Acoustic monitoring of wildlife in inaccessible areas and automatic detection of bird songs from continuous recordings

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Emily Vera Hockman

May 2018

Acknowledgments

I would like to express my appreciation to the Department of Defense's Environmental Technology Demonstration and Validation Program (ESTCP) for funding this project. Additional thanks go to the range control operations and environmental division staff at Fort Riley, Fort Bragg, and Jefferson Proving Grounds. We always felt welcome and supported in otherwise intimidating field situations.

I would also like to thank the Forestry, Wildlife, and Fisheries Department at UT for their continued support of this project and graduate students involved. We could not have completed our work without the additional help and support the entire department provided. John Wilkerson, Stacy Worley, and David Smith from the UT BESS Sensors and Controls Lab also did an amazing job updating our AAARS platform and an even more amazing job teaching us how to repair them on the fly. Much thanks go out to the lab for their help and patience.

Lastly, a special thank you to my major professor Dr. David Buehler for the encouragement and patience he has shown me through the years at UT. I truly appreciate your guidance, and hope I can make you proud in the years to come.

Abstract

The use of new technology for wildlife monitoring comes with both possible benefits and challenges. Unmanned aerial vehicles (UAVs) and automatic recording units (ARUs) can allow researchers to automatically record videos, photographs, and audio recordings of animals in unusual or inaccessible locations. However, new acoustic monitoring techniques require innovative methods to extract and utilize data from acoustic recordings. In this project we developed novel technology to record bird songs in inaccessible areas and demonstrated a useful method for extracting and classifying songs from continuous recordings. The autonomous aerial acoustic recording system (AAARS) was a UAV developed at the University of Tennessee capable of generating high-quality WAV recordings of bird songs in a variety of landscapes. The AAARS was completely silent in flight controlled by a ground-based computer monitoring station. I developed a model to convert the AAARS GPS-based flight path into a microphone exposure surface to relate species-specific acoustic signals recorded to area of microphone coverage. The vocalizations per unit area per unit time for a given focal species could then be used as an index of relative abundance or as an input in density estimation. Once collected, extraction and classification of birdsongs from acoustic recordings remains a major technological challenge. I used quadratic discrimination analysis to differentiate between inter- and intra-specific bird songs using up to sixteen acoustic measurements on human-extracted signals from audio spectrograms of five focal songbird species. Measurement-based classification was successful at separating the five species apart with only ≤5% classification error. I

then used a template-matching model to extract target birdsongs from continuous field recordings and investigated the efficiency of different analytical options for classification of five focal songbird species. Decision trees, neural networks, and quadratic discriminant analysis all produced similar classification results. The means to optimize the analytical approach varied by species. I concluded that a species-specific approach should be used to accurately extract and classify songs from continuous recordings.

Table of Contents

List of Tables

List of Figures

Introduction

U. S. military lands contain more threatened and endangered species than any other federal agency, nearly three times more than on lands managed by the U. S. National Park Service. Large areas of many military installations are designated impact zones, which largely prohibit human access and often are subject to periodic burning. The lack of human development and activity together with periodic burning can lead to ideal conditions for disturbance-loving wildlife that are otherwise declining in the area. Because access to these impact zones is limited to non-existent, we developed and validated an unmanned aerial acoustic vehicle (UAV) capable of recording acoustic data from animal vocalizations to monitor wildlife populations in inaccessible areas. We demonstrated a method for generating density estimates using songbird cue counts and showed that our approach attained density estimates of comparable accuracy and precision as human-based methods. In Chapter One of this dissertation, I describe the components of the UAV technology that we developed and present the results from our validation trials that document system performance in field applications at three U. S. military installations.

Technology such as our UAV and other autonomous recording devices can facilitate the collection of large volumes of acoustic monitoring data. However, extensive audio recordings in natural soundscapes often contain overlapping anthropomorphic or biotic signals which make analysis problematic. There are automated methods for extraction and classification of songs and calls for different groups of species from such recordings, but the variation in vocal signals and the

presence of “noise” in natural soundscapes make it difficult for one method to work accurately for all species.

I developed an analytic pathway for extraction and classification of birdsongs that was versatile enough to work on multiple avian species with different song types in typical field recordings collected from three U. S. military installations. In Chapter Two of the dissertation, I describe preliminary steps for the pathway in which I used quadratic discriminant analysis to identify a suite of 11 to 19 uncorrelated acoustic measurements. These measurements yielded ≥95% correct classification of five grassland and shrub-scrub bird species songs from human-extracted song clips.

In Chapter Three of the dissertation, I describe the results of the two-step analytic pathway. For the first step, I used a template matching extraction method to identify signals from continuous recordings. I used the same 11-19 measurements from Chapter Two to classify the bird songs to species in the second step. I evaluated how different options in this analysis affected classification accuracy for the five focal song bird species. Finally, in the final chapter of the dissertation, I draw conclusions about the utility of our approach in collecting acoustic data on wildlife populations in inaccessible areas and efficiently analyzing such data.

Chapter 1: Description and functionality of the autonomous aerial acoustic recording system for monitoring vocal wildlife in inaccessible areas

Abstract

Recent studies have implemented passive acoustic recording to monitor vocal wildlife populations either in conjunction with traditional methods or as a stand-alone approach. However, applying these new acoustic-based techniques becomes difficult when species of interest don't vocalize frequently, are rarely encountered, or are found in remote, inaccessible areas. We developed an autonomous aerial acoustic recording system (AAARS) for monitoring avian populations in inaccessible areas. We designed an instrument payload which incorporated an acoustic recorder and microphone. The system was transported by a helium weather balloon along an aerial transect determined by the prevailing winds while we monitored the system altitude and location from ground-based computer base stations. In 2012-2014 we tested the system in different landscape and environmental conditions by conducting 320 free flights and 582 tethered flights which recorded avian vocalizations. Free flights were successful 85% of the time and only three units were unrecoverable. We developed an algorithm for density estimation using avian cue counts from recordings of five different songbird species with varying song characteristics. Validation of the technology has shown that the AAARS provides an accurate and economical means to monitor vocal wildlife populations in remote or otherwise inaccessible areas.

Introduction

Conservation of wildlife populations requires information on actual density of the population or an index to relative abundance, so that the success of the conservation efforts can be assessed as populations respond to management. One method of wildlife monitoring currently increasing in application uses passive audio recording of vocal species (Marques et al. 2013). Passive recorders can be installed in an area and left for a period of time to create a permanent record of the soundscape, and can record massive amounts of data with minimal effort and limited expense. Passive approaches are especially useful for species which may emit more vocal cues than visual ones or are otherwise rarely available for counting using traditional methods (Lambert and McDonald 2014, Zwart et al. 2014, Nowacek et al. 2016). Stationary passive acoustic recordings can also maximize the amount of field data obtained while minimizing human effort when time is limited for traditional observer methods for even common or highly vocal species.

As acoustic monitoring technology has become more widely available, more complex monitoring approaches have developed, including the use of multiple microphone arrays (Yu et al. 2013, Stepanian et al. 2016) and unmanned motorized or moveable platforms (e.g., drones) (Klapel 2014, Chabot and Bird 2015, Ivošević et al. 2015). These systems have the potential to take passive recording to the next level in terms of total monitoring area covered, accurate density estimates, and monitoring inaccessible locations. For example, cetacean studies now commonly use large grids of hydrophones to pinpoint call locations of individuals in 3D space over large areas

(Roy et al. 2010, von Benda-Beckmann et al. 2010, Martin et al. 2013). However, many of these complex methods are not as useful in terrestrial systems due to the limited distance that sound waves travel in air and also because many target species maintain relatively small home-ranges. Unmanned Aerial Vehicles (UAV's) like commercial drones have been used primarily for remote sensing and photographic population monitoring of large mammals or sea birds to date (Chabot and Bird 2015, Goebel et al. 2015, Ivošević et al. 2015). These types of recording platforms aren't currently used for acoustic monitoring because they create large amounts of noise, are expensive, sometimes require extensive man-hours to deploy, and may require special FAA permits to operate. Commercially available drones are often limited by battery life, effective range, and are required to be flown strictly within the pilot's line-of-sight. To date, drones have only been tested to cover small-scale point counts for avian species (Wilson et al. 2016).

We took proven concepts from these stationary, towed, and flying acoustic monitoring methods to develop a free-flying acoustic monitoring platform (Figure 1.1). The autonomous aerial acoustic recording system (AAARS) was originally conceived by researchers with the Cornell Lab of Ornithology to fly over inaccessible areas on Fort Hood, TX and record vocalizations of two endangered songbirds (Fristrup and Clark 2009). We updated the system to take advantage of advances in associated technology and developed a novel approach for estimating target bird population densities based on AAARS flight area-exposure time and avian cue counts.
The AAARS has several advantages over other aerial designs, primarily in that it doesn't use a noisy, self-propelling motor but simply drifts quietly with the prevailing

winds. This silent flight allows for the collection of high-quality audio recordings free from unwanted external noise which could otherwise interfere with target signals from the birds. The AAARS is also simple and less expensive to operate than alternative designs. Lastly, the AAARS weighs less than 2 kg, is designed to operate below 300 m altitude, incorporates a weather balloon in its design, and is deployable without typical FAA restrictions that cover self-propelled unmanned aircraft. The four major functions of the AAARS payload are to record audio recordings of target avian species, track the 3D location and speed of the system, communicate effectively between the payload and computer-based ground monitoring stations, and effectively retrieve the system at the termination of a flight.

In this chapter, then I provide a general introduction to the AAARS technology and address the following specific objectives:

1. Document main components and system performance of the AAARS
2. Describe the AAARS technology for collecting acoustic data along a moving transect
3. Discuss constraints of the system
4. Describe our approach for data analysis related to density estimation

AAARS Description and Functionality

AAARS technology has been developed and improved through funding provided by the Department of Defense Environmental Security Technology Certification Program (ESTCP Project # RC-201112) and the University of Tennessee Knoxville. The main purpose of this project was to develop a means to monitor threatened and

endangered birds within inaccessible areas. Demonstration and validation flights were conducted from May 2011 – August 2013 in Fort Bragg, NC, Jefferson Proving Grounds, IN, and Fort Riley, KS.

Variation in both vegetation and bird species across installations facilitated in-depth testing of the communication, recording, and retrieval methods related to the AAARS payload. Fort Bragg is characterized by open longleaf pine (*Pinus palustris*) savanna, Big Oaks is a mosaic of open fields and deciduous woodlands, and Fort Riley is largely native tallgrass prairie in an open agricultural landscape. We selected avian species with a variety of song/call characteristics to validate the recording system, including Bachman's Sparrow (*Peucaea aestivalis*), Field Sparrow (*Spizella pusilla*), Grasshopper Sparrow (*Ammodramus savannarum*), Henslow's Sparrow (*Ammodramus henslowi*), Northern Bobwhite (*Colinus virginianus*), and Prairie Warbler (*Setophaga discolor*). All but two of these species occurred on a least two of the study sites so that we could compare the ability of the AAARS to detect regional dialects in different landscape contexts.

System Components

The AAARS was a free-floating device with a helium weather balloon lifting system that was controlled from the ground via a laptop computer interface with a radio-communication link to the payload. The four major components included altitude controls, GPS location system, ground communication, and audio recording equipment (Figure 1.2, Figure 1.3). GPS, battery, and micro-processing components were updated

in 2014 to take advantage of greater processing power, lightweight components, and efficiency optimization.

The onboard microprocessor (Parallax Propeller, Parallax Inc., Rocklin, CA) was an 8-core, 32-bit processor with 32 digital I/O pins and 64 KB of RAM/ROM memory. Each of the processor cores had access to the shared RAM, ROM, and I/O pins while executing independent programs. This microprocessor allowed each core to be assigned to a different task: e.g., one core each for payload-to-ground station communication, GPS module communications, servo valve control, ballast system control, and a watchdog to reboot the entire system in the unlikely event the microprocessor malfunctioned. Multiple cores allowed for uninterrupted communication with the PC-based ground station while other operations were being performed. The additional memory, processing power, and I/O pins allowed complete GPS data strings to be collected each second, robust communications protocols to ensure accurate and complete data transmission, and additional functions such as a watchdog program, servo valve position monitoring, ballast dropping system, and automatic recovery based on GPS position. The 8-core microprocessor expanded the system capability substantially, and allowed for considerable flexibility in future improvements with little or no changes to the hardware.

Communication between the AAARS and ground station were based on 100 mW – 1-Watt XTend 900 MHz RF modules (Digi International Inc., Minnetonka, MN). The ground module was attached to a 10-m telescoping mast with an attached directional Yagi antenna that rotated to follow the balloon in flight, while the AAARS payload module used an omni-directional ‘rubber duck’ style RF antenna. The ground module

allowed the user to record real-time location data, manipulate the on-board altitude controls, track the AAARS during flight, and initiate the recovery protocol from a PC-based laptop computer with a custom control program developed with Labview systems engineering software (National Instrument, Austin, TX).

A Trimble Copernicus IIA module (Trimble Inc., Sunnyvale, CA) with active patch antenna generated position and altitude data used to provide spatial correlation of the audio data. GPS data strings were collected and stored every second while the AAARS was in flight. The module had WAAS correction capability which improved positional accuracy and was programmed to maximize elevational accuracy. The TSIP protocol between the processor and GPS module improved the quality of the altitude measurements, and the updated microprocessor also allowed for more detailed GPS information to be relayed to the ground stations (i.e. heading, vertical velocity, horizontal velocity, and fix quality information).

We controlled altitude of the AAARS by venting helium through a custom valve developed by T. Fowler (Cornell Lab of Ornithology) and dropping ballast through a custom ballast dropper system we designed. Venting helium and dropping ballast allowed for controlled rise of the AAARS to a target altitude and the means to adjust the buoyancy of the system while in the air to maintain that altitude over the duration of the flight. Both systems were controlled from the ground base station with commands contained in the Labview control program.

AAARS lift was provided from a 300-g meteorological balloon attached to a custom helium valve at the top of the AAARS payload. Valve positions (fill, close, vent, or full dump) were controlled by an analog servo motor (Hitec RCD USA, Inc., Poway,

CA). Free flight preparation involved filling the balloon until it retained 2-4 oz of lift. The extra helium was vented when the AAARS was 50-100 m from the ground, slowing the ascent until it achieved neutral buoyancy and making it possible to achieve a consistent flight at the target altitude. Altitude was maintained at the target value by incrementally venting helium and/or dropping ballast as needed and monitoring balloon response in the flight tracking software on the laptop computer. At the terminus of each flight a command was given from the base station to move the valve into the 'dump' position to vent helium and rapidly bring the AAARS down to enable recovery.

The ballast dropper consisted of a custom PCB board populated with a circle of pegs connected to each other by nichrome wire. Nontoxic ¼-2 oz. weights were attached to each nichrome wire loop with 6x monofilament line (Figure 1.4). A ballast command from the base station would send electric current to heat up a specific nichrome wire loop, burn through the monofilament line at that loop and drop the specific ballast desired to increase lift of the AAARS during a flight.

Audio recordings were created using a commercially available digital Zoom H2 digital recorder (Zoom North America, Hauppauge, NY) and an active directional microphone (PA3-IL, Super Circuits Inc., Austin, TX) with an attached directional cone. The cone surrounded the microphone and provided downward directionality and significant noise-free amplification (~20 dB). The H2 recorded uncompressed WAV files directly onto SD memory cards that were removable for download to the ground computers upon successful recovery of the AAARS. The H2 recorder also had ten levels of input amplification that provided flexibility for recording vocalizations with either different call intensities or recording at different flight altitudes while maintaining the

ability to detect signals from target species. We kept the amplification at level 5 for all flights for consistency during validation flights.

Flight tracking and control were achieved through a custom program developed by S. K. Worley in LabVIEW software (National Instruments, Austin, TX). The program was designed to run on a laptop PC computer running Windows 7 or more recent operating systems. The control program contained four main user interface screens: pre-flight, altitude, flight path visualization, and communication data. The pre-flight interface was used to set all flight parameters before launch. The altitude interface plotted the real-time altitude (Figure 1.5). The flight path interface tracked the real-time position of the AAARS against an aerial basemap that was loaded on the laptop prior to the flight. The communications interface allowed the user to view the communication data that were being transmitted to and from the AAARS. All four user interface screens included the main AAARS valve controls, flight velocity data, and heading data and were accessible to the user for flight control.

AAARS Recovery System

The 2014 AAARS design required the payload to be recovered for the acoustic data to be downloaded from the H2 recorder. Several redundant systems were designed to ensure a safe recovery of the AAARS at the end of the flight or to bring it down if communication was compromised or lost. The simplest way of putting an AAARS into recovery mode was to use the computer command from the base station. Putting the payload into recovery mode automatically dumped all helium through the

valve and initiated an audible alarm ‘beep’ every second which could be heard up to 2 km in open grasslands. However, potential issues related to computer malfunction, loss of radio communication, and signal power warranted alternate methods for balloon control and/or data storage.

Back-up systems were established to automatically control the flight and save GPS data in the event of loss of communication between the computer and the AAARS. We developed a program for the microprocessor that defined a coordinate ‘box’ for each flight which represented the extreme boundaries of the target area for the AAARS. Once the AAARS location exceeded the coordinates of the box in any direction, it was programmed to go into recovery mode, dump helium, and descend regardless of any command from the ground computer. All incoming GPS data were additionally recorded internally onto a mini SD card as back-up if communications were lost with the base station computer. Finally, a nichrome wire was placed around the neck of the balloon prior to flight. A command from the base station computer allowed the wire to heat up and burn a 1-cm hole through a portion of the neck, accelerating the descent of the AAARS if the helium release valve malfunctioned.

