittariidae, Secretary-birds |
| 205 | Accipitridae, Hawks, Old World Vultures, Harriers |
| 1 | Pandionidae, Ospreys |
| 58 | Falconidae, Falcons, Caracaras |
| 63 | Charadriidae, Plovers, Turnstones, Surf-birds |
| 77 | Scolopacidae, Snipe, Woodcock, Sandpipers |
| 7 | Recurvirostridae, Avocets, Stilts |
| 3 | Phalaropodidae, Phalaropes |
| 1 | Dromadidae, Crab-plovers |
| 9 | Burhinidae, Thick-knees |
| 16 | Glareolidae, Pratincoles, Coursers |
| 4 | Thinocoridae, Seed-snipe |
| 2 | Chionidae, Sheath-bills |
| 4 | Stercorariidae, Skuas, Jaegers |
| 82 | Laridae, Gulls, Terns |
| 3 | Rynchopidae, Skimmers |
| 22 | Alcidae, Auks, Auklets, Murres |
| 16 | Pteroclidae, Sand-grouse |
| 3 | Raphidae, Dodos, Solitaires |
| 289 | Columbidae, Pigeons, Doves |
| 315 | Psittacidae, Lories, Parrots, Macaws |
| 19 | Musophagidae, Plantain-eaters |
| 127 | Cuculidae, Cuckoos, Roadrunners, Anis |
| 11 | Tytonidae, Barn-owls |
| 123 | Strigidae, Owls |
| 1 | Steatornithidae, Oil-birds |
| 12 | Podargidae, Frogmouths |
| 5 | Nyctibiidae, Potoos |
| 7 | Aegothelidae, Owlet-frogmouths |
| 67 | Caprimulgidae, Goatsuckers |
| 76 | Micropodidae, Swifts |
| 3 | Hemiprocnidae, Crested Swifts |
| 319 | Trochilidae, Hummingbirds |
| 6 | Coliidae, Colies |
| 34 | Trogonidae, Trogons |
| 87 | Alcedinidae, Kingfishers |
| 1 | Oxyruncidae, Sharp-bills |
| 3 | Phytotomidae, Plant-cutters |
| 23 | Pittidae, Pittas |
| 4 | Acanthisittidae, New Zealand Wrens |
| 2 | Philepittidae, Asities |
| 2 | Menuridae, Lyre-birds |
| 2 | Atrichornithidae, Scrub-birds |
| 74 | Alaudidae, Larks |
| 75 | Hirundinidae, Swallows |
| 48 | Motacillidae, Pipits, Wagtails |
| 58 | Campephagidae, Cuckoo-shrikes |
| 72 | Laniidae, Shrikes |
| 11 | Vangidae, Vanga Shrikes |
| 10 | Artamidae, Wood-swallows |
| 2 | Grallinidae, Magpie-larks |
| 13 | Prionopidae, Wood-shrikes |
| 10 | Aegithinidae, Leafbirds |
| 109 | Pycnonotidae, Bulbuls |
| 5 | Cinclidae, Dippers |
| 63 | Troglodytidae, Wrens |
| 30 | Mimidae, Mockingbirds |
| 1360 | Muscicapidae, Old World Insect-eaters |
| 304 | Turdinae (incl. *Zeledonia*), Thrushes |
| 261 | Timaliinae, Babblers |
| 19 | Paradoxornithinae (incl. *Chamaea*), Parrotbills |
| 12 | Polioptilinae (incl. *Rhamphocaenus* and *Microbates*), Gnatcatchers |
| 386 | Sylviinae (incl. *Regulus*) |
| 328 | Muscicapinae, Flycatchers |
| 50 | Pachycephalinae, Whistlers |
| 65 | Paridae, Titmice |
| 17 | Sittidae, Nuthatches |
| 1 | Hyposittidae, Coral-billed Nuthatches |
| 5 | Neosittidae, Australian Nuthatches |
| 17 | Certhiidae, Creepers |
| 54 | Dicaeidae, Flowerpeckers |
| 106 | Nectariniidae, Sunbirds |
| 80 | Zosteropidae, White-eyes |
| 160 | Meliphagidae (incl. *Promerops*) Honey-eaters |
| 12 | Prunellidae, Hedge-sparrows |
| 426 | Fringillidae, Finches, Buntings |
| 197 | Thraupidae, Tanagers |
| 36 | Coerebidae, Honeycreepers |
| 109 | Parulidae, American Warblers |
| 41 | Vireonidae, Vireos |
| 22 | Drepanididae, Hawaiian Honey-creepers |
| 88 | Icteridae, Troupials |
| 1 | Dulidae, Palm-chats |
| 4 | Ptilogonatidae, Silky Flycatchers |
| 3 | Bombycillidae, Waxwings |
| 263 | Ploceidae, Weavers |
| 103 | Sturnidae, Starlings |
| 13 | Cracticidae, Bell-magpies |
| 3 | Callaeidae, Wattle-birds |
| 20 | Dicruridae, Drongos |
| 32 | Oriolidae, Orioles |
| 17 | Ptilonorhynchidae, Bowerbirds |
| 100 | Corvidae, Crows, Magpies, Jays |
| 43 | Paradisaeidae, Birds of Paradise |
| 5093 | Total Passeres |
| 3523 | Total non-Passeres |
| 8616 | Total birds |

Figure 1: Visual providing a comprehensive overview of the number of bird species within each family.

Compared to other species, research studying birds presents ornithologists with unique challenges due to the migratory nature of most avian species. In general biologists, need to monitor an animal's habitat in order to properly observe them. When studying other species, such as fish, livestock, and even reptiles, many mannerisms can be observed within controlled environments devised by the researchers. Such examples include fisheries, farms, and nature preserves. Birds, on the other hand, typically use flight as their main form of travel creating a much larger environment that challenges researchers' abilities to track and monitor them. Although aviaries exist, the observations found in captive environments may not reflect the species natural behavior. This makes it much more difficult to study different species and understand their authentic mannerisms without traveling to different parts of the world for the purpose of observation. Technology to track birds is possible as well and is a useful tool to study their movements and behavior over long distances, but it can miss important details that could be observed with direct research in their environment. For example, tracking data can tell us where a bird travels, but it cannot provide information on the specific types of food it eats or the social interactions it has with other birds. Direct research in the field allows for a more comprehensive understanding of a bird's ecology and behavior.

The study of animal sounds is known as zoo-musicology. There are many different challenges for researchers embarking in this line of bioacoustic analysis. As an example, researchers typically perform field work in the wild to capture recordings for different species of birds in order to further study these calls in their respective labs. The collection of audio recordings in the field is expensive and time-consuming, and the data collected is arduous to analyze and annotate. There can be hundreds or even thousands of recordings ornithologists have to sift through in order to find high quality examples of the species of interest.

In recent years, advancements in the field of zoo-musicology have propelled human understanding of bird calls. There are websites and mobile applications capable of identifying a bird species based on their vocalization. Merlin Bird ID[1], developed by Cornell Lab, is an application that allows users to snap a photo or record a bird sound and the application utilizes machine learning components to identity the specific species associated with the photo or call. Merlin learns to recognize these species based off a giant database, eBird[2], which has more than one million species in its system. Despite the availability of artificially intelligent systems like Merlin, the accuracy in the identification has never been quite perfect, leaving room for improvement [36]. We hypothesize that the inconsistencies between the audio fidelity and recording conditions of the various recordings uploaded to applications are one of the main reasons that autonomous system continue struggle to differentiate between the calls of different avian species. In order to accurately predict a species by just its call, the audio example provided must be of sufficient quality, have long enough of a duration, and sufficient clarity to be compatible with the algorithm used to identify a call. If a machine learning system trained to identify birds by their calls is not implemented with the considerations previously stated, insufficient training data covering a variety of noise and other acoustic conditions would make these systems less robust against low quality or noisy examples and have difficulty generalizing to any example

[1] https://merlin.allaboutbirds.org/

[2] https://ebird.org/home

provided by a user.

In this work, we present an application targeted specifically towards ornithologists with potential to release to the general public at a later time. Our application will allow research scientists to upload audio clips, label examples, and annotate any unlabeled examples in the data set. There will also be a feature to allow users to label a bird call if there is audio but no label within our database. The uploaded audio file will enter a preprocessing algorithm responsible for splitting the file based on the automatic detection of events using only signal processing. We keep the events detected and discard the rest of the the irrelevant audio sections. Our application stands apart from others on the market given its comprehensive features which have yet to be found collectively in any other competing product. This novelty is the result of extensive research in consultation with an expert in the field and paves the way for a heightened user experience. We hypothesize that our system can lead to better labeled data and a more robust dataset by analyzing and considering examples meeting a range of acoustic conditions and background noise levels. The work we present here is part of a larger research ecosystem that seeks to understand the behavior, ecology, and conservation of avian species.

Bird calls are unique from species to species. Furthermore there are often a variety of differences and variances between individuals of the same species. Navigating these minuscule differences will be the most challenging component of this work as we will need to handle these differences in our event detection algorithm.

# 2 Avian Acoustics and Artificial Intelligence

In this section, we discuss many types of birds and some of the ways in which they differ. Although birds share many similarities, there is a lot of data analyzing the differences with respect to their calls, flights, and unique mannerisms by species [7]. We will continue the conversation with an overview of the types of different tools used to develop our application.

## 2.1 Bird Communication and Social Interactions

Bird calls are produced in the vocal apparatus of birds and the resonance in the vocal tract is responsible for modifying the tonal quality of birdsong [32]. Before we examine the different types of bird calls, we must first look at why birds make any type of sound in the first place.

Many birds are known for chirping at the crack of dawn, much to the dismay of humans who desire to sleep in. The prevalence of bird calls at dawn across species is hypothesized to result from the acoustic benefit during that time as a result of low atmospheric turbulence, allowing bird calls to be transmitted anywhere from 25 to 100 meters [8]. The main purpose for their singing is to defend themselves or their territory and to impress or rival other birds and therefore the range in which their sounds are heard are crucially important to their survival. Ensuring other birds and predators around them can hear their song is essential to both their safety and their ability to find a mate. Females are known to be attracted to males who display complex songs with certain types of local dialect structure so this accounts for part of the reason why birds use different tones, beats, and sound levels for their song [31].

Not only do vocalization differ greatly between different species, but birds of the same species also have unique vocal dialects in both their speech and song which are the result of vocal learning [44]. The emergence of vocal learning is integral to the human language but this ability may not be limited to humans. Birds exhibit similar developmental and behavioral features that result from vocal learning and the potential parallels to human learning are not yet well understood [44]. Despite research exploring this area, we have only rudimentary understanding of their communication.

In this project, we analyze examples from different avian species and specimens in order to identify the specific species that produced a call. In the wild, primates recognize other members of their group recognizing the individual differences in their voice [28]. This same method is how humans recognize familiar people in their own lives. In order to train machine learning models to identify the specific species that produced a given call, we must first break apart the algorithm into its respective individual sections. The algorithmic process starts with the data file itself.

## 2.2 Programming Tools

In this section, we discuss the tools and techniques used for achieving the objectives outlined in the previous section. Python is a programming language that supports imperative programming styles and object-oriented programming paradigms, making it one of the most versatile languages available today [20]. This language is one of the most popular for data science specifically due to its tools and vast extent of open source libraries [41]. We utilize the library packages `NumPy`[3] and `pandas`[4], both known for their ability dissect data tables and manipulate data [25]. These libraries can perform a range of functionalities, including slicing and dicing, reshaping, and grouping data.

### 2.2.1 Librosa

Librosa is a useful python package specializing in the analysis of audio data within audio files. This package is also useful to visualize an acoustic signal in an audio clip, such as a spectrogram which displays the frequency content of the signal [24]. In Figure 2 we show an example of spectrogram generated in librosa exemplifying two instances of a bird call.

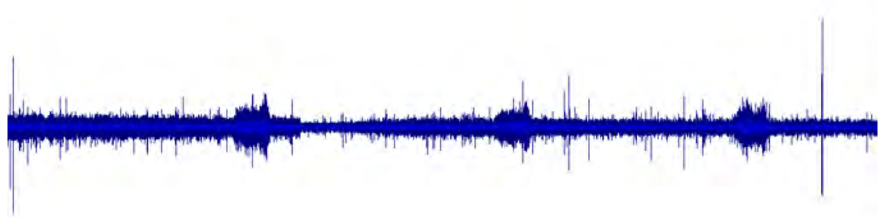

Figure 2: A visual representation showcasing a spectrogram displaying two instances of a bird call, providing a look into the acoustic structure of the sound.

[3] https://numpy.org/

[4] https://pandas.pydata.org/

With Librosa, one can load an audio clip and programmatically analyze and manipulate the data using the various built in functions of the package. Librosa is frequently employed in research studies related to speech emotion recognition. In one example, the authors use a Python dictionary to define emotions, and then fit and train the model created using machine learning elements [3].

### 2.2.2 Google Colab

Google Colab is a flexible tool used to collaborate effectively with other programmers on a cloud based programming project. It is also a convenient way to visualize, test, and explain code to programmers and non programmers alike. Specifically, Colab is used to prototype machine learning models on a powerful Graphics Processing Unit (GPU) and allows for excellent readability and interactive development [5]. Since each coding process is executed through the use of short code snippets, it is often easy to follow the flow of the code when run and the added visualization allows for a more comprehensive understanding for visual-oriented learners. It is important that the usability aspect of our application is exceptionally well-designed since we should not assume all ornithologists are particularly tech savvy and we want to make our application accessible to many users.

### 2.2.3 Amazon Web Services

Amazon Web Services (AWS) Lambda is a tool developed by Amazon that facilitates executing code in the cloud without the need to directly manage a dedicated server [38]. A programmer can run code for any back-end service by including infrastructure maintenance. For smaller workloads, utilizing cloud services, such as AWS Lambda, can be a more cost-effective alternative than using an EC2 server instance [43]. In order to use AWS Lambda, developers simply upload their code to Amazon's service and determine what conditions will trigger their code to run. This service has become popular over the years because it does not cost anything when the server is inactive [35]. AWS Lambda is an ideal solution for many tasks, such as file processing, web applications, stream processing, and mobile back ends [18].

## 2.3 Event Detection

Intelligent audio detection is necessary component of many systems, including audio streams, news broadcasts, and online meetings. These detection strategies extend beyond just volume based events, but also can extract features based on ranges of frequencies in order to better isolate events of interest [2]. Sound event detection is the act of recognizing certain events and differentiating their start and end from other sounds occurring at or near the same time [27]. Research has shown that a mix of both weak event detection and strong event detection can lead to the improvement of classifier performance [15]. This is resulting from the wide range of data points that target different frequency bands and and can lead to better recognition of events overall.

### 2.3.1 Machine Learning

The various repositories of information available to researchers are integral to machine learning. Because machine learning is a process in which a computer learns from the environment it is given, large amounts of past data are often crucial for a model to learn and improve further [13]. Machine learning techniques are ubiquitous in modern life, but here, the focus of our investigation is pattern and event detection from acoustic signals. In event detection machine learning algorithms are looking for 'interesting' events that occur based on certain thresholds a user sets, depending on the task at hand[23]. Detecting events that are anomalous and not relevant to a researcher's interest is one of the biggest challenges with this method. There are ways to combat this by setting thresholds for applicable events so an algorithm will be able to detect what data should be discarded versus what should be retained. One method to achieve this level of machine understanding is by training deep learning models to perform the event detection directly [14].

### 2.3.2 Deep Learning

Deep learning approaches enable automatic learning for large volumes of high dimensional data [16]. As increasingly more information is fed to an model, the better it understands the relationships between the input and the desire output. A trained model can potentially discover intricate structures within the data they are provided in order to discover complex nonlinear relationships. In the realm of acoustic understanding, an audio clip is fed into a data processor and the data is split using stratified sampling on the original audio clips. The data is partitioned into sets to train, validate, and test the model. Certain audio features are retracted from each of the clips and then combined into different abstracts to help train the machine learning model further [37]. While training deep neural networks can be a long and tedious process, the benefits of machine learning often outweigh these challenges.

## 2.4 Machine Learning Techniques

Prior research has thoroughly explored various machine learning models, including a variety of machine paradigms and approaches, such as supervised learning, unsupervised learning, reinforcement learning, and learning to learn [1]. These methods differ in how the models learn and respond to different types of learning. The goal of supervised machine learning is to learn a map for the provided inputs and then maps the inputs to the desired output [1]. This approach is the most prevalent when training neural networks, decision trees, and other methods where the output is highly dependent on the input [10]. Unsupervised learning algorithms work to automatically recognize patterns in a pool of data and are often used by researchers and practitioners [11]. This method is typically much more difficult since we are trying to train a model without directly providing it examples of the expected output. Reinforcement learning focuses on how to act given the environment and observation of the world around [33]. The learning to learn method is an approach in which an algorithm will learn continuously based on the prior experiences it has picked up from previous learning [1]. These methods, despite having different strategies, have the same result which is training intelligent models to make decisions based on data or observations.

The selection of the appropriate model for a specific task can be difficult. There is no one size fits all method to go about deciding which to solve a specific problem. Selecting the right algorithm can take a lot of trial and error, as one must figure out which one works best with the data set. The factors that influence the model we choose are the amount of data available to the algorithm, the type and statistical distributions of the data itself, and what needs to be learned from the data. Supervised learning and unsupervised learning are the two frequently employed machine learning approaches, and there are many algorithms and models that fall under these two categories. To train a model for making future predictions based on the data it has been given, supervised learning is usually the preferred method [9]. Tasks appropriate for supervised machine learning include predicting stock prices, classifications of different items, spam detection, face detection, weather forecasting, customer discovery, and identifying bird calls. If the purpose is to interpret a set of data and train a model for internal representation, then unsupervised learning is the method to choose [9]. Tasks appropriate for unsupervised machine learning include audience segmentation, customer persona investigation, and pattern recognition. Each method of machine learning has their own strengths and weaknesses, but the best method for a set of data most depends on the domain, purpose, and nature of the data being analyzed.

# 3 Examining Prior Research in the Field

In this section, we review previous studies and research related to animal vocalization, bird call applications already on the market, and the data sources in use within our application.

## 3.1 Animal Vocalizations

We begin by discussing animal calls across all different species. As an animal frequently under threat, whales have been at the center of research in animal communication for many years [34]. There are three main types of whale sounds: clicks, whistles, and pulsed calls [17]. Researchers pose that clicks are most likely for navigation purposes, to orientate the whales body in relation to their physical surroundings. Often longer clicks are said to estimate farther distances while shorter clicks are meant for short range movements. In Figure 3 we see a figure comparing the sequence of whale clicks compared to the intervals between clicks.