Grass and shrubland vegetation had little effect on recovery, however retrieval of the AAARS in a wooded area was more difficult (Figure 1.6). The longleaf pine at Fort Bragg, NC often ‘caught’ the AAARS in the treetops during decent. We used a variety of approaches to retrieve the payload if the payload was lodged in a tall tree. Use of a thinner-walled helium balloon also aided in recovery, as the thinner balloons were more likely to pop on entry into the canopy. We used a separate power source for the Zoom H2 recorder and microphone (i.e.., not connected to the main PCB board) to ensure that

the audio data were not lost if the main power was interrupted during a rough landing. A foam section in the bottom of the AAARS absorbed shock during landing, and minimized damage to more sensitive internal components.

Microphone Exposure

During flight the AAARS defined an acoustic monitoring transect (footprint): the outlined area where potential acoustic signals were recorded. The acoustic footprint was defined by the flight path, the altitude of the flight, combined with the cone angle of the microphone. We measured the acoustic footprint empirically by deploying a grid of remote-controlled speakers that played unique bird songs at realistic amplitudes. Song clips for each species were chosen from random from the Macaulay Library of the Cornell Lab of Ornithology. Each clip had a unique 0.5-sec tone added after the song so that individual songs could be differentiated during analysis. We flew the AAARS through the speaker grid at various altitudes while the speakers were broadcasting unique bird songs to determine the functional edge of the recording for the AAARS during free-flights for each target species. Audio recordings from flights over speaker grids were analyzed individually by species to determine the maximum horizontal distance from the source of each speaker to the AAARS at a variety of flight altitudes. This relationship between the horizontal to vertical distance at flight altitudes differed by species due to song characteristics, vegetation, and sound attenuation; higher frequency sounds and trills don't carry as far as sustained tones or lower frequency sounds do (Aylor 1972, Marten and Marler 1977).

The aggregate footprint for each species resulting from microphone testing was the basis of all area-based flight exposure calculations. A footprint ratio for each flight was first calculated by dividing the radius (m) of the footprint on the ground representing the extent of the area that the microphone could pick up target songs given the average altitude of the AAARS average altitude (m) for the flight. This relationship changed as the microphone moved closer or further away from a sound source because both loud and very close vocalizations 'bleed' through the microphone cone at low altitudes, disproportionally increasing the radius. At greater altitudes, the radius decreased as the edges of the exposure circle were no longer available to the microphone.

I developed a flight-exposure program in ArcMap 10.1 (ESRI, Redlands, CA) that took the GPS flight data together with microphone exposure time and calculated the flight-exposure area. Flight area, exposure time, and species-specific cue count rates were all inputs into our calculations of target species density for each flight (Prevost 2016). Cue counts were used to estimate total number of male songbirds of a species by incorporating singing rates in the presence of interspecific competition as well as time of season and weather conditions. Abundance was then calculated using negative binomial regression modeling of song activity and produced the following equation documented in a different portion of this study:

$$\text{Abundance} = \exp\ \hat{}\ (\beta_s x_s + \beta_1 x_1 + \ldots \beta_k x_k)$$

In this equation, β_s is the regression coefficient for the total number of songs observed in a 5-minute time period, x_s is the total number of songs observed in the 5-minute time period, $\beta_1 \ldots \beta_k$ represent additional coefficients (abundance index, day of season, time

of day, temperature, wind speed, atmospheric pressure, study area, and year), and $x_1 \ldots x_k$ are additional covariate values (Prevost 2016).

I used the average altitude for each flight to determine the footprint ratio for each species and calculate the flight exposure pathway. For every point in an AAARS flightpath, I created an individual polygon that represented how much area the AAARS was recording with the microphone for that second in time, representing a measure of 'hectare seconds' on a species-specific basis (Figure 1.7). The polygons were converted to individual raster files with a constant value of one, and then joined to create a composite coverage of both the total area covered and the combined exposure times covered by the AAARS on a 1-m pixel by 1-m pixel basis for an entire flight. This approach accounted for the fact that some pixels on the periphery of the flight path had only limited microphone exposure time whereas pixels in the center of the flight path had maximum exposure time. Summing all these pixel exposure times allowed for the calculation of the area covered and the total exposure time per pixel recorded.

Results and system specifications

We flew n = 320 free flights over a variety of 25-ha target grids during summer 2012-2014. An additional 223 tethered 500-m line transect flights and 359 tethered 10-min point counts flights were conducted. Useable audio data were collected 85% of the time, with the major failures stemming from the H2 audio recorder losing power (n = 54), wind noise overlapping bird song frequencies (n = 31), human equipment error during pre-flight (n = 29), general equipment failure (n = 16), and other biological interference

(n = 6). Two payloads were lost during this period when tethers broke unexpectedly, and one was lost due to unexpected loss of communication with the base station. We implemented the GPS coordinate box during 109 free flights and experienced 4 errors (4.4% error rate) due to human failure in correctly delineating the coordinate box.

Launching, flying and recovery of the AAARS was limited to weather conditions where wind speeds were generally <15 km/h and the likelihood of precipitation was minimal, which were consistent with weather conditions typically used for bird monitoring (Bibby et al. 2000). Weather conditions in all three locations from May-July were variable and limited the total number of days available for monitoring. High wind speeds (i.e.., >15 km/h, heavy gusts, or precipitation prevented potential flights on 11% of the mornings at Big Oaks NWR (n = 30), 8% of the mornings at Fort Bragg (n = 21), and 34% of the mornings at Fort Riley (n = 92). Weather limitations were not prevalent in any one particular month (May-July), although 2013 had slightly fewer useable days compared to the other two years. Average AAARS velocity in the air varied by location, with the fastest velocity at Fort Riley, Kansas (9.2 m/s), then Fort Bragg, North Carolina (5.5 m/s), and finally Big Oaks, Indiana (3.8 m/s). The fastest horizontal velocity recorded during a successful flight was 18.6 m/s.

In this study we validated the AAARS exclusively on Department of Defense (DoD) property. As such, the access to study sites and the impact zones was a limitation for implementation of the technology. We conducted system validation in seldom-used areas to avoid conflict with military training. The ultimate use of the technology, however, will be based on flights over inaccessible areas/impact zones. With access and weather combined, Jefferson Proving Grounds had the best average

opportunity for flights over all three years (31 days out of a 90-day field season/year) (Table 1.1). Access to the impact zones had the greatest impact on total availability for AAARS monitoring. With weather alone, Fort Riley had the least amount of access with favorable days at 51% averaged over all three years (n = 137).

We tested AAARS communications capability in line-of sight tests at Fort Riley, KS in a relatively open prairie landscape. We monitored GPS data streams communicated between the AAARS and the base station while the AAARS was at 100-m altitude on tether and we increased the distance to the base station out to 20 km. Adequate communications in which complete GPS data strings were recorded were achieved at ≥16 km. Communication range can be extended by elevating the base station antenna. Reliable communication on flights was achieved in the pine savannas of North Carolina simply by elevating the Yagi antenna on a 10-m mast in small forest canopy openings. It is also possible to 'hand off' control of the AAARS between multiple base stations. Thus, there is potential to fly very long flights (e.g., >30 km) with the addition of one or more base stations strategically placed along an expected flight path.

Median altitude of all free flights was 143 m. Using this average altitude ± 10 m, the resulting area covered (footprint) differed widely by species. The footprint at 143 m of altitude was least for Henslow's Sparrows at 9.7 ha and 171 ha-sec of exposure, whereas the footprint was greatest for Northern Bobwhite at 61.8 ha with 2,170 ha-sec exposure, and the other species falling in between (Table 1.2). Results for all species showed a decreasing relationship between the horizontal footprint ratio on the ground and altitude (Figure 1.8). In other words, the radius of coverage decreased in relation to height as the average flight altitude increased. Some species' songs showed a greater

negative relationship between altitude and horizontal distance covered (footprint), such as the high-pitched trill of the Grasshopper Sparrow, whereas other species didn't have the same signal attenuation, like the sharp, hiccup-like chirp of the Henslow's Sparrow.

We designed the AAARS to take the place of humans when access, time, or observer ability was an issue. Using the results presented above assuming successful flights, in one morning the AAARS could complete four 10-km flights at the average 250-m altitude and 10 kph speed, surveying a total of 2000 ha and generating roughly 300 min of audio recordings. Three field assistants would be needed to complete this work. Those same three field assistants could complete ten 10-minute point counts each in the same morning using traditional methods, generating roughly the same amount of audio data. However, assuming a 100-m radius point count, this effort would only cover a total of 94 ha (3.14 ha/count * 10 point counts * 3 observers). Thus, the AAARS is capable of monitoring 21x the area of traditional point counts with the same amount of effort under ideal circumstances.

Discussion

The AAARS is a novel and effective way to monitor birds and other vocal species via aerial acoustic transects. The updated components of the system are effective over long ranges, and collect high quality audio data. Methods described here combine cue count rates and area estimations to generate relatively accurate density estimations of target populations. This aerial method fills in a gap not currently being explored with

other acoustic techniques, and opens valuable potential for studying otherwise inaccessible wildlife populations in the future.

The validation results demonstrated the effectiveness of the AAARS to collect useable data and generate population density estimates for a wide range of bird song types in a variety of landscapes and environmental conditions. The unique design allowed for silent flight across potentially long distances and collection of songs useful for further population estimation. Unlike drones or kites, the AAARS could be controlled far beyond normal line-of-sight and other regulatory restrictions. By keeping the payload weight below 2 kg and using a weather balloon as a lifting system, the AAARS can operate without FAA restrictions, where other unmanned devices (i.e. drones) cannot. The AAARS can be controlled for much longer flights than typical remote-controlled unmanned aerial vehicles, and could be deployed for long flights over inaccessible areas using multiple base stations. The single microphone used by the AAARS isn't compatible with studies that rely on animal movement or direction to generate density estimation (Barlow and Taylor 2005, Roy et al. 2010, Yu et al. 2013, Stepanian et al. 2016). However, the AAARS could be combined with current photography methods (Thome and Thome 2000, Goebel et al. 2015) to monitor individuals or colonies of species in the open in addition to audio recordings. We developed the AAARS to be modular in design to accommodate additional sensors as desired to facilitate data collection for different taxa; there is potential for further adaptations, such as the addition of a bat-sensor or streaming data collection in real-time.

The majority of flight failures in 2011-2013 stemmed from superficial issues that were addressed with slight modifications of the AAARS design during the course of the study. For example, the Zoom H2 recorders occasionally lost power on hard landings, which erased all audio data. Recorders were subsequently hard-wired into a stand-alone 9v battery source which fixed the problem. The majority of audio interference problems stemmed from using the system on tether, which was needed for validation purposes but was not necessarily an intended use of the technology. Cicada and some frog call interference accounted for the rest of the overlap, but was not a major cause of flight/audio failure. Removing the H2 data loss and tether frequency issues from the calculations revealed a 94% flight success rate overall.

Use of free-flying acoustic recording units can be extremely valuable in recording species in inaccessible areas or in maximizing the total surveyed area in a single day in accessible areas. For example, four 10-km flights could be flown in a morning (sunrise to 4 h after sunrise). Using the Prairie Warbler footprint ratio and average flight altitude as an example, these flights would cover ~2000 ha. By contrast five ~ 30 minute, 500-m line transects run by technicians with traditional distance sampling would cover only 50 ha if five could even be run in a single morning. Based on the example illustrated above, the AAARS may cover over 20 times the effective area that traditional human-based methods would cover given the same amount of effort. In both cases, additional labor must be expended to enter bird count data for human-based count data and to either transcribe the acoustic data or use an automated extraction and classification approach (see Chapter 3 below). In most cases, the time spent analyzing acoustic data recording far exceeds the time spent simply inputting data from human-based methods.

The added advantage of the AAARS is that a permanent record of the acoustic data can be archived for future potential use.

We have demonstrated how the AAARS technology can be used for producing an index to relative abundance (# of detections/unit area covered) or estimation of actual breeding bird densities based on incorporation of cue counts. To estimate actual densities, however, a field study of song behavior must be conducted to establish the relationship between song rates and number of singing males (abundance) for each target species. This cue rate study would ideally be completed during the period of deployment of the AAARS. Additional field validation is required to establish the effective footprint for target species prior to or during deployment of AAARS for monitoring.

Appendix

Figure 1. 1: The Autonomous Aerial Acoustic Recording System in flight, Fort Bragg, NC, 2012.

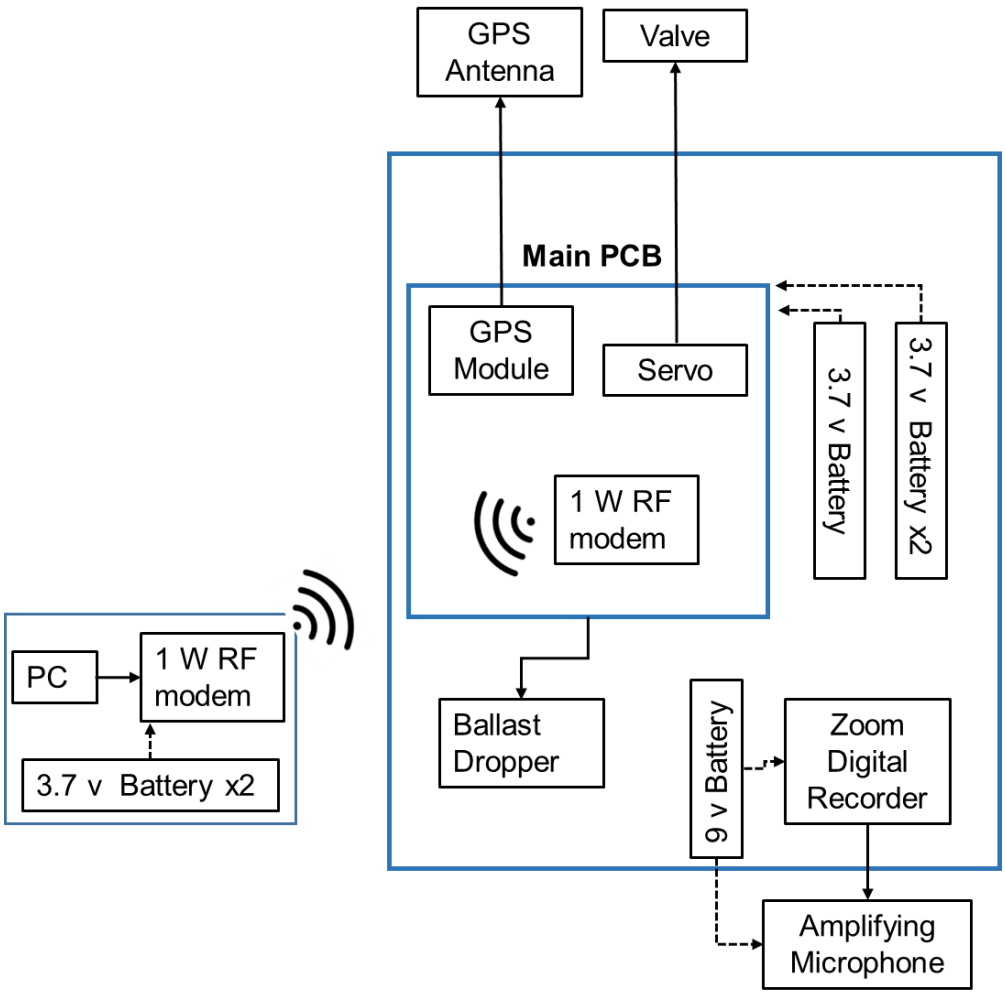

Figure 1. 2: Schematic of the Autonomous Aerial Acoustic Recording System showing major components and power sources of the payload (right) and the base station (left).