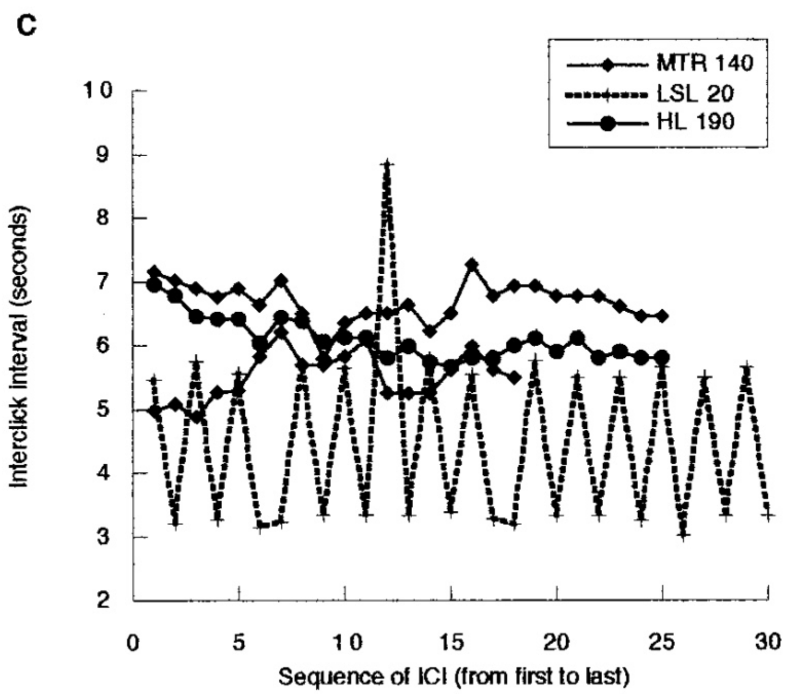


Figure 3: A visual comparison of whale click sequences and intervals, displaying the unique patterns and variations in these vocalizations.

More in depth studies like ones performed by David K. Mellinger involve the detection of specific calls like the up call, a frequency-modulates upsweep in the 50-

200 Hz range [26]. The methodology utilized for tracking these calls is performed by comparing spectrogram correlation with manual parameter choice and optimized parameter choice [26]. This method has been successfully applied in several research tasks, such as detecting ultrasonic harbor porpoise clicks [12]. In Figure 4 we see the comparison of spectrogram correlation with the manual parameter choice exemplified.

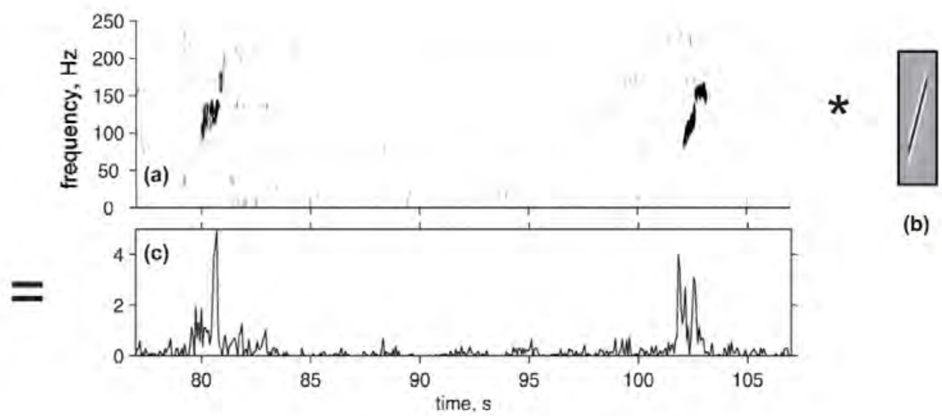


Figure 4: Comparison of spectrogram correlation: manual vs optimized parameter choice for whale clicks [26]

Whales are very social creatures, frequently travelling in pods which enable researchers to capture a substantial amount of calls in a short period of time. This research is of paramount importance in comprehending the behavior of whales and other cetaceans, while also advancing the study of endangered species. As we can further our understanding of each species, the higher chance we will have a better idea of their needs and we can work to support populations to continue to reproduce and increase their numbers.

There are animals like apes and other primates who put stronger emphasis on body language when communicating. Although apes have common evolutionary lineage to humans, humans rely principally on oral communication while other primates tend to lean on different mannerisms like body language. This information is gathered from researchers who observe their behavior in both captivity and out in nature. Despite body language being such a prevalent part of their communication, there are similarities in the way chimpanzee's communicate using different vowel-like sounds to suggest that their oral communication may be more similar to humans than we initially thought. Researchers once thought simian vocal tracts took on a uniform tube like structure with little to no capacity to form patterns as humans might, but that has been proven to be untrue. Monkeys as a species have vocal tracts that are capable of producing their own dialect, but not to the same capacity as human speech [22]. Continued research of primate communications could one day lead to a breakthroughs in the understanding of vocal communication of non-hominids.

There have also been research studies comparing the relation of sheep vocalisation to the overall well-being of sheep in New Zealand [19]. The vocalisations were first recorded and with the utilization of machine learning elements, classified into happy and unhappy classes. Agriculture is a critical economic industry in New Zealand. In particular, sheep are one of the most important farming sectors and places New Zealand among the top five countries that export sheep [19]. As an integral part of their economy, it is important for researchers to understand what environments these animals best thrive. Analysis of sheep vocalizations can be used to predict moods, general well-being, and a plethora of yet-undiscovered phenomena as research advances. In general, collecting audio data with the hopes of capturing certain mannerisms and methods is only the first step, but what researchers can do with that data with the latest state-of-the-art technology may potentially open a whole new world into understanding animals that could not have been envisioned a decade ago. Although sheep may not be popular enough to warrant the development of specific applications, there are other animals that have gained enough attention to have dedicated applications already on the market.

## 3.2 Bird Calls Applications

There are several application on the market focused specifically on bird calls. First, we examine Merlin Bird ID[5] which was developed by the Cornell Lab of Ornithology and is one of the most popular bird identification apps available today. The application helps users identify a specific bird by either uploading a photo or answering five questions. The application utilizes crowd sourcing to incorporate user input and verify if a bird was correctly identified. The more input the app data receives related to a specific photo, the higher likelihood the same species will be correctly identified in the future. Merlin Bird ID relies on the quality of the users photo along with the expertise and ability of the users to answer questions accurately. There is also a feature which allows users to record bird calls and identify the species based on the audio file. Merlin is currently powered by eBird[6], the world's largest database of bird depictions, calls, and photos which sets it apart from anything else currently on the market.

Similarly, the Audubon application also identifies bird species based on user input. Users can enter all features they see on a bird like its feather color, length, and tail length and the database will search its entries and identify the species that matches

[5] https://merlin.allaboutbirds.org/

[6] https://ebird.org/home

most closely. Audubon not only identifies the species, but delivers educational information related to the birds appearance, habitat, and behavior. Nearly anything a user might want to learn about a species in a limited amount of time, is described or referenced by the app. Similar to the Merlin Bird ID application, Audubon utilizes machine learning components to correctly identify a bird species and utilizes user input to continuously improve its accuracy. Audubon has a more broad purpose than just a singular application, they are a large and established organization that focuses on learning more about birds, but also conservation of birds and the environments they inhabit. The information gained from their application directly relates to their charitable, environmental, and educational missions.

## 3.3 Crowd-sourcing Techniques

Crowd sourcing techniques are an increasingly popular method to collect data for machine learning. For an application using trained machine learning models to make identifications, the collection of data from users can help model generalizations across diverse conditions and help the model improve prediction accuracy. Crowd sourcing techniques are used in Merlin Bird ID, the Audubon[7] application, and countless others. It is important to remember data collected by human users has the risk of the Human Factor [21]. If an application has the target audience of researchers, the effect of the Human Factor might diminish due to their expertise and experience when it comes to handling data. Low quality data can be harmful for any machine learning system and can put an application at risk of providing erroneous predictions. When a machine learns from wrong information, it becomes more likely to make mistakes in its predictions. It will keep making those same mistakes until it is given the right information to learn from. It is important to continuously check trends in the data and ensure that the data is reliable.

One might ask why crowd sourcing is used when it is prone to so much error? The advantages of obtaining large quantities of labeled data is essential to train accurate machine learning models, and crowd sourcing allows a relatively quick and inexpensive way for researchers to collect this data [42]. The method integrating both artificial intelligence(AI) algorithms with human input is known as "hybrid intelligence systems" and is developed with the goals of making AI even smarter with the capabilities to continuously be learning [4]. Crowd sourcing confers numerous benefits such as data generation, evaluating and debugging models, hybrid intelligence models, and behavioral studies to inform machine learning models [42]. Crowd sourcing provides something that computers cannot, which is access to data sourced from individuals who share a common interest in a particular topic. Researchers have only just started exploring the benefits of combining human data with machine learning, such as the

[7] https://www.audubon.org/

popular ChatGPT language model. Although there are some opportunities for error, in most cases, the advantages outweigh the risks. With sufficient volume of data, the human errors mentioned and any outliers tend to balance themselves out statistically.