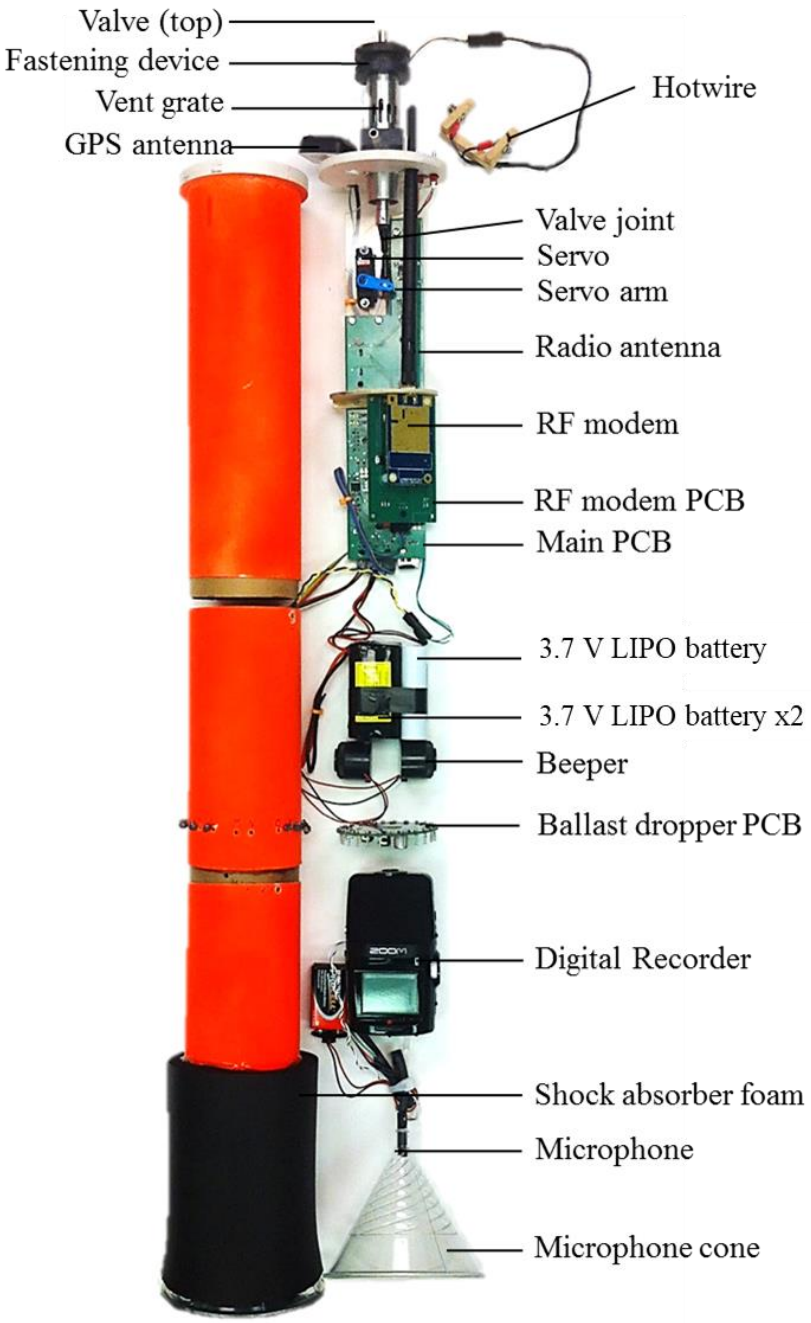

Figure 1. 3: The Autonomous Aerial Acoustic Recording System and its current components as of May 2014.

The components are inserted into the orange payload tube prior to launching.

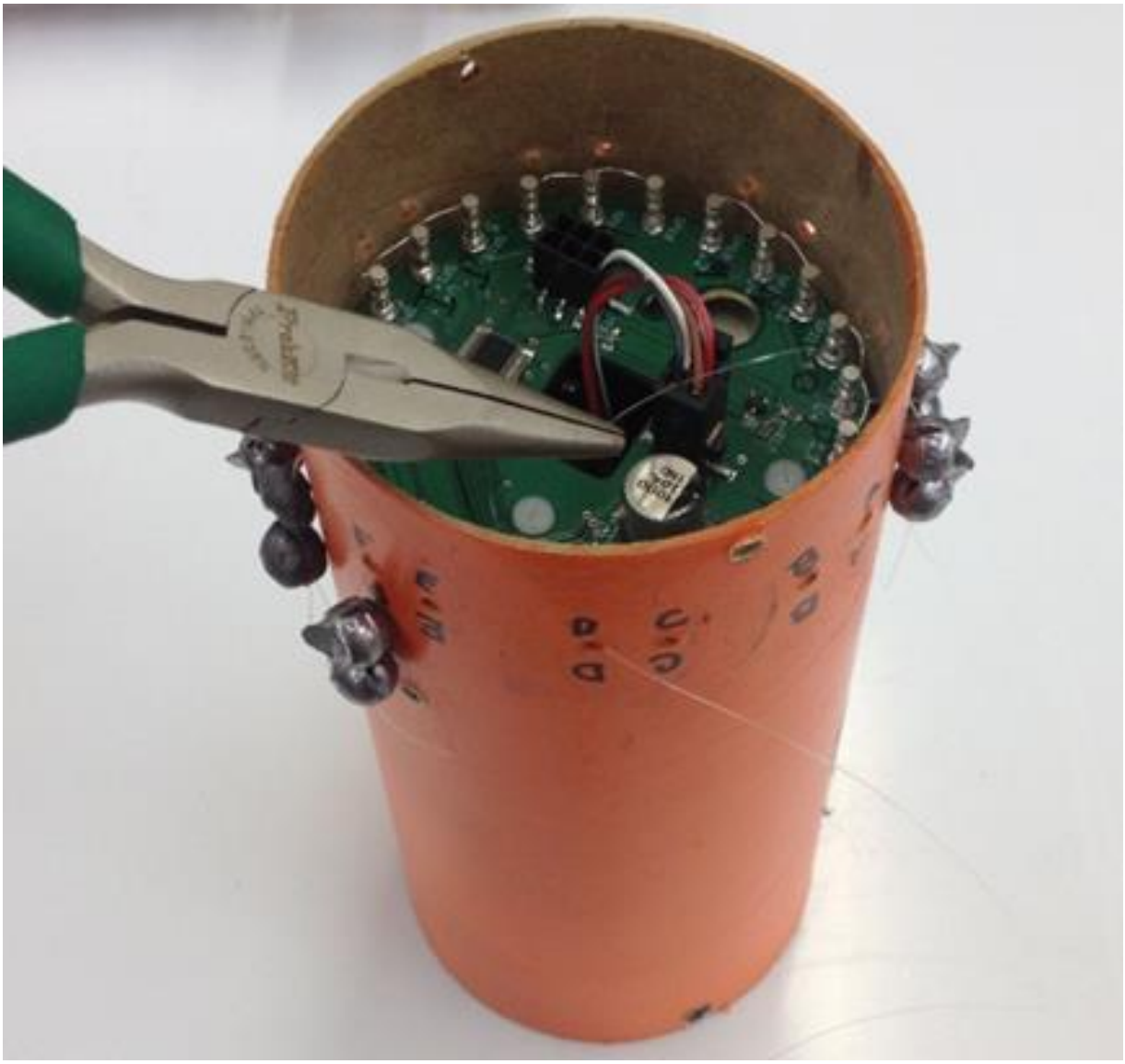

Figure 1. 4: Autonomous Aerial Acoustic Recording System ballast dropper system, showing nichrome wire and ballast weights attached with 6x monofilament line.

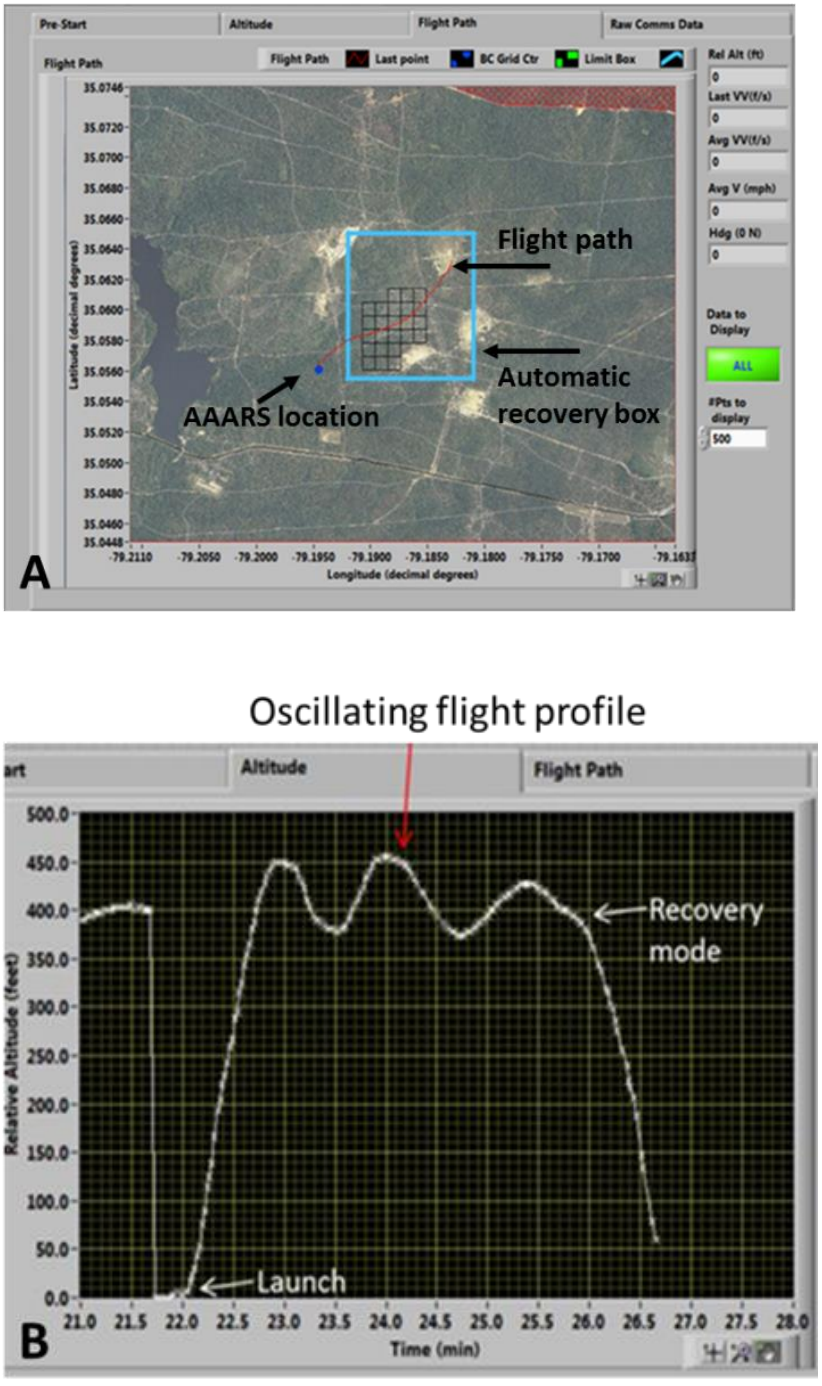

Figure 1. 5: Screen view of the base station laptop showing an example of real-time flight (A) and altitude (B) mapping in LabVIEW software of the Autonomous Aerial Acoustic Recording System.

Figure 1. 6: Autonomous Aerial Acoustic Recording System recovery in longleaf pine at Fort Bragg, NC 2012.

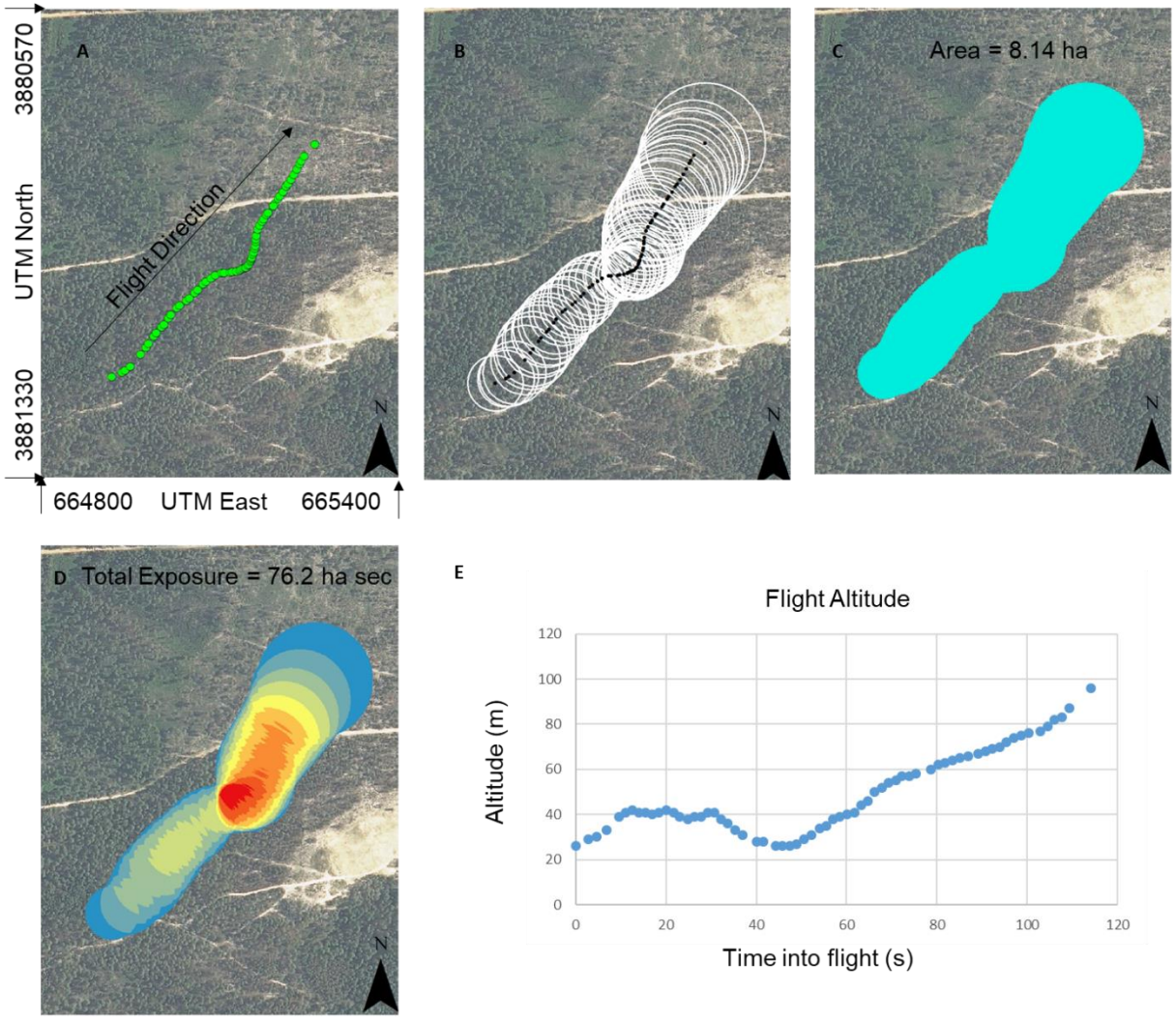

Figure 1. 7: Post-processing of flight GPS files including the 500-m flight path (A), individual buffers for each recorded Autonomous Aerial Acoustic Recording System location (B), total flight area covered (C), total flight exposure in ha sec (D), and altitude profile (E).

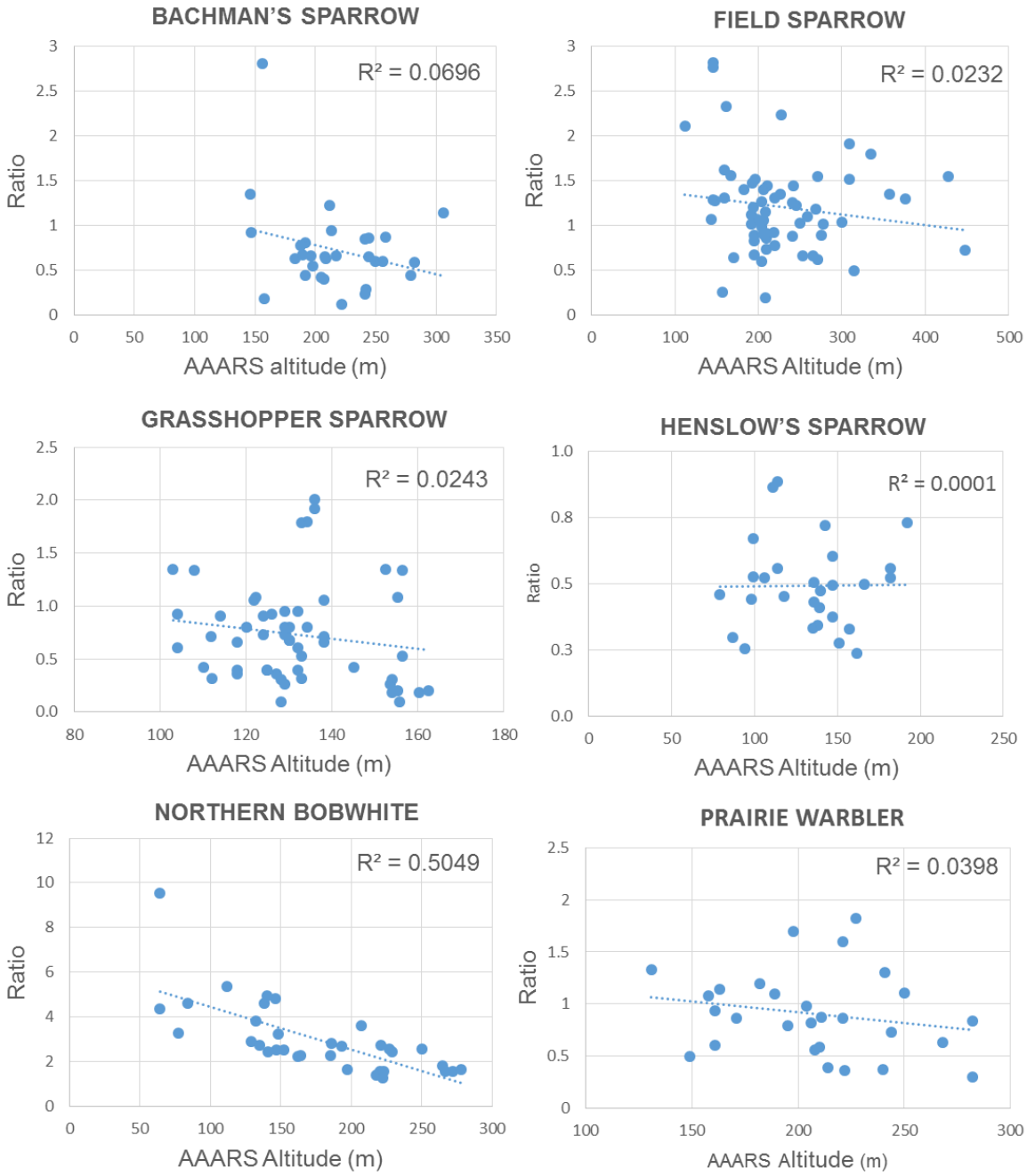

Figure 1. 8: Relationship between each target species' footprint ratio (horizontal footprint radius / Autonomous Aerial Acoustic Recording System altitude) and average altitude for the entire flight used in footprint analysis.