## 3.4 Data Sources of Bird Calls

For many applications relating to birds, crowd sourcing is employed in conjunction with other established data repositories. One of the largest of these repositories is eBird. Originally created by the Cornell lab of Ornithology, eBird collects information about different bird species taking advantage of information provided by bird watching volunteers from across the world [39]. This large and established database of knowledge is used not only by the ornithology community but also by researchers in landscape ecology, macroecology, computer science, statistics, computational sustainability, human computation, informal science education, conservation science, and public policy [39]. eBird meets the needs of bird watchers and provides sufficient data for future research. Another such data source is Avibase[8] which is another large database containing information of bird species across 20,000 geographic regions of the world. Avibase accepts information from their thousands of users to continuously update pictures, regions, and important statistics regarding their bird species. Although these databases are substantial already, they rely on user-provided information to stay updated and expand their library.

## 3.5 Data Sources Used to Train Models In This Study

There are are existing large online research data sets of bird calls for different species that can be used to train different types of models. These data-sets typically contain audio recordings of various bird species, along with accompanying metadata such as the species name, location, and time of recording. Some commonly used data-sets include:

**Xeno-Canto** : Xeno-Canto[9] is a community-driven database of bird sounds from around the world. It has over 480,000 recordings from over 10,000 species, and is a popular resource for birders, researchers, and conservationists.

**Macaulay** : The Macaulay Library[10] is a collection of natural sounds and video recordings, with a focus on birds, housed at the Cornell Lab of Ornithology. It is one of the largest and most comprehensive collections of its kind in the world, containing millions of recordings of thousands of species from all over the world.

[8] https://avibase.bsc-eoc.org/avibase.jsp?lang=EN
[9] https://xeno-canto.org/
[10] https://www.macaulaylibrary.org/

**Borror** : The Borror Laboratory of Bioacoustics at The Ohio State University maintains a database[11] of natural sounds, with a focus on the sounds of animals and their environments. This type of resource is invaluable material to researchers studying animals in their natural habitats.

**Florida** : The Florida Shorebird Database[12] is a collection of data on shorebirds and seabirds that nest or migrate through Florida. The database contains information on the location and abundance of bird populations, as well as details on breeding success and survival rates.

Audio examples from these datasets can be used to train various types of machine learning models, such as neural networks or support vector machines, to identify different bird species based on their calls. Such systems could be very helpful to study the patterns and behaviors of birds in different environments, and to track changes in bird populations over time.

[11] https://www.gbif.org/dataset/f11db245-3f9f-4fc6-a0cc-12b4124d081b

[12] https://geodata.myfwc.com/datasets/myfwc::florida-shorebird-database/explore

# 4 Avian Audio Segmentation (AAS)

In this section, we discuss the process of developing the Avian Audio Segmentation Algorithm, along with the challenges and limitations we encountered. Two different approaches were taken to achieve our desired output: split by silence and event detection. While the first approach of splitting by silence did not produce the desired result, we learned from these missteps and we improved our approach with a second version of the algorithm to detect target events relatively accurately.

## 4.1 Motivation and Approach

In order to design an application that will identify bird calls to their species, we must first analyze the data set provided to us by an expert in the field, Kurt Trzcinski, Ph.D.[13]. Trzcinski is an ecologist specializing in the studies of woodpeckers and avian diversity, aquatic micro-worlds, fisheries and sustainable harvest, and other related areas. His work extends across many ecosystems and apart from collecting research on site, he focuses on study design, statistical quantitative analysis, and report writing. The motivation behind this project was Trzcinski's vision for an application designed to analyze his field recordings. There can be many hours of footage to sort through from just one field recording, and an application able to sift through the content and identify what pieces are data points would save invaluable time.

### 4.1.1 Approach One: Silence Detection

We started off our research stage by analyzing a small sample of five audio files from Trzcinski's data repository. We hypothesized that a split by silence algorithm would be an appropriate first step in identifying the specific pieces of audio needed since most files contained occasional bird calls between long periods of silence. This phenomenon arises from the fact that birds do not vocalize continuously, rendering it impossible for researchers to accurately predict when vocalizations will occur, resulting in hour-long recordings that often contain extended periods of silence interspersed with sporadic bird calls. Some other extraneous noises in these files included voices, rain, and other environmental disturbances. In order to improve long audio data into useable pieces of relevant data, algorithms that split by silence are a common since they do not involve any training and can be deployed rather easily [29].

[13] https://www.trzcinskiconsulting.com/

Here is an overview of our starting procedure, which is also illustrated in Figure 5.

1. Read in field recording using `pydub` and `librosa`.
2. Identify relevant events using a split by silence with `pydub`.
3. Discard silent portions of file.
4. Each piece of relevant audio event data is combined and outputted as one file.

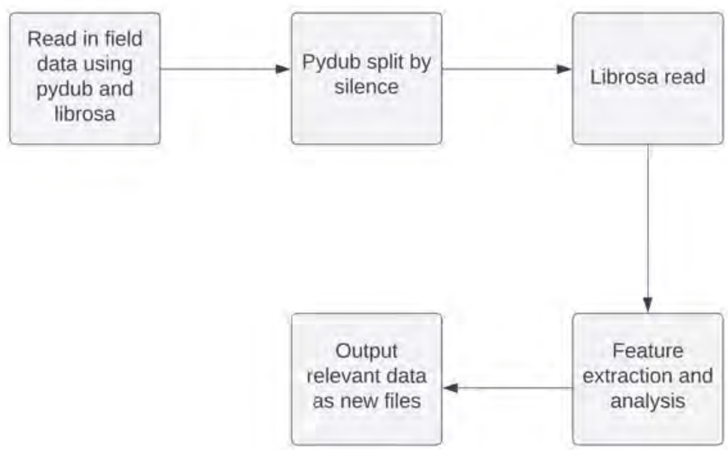


Figure 5: Block diagram illustrating the different stages of the split by silence algorithm, providing a visual representation of the system's architecture and the flow of data

This approach involved receiving an audio file, splitting the file based on acoustic events, and delivering output files with only relevant data of interest in the acoustic signal. Using `pydub` and `librosa` we first read in the provide audio file, using `pydub` to identify the periods of silence using an appropriate silence threshold which leaves out the frequency ranges bird calls are found.

We outputted new audio files, each of which were made up of the relevant audio events found in the original file. This method worked rather well for the our high quality library files we used during development, which featured very little background noise and prominent and distinct bird calls throughout. Unfortunately, upon analyzing real-world data-sets of noisy field recordings, it quickly became apparent that many of the files had background noise of rain, wind, and talking, occasionally overlapping with the bird calls, and at times, louder than the target call itself. Figure 6 is a spectrogram illustrating an audio clip from Trzcinski's field recording data with constant rain.

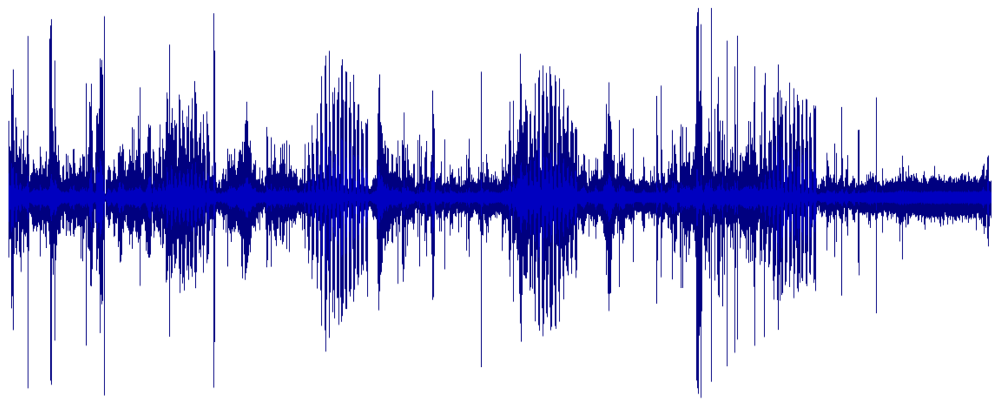

Figure 6: Visual providing an illustration of an audio clip with the sound of rain, highlighting the unique sound profile of rain, showcasing how it dominates the acoustic environment and creates a distinct soundscape

In these figures, we can clearly observe the continuous extraneous noise such that the noise level never diminishes to zero. We can also see the three spikes of audio which are bird calls also present in the audio signal.

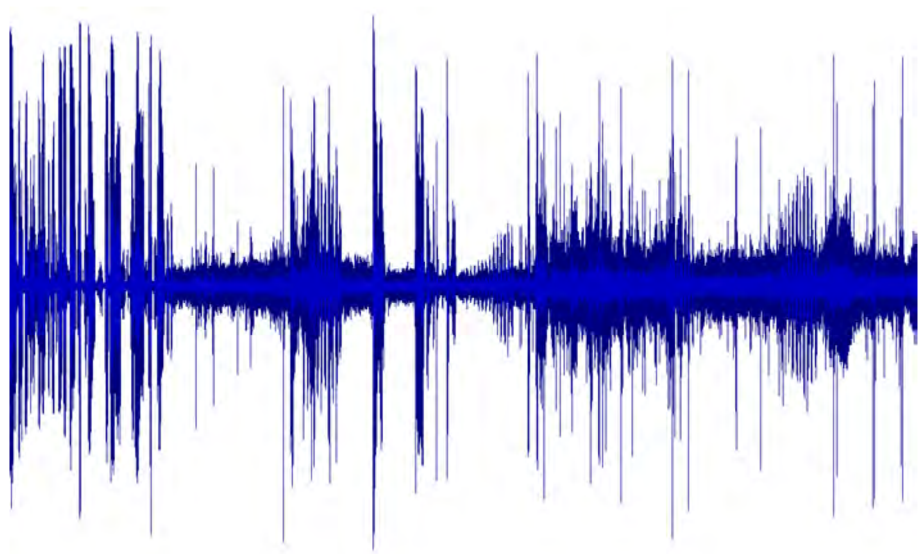

Figure 7: Visual providing an illustration of an audio clip with human speech or talking.

Figure 7 demonstrates a visualization of an audio signal with someone speaking at the start. It is clear to see that the frequencies are much stronger when the talking is occurring and diminishes back to the constant rain in the background. This spectrogram is a different segment of the audio signal shown in Figure 6.

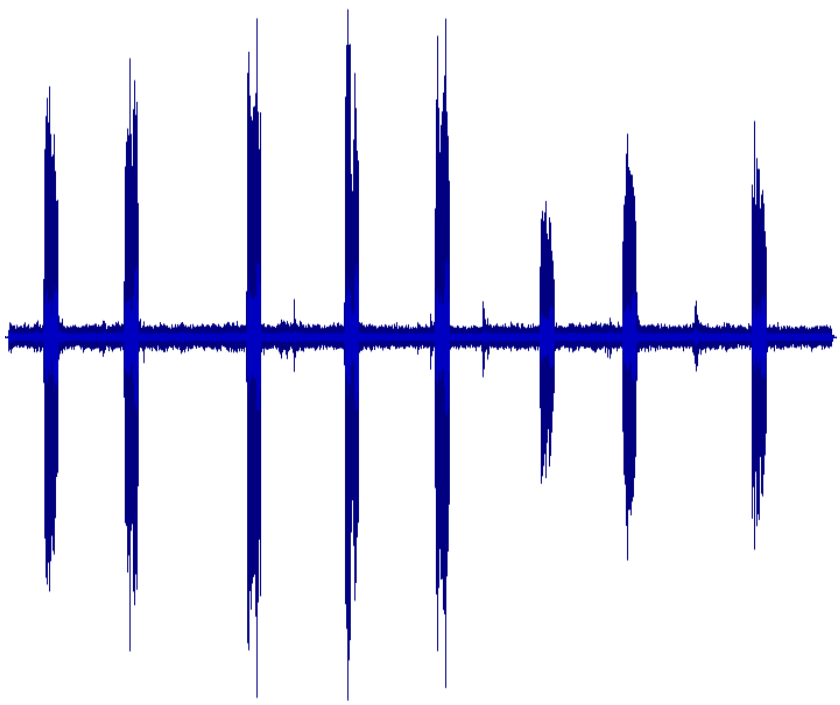

Figure 8: Visual highlighting the unique sound profile of bird vocalizations, showcasing how they stand out amidst periods of relative quiet

In Figure 8, we can observe an exemplary audio file that perhaps represents an ideal standard. There is very little noise except for the bird calls that are occurring at a fairly steady rhythm. We can see that the amplitude of the bird call signal itself is identifiable, rising well above the quieter periods of relative silence in the signal.

#### 4.1.2 Approach Two: Event Detection

Learning from our first design, we had to revise our primary approach to account for the variability between audio files in the real world data-set of field recordings. From our original design, based on silence detection, we shifted to an event detection method designed to notify us of the areas in the signals where major events are detected. Here is an overview of this new procedure:

1. Read in field recording using `pydub` and `librosa`.
2. Identify areas of interest in the audio file looking for spikes in amplitude that fall under a certain range.
3. Include a margin of time around the discoverd event and split the audio.
4. Output each piece of relevant audio detected as a new file.

This approach allowed us to change the noise levels in which audio was detected and we specifically focused on the range of frequencies of bird calls. We found that frequencies below 1 kHz typically contain few bird calls and most calls are between 1000kHz and 8000kHz [30]. We adjusted the frequency thresholds in our code accordingly. In doing so, we were able to focus on the major events and eliminate a large amount of the background noise. Despite allowing for the identification of more relevant events in the audio signal, the algorithm still detected a large portion of irrelevant audio, such as rain, people talking, or other extraneous noises that fell into the same frequency range as the bird songs. It was not possible to minimize all types of errors, so we designed this method in a flexible way that would factor in user input to account for relevancy of audio clips before being fed to the machine learning portion of the algorithm. All smaller clips that were generated from the original audio clip will be presented to the user using the application interface where the user will be able to select which audio files should be kept and which should be discarded. Following the creation of relevant files, the machine learning algorithms are subsequently employed. The diagram shown in Figure 9 exemplifies this process.

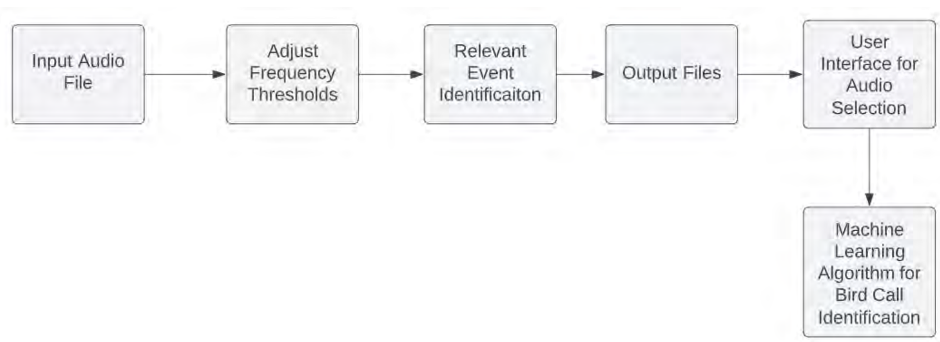


Figure 9: Block diagram presenting a visual representation of the system's architecture and the data flow, illustrating the different stages of the event detection algorithm.

Before attempting the machine learning portion of the algorithm, we needed a way to establish connectivity between the data science back end and the front end. In order to accomplish this, we implemented the AWS Lambda provided by Amazon to establish connectivity between our two computer programs. The steps followed for this method was for the clip the user inputs to be stored temporarily for processing and for the AWS API to act as a gateway to the AWS Lambda where the audio splitting occurs. The first step of this sequence acts as the interface for user requests while the second half is the area that holds the code for the audio split. Once the audio split takes place, we move into migrating the data into the database. The user would have to first approve the array of split audio clips, discarding any that might have been incorrectly interpreted as bird sounds by the algorithm. The saved audio clips would then be uploaded to the database using the AWS API as a gateway once again. By leveraging the API, there is a lot of complicated manual work saved because of the automatic integration with our application.

**Algorithm 1** Event Detection

**procedure** SPLIT($a, b$)<br>
    Read in the audio file<br>
    For $x$ in range(number of events):<br>
    **if** Frequency range = Bird call range **then**<br>
        Split on notable events found by librosa onset detect<br>
        Add three second buffer to each event<br>
        Create a new audio file for each event<br>
        **if** Format of audio file $\neq$ mp3 **then**<br>
            Convert to mp3<br>
            Output all new audio files

### 4.2 Serving As a Pipeline

This project consisted of three distinct components that were developed and executed by separate project teams.

1. The first portion consists of the front end design that was completed by my capstone team as our final project for the CS 46X series. In this three term project, I was assisted by three fellow computer science seniors at Oregon State University.

2. The second portion of this larger project was my individual contribution undertaken as my Honors College thesis research, consisting of the event detection algorithm: Avian Audio Segmentation (AAS). AAS is responsible for splicing long audio files into smaller clips with only bird calls and no extraneous noises.

3. The final section of this project was developed by Roopesh Ravishankar. As part of his Masters thesis, he developed a machine learning approach capable of identifying some bird species based on the specific bird call.

Serving as a pipeline, my purpose is to connect the sub-sections of the project together to complete the full picture application as shown in Figure 10.

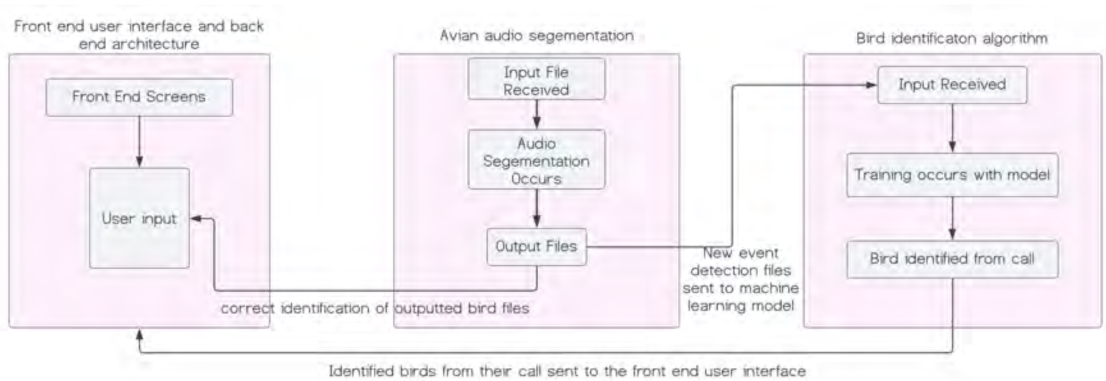


Figure 10: Representation depicting the Avian Audio Segmentation as a pipeline interconnecting stages. This pipeline serves as a tool for segmenting audio data into distinct parts, enabling the analysis of specific features or patterns in bird vocalizations.