Table 1. 1: Total favorable days with weather and access necessary for flying the Autonomous Aerial Acoustic Recording System over impact areas on three military bases during 90-day field seasons (2011-2013).

Location	Year	Days with Favorable Wind Speed & Precipitation	Days with Favorable Wind Direction	Days with access	Total Favorable Days
Jefferson Proving Ground, IN	2011	89	90	36	34
	2012	89	90	33	33
	2013	67	90	34	20
Fort Bragg, NC	2011	90	84	12	12
	2012	89	86	12	12
	2013	76	78	12	12
Fort Riley, KS	2011	64	41	73	34
	2012	54	36	57	26
	2013	63	90	7	5

Table 1. 2: Average species-specific footprint area and exposure covered by the Autonomous Aerial Acoustic Recording System for 500 m flights with the overall average altitude (143m ±10m).

Species	**Area (ha)**	**Exposure (ha sec)**
Bachman's Sparrow	21.4 ± 1.1	628.6 ± 137.3
Field Sparrow	32.4 ± 1.7	1098.3 ± 242.4
Grasshopper Sparrow	13.4 ± 0.7	283.9 ± 59.35
Henslow's Sparrow	9.7 ± 0.5	171.5 ± 38.2
Northern Bobwhite	61.8 ± 3.2	2671.0 ± 595.4
Prairie Warbler	23.9 ± 1.2	700.6 ± 152.9

Chapter 2: Impact of dialects, song type, and species on classification of full songs extracted from continuous acoustic recordings of five grassland songbird species breeding in the eastern United States

Abstract

New acoustic monitoring technology has allowed for the collection of massive amounts of passive acoustic data. Variation inherent in passerine songs creates difficulties for automated extraction and classification of target songs for this group of highly vocal species. Annotating bird songs in recordings manually is time-consuming and negates many of the benefits of using passive recorders in the first place. We conducted a study to document the sources of variation in acoustic characteristics of birdsong among and within five sympatric grassland songbird species breeding in Kansas, Indiana, and North Carolina. Thirty-eight individual acoustic characteristics were measured on target songs in Raven Pro 1.5 (Cornell Lab of Ornithology) software. We conducted correlation analysis in program R on those 38 metrics to eliminate collinearity. We then used the reduced explanatory variable set of 11-19 song metrics in quadratic discrimination analysis in program R to discriminate among and within species' songs based on acoustic characteristics. Acoustic characteristics successfully discriminated among 95.3% of all songs at the species level, with the greatest loadings on metrics based on entropy and spatial distribution of energy within song boundaries. Intra-specific analyses correctly classified 88% of Bachman's Sparrow song types, 76% of Field Sparrow song types, 95% of Grasshopper Sparrow song types, and 84% of Prairie Warbler song types. Time 5% (the point into a selection that divides it into intervals containing 5% and 95% of the energy of the selection) and aggregate entropy (measurement of the disorder in a selection) were the only acoustic metrics which heavily loaded in many discriminant function analyses and at the population level for a given species. We conclude that acoustic metric-based bird song classification can accurately discriminate among songs at the species and population level and this approach could improve the automatic detection of birdsong process.

Introduction

Using automatic detection of bird songs from acoustic recordings is an increasingly necessary process as the use of passive, automatic recording technology continues to grow (Brandes 2008, Frommolt and Tauchert 2014, Zwart et al. 2014). Availability of relatively inexpensive autonomous recording equipment has led to the ability to collect large, permanent acoustic databases. The majority of traditional avian monitoring techniques rely on bird song for population monitoring (Lack 1964, Jančovič and Köküer 2011). Acoustic monitoring of avian species through autonomous recording methods can produce accurate abundance estimates for a number of territorial species (Hutto and Stutzman 2009, Digby et al. 2013, Borker et al. 2014, Prevost 2016, Darras et al. 2017). However, relatively few well-documented and universally applicable methods exist for automated song extraction and classification within extensive acoustic datasets. Any advances in song extraction and classification methodology could improve the utility of acoustic monitoring methods.

Sources of variation in bird song

Variation in bird song repertoires or dialects adds complexity to automated detection that is difficult to accommodate (Anderson et al. 1996). There can be vast differences in bird songs at the species level, at the regional/dialect level, and at the individual level for most widespread avian species, particularly in passerines (Marler and Tamura 1964, Kroodsma et al. 1982). Flexibility in bird song reflects an individual's

ability to learn variants of their songs and respond appropriately to the behavioral context for communication (Baker and Cunningham 1985). Bird song plasticity can come from the species' immediate community structure, sexual selection, or song learning type, and this plasticity can lead to distinct geographic dialects (Kroodsma et al. 1982, Podos and Warren 2007).

Many hypotheses exist to explain how a juvenile male may learn and potentially adapt songs including the genetic adaptation hypothesis, the social adaptation hypothesis, and the founder effect hypothesis (Payne 1981, Kroodsma et al. 1982, MacDougall-Shackleton and MacDougall-Shackleton 2001). The genetic adaptation hypothesis proposes that birds learn their songs early while still on their breeding grounds and as adults use the cues to find similar songs in subsequent years. This leads to relatively low genetic diversity and potentially strong regional dialects in a population. The social adaptation hypothesis suggests that adults can still adjust their songs and acquire new ones, learning from their neighbors and constantly tweaking their songs. This would lead to a more fluid song structure within a population at any level. The founder's effect could affect song structure by limiting the total potential variation in songs heard from both neighbors and parents for numerous generations. Podos and Warren (2007) further classify learning types like those listed above into broader local and social adaptation groups, with a third category that classifies song variation as simply by-products of learning, dispersal, and energetic restraints. They argue that no one learning style fits all species, and there is too much variation to support any specific theory as a rule. They instead favor the by-product theory and focus on restraints caused by metabolic factors and sound propagation through local

vegetation as reasons for dialects and geographic variation (Podos and Warren 2007). Each of these song learning types can explain how different dialects form, but they don't necessarily explain how they affect measurable characteristics of bird songs. Song variations created by learning type could include changes in pitch, speed, repetition, and individual note characteristics (Hunter and Krebs 1979, Morton et al. 1998, Leger et al. 2003). A better understanding of the nature of these differences in song characteristics and how to categorize the dissimilarities may improve automated bird song detection and classification.

Community structure on a local scale can influence singing characteristics because of song overlap or interference with song attenuation related to vegetative cover. A complex avian community with competing or overlapping songs makes it hard for an individual's transmitted signal to be heard. Birds in a busy acoustic community may modulate the temporal or pitch characteristics of their song as well as increase the complexity to be heard (Shiovitz and Thompson 1970, Kroodsma 1981, Catchpole and Slater 2003). Species have also been shown to introduce variation into their songs to facilitate transmission through different vegetation. For example, Great Tits (*Parus major*) in different European countries vary songs to match the structure of open woodlands vs closed-canopy forests, not physical geographic regions (Hunter and Krebs 1979).

Most passerine males have multiple song types that are used to convey different types of information. Via playback tests and lab experiments, researchers have shown which songs are used for territorial displays, mating, contact calls, flight calls, etc. For example, wood warblers typically have a Type A song for male-male interactions and

Type B songs used to communicate with females (Kroodsma 1981, Nolan et al. 1999, Houlihan 2000). Grasshopper Sparrows (*Ammodramus savannarum*) similarly have distinct songs for defending territories, for communicating with females, and also a low trill used when approaching their nests (Smith 1959a, Thorpe and Lade 1961). These songs can vary on a species by species basis and depending on the intended recipient, may not be even be easily heard by human observers during traditional population monitoring.

Conspecific interactions between males and females can also affect song repertoire and type in passerines. Field Sparrows (*Spizella pusilla*), for example, typically learn a wide repertoire for each of their songs when young, then match one song with their neighbors during the first breeding season and stick with one simple and one complex song for the rest of their lives (Nelson 1992). By copying their neighbor's song and proving they belong, first-time breeding males increase their chances of obtaining a mate (Podos and Warren 2007). Females may either respond to natal-learned dialects and show little interest in different sub-species or current neighbors, leading to very similar song types and future speciation (Marler and Tamura 1962, Baker et al. 1987, Leger et al. 2003, Toews and Irwin 2008). Conversely, females may respond more positively to conspecific songs with the greatest complexity or largest repertoire size, affectively selecting for highly variable songs within the same population (Smith 1959a, Kroodsma 1981, Searcy 1992).

Knowledge of the different song types, the source and extent of variation, and intended use are vital for understanding how to approach song extraction and classification for a target species in analysis of acoustic recordings. Accounting for the

unique characteristics that describe and define the different songs within and among similar species may lead to improved accuracy for automatic detection analysis.

Acoustic characterization of bird songs

Many different types of acoustic measurements have been used to discriminate between different bird songs at the species, population or individual level. Different approaches can be broken up into those that characterize entire songs vs, those that measure individual notes. The earliest studies involved classifying songs using measurements that spanned the entirety of the signal. Shape of total song amplitude (Thorpe and Lade 1961), and occupancy of song amplitude (the percentage of sustained noise over the entire song, Shiovitz and Lemon 1980, Isler et al. 1998) were successfully used to tell populations of Indigo Buntings (*Passerina cyanea*) apart. Other successful measurements used to distinguish populations and/or communities included the minimum and maximum frequencies (Hunter and Krebs 1979, Leger et al. 2003), loudest overall frequency band (Brandes 2008), and total song length (Borror and Gunn 1965, Nelson 1992). Even more complicated computer-based measurements successful in differentiating groups included quartile frequencies and the presence or absence of harmonics (Isler et al. 1998). Some studies concluded that body size and local vegetation (i.e. thick forest vs open woodland) were related to differences in songs across populations, supporting the by-product theory of song variation (Hunter and Krebs 1979, Nicholls et al. 2006).

Some studies targeted individual notes within a song, taking more time to individually extract precise measurements which differentiated between populations or individuals. The number of notes or phrases (Hunter and Krebs 1979), shape of individual notes (Baker et al. 1987), change in frequency between notes (Weisman and Ratcliffe 1989, Kershenbaum and Garland 2015), phrase order (Isler et al. 1998), and number of notes per phrase (Hunter and Krebs 1979, Catchpole and Slater 2003) were all used successfully. However, these types of measurements are time-consuming to generate for large datasets.

Improvements in technology allowed for the use of more precise measurements such as those quantifying the distribution of total energy within waveforms or spectrograms and harmonic elements to differentiate between groups (Shiovitz and Lemon 1980, Isler et al. 1998, Brandes 2008, Frommolt and Tauchert 2014). While successful, the results for the majority of these studies were species-specific and often based on small, local datasets.

Song variety across a species' entire range, along with the complexity of information communicated in passerine songs, can complicate classification process by requiring larger training sets, massive amounts of hand-labeled song clips, and/or increasingly complicated techniques. Using continuous recordings of passerines recorded during typical monitoring periods also introduces more difficulties due to overlapping noises and poor-quality songs. An increase in analysis complexity to accommodate these complications is not necessarily the best option as it requires more expertise and time to conduct. Discrepancy in results and variation in methods demonstrate the benefit re-evaluating ways to investigate which specific measurements

can capture variation needed to accurately categorize bird songs from continuous recordings using economical methods across multiple fundamental song types.

In this chapter I present the results of a study of avian song characteristics for a suite of five sympatric grassland songbird species from eastern North America and address the following specific research questions:

1) Which spectrograph and waveform-based measurements capture variation across multiple bird songs for different species,
2) Can similar spectrograph and waveform-based measurements capture variation in individual song types within a given species,
3) Can similar spectrograph and waveform-based measurements capture variation in individual songs within different populations of a given species, and
4) Does the quality of the songs extracted from continuous recordings effect the ability to classify songs at the species, song type and population level?

Methods

Song Type

We chose five different grassland and shrub/scrub species breeding in the eastern United States from three individual study sites that represented a variety of song characteristics for this study (Figure 2.1). These species generally co-occur on the same study sites and thus would require discrimination during automatic detection

analyses. Using an assortment of songs with short chips, tonal notes, buzzy notes, and trills we investigated which acoustic measurements were important to characterize each group. Song types for each species were also divided into functional behavior groups as described in the literature for intraspecific song analysis using unique measurements.

Bachman's Sparrow (*Peucaea aestivalis*) songs typically involve a sustained note followed by a short trill, but are extremely variable (Figure 2.2, Borror 1961, Bent and Austin 1968). A very limited study of four males in Ohio and nine males in Florida produced a repertoire of 528 songs with 244 different song patterns (Borror 1971). While variable, Bachman's Sparrow songs can be broken down by defining tonal notes (W) or buzzy notes (Z) followed by trills of varying frequency (T). Borror (1971) described the distribution of common song types as primarily WT (56%), ZT (17%), and WWT (10%), with all others song types representing 5% of the total. Descriptive studies of Bachman Sparrow songs have not included differences based on the frequency (kHz) of introductory notes in comparison to trill/trills, and this may be important when describing the song types. For this reason, each annotated selection in this study was designated as either WT, WWT, or ZT but also included the relation of the introductory note to the center frequency of the trill (either lower than, even with, or higher than the trill).

Field Sparrows (*Spizella pusilla*) have many variable and hard to categorize song types. Song types are defined by designating the song as either simple or complex songs (Nelson and Croner 1991). The simple song is typically described as a 'bouncing ping-pong ball', with the pitch and speed of each note increasing in the second half of

the unique song. The notes in a complex song slow down again after the initial increase, and may repeat this pattern more than once (Figure 2.3). For the simple songs I added further classified the song types as either increasing in pitch over the front third of the call, decreasing in pitch, or maintaining a constant pitch.

Grasshopper Sparrow songs are simple and consist of short introductory notes followed by an insect-like trill. A longer or sustained song may be sung occasionally by males while either doing a display in the air to attract females or when particularly aggravated (Figure 2.4, Borror 1961). The sustained song includes the recognizable trill followed by a fast jumble of notes, although the trill can be dropped after territories are established. Another lower, softer trill has been described but is primarily used by paired birds approaching a nest and is not very loud (Smith 1959b). I didn't detect the lower frequency, softer trills in our recordings, so this song type was not included in the analysis. Grasshopper Sparrow songs were split into either simple or excited songs for analysis, with the latter broken up into separate songs for the trill and jumbled portions.

Henslow's Sparrow (*Ammodramus henslowii*) songs were the simplest songs included in this study. They are referred to as 'hiccup' due to their short, staccato sound. Each song is actually a series of brief, tonal chips in close proximity (Figure 2.5). There is not enough variation in these calls to classify them into different types, and therefore Henslow's Sparrow songs were only used in interspecies analysis.

Prairie Warblers (*Setophaga discolor*) have two distinct song types. The more typical Type A song consists of ascending buzzy notes used for territory establishment and mate attraction, whereas the Type B song is used in male to male confrontation, changes very little in pitch, and is more variable in regards to the 'buzziness' and length

of individual notes within the vocalization (Figure 2.6, Nolan 1978, Kroodsma 1981, Houlihan 2000). A second Type B song has been described as a “sing-song” variant, with a shorter overall song length and longer, buzzier individual notes (Houlihan 2000). Nolan (1978) also reported a Type B song that closely resembled Field Sparrow songs. These two song types were included in my analysis because they were observed on our study sties, albeit infrequently. Prairie Warblers have many intermediate songs between these groups, so a the Type B category was also used for intermediate songs between Type A and B (Prevost 2016). A fifth category was used to capture any truly aberrant song that didn’t match the types described above.

Study Areas

Recordings were made at Fort Bragg, NC (2012-2013), Jefferson Proving Grounds, IN (2012), and Fort Riley, KS (2013). Bachman’s Sparrows and Prairie Warblers were recorded in North Carolina, Field Sparrows, Henslow’s Sparrows, and Prairie Warblers were recorded in Indiana, and Grasshoppers Sparrows, Henslow’s Sparrows, and Prairie Warblers were recorded in Kansas.

Jefferson Proving Grounds (JPG) was decommissioned in 1994 and is now designated as the Big Oaks National Wildlife Refuge, although the Department of Defense still utilizes some areas for air national guard training on a limited basis. Much of the 20,234 ha are dominated by oak-hickory (*Quercus* and *Carya* spp.) woodlands (Meretsky et al. 2006, Roth and Islam 2008). Additional patches of native warm-season

grasses are maintained by prescribed fire within the larger mosaic to promote habitat for grassland birds (Nott et al. 2003, Hourdequin and Havlick 2011).

Fort Bragg Military Reservation is located within the Sandhills region of North Carolina, and is part of the largest block of historical longleaf pine-wiregrass (*Pinus palustrus-Aristida stricta*) vegetation in the area (Sorrie et al. 2006). The installation is home to a population of federally endangered Red-cockaded Woodpeckers (*Picoides borealis*) and other savanna obligates despite being heavily used for training exercises. The pine-wiregrass savanna covers the majority of the rear area on the 73,500-ha instillation, and is maintained with prescribed fire on a 2-4 year rotation.