#### 4.2.1 The Capstone

The Capstone project was the impetus for the broader investigation and scope of the research endeavor. The main purpose of the capstone application was to create a user interface to which ornithologists could upload data they collect within their research to a database that could be referenced later. Someone using this application would have the ability to add a sound, label a sound, find information about a specific bird, and other features potentially beneficial to the work of an ornithologist. The user interface along with a connection to the back end hosting the database was the main purpose this project served to complete. The capstone serves as a baseline architecture for the three different parts this mapped out project contains. The front end of the application was developed with Flutter, an open source framework developed by Google allowing for the creation of multi-level application interfaces [40]. With the connection of the Flutter front end to the back end containing the database utilizing an AWS API, the base structural components were complete and the next phase of development could begin.

### 4.2.2 Event Detection Algorithm

The application accepts audio clips of different lengths that a user chooses to upload a sounds from their local machine. This audio can vary in length, anywhere from ten seconds to one hour. Sometimes the audio submitted is not always relevant clips; instead there might be clips with long periods of silence, voices, or noises from the environment the birds are captured within. In order for researchers to expedite their scientific goals, we desire to collect only the relevant parts of the audio which are the bird call. In order to accomplish this, the algorithm developed had to detect both the start and end of relevant audio events of interest while also knowing which segments of the audio to ignore. To accomplish this goal we chose to use an event detection algorithm since it is meant to recognize temporal difference within an audio signal [27]. Utilizing `pydub` and `librosa`, we were able to develop an algorithm that could differentiate bird calls from extraneous noises and spit out a shortened audio clip with only the relevant pieces of data. This algorithm can potentially save ornithologists long hours of sifting through data in order to find the pieces relevant to their study. With the ultimate goal of the application being the identification of the bird species directly from its song, the event detection piece acts as a pipeline connecting the capstone architecture to the identification algorithm developed as part of a graduate students research. This algorithm, developed individually as my Honors College thesis research project, served as a vital contribution in bridging the gap between the graduate student's project and the capstone project, providing the necessary connectivity for the successful completion of the overall application.

### 4.2.3 Machine Learning Algorithm

The event detection algorithm feeds the newly split clips of data into the machine learning algorithm which analyzes the audio and attempts to predict which species of bird species may have made the call. The model used in this study was trained using four datasets: Borror[14], Florida[15], Xeno-Canto[16], and Macaulay[17] ; all repositories known for their invaluable data related to birds and other species. In this work, we consider nine species of owls and two other noctural birds commonly found in the Sisters Wilderness area in Central Oregon where our partner Kurt currently conducts his fieldwork.

| | | **B** | **F** | **X** | **M** |
|---|---|---|---|---|---|
| **Barred owl** | *Strix varia* | 16 | 66 | 201 | 5595 |
| **Spotted owl** | *Strix occidentalis* | 5 | 3 | 132 | 552 |
| **Western screech owl** | *Megascops kennicottii* | 20 | 28 | 196 | 1773 |
| **Great horned owl** | *Bubo virginianus* | 24 | 28 | 307 | 7691 |
| **Boreal owl** | *Aegolius funereus* | 11 | 11 | 25 | 362 |
| **Northern saw-whet owl** | *Aegolius acadicus* | 7 | 10 | 299 | 3358 |
| **Great grey owl** | *Strix nebulosa* | – | 1 | 10 | 117 |
| **Flammulated owl** | *Psiloscops flammeolus* | – | – | 109 | 520 |
| **Northern pygmy owl** | *Glaucidium californicum* | – | – | 188 | 161 |
| **Common poorwill** | *Phalaenoptilus nuttallii* | 7 | 11 | 160 | 1339 |
| **Common nighthawk** | *Chordeiles minor* | 22 | 19 | 127 | 2347 |
| | | **112** | **177** | **1754** | **23,815** |

Table 1: List of the nine species of owl and two nocturnal song birds considered in this study alongside counts for the each of the data sources: *Borror* (B), *Florida* (F), *Xeno-Canto* (X), and *Macauly* (M).

Table 1 presents a visual representation of the data repositories, their corresponding species, and the number of audio recordings available for each species. The information is organized in a tabular format, where each row represents a species and displays the name of the repository, the species' scientific name, and the number of audio recordings available for that species within that repository.

Figure 11 presents a diagram highlighting the connectivity of the entire system. We see the various ways the Capstone project, thesis, and graduate machine learning project are interconnected, with the Flutter front end, Avian Audio Segmentation, and bird identification machine learning models all playing their respective role. The AWS Infrastructure is the element tying AAS to the Capstone front end, and the resulting files from AAS are to be fed to the bird identification algorithm for the result to be displayed back on the Flutter front end.

[14] https://mbd.osu.edu/collections/borror-laboratory-bioacoustics

[15] https://geodata.myfwc.com/maps/florida-shorebird-database

[16] https://xeno-canto.org/

[17] https://www.macaulaylibrary.org/

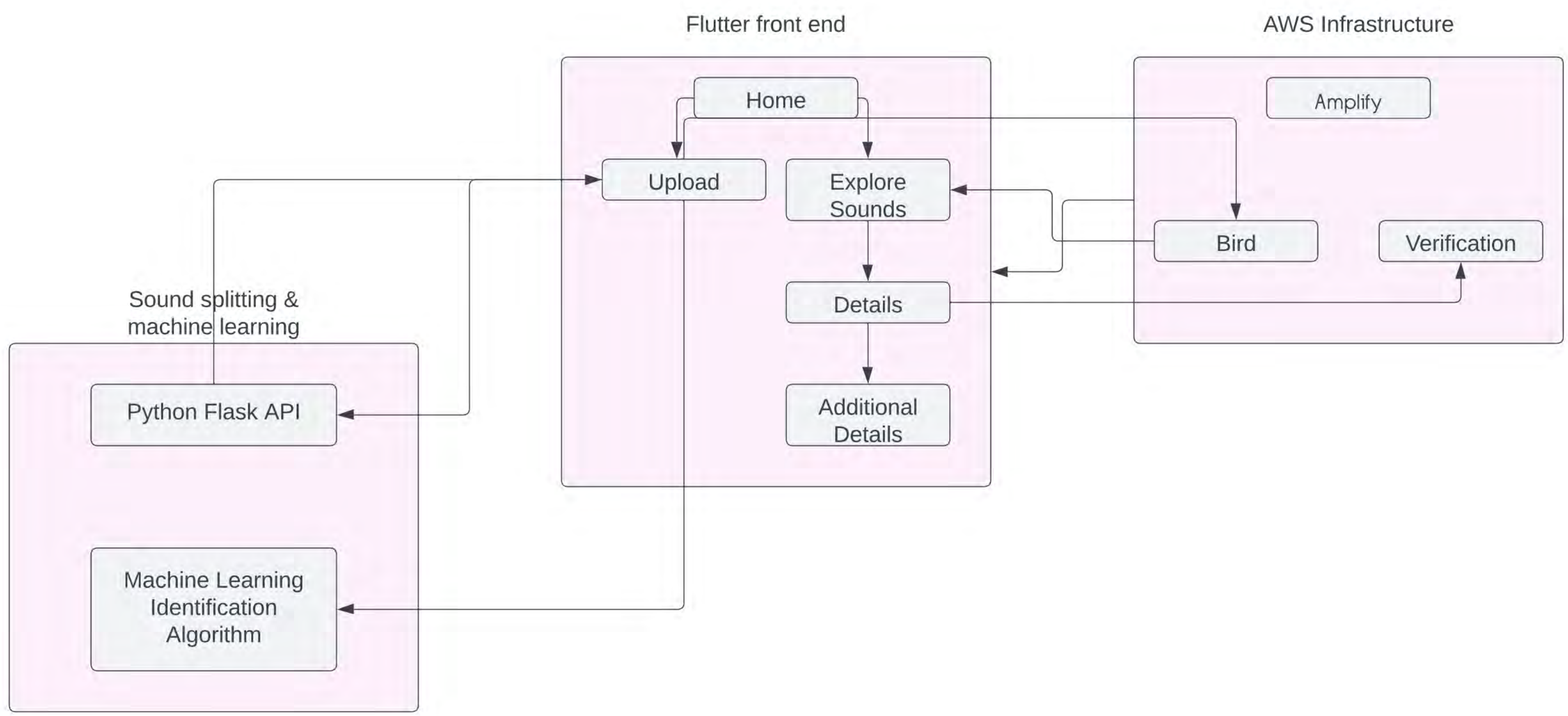


Figure 11: Diagram serving as an exemplification of the system, showcasing the different components and their interactions with each other. This representation is designed to provide a clear and concise overview of the system's architecture, highlighting the flow of data and the different stages of processing.