Fort Riley is a 100,500-ha installation located in the Flint Hills region of eastern Kansas and is primarily covered by native upland prairies, former agricultural fields, and bottomland flood plains (Althoff et al. 2006). Fort Riley provides some of the largest tracts of grassland bird habitat in the region (Cully and Michaels 2000, Eberly and Keating 2006). A three-year schedule of prescribed burning maintained the grasslands over much of the installation. The native grasslands facilitated training maneuvers as well as supported diverse grassland bird populations.

Data Collection and Classification

Three training areas per study site were chosen as replicates during each year of the study. Training areas were chosen based on the number of target species present, day-to-day accessibility, and lack of active military training. Three males of each available target species in each training area were captured via target mist-netting,

color-banded, and territory-mapped during the breeding season. SM2 song meters (Wildlife Acoustics Inc., Maynard, MA) were placed in the center of each delineated territory to record bird songs. We assumed such an approach captured the vocalizations of the color-banded individual and his nearest neighbors within ~100 m. Given the territory densities of our focal species, 1-4 additional males were recorded of each focal species at each monitoring site (Prevost 2016).

Acoustic data were recorded using both right and left speakers at a gain setting suitable to record uncompressed WAV files of each species from up to 100 m (either 48 or 60 dB gain) at a sampling rate of 24 kHz. Song meters recorded audio from sunrise to four hours past sunrise every day from mid-May to July; each individual male was recorded 2-3 days per 7-day cycle. Approximately 9,720 hours of passive recordings were collected during this study (3 years * 2 locations per year * 3 training areas per location * 3 SM2 recorders per training area * 45 4-hr recordings per SM2). Occasional equipment failure (e. g., battery failure or corrupt SD cards) and other logistic constraints reduced the total number of hours monitored by approximately 15%.

We selected a random subset of recordings which were evenly distributed across the entire recording period for each focal color-banded male. Five minute song clips were randomly selected from each hour segment within these recordings, and technicians manually extracted songs of the target species from each five-minute clip using Raven Pro 1.5 software (Bioacoustics Research Program, Cornell Lab of Ornithology). All files were double-checked for accuracy by experts after technicians extracted target songs. Recordings were randomly chosen until 1000-1200 individual bird songs were extracted for each species. A 450-550 sample Haan window with 50%

overlap was used to visualize the spectrogram and extract song measurements. Each song selection was classified as either clean, quiet, or overlapping a different bird song in the recording to document of sound quality.

Song selections from each bird were assigned to the basic song types described above. However, we encountered song variations not encompassed in traditional song descriptions which had the potential to influence classification results. As a result, we added additional song categories for Bachman's Sparrow and Field Sparrow songs to more accurately capture variation not encompassed by basic song type previously described in the literature. The most frequently recorded Bachman's Sparrow song types were divided into groups determined by the relationship of the first W note to the following trill. The notes either had a lower frequency, approximately the same frequency, or a higher frequency than the trill. Field Sparrow simple songs were divided into three groups (up, down, and flat) describing the change in frequency for the first 5-10 notes in each song.

Data Analysis

Thirty-eight unique descriptive measurements on species' songs were made using Raven Pro 1.5 (Table 2.1). These metrics captured information on song length and frequency bounds, energy within the song window, amplitude, and pitch tracking of labeled songs and were easily calculated in Raven Pro for a large dataset. Variables with a correlation coefficient (r) absolute value greater than 0.7 were dropped on a species by species basis to reduce collinearity in the data.

All analyses were run in Program MASS using R Studio version 1.0.136 (Venables and Ripley 2013). I performed quadratic discriminant analysis (QDA) for data with unequal variance and covariance matrices between groups on three different levels of bird song using leave-one-out cross validation and equal prior probabilities (Keen et al. 2014). Levels analyzed included interspecific, intraspecific, between song types (except Henslow's Sparrows), and interpopulation (for Field Sparrows, Prairie Warblers, and Henslow's Sparrows).

Results

Metrics retained after correlation analysis were not the same for all species, with sixteen metrics retained for Bachman's Sparrows, eleven for Field Sparrows, fourteen for both Grasshopper Sparrows and Prairie Warblers, and nineteen for Henslow's Sparrows (Table 2.1). Only six measurements were retained for all five species and interspecific analysis. Retained metrics covered all main groups of potential measurements including those based on spectrogram values, waveform values, and pitch tracking.

Distribution of song types within each species were similar to reports in the literature (Table 2.2). Not all song types were included in the discriminant function analysis because of low sample size. Bachman's Sparrow songs were categorized as WT (85%), WWT (3%), ZT (10%), and other (2%). The distribution of W note frequency in relation to trill frequency (lower than, equal to, or higher than) was balanced relatively evenly within both the WT and ZT song types. Ninety-nine percent of extracted Field

Sparrow songs were categorized as simple (individual notes increasing in tempo for the duration of the song) with the majority comprised of flat or downward sloping notes (48% and 44% of total songs, respectively). Excited songs comprised only 6% of all Grasshopper Sparrow recordings. The most frequently encountered Prairie Warbler songs were Type A (47%), Type B (33%) and the sing-song version of the Type B song (23%).

Interspecific discriminate function analysis of all bird songs correctly classified song samples to species 95.3% of the time using 11 measurements and equal prior probabilities ($F = 989.13$, $P < 0.001$, $n = 5985$, Figure 2.7). Out of 286 total misclassifications, 107 (37%) were between Bachman's Sparrow and Field Sparrow, 84 (29%) were between Bachman's Sparrow and Prairie Warbler, and 52 (18%) were between Field Sparrow and Prairie Warbler. Loadings from quadratic discriminant analysis were greatest on aggregate energy, IQR duration (s), minimum entropy (dB), and 5% time (s) (Table 2.4).

Using only 'clean' measurements (those that weren't very faint or overlapped by a stronger signal), the correct classification rate increased only marginally to 95.9%. Henslow's Sparrows had the least percentage of faint or overlapping song clips (22%), whereas Field Sparrows had the greatest (33%, Table 2.3). Most misclassifications using clean samples occurred between Bachman's Sparrow and Prairie Warbler (56 of 172, 33%).

Bachman's Sparrow WT, WWT, and ZT songs were correctly classified 88% of the time using 16 metrics ($F = 4.82$, $n = 1222$, $P < 0.001$, Figure 2.8). Discriminant function loadings were greatest for average amplitude, IQR duration (s), time 5% (s),

and time 95% (s) (Table 2.5). Of all Bachman's Sparrow songs, 32% were of low quality, however misclassifications of these song clips represented only 24% of total incorrect classifications. Inter-song type analysis had a correct classification rate of 58.16% between WT songs ($F = 6.74$, $n = 1054$, $P < 0.001$) and 89.15% between ZT songs ($F = 2.27$, $n = 129$, $P = 0.003$). WT song classification was primarily based on values of average amplitude, average entropy, IQR duration, time 5% and time 95% (Table 2.6). ZT song types were separated using average amplitude, center time, IQR duration, time 5% and time 95%.

Field Sparrow song types were correctly classified 76% of the time ($F = 32.94$, $n = 1617$, $P < 0.001$, Table 2.7, Figure 2.9). Poor quality songs made up 35% of all Field Sparrow song clips and 41% of all misclassifications. Despite only labeling 21 individual complex songs, 11 were classified incorrectly during analysis. Most misclassifications occurred between the simple songs starting with downward sloping and flat notes (271 of 388 [70%] total incorrectly classified songs). Average amplitude, average entropy (dB), IQR duration (s) minimum entropy (dB) and time 5% (s) had the greatest scores on the discriminant loadings.

Grasshopper Sparrow discriminant analysis was run using 15 measurements between typical and excited songs (Figure 2.10). Aggregate and average entropy were the most important in discriminating between the two song types. The analysis incorrectly labelled songs only 5% of the time ($F = 16.01$, $n = 1148$, $P < 0.001$) (Table 2.8). Faint or overlapping songs made up 33% of all song clips and represented 46% of the total errors.

Discriminant function analysis accurately classified five different Prairie Warbler song types 84% of the time using 14 unique measurements ($F = 38.41$, $n = 1220$, $P < 0.001$) (Figure 2.11, Table 2.9). Aggregate entropy, duration 90% (s) and time 5% (s) had the greatest values on all four discriminant loadings. Of 197 total misclassifications, 99 (50%) occurred between the typical Type A and Type B songs. Thirty percent of all Prairie Warbler song clips were categorized as quiet or overlapping, and 33% of all misclassifications involved these poor-quality songs.

Discriminant analysis showed strong differences between the average values of individual song metrics for Field Sparrows, Henslow's Sparrows, and Prairie Warblers at the population level (different study sites). Field Sparrow songs were correctly assigned to populations 71.61% of the time, with length, IQR bandwidth, IQR duration, and minimum entropy loading heavily in the analysis ($F = 53.0953$, $n = 1617$, $P < 0.001$). Henslow's Sparrow songs were correctly assigned to populations 87.67% of the time, with maximum entropy, 1st quartile frequency, and maximum slope loading heavily in the analysis ($F = 90.97$, $n = 981$, $P < 0.001$). Prairie Warbler songs were correctly classified to the population level 95.28% of the time ($F = 232.82$, $n = 1272$, $P < 0.001$), with PFC max slope, aggregate entropy, and high frequency loading heavily in the analysis. Not all the metrics important for discriminating between populations for the three species were similarly important in song type classification. Only aggregate entropy was important in interspecies song classification and population for Henslow's Sparrow. Metrics important to both song and population classification for Field Sparrows were inter-quartile range duration and minimum entropy, while only aggregate entropy was influential for both analysis for Prairie Warblers.

Discussion

Song Variation among Species, Song Types, and Populations

Passerines use song for the majority of their communication (Kroodsma et al. 1982), and as such have evolved complexity to convey different meaning. The variability of these songs can impede our ability to monitor their populations with passive acoustic techniques. Song variation was characterized for five sympatric grassland species by a suite of 38 acoustic measurements that were easily attainable in Raven Pro sound analysis software. This approach was highly successful in discriminating between songs at the interspecific level with correct classification >95%, and was also successful at discriminating among intraspecific song types and intraspecific songs at the population level.

Three main groups of measurement types were important across all discriminant analyses, including the magnitude or type of entropy, the average amplitude, and descriptions of energy partitioning within song clips. These three groups of measurement were the most important in interspecies song type classification, although the importance of each in intraspecific analysis varied depending on the species. The entropy group described the average and aggregate levels of disorder within song clips as measured in bits. Aggregate entropy measures the tonal disorder in a sound by looking at the distribution of energy within the clips' spectrogram. A single pure tone would have a lesser value, whereas a more complex song with 'buzzy' notes across a wide frequency range would have a greater value. The average entropy takes the same

measurement but averages it across the number of data frames. While these two different measurements differ in some respects, they were frequently correlated enough for one or the other to be dropped from analysis. Average amplitude for the totality of each song clip was included in all levels of analysis as it had low correlation with all other measurements. Average amplitude was never the most important variable on the first discriminant loading for any analysis but was highly weighted on later loadings for most of the analyses. The third group of measurement types important in classifying songs clips included time 5%, time 95%, duration 90%, and IQR duration. These four metrics all measure how much time into a song clip has passed before different levels of total energy of the song has been sung. These measurements are also more sensitive to song length, as they represent time in seconds into a recording, and not % of total time.

The first discriminant loading in interspecific analysis was most influenced by the difference between the entropy measurements and energy partitioning. Energy partitioning measurements ranked strongly on this loading because of the differences in average song length between each species: Henslow's Sparrow and Grasshopper Sparrow songs were much shorter on average than the other three. Previous studies have used song length and interquartile frequencies to differentiate between groups of the same species, where averages were much more similar between groups (Borror and Gunn 1965, Hunter and Krebs 1979, Houlihan 2000). Differences in aggregate entropy means between groups showed that Henslow's Sparrows had the greatest average (4.86 ± 0.05) and Prairie Warblers the least (3.73 ± 0.04). Aggregate entropy quantifies the complexity of a song within the defined song clip boundaries. Counter-

intuitively, the short, sharp hiccup of a Henslow's Sparrow is more acoustically 'complex' than the slower, more tonal songs of the other species. Grasshopper Sparrows (4.79 ± 0.02) also showed high entropy across its longer song, reflecting the wide frequency range occupied by the trills.

The acoustic characteristics that separated intraspecific song types were similar in nature among species although the specific loadings in the discrimination analyses differed. Bachman's Sparrow songs were best told apart by entropy vs energy partitioning, with the WTT songs having greater averages for both types of measurements. The second loading relied on entropy, amplitude, and differences within the energy partitioning measurement types. The first loading for Field Sparrows was also dependent on differences between the energy partitioning. Complex songs (i.e., longer songs) had a greater IQR duration, whereas time 5% was greatest for the simple songs with downward sloping notes. Prairie Warbler songs also were best separated by using entropy vs energy partitioning, with average amplitude only important on the second loading. QDA between Grasshopper Sparrow songs only used one loading because there were only two song types. Aggregate and average energy had the greatest values on the loading; the excited songs had greater average aggregate entropy (4.95 ± 0.06 vs 4.78 ± 0.02) and lesser average entropy (4.09 ± 0.07 vs 4.16 ± 0.01) than the regular song/trill.

Discriminant analysis of Field Sparrow songs, which included novel song types not previously described in the literature, was relatively accurate (76% correct classification). Classification rates based on discriminant analysis on the WT and ZT song types were more variable (WT = 58%, ZT = 89%). Brandes (2008) proposed

dividing all songs into five unique syllable types (constant frequency, frequency modulated whistles, broadband pulses, broadband with varying frequency components, and harmonic elements), each syllable type with a different approach to measurement and classification to differentiate acoustic signals at all analysis levels. However, our approach demonstrated that relatively high classification rates can be achieved by using easily measurable and available metrics, even at this sub-song level, using entire songs. The benefit of an analysis such as ours with more limited data processing time may outweigh small gains in classification accuracy.

Inter and intra-species classification did very well overall in this study using these 38 easily calculated metrics. We were able to differentiate groups with high degrees of success even when using songs extracted from realistic soundscapes. This study suggests that a highly complex level of analysis isn't necessarily needed for realistic song/species classification and that simple solutions can be effective. We were only concerned with classification in this step, however, and results may differ when incorporating extractions of songs as well. Still, these methods are easily calculated and reproduceable: easy methods to use for a user who is faced with tackling hundreds of hours of automatic recordings generated from ARUs.

Population-level song variation

Through the discriminant analysis, I identified numerous intraspecific differences in song characteristics between different populations for the three species in which we had data from two distinct geographic regions (study sites). However, not all of the

measurements which differed by study site had high loadings on the QDA analysis and therefore weren't important in classification analysis. Metrics important in capturing geographic variation in song may not be similar to the ones heavily loaded in inter and intraspecies QDA. For example, we did not expect any measurements based on song amplitude to be loaded heavily in classification because this metric reflected both the distance of the individual to the SM2 recorder and the signal to noise ratio based on the background "noise" of the soundscape at each location. Average amplitude could differ between locations due to factors unrelated to the target species, such as total number of birds singing, frequency of the target song in question, and insect noise. However, amplitude could possibly be attributed to the soundscape itself at different locations. Different vegetation types, conspecifics, and neighboring species vocalizing in the same frequency band could affect the frequency and/or amplitude of a bird (Hunter and Krebs 1979, Nicholls et al. 2006); this could be reflected in the amplitude metrics and illustrates why amplitude may be useful at discrimination among populations. Overall, the acoustic metrics we calculated were effective at discrimination among populations for all three species.

Effects of song quality

Correlation-based automated song analysis techniques, such as template matching and dynamic time-warping often encounter problems with song extraction from continuous recordings because of overlapping sounds, low-amplitude signals from distant birds, and other natural interference (Anderson et al. 1996, Jančovič and Köküer

2011). However, our data suggest that understanding the nature of bird songs for a species can be beneficial in correctly classifying extractions made from noisy environments. Interspecific classification among the five focal species was improved by <1% after excluding all songs that were either very quiet or were overlapped by another song recorded simultaneously in the same frequency and time span as the target species. The variation between species' songs for this suite of species was effectively greater than the variation created by measurements on poor quality song recordings. Similarly, the poor-quality songs of Bachman's Sparrows made up a third of all audio clips but were associated with proportionately fewer classification errors than were classifications errors on high quality songs (24% vs 76%, respectively). This species' songs have been described as highly variable even within a single individual, which may be more important in limiting classification accuracy than the quality of the individual songs.

The remaining species showed slightly greater classification error rates for poor-quality recordings compared to error rates for high-quality recordings, despite very similar proportions of poor- quality recordings across all species. The most extreme was in Grasshopper Sparrows, where poor-quality songs made up a third of the total recordings but were associated with 46% of misclassifications between regular and excited song types. More than any other species' song, the Grasshopper Sparrow's trill was frequently indistinct around the edges for songs recorded distant from the SM2 recorders. This degradation of the signal may be accountable for the greater misclassification rate on faint and overlapped songs, and the result may be unique to sustained trills. Designation of an amplitude cut-off for analysis purposes could

potentially reduce the classification errors encountered for Grasshopper Sparrows and other species with similar call structure. Like the Bachman's Sparrow, these classification results demonstrate that understanding the nature of a bird song (either biologically or structurally) is important when classifying them. It may be beneficial to exclude poor-quality songs with an amplitude filter before classification and use the results as an index instead of a true song count, particularly if the proportion of poor-quality songs is consistent across recordings.