## 4.3 Advancing Research Through Collaborative Effort

This collaborative project aimed to create an application that assists ornithologists in their research by detecting relevant bird calls within long audio files and identifying bird species from their calls. The project consists of a capstone project building the interactive website and dataset, our contribution of the Avian Audio Segmentation (AAS) algorithm, and a machine learning algorithm to identify the avian species making the call. The capstone project serves as the primary user platform, while the role of the AAS algorithm is to splice audio files into smaller clips and ignore extraneous noises. The machine learning algorithm identifies bird species from their calls. The project exemplifies the benefits of collaboration between different groups across Oregon State University, brought together to explore various techniques and technologies to develop a comprehensive solution for ornithologists to their research more efficiently.

# 5 AAS Performance Evaluation Results

We aim to develop an application to be a resource for ornithologists and others interested in bird song to be able to record, store, and interpret the data they collect. This application will allow for input from the user throughout the process, specifically, in relation to the addition of data and verifying certain results. The application will first identify relevant acoustic events and ask users to verify the quality and importance of the detected events. The computational intelligence portion of the application aims to extract events from the audio files provided by the user for the purpose of identifying the species of bird present in the audio file.

## 5.1 Refining AAS

Our final algorithm, which extracts high-quality data from audio files, is comprised of components derived from our previous design. We applied event detection approaches to refine our focus on specific timestamps when it was clear bird related activity was occurring in the signal. We were also able to split the inputted audio into multiple files utilizing what we learned in our split by silence algorithm. The only difference being instead of splitting by silence, we split so there was a consistent margin of time before and after the event detection occurred. By adopting this methodology, we effectively utilized the primary sound source and captured any potentially significant events occurring in the seconds preceding or succeeding each occurrence. Although we were able to get pieces of events that we wanted from the audio clips, there was still extraneous information that would get factored in as a bird call because of overlapping frequencies. Since most audio clips contained calls of the same species or just a varying one or two, we set a limit on the number of clips collected to ten randomly chosen events from the original audio clip. Setting a limit on the number of clips collected to ten randomly chosen events from the original audio clip helps to ensure that the dataset remains manageable and focused on the specific species or variations being studied, while still providing enough data points for analysis and accurate representation of the species' vocalizations. Although we retrieved accurate bird call events for most of the clips collected, there were still some cases in which the frequencies of the bird calls and extraneous sounds matched too closely and some clips would pick up only background noises and no bird calls.

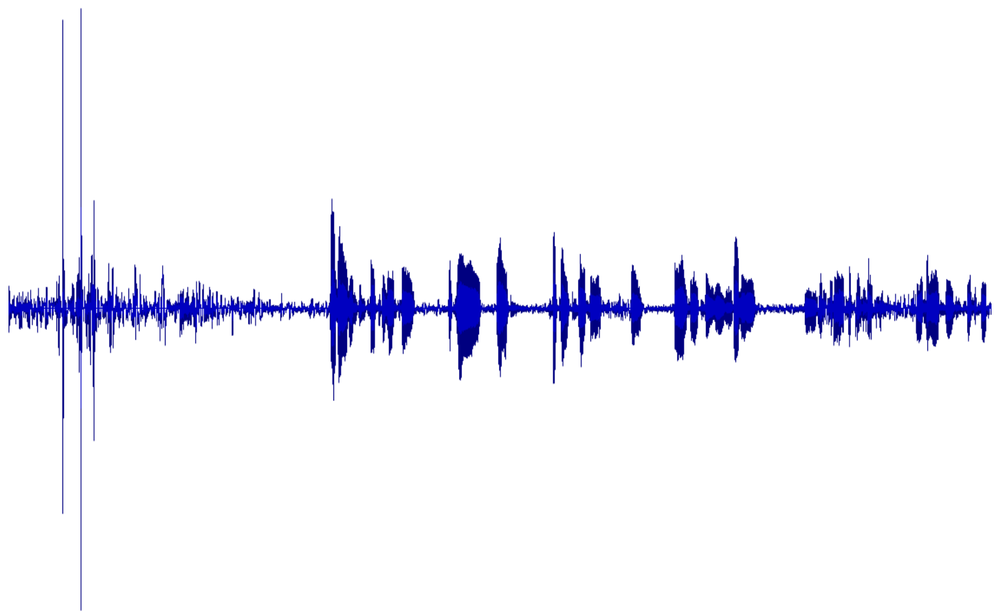

Figure 12: Visual illustrating the background noise generated during walking, displaying the different sound components that contribute to the overall noise profile.

Figure 7 is an example of one of these clips where an event was presumed to be a bird call, but instead was the sound of walking and snapping branches. In order to resolve this issue, we decided to turn to the users of our application for their input.

After the algorithm has identified potential bird calls found in the audio clip, the user will be presented the example and asked to click yes, confirming that the clip is in fact a bird call, or no, if the clip is not a bird call. The no-response will be discarded, and the yes-response will be marked for use in the machine learning portion of our algorithm. By adopting this approach, we can have a high level of confidence in the relevance of the audio clips and avoid expending resources on the identification of non-bird call sounds. We assume our target audience is composed of professionals in the field or ornithology and that they will accurately be able to identify which audio clips are relevant. Additionally, this approach hinders the machine learning portion of the algorithm from becoming corrupted by learning from audio clips presented as bird calls when they are not. Although we would have preferred to ensure certainty of the audio clips without user input, the range of frequencies from 1000 Hz to 8000 Hz was too wide to uniquely identify only bird sounds without a possibility of accidentally picking up some extraneous data. If we made the range smaller, we would risk not picking up certain birds in the audio and with a large range that leaves room for other sounds within that same range to be accidentally retained.

## 5.2 Results

In order to understand how well our algorithm worked, we conducted a few tests to analyze the results. We used ten field recordings, which included a mix of pure bird call audio clips and clips with extraneous noises. All of these recordings were processed through the AAS algorithm. To evaluate the algorithm's performance, we applied a random sampling method and selected five newly split audio files from each of the ten field recording, resulting in a total of 50 recording samples. Out of these 50 samples, the algorithm successfully identified bird noises in 43 clips, achieving an overall success rate of 86%. The average length of the audio clips was 8 seconds, with a standard deviation of approximately 2 seconds. We observed the highest success rate for identifying bird noises in 18 audio clips recorded in quiet environments, with a 100% success rate. However, for 32 audio clips recorded in noisy environments, the success rate was lower at 78.13%. These findings confirm that the algorithm's performance is significantly influenced by the background noise level of the recording environment. Further investigation is necessary to optimize the algorithm's generalization across different noisy environments. These findings suggest that machine learning algorithms can be effective tools for identifying specific sounds in audio recordings, however, the complexity and variability of field data can remain significant challenges to data analysis. By using user identification in the clips, we can ensure the accuracy of the audio recording that will be used for the machine learning model.

## 5.3 Future Work

There is a great deal of potential for this application to advance bioacoustics and the larger ornithology community in the future. One area of future work is to continue refining the algorithm's accuracy and efficiency, by incorporating machine learning techniques and optimizing the event detection and audio segmentation processes. The machine learning component of the system will be able to identify bird species and possibly even individual birds with increasing accuracy as more audio data is gathered and verified. This could lead to new insights into the behavior and distribution of bird species and help inform conservation efforts. Additionally, the user input component of the application has the potential to facilitate collaboration among ornithologists and other researchers, enabling them to share and verify data more easily. Overall, this application has the potential to streamline the process of collecting and analyzing bird vocalization data, ultimately contributing to a better understanding of avian biodiversity and ecology, and aiding in conservation efforts to protect these important species in Central Oregon and elsewhere.

## 5.4 Reflections and Insights

We learned many important lessons in our first and unsuccessful approach developing the first version of the AAS algorithm. Making assumptions about data without conducting a thorough investigation can be a costly mistake. The complexity and variability of field data can present significant challenges to data analysis, and any premature assumptions may result in errors that can have serious implications. Therefore, it is crucial to conduct thorough research and analysis on field data to ensure its accuracy, reliability, and generalizability as this information serves as a foundation for informed decision-making and practical applications in various fields. Had we initially recognized the presence of background noise in nearly all of the audio clips, we could have allocated our research efforts towards developing an effective method to mitigate this issue instead of expending valuable time on a method that ultimately proved to be ineffective.

The AAS algorithm we eventually came up with was effective in identifying bird call events accurately around 86% of the time. Although this algorithm could not serve as a stand alone solution, the user input provided to correctly identify the accurate audio clips is an effective method to input only relevant data to the machine learning algorithm. From this experience, we can also learn the importance of adaptability and flexibility in data analysis. As we encountered unexpected obstacles, we had to adjust our methods and strategies to effectively address the challenges at hand. This required a willingness to reevaluate our assumptions and preconceptions, as well as the ability to think creatively and innovatively to develop alternative solutions. By being adaptable, we were ultimately able to improve the accuracy and reliability of our analysis, and produce more meaningful insights and findings. As technology continues to evolve, the potential applications for these fields will only continue to expand, and the insights gained from this project will be invaluable for future research and development.

## References


[1] Taiwo Oladipupo Ayodele. "Types of machine learning algorithms". In: *New advances in machine learning* 3 (2010), pp. 19–48. DOI: 10.5772/9385.

[2] Elham Babaee et al. "An overview of audio event detection methods from feature extraction to classification". In: *Applied Artificial Intelligence* 31.9-10 (2017), pp. 661–714. DOI: 10.1080/08839514.2018.1430469.