We used three different types of songs for this classification study: a short burst (Henslow's Sparrow), extended trill (Bachman's Sparrow and Grasshopper Sparrow), and traditional sing-song/tonal songs (Prairie Warbler and Field Sparrow). Three main groups of relatively easily calculated measurements achieved effective classification both between species and between song types and populations of individual species. We only looked at functional types of full songs and were still able to achieve good classification rates. Song clips recorded in poor conditions had lower impact for the classification of highly-variable songs, and geographic variation between songs of the same species was most apparent in relation to the energy partitioning within a song. These results show that measurements which are easily calculated from complete song clips can effectively differentiate between different songs and song types, and suggest that they should be the focus of future studies using automatic detection and classification of bird songs from continuous recordings.

Appendix

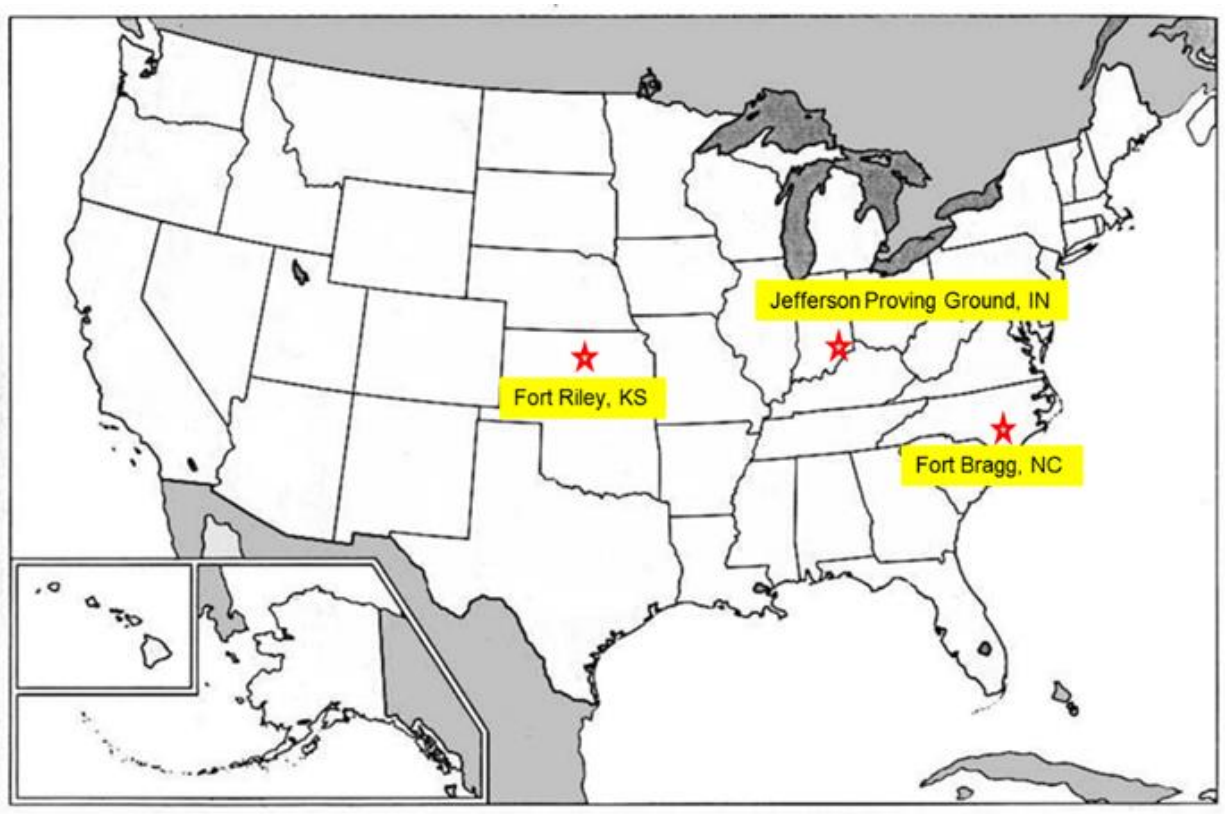

Figure 2. 1: Study areas used for recording grassland and shrub/scrub bird songs (2011-2013).

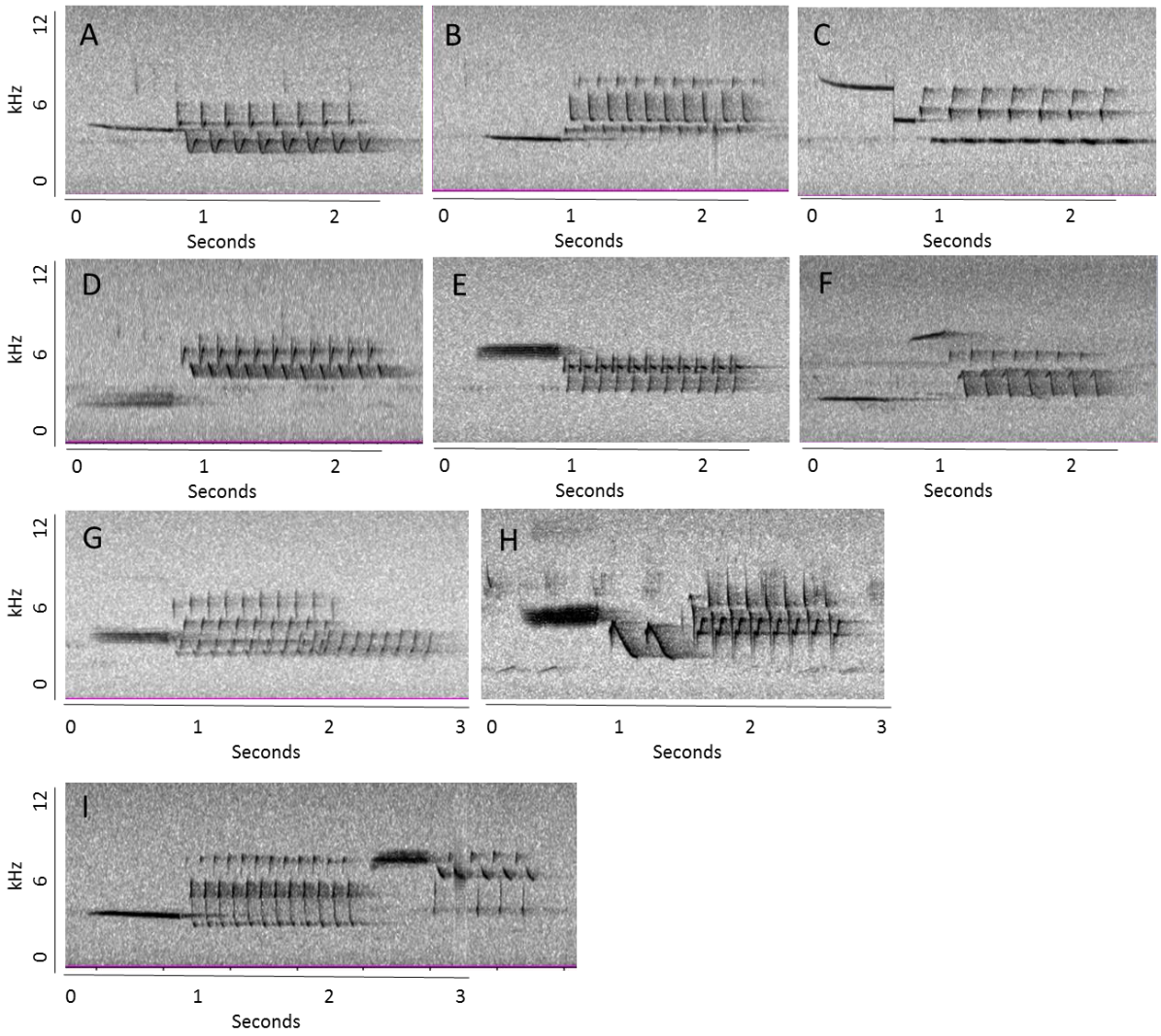

Figure 2. 2: Bachman's Sparrow songs recorded on Fort Bragg, NC (2012-2013).

Song types are different combinations of single tones (W), buzzy tones (Z), and trills (T) described in Borror (1961)

A – C: WT songs with a W note either below, even with, or above the frequency of the subsequent single trill

D -E: ZT songs with a buzzy introductory note and single trill

F: WWT song with two tonal notes of different frequencies before the trill

G: WTT song with two varying trills

H: ZTT song type with a short trill section

I: WTWT song type with two different trill notes

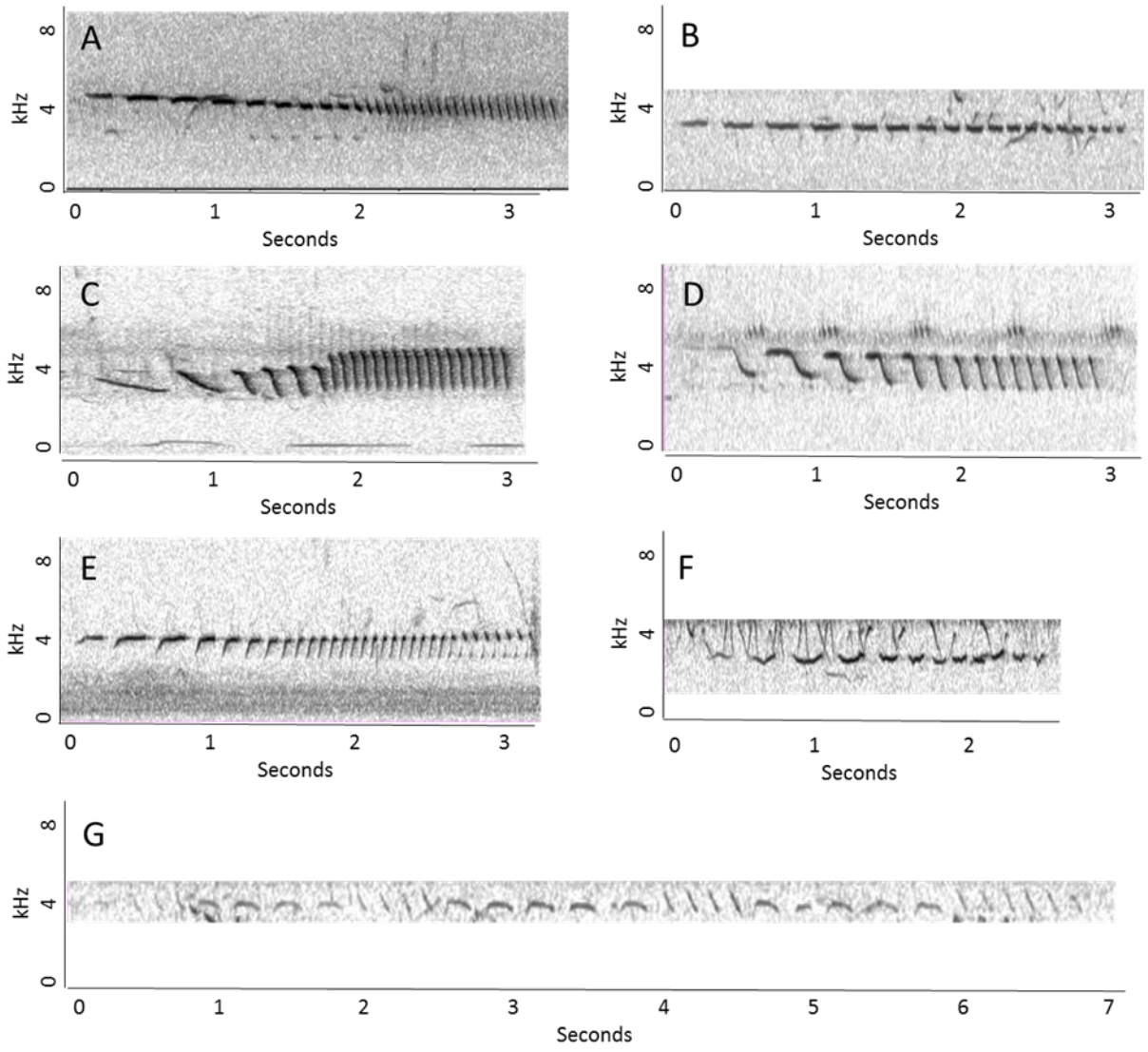

Figure 2. 3: Field Sparrow songs recorded on Jefferson Proving Grounds, IN and Fort Riley, KS (2011-2013).

A – F: Simple songs characterized by notes that get shorter during song progression. Syllables in the beginning of the song either stay flat in pitch (A and B), decrease in pitch (C and D), increase in pitch (E), or form a 'w' (F).

G: An example of a complex Field Sparrow song that slows and increases in pitch multiple times during its duration.

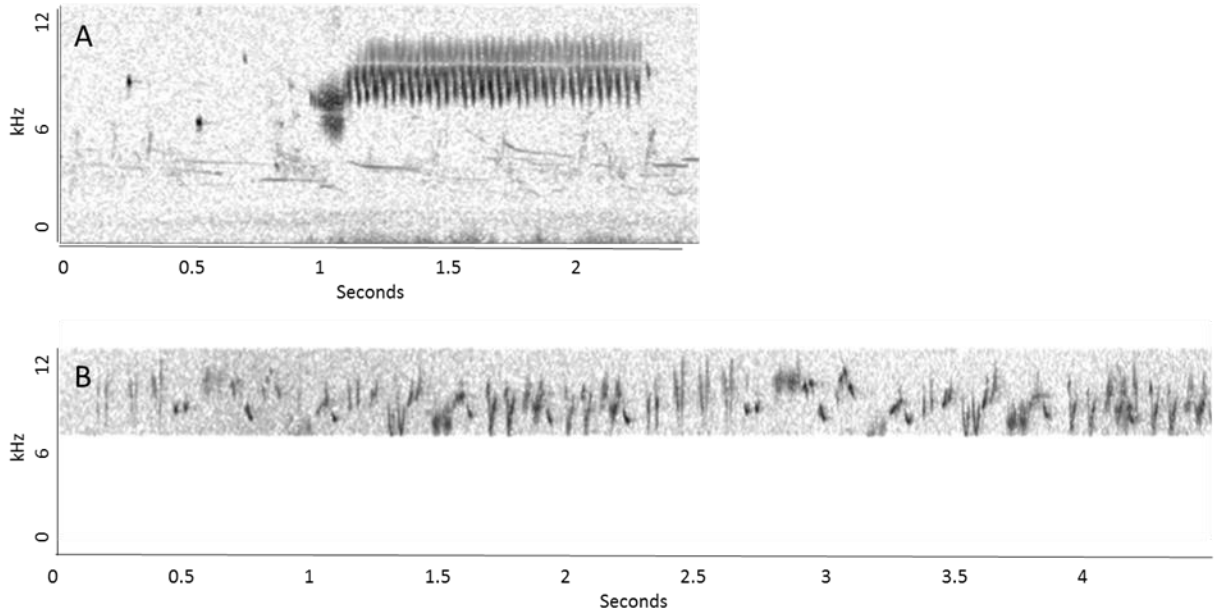

Figure 2. 4: Grasshopper Sparrow songs recorded on Fort Riley, 2013.

The typical song consists of faint introductory notes and a sustained trill (A). The excited song also includes an intense, fast combination of short notes (B) either after the regular song or by itself.

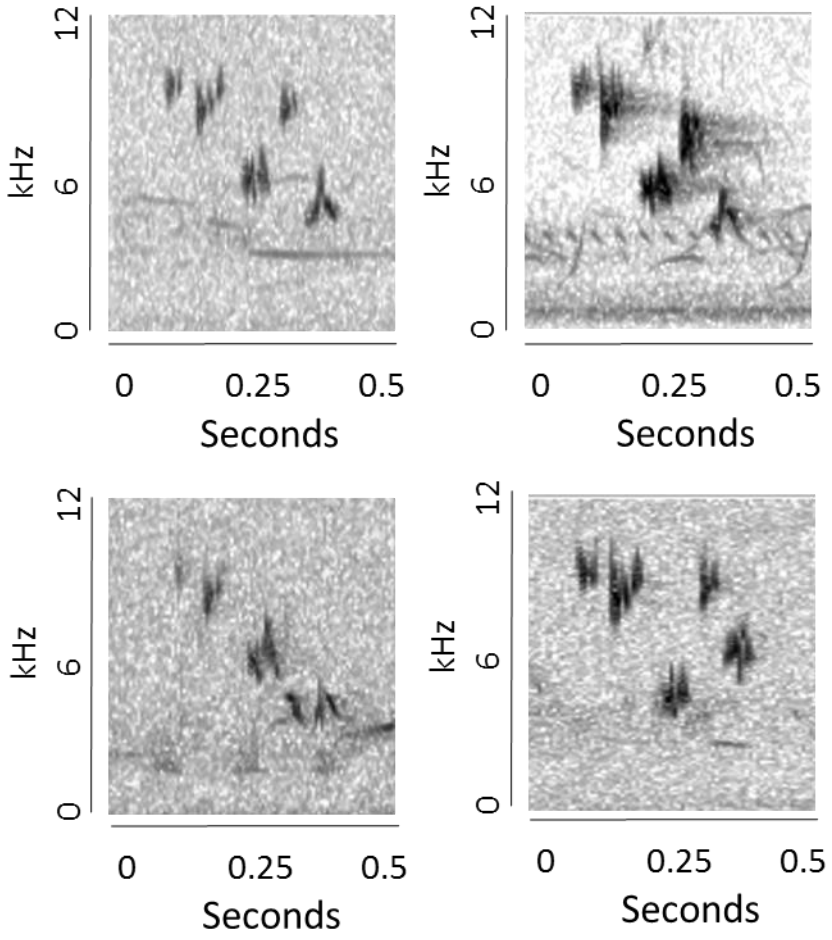

Figure 2. 5: Variation in Henslow's Sparrow songs recorded on Jefferson Proving Grounds, IN and Fort Riley, KS, 2011-2013.

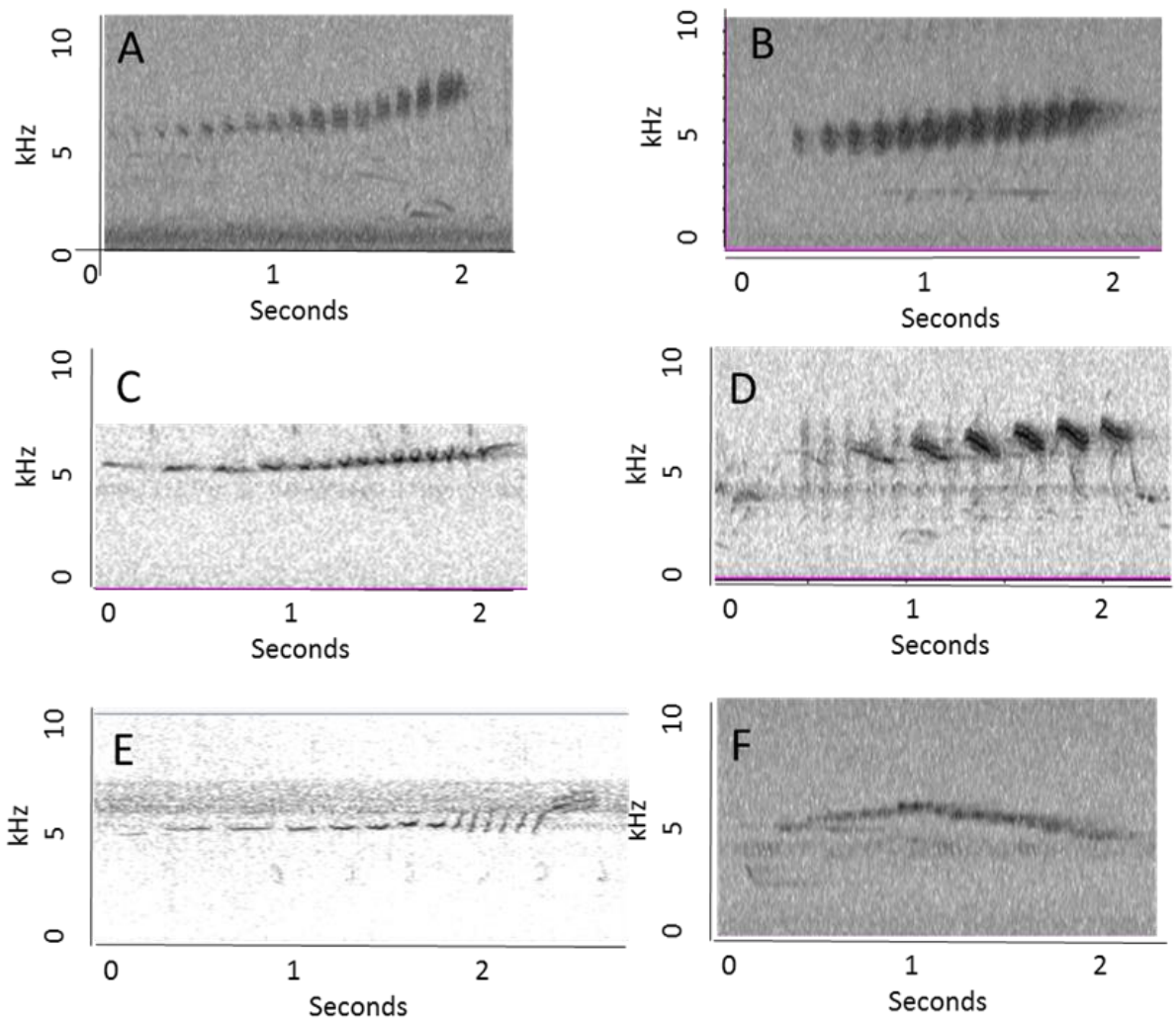

Figure 2. 6: Prairie Warbler songs recorded at Jefferson Proving Grounds, IN and Fort Bragg, NC (2011-2013).

A – B: Type A songs with buzzy notes

C: Type B song with flatter notes

D: A variant on the B type song with a 'sing song' quality as described by Nolan (1978)

E: An unusual Type B song that closely resembles a typical Field Sparrow song

F: An abnormal Prairie Warbler song which resembles a Type B but decreases in pitch

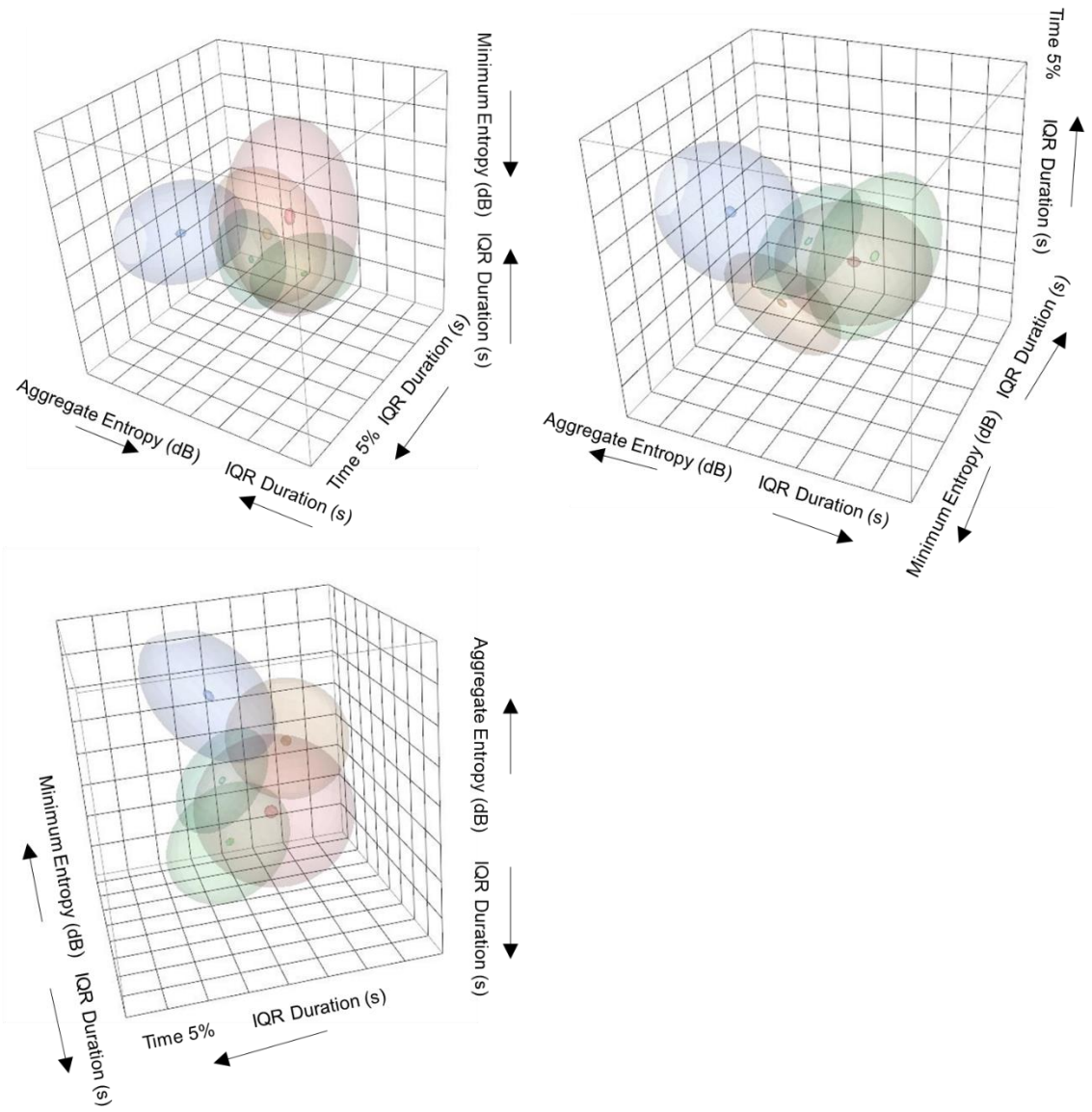

Figure 2. 7: Canonical plot for quadratic discriminant analysis between songs of all species recorded at Fort Riley, KS, Fort Bragg, NC, and Jefferson Proving Grounds, IN (2011-2013).

Means (dark ovals) and 50% contours (light ovals) are shown for each species Bachman's Sparrow (pink), Field Sparrow (light green), Grasshopper Sparrow (blue), Henslow's Sparrow (orange), Prairie Warbler (dark green)

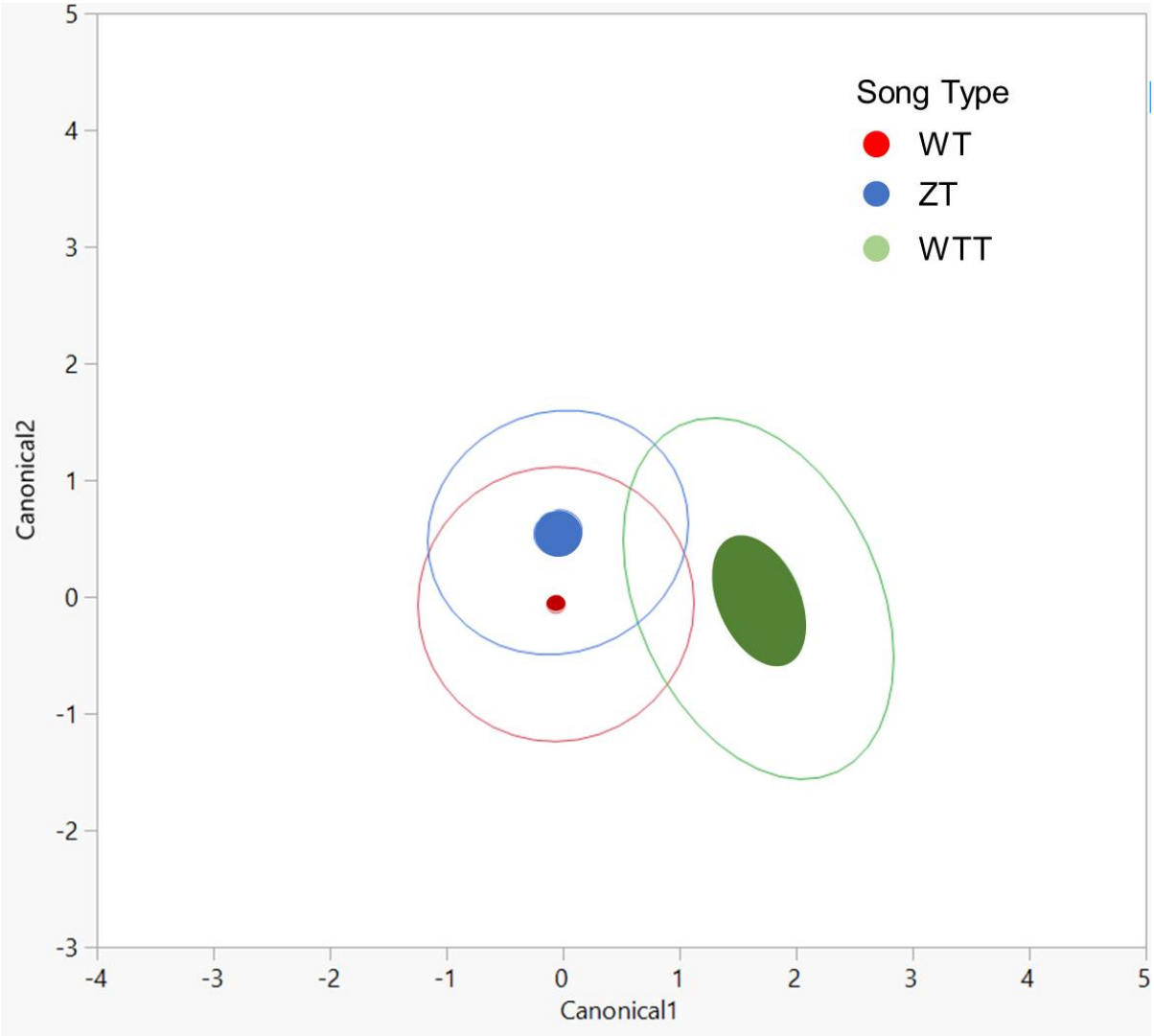

Figure 2. 8: Canonical plot for quadratic discriminant analysis between the three most common Bachman's Sparrow songs recorded at Fort Bragg, NC (2012-2013).

Circles show the average and 50% contours for each song type

WT songs have a tonal note followed by a trill

ZT songs have a buzzy introductory note followed by a trill

WTT songs have a tonal note followed by two trills

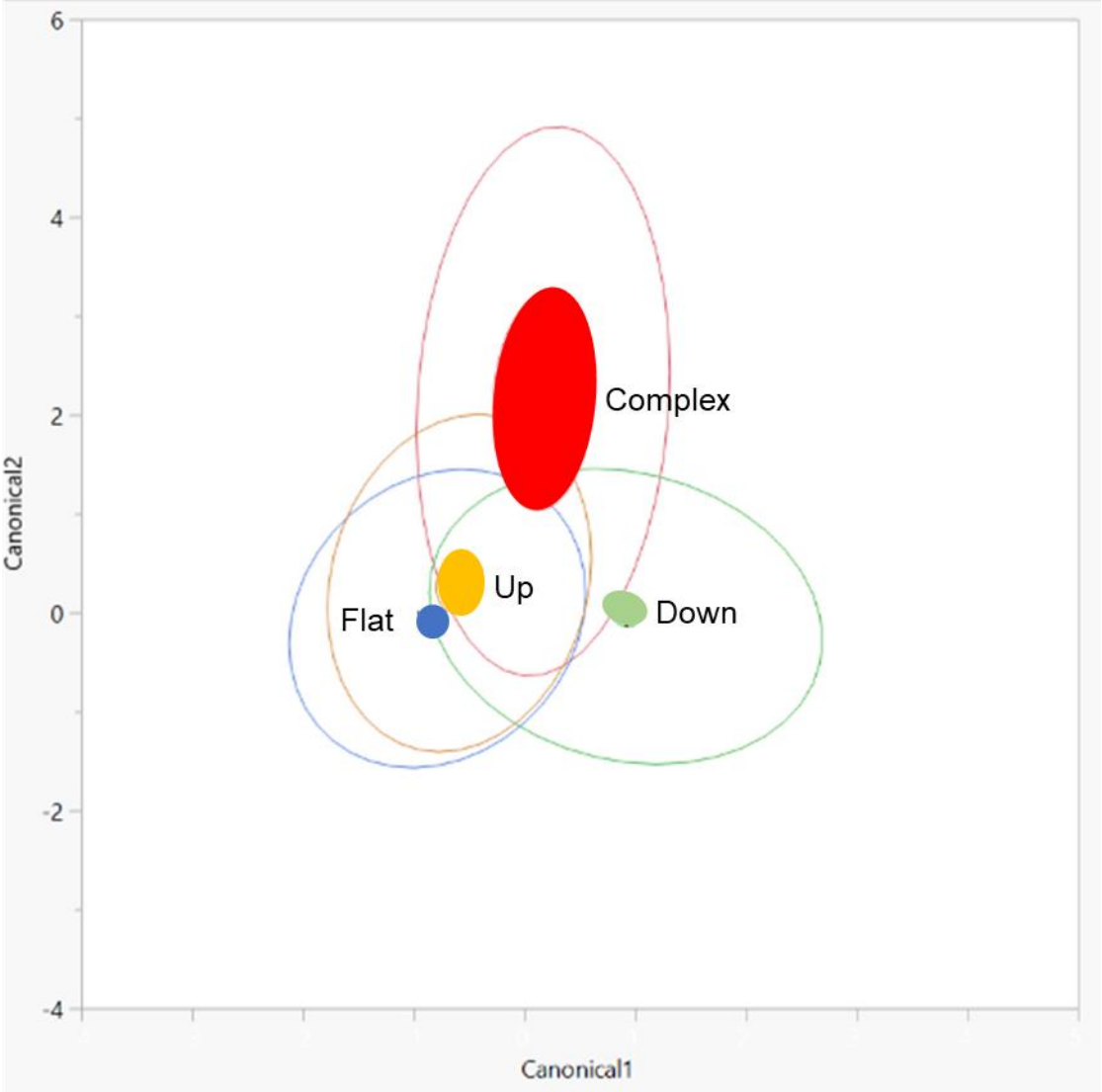

Figure 2. 9: Canonical plot for quadratic discriminant analysis between the four most common Field Sparrow songs recorded at Fort Riley, KS and Jefferson Proving Grounds, IN (2011-2013).