[3] P Ashok Babu, V Siva Nagaraju, and Rajeev Ratna Vallabhuni. "Speech Emotion Recognition System With Librosa". In: *2021 10th IEEE International Conference on Communication Systems and Network Technologies (CSNT)*. IEEE. 2021, pp. 421–424. DOI: 10.1109/CSNT51715.2021.9509714.

[4] Michael S Bernstein et al. "Soylent: a word processor with a crowd inside". In: *Proceedings of the 23nd annual ACM symposium on User interface software and technology*. 2010, pp. 313–322. DOI: 10.1145/1866029.1866078.

[5] Ekaba Bisong and Ekaba Bisong. "Google colaboratory". In: *Building machine learning and deep learning models on google cloud platform: a comprehensive guide for beginners* (2019), pp. 59–64. DOI: 10.1007/978-1-4842-4470-8.

[6] Tim M Blackburn and Kevin J Gaston. "The distribution of body sizes of the world's bird species". In: *Oikos* (1994), pp. 127–130. DOI: 10.2307/3545707.

[7] RHJ Brown. "The flight of birds". In: *Biological Reviews* 38.4 (1963), pp. 460–489. DOI: 10.1111/j.1469-185X.1963.tb00790.x.

[8] Timothy J Brown and Paul Handford. "Why birds sing at dawn: the role of consistent song transmission". In: *Ibis* 145.1 (2003), pp. 120–129. DOI: 10.1046/j.1474-919X.2003.00130.x.

[9] Fabrizio Carcillo et al. "Combining unsupervised and supervised learning in credit card fraud detection". In: *Information Sciences* 557 (2021), pp. 317–331. ISSN: 0020-0255. DOI: 10.1016/j.ins.2019.05.042.

[10] Rich Caruana and Alexandru Niculescu-Mizil. "An empirical comparison of supervised learning algorithms". In: *Proceedings of the 23rd International Conference on Machine Learning*. New York, NY, USA: Association for Computing Machinery, 2006, pp. 161–168. ISBN: 1595933832. DOI: 10.1145/1143844.1143865.

[11] M Emre Celebi and Kemal Aydin. *Unsupervised learning algorithms*. Vol. 9. Springer, 2016. DOI: 10.1007/978-3-319-24211-8.

[12] Patrick J Donnelly, Daniel De Leon, and David K Mellinger. "Detecting Ultrasonic Harbor Porpoise Clicks with Hetrodyning and Teager–Kaiser Energy Operator". In: *Proceedings of the 7th International Conference on Frontiers of Signal Processing (ICFSP)*. IEEE. 2022, pp. 142–146. DOI: 10.1109/ICFSP55781.2022.9924846.

[13] Issam El Naqa and Martin J Murphy. *What is machine learning?* Springer, 2015. DOI: 10.1007/978-3-319-18305-3_1.

[14] Frank Emmert-Streib et al. “An introductory review of deep learning for prediction models with big data”. In: *Frontiers in Artificial Intelligence* 3 (2020), p. 4. DOI: 10.3389/frai.2020.00004.

[15] Shawn Hershey et al. “The benefit of temporally-strong labels in audio event classification”. In: *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2021, pp. 366–370. DOI: 10.1109/ICASSP39728.2021.9414579.

[16] Geoffrey Hinton, Yann LeCun, and Yoshua Bengio. “Deep learning”. In: *Nature* 521.7553 (2015), pp. 436–444. DOI: 10.1038/nature14539.

[17] Nathalie Jaquet, Stephen Dawson, and Lesley Douglas. “Vocal behavior of male sperm whales: Why do they click?” In: *The Journal of the Acoustical Society of America* 109.5 (2001), pp. 2254–2259. DOI: 10.1121/1.1360718.

[18] Peter Joosten and Christian Tiberius. “Lambda: Faqs”. In: *GPS Solutions* 6 (2002), pp. 109–114. DOI: doi.org/10.1007/s10291-002-0004-8.

[19] Bilal Nawaz Kayani. “Characterising sheep vocals using a machine learning algorithm: A thesis submitted in partial fulfilment of the requirements for the Degree of Master of Applied Science at Lincoln University”. PhD thesis. Lincoln University, 2021.

[20] Sloan Kelly and Sloan Kelly. “What Is Python?” In: *Python, PyGame and Raspberry Pi Game Development* (2016), pp. 3–5. DOI: 10.1007/978-1-4842-2517-2_2.

[21] Matthew Lease. “On quality control and machine learning in crowdsourcing”. In: *Proceedings of the 11th AAAI Conference on Human Computation*. AAAIWS’11-11. 2011, pp. 97–102.

[22] Philip Lieberman. “Comment on “Monkey vocal tracts are speech-ready””. In: *Science Advances* 3.7 (2017), e1700442. DOI: 10.1126/sciadv.1700442.

[23] Dragos Margineantu, Weng-Keen Wong, and Denver Dash. “Machine learning algorithms for event detection”. In: *Machine Learning* 79.3 (2010), p. 257. DOI: 10.1007/s10994-010-5184-9.

[24] Brian McFee et al. “librosa: Audio and music signal analysis in python”. In: *Proceedings of the 14th python in science conference*. Vol. 8. 2015, pp. 18–25. DOI: 10.25080/Majora-7b98e3ed-003.

[25] Wes McKinney. *Python for data analysis: Data wrangling with Pandas, NumPy, and IPython*. “ O’Reilly Media, Inc.”, 2012.

[26] David K Mellinger. “A comparison of methods for detecting right whale calls”. In: *Canadian Acoustics* 32.2 (2004), pp. 55–65.

[27] Annamaria Mesaros et al. "Sound event detection: A tutorial". In: *IEEE Signal Processing Magazine* 38.5 (2021), pp. 67–83. DOI: 10.1109/MSP.2021.3090678.

[28] Alexander Mielke and Klaus Zuberbühler. "A method for automated individual, species and call type recognition in free-ranging animals". In: *Animal Behaviour* 86.2 (2013), pp. 475–482. DOI: 10.1016/j.anbehav.2013.04.017.

[29] Claude Montacié and Marie-José Caraty. "A silence/noise/music/speech splitting algorithm". In: *Fifth International Conference on Spoken Language Processing*. 1998, paper 1141. DOI: 10.21437/ICSLP.1998-597.

[30] Lawrence Neal et al. "Time-frequency segmentation of bird song in noisy acoustic environments". In: *2011 ieee international conference on acoustics, speech and signal processing (icassp)*. IEEE. 2011, pp. 2012–2015. DOI: 10.1109/ICASSP.2011.5946906.

[31] Searcy Nowicki, William Searcy, and S Peters. "Brain development, song learning and mate choice in birds: a review and experimental test of the" nutritional stress hypothesis"". In: *Journal of Comparative Physiology A* 188 (2002), pp. 1003–1014. DOI: 10.1007/s00359-002-0361-3.

[32] Stephen Nowicki and Peter Marler. "How do birds sing?" In: *Music Perception* 5.4 (1988), pp. 391–426. DOI: 10.2307/40285408.

[33] Junhyuk Oh et al. "Discovering reinforcement learning algorithms". In: *Advances in Neural Information Processing Systems* 33 (2020), pp. 1060–1070. DOI: 10.5555/3495724.3495814.

[34] Simona L Perry, DOUGLAS P DeMASTER, and Gregory K Silber. "The great whales: History and status of six species listed as endangered under the US Endangered Species Act of 1973". In: (1999).

[35] Danilo Poccia. *AWS Lambda in Action: Event-driven serverless applications*. Simon and Schuster, 2016.

[36] CR Raquel, KMA Alarcon, and LL Figueroa. "Image classification of Philippine bird species using deep learning". In: *Theory and Practice of Computation*. CRC Press, 2019, pp. 92–98. DOI: 10.1201/9780429261350-10.

[37] Saman Sarraf and Mehdi Noori. "Multimodal deep learning approach for event detection in sports using Amazon SageMaker". In: *AWS Machine Learning Blog* (2021).

[38] Peter Sbarski and Sam Kroonenburg. *Serverless architectures on AWS: with examples using Aws Lambda*. Simon and Schuster, 2017.

[39] Brian L Sullivan et al. "The eBird enterprise: An integrated approach to development and application of citizen science". In: *Biological conservation* 169 (2014), pp. 31–40. DOI: 10.1016/j.biocon.2013.11.003.

[40] Aakanksha Tashildar et al. “Application development using flutter”. In: *International Research Journal of Modernization in Engineering Technology and Science* 2.8 (2020), pp. 1262–1266. DOI: 10.5121/csit.2023.130509.

[41] Jake VanderPlas. *Python data science handbook: Essential tools for working with data*. “ O’Reilly Media, Inc.”, 2016.

[42] Jennifer Wortman Vaughan. “Making Better Use of the Crowd: How Crowdsourcing Can Advance Machine Learning Research.” In: *J. Mach. Learn. Res.* 18.1 (2017), pp. 7026–7071. DOI: 10.5555/3122009.3242050.

[43] Mario Villamizar et al. “Infrastructure cost comparison of running web applications in the cloud using AWS lambda and monolithic and microservice architectures”. In: *2016 16th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid)*. IEEE. 2016, pp. 179–182. DOI: 10.1109/CCGrid.2016.37.

[44] Linda Wilbrecht and Fernando Nottebohm. “Vocal learning in birds and humans”. In: *Mental retardation and developmental disabilities research reviews* 9.3 (2003), pp. 135–148. DOI: 10.1002/mrdd.10073.