Circles show the average and 50% contours for each song type

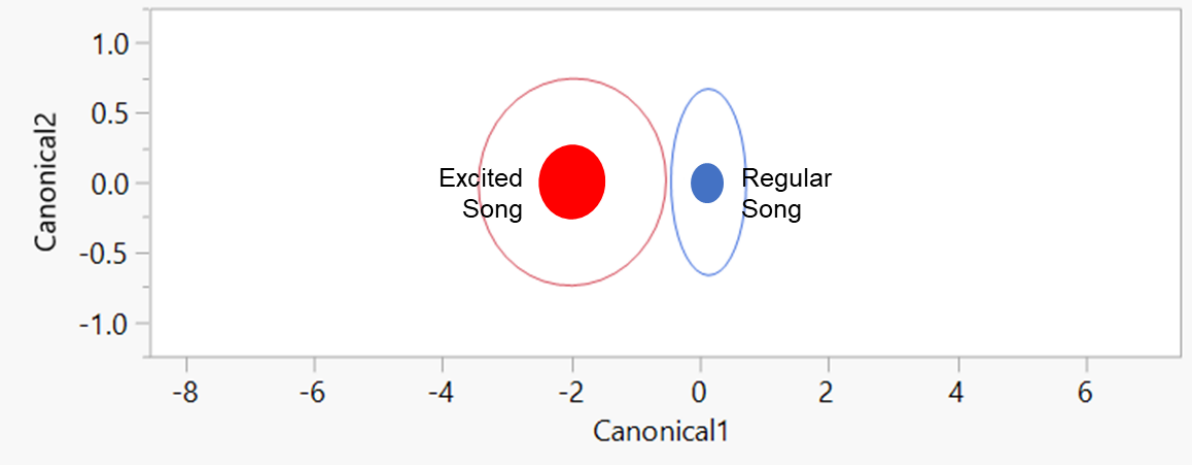

Figure 2. 10: Canonical plot for quadratic discriminant analysis between Grasshopper Sparrow songs recorded at Fort Riley, KS (2011 & 2013).

Closed circles show the average and open circles show the 50% contours for each song type

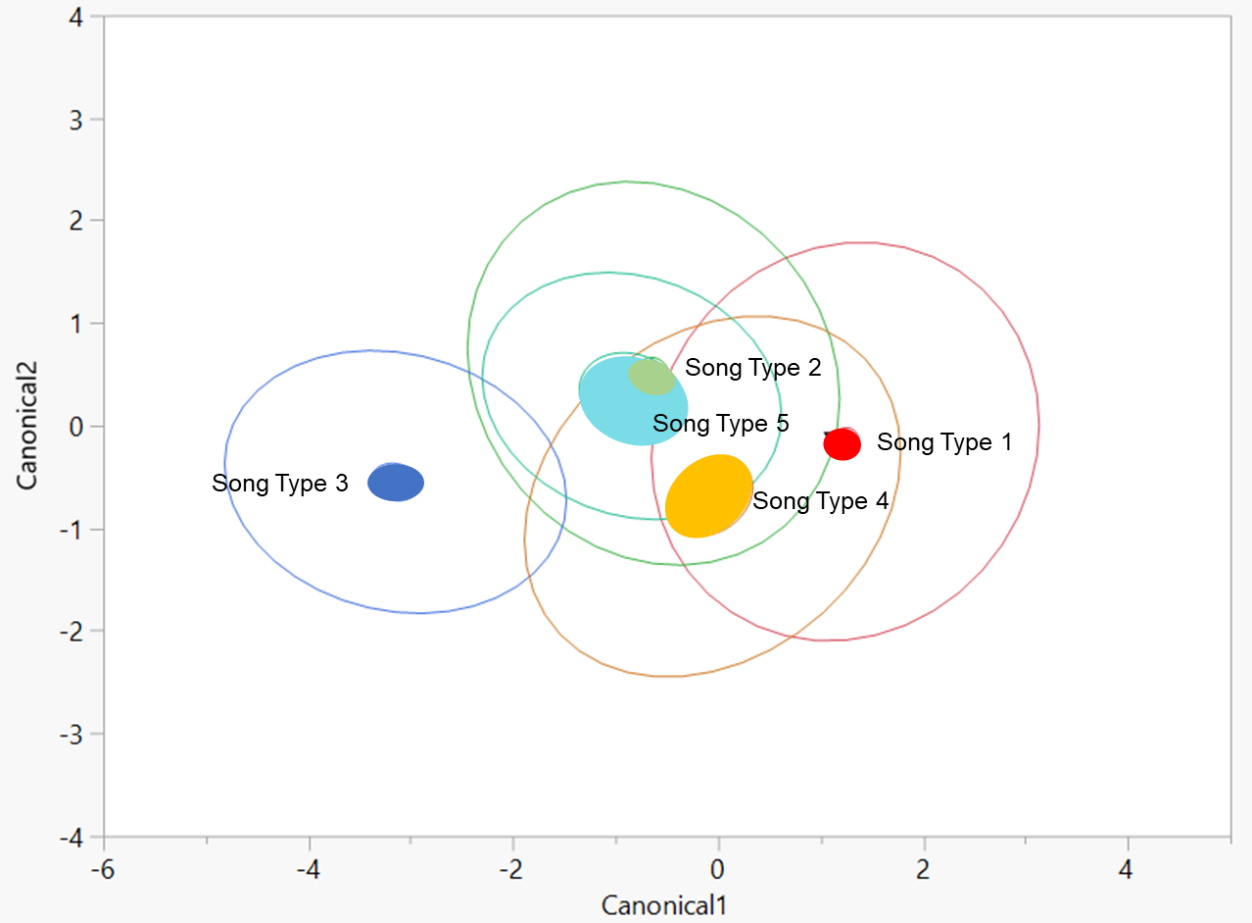

Figure 2. 11: Canonical plot for quadratic discriminant analysis between the five most common of Prairie Warblers songs recorded at Fort Bragg, NC and Jefferson Proving Grounds, IN (2011-2013).

Closed circles show the average and open circles show the 50% contours for each song type

Table 2. 1: Raven Pro 1.5 measurements used to characterize bird songs and retained in QDA analysis.

Measurement	Description	Species				
	Measurements Based on Spectrogram Values	BACS	FISP	GRSP	HESP	PRAV
Bandwidth 90%	Difference between the 5% and 95% frequencies (Hz)					
Center Frequency	Frequency that divides the selection into two frequency intervals of equal energy (Hz)					
Center Time	Point in time when the selection is divided into two time intervas of equal energy (s)	x			x	
Duration 90%	Difference between the 5% and 95% times (s)			x	x	x
Frequency 5%	Frequency that divides the selection into frequency intervals contaning 5% and 95% of the energy in the selection (Hz)				x	
Frequency 95%	Frequency that divides the selection into freqency intervals contaning 95% and 5% of the energy in the selection (Hz)				x	
IQR Bandwidth	Difference between 1st and 3rd quartile frequencies (Hz)	x	x		x	
IQR Duration	Difference between the 1st and 3rd quartile times (s)	x	x		x	
Peak Frequency	Frequency where the maximum power occurs (Hz)	x				
Time 5%	Point in time that divides the selection into time intervals containing 5% and 95% of the energy of the selection (s)	x	x	x	x	x
Time 95%	Point in time that divides the selection into time intervals containing 95% and 5% of the energy of the selection (s)	x			x	
1st Quartile Frequency	Frequency that divides the selection into frequency intervals contaning 25% and 75% of the energy in the selection (Hz)	x	x	x	x	x
3rd Quartile Frequency	Frequency that divides the selection into frequency intervals contaning 75% and 25% of the energy in the selection (Hz)			x		x
1st Quartile Time	Point in time that divides the selection into time intervals containing 75% and 25% of the energy in the selection (s)					
3rd Quartile Time	Point in time that divides the selection into time intervals containing 25% and 75% of the energy in the selection (s)				x	
	Measurements based on Waveform Values					
Average Amplitude	Average amplitude across the entire selection (dB)	x	x	x	x	x
Average Power	The value of the spectrogram's power spectral density for each pixel averaged over the entire selection (dB)				x	
Aggregate Entropy	Measure of disorder in the selection (pure tones = low) (bits)			x		x
Average Entropy	Amount of disorder for a frequency bin and averaged over the entire selection (dB)	x	x			
Delta Power	Difference between the value of the power value at the upper and lower frequency limits of the selection (dB)					
Energy	Total energy within the selection bounds (dB)					
Maximum Amplitude	Maximum of all the values in the selection	x				
Minimum Amplitude	Minimum of all the values in the selction					
Maximum Entropy	Measurement of the highest entropy for any frame within selection (bits)	x				
Minimum Entropy	Measurement of the lowest entropy for any frame within selection (bits)	x	x	x	x	x
Peak Amplitude	Greater of the absolute values of maximum and minimum amplitudes					

Table 2.1 (continued)

Peak Power	Maximum power in the selection, or the value of the darkest point in the selection (dB)					
Peak Time	First time in the selection when the amplitude is equal to the peak amplitde (s)	x	x	x	x	x
RMS Amplitude	Root-mean-square or effective amplitude of the selection		x	x		x
Basic selection based measurements						
Delta Frequency	Difference between the upper and lower frequency limits of the selection (Hz)			x		x
Delta Time	Length of the selection (s)				x	
High Frequency	Highest frequency of the selection (Hz)			x	x	x
Length	Number of frames contained in a selection		x		x	
Low Frequency	Lowest frequency of the selection (Hz)	x		x		x
Measurements derived from pitch tracking						
PFC Average Slope	Measurement where the slope is at its average throughout the selection (Hz/ms)	x	x	x	x	x
PFC Maximum Frequency	Measurement that traces the peak frequency throughout the selection (Hz/ms)					
PFC Maximum Slope	Measurement where the slope is at its maximum throughot the selection (Hz/ms)	x		x	x	x
PFC Minimum Frequency	Measurement that traces the peak frequency throughout the selection (Hz/ms)					
PFC Minimum Slope	Measurement where the slope is at its minimum throughot the selection (Hz/ms)					

X's indicate measurements which were retained for each species after a correlation analysis with a cutoff of r = 0.7

BACS = Bachman's Sparrow

FISP = Field Sparrow

GRSP = Grasshopper Sparrow

HESP = Henslow's Sparrow

PRAW = Prairie Warbler

IQR = Inter-quartile Range

PFC = Peak Frequency Contour

Table 2. 2: Breakdown of song type by species used in discriminant function analysis recorded in Fort Bragg, NC, Fort Riley, KS, and Jefferson Proving Grounds, IN (2011-2013).

Species	Song Type	Percent
Bachman's Sparrow (n=1245)	WT (Center)	25%
	WT (High)	27%
	WT (Low)	33%
	WTT	3%
	ZT (Center)	5%
	ZT (High)	3%
	ZT (Low)	2%
	WTWT *	<1%
	WWT *	<1%
	ZTT *	<1%
	ZTWT *	<1%
Field Sparrow (n=1626)	Complex	1%
	Simple (Down)	44%
	Simple (Flat)	48%
	Simple (Up)	7%
	Simple (Wavy) *	1%
Grashopper Sparrow (n=1126)	Simple	94%
	Excited	6%
Henslow's Sparrow (n=966)	Hic-up	100%
Prairie Warbler (n=1247)	Type A	47%
	Type B	33%
	Sing-song	23%
	FISP-like	5%
	Other	3%

** = Song Types with low occurrences which were excluded from analysis*

Table 2. 3: Quality of individual song clips from continuous recordings of five target songbirds recorded in Fort Riley, KS, Fort Bragg, NC, and Jefferson Proving Grounds, IN (2011-2013).

Species	Song Quality Category		
	Faint	**Overlapped**	**Clean**
Bachman's Sparrow (n = 1253)	15%	17%	68%
Field Sparrow (n = 1533)	12%	23%	65%
Grasshopper Sparrow (n = 1145)	23%	10%	67%
Henslow's Sparrow (n = 1067)	12%	10%	78%
Prairie Warbler (n = 1251)	15%	15%	70%

Table 2. 4: Means and discriminate function loadings between five different bird species recorded at Fort Riley, KS, Fort Bragg, NC, and Jefferson Proving Grounds, IN (2011-2013).

Measurement	Bachman's Sparrow	Field Sparrow	Grasshopper Sparrow	Henslow's Sparrow	Prairie Warbler	Discriminant Function Loadings			
						QDA 1	QDA 2	QDA 3	QDA 4
Aggrogate Entropy (dB)	3.97 ± 0.04	3.52 ± 0.03	4.79 ± 0.02	4.86 ± 0.05	3.73 ± 0.04	**0.69**	-0.08	**0.73**	**-1.36**
Average Amplitude	-0.19 ± 0.01	-0.18 ± 0.03	-0.17 ± 0.02	-0.19 ± 0.04	-0.19 ± 0.01	0.01	0.02	-0.08	-0.11
IQR Duration (s)	0.86 ± 0.02	0.97 ± 0.02	0.58 ± 0.01	0.12 ± 0.003	0.71 ± 0.01	**-0.68**	**-2.47**	**-1.97**	**-1.09**
Minimum Entropy (dB)	1.94 ± 0.04	1.40 ± 0.03	2.05 ± 0.03	2.02 ± 0.04	1.62 ± 0.03	0.06	0.34	**-0.62**	0.4
PFC Average Slope (Hz/ms)	-0.10 ± 0.04	0.01 ± 0.01	-0.02 ± 0.06	-0.04 ± 0.27	0.05 ± 0.02	< -0.01	-0.02	0.02	0.02
PFC Maximum Slope (Hz/ms)	675.81 ± 15.29	218.54 ± 4.92	632.19 ± 9.29	692.82 ± 14.54	354.48 ± 6.06	< 0.01	-0.001	< 0.01	< 0.01
PFC Minimum Frequency (Hz)	3149.79 ± 33.13	3081.61 ± 22.95	6156.55 ± 42.93	4017.53 ± 37.38	4561.35 ± 16.30	< 0.01	< -0.01	< 0.01	< 0.01
Peak Power (dB)	77.59 ± 0.72	74.45 ± 0.41	72.90 ± 0.59	75.20 ± 0.51	70.05 ± 0.54	0.01	0.01	0.02	-0.04
Peak Time (s)	1.09 ± 0.03	1.37 ± 0.04	0.80 ± 0.03	0.19 ± 0.01	1.06 ± 0.03	-0.19	-0.09	-0.15	0.35
RMS Amplitude	776.05 ± 82.42	311.94 ± 12.66	772.13 ± 48.40	411.29 ± 30.80	294.34 ± 21.63	< 0.01	< 0.01	< 0.01	< 0.01
Time 5% (s)	0.25 ± 0.01	0.36 ± 0.01	0.30 ± 0.02	0.06 ± 0.003	0.27 ± 0.01	-0.03	**-2.78**	-0.5	**-1.27**

Table 2. 5: Means and discriminate function loadings for Bachman's Sparrow songs recorded at Fort Bragg, NC (2011-2013).

Measurements	Song Type Mean Values			DFA Loadings	
	WT	**WTT**	**ZT**	**QDA 1**	**QDA 2**
Average Amplitude	-0.18 ± 0.01	-0.21 ± 0.06	-0.22 ± 0.05	-0.46	**-1.26**
Average Entropy	3.37 ± 0.04	3.24 ± 0.17	3.44 ± 0.08	-0.28	0.93
Center Time (s)	1.05 ± 0.02	1.22 ± 0.12	1.10 ± 0.05	-0.88	-0.36
IQR Bandwidth (Hz)	1045.07 ± 40.55	1018.75 ± 196.17	1031.25 ± 91.12	< -0.01	< -0.01
IQR Duration (s)	0.83 ± 0.01	1.09 ± 0.10	0.79 ± 0.04	**1.44**	0.03
Low Frequency (Hz)	3013.55 ± 37.65	2583.48 ± 85.67	3058.49 ± 78.56	< -0.01	< -0.01
Maximum Amplitude	3459.23 ± 394.13	8373.83 ± 323.85	3315.33 ± 867.42	< -0.01	< -0.01
Maximum Entropy (dB)	4.49 ± 0.03	4.63 ± 0.122	4.50 ± 0.09	0.68	-0.95
Minimum Entropy (dB)	1.92 ± 0.04	1.95 ± 0.21	2.08 ± 0.11	0.03	0.46
Peak Frequency (Hz)	4496.41 ± 56.80	4342.97 ± 304.59	4599.85 ± 153.42	< -0.01	< -0.01
Peak Time (s)	1.04 ± 0.03	1.12 ± 0.21	1.18 ± 0.08	-0.25	0.77
PFC Average Slope (Hz/ms)	-0.13 ± 0.04	-0.07 ± 0.19	-0.10 ± 0.12	0.11	0.20
PFC Max Slope (Hz/ms)	641.12 ± 14.31	811.17 ± 84.27	712.99 ± 42.02	< -0.01	< -0.01
Time 5% (s)	0.24 ± 0.01	0.28 ± 0.06	0.27 ± 0.02	**1.20**	**2.67**
Time 95% (s)	1.86 ± 0.02	2.27 ± 0.12	1.86 ± 0.05	**1.82**	**-1.00**
1st Quartile Frequency (Hz)	4013.13 ± 37.83	3840.62 ± 145.56	4144.78 ± 103.87	< -0.01	< -0.01

Table 2. 6: Means and discriminate function loadings between Bachman's Sparrow WT and ZT songs recorded at Fort Bragg, NC (2012-2013).

	Measurement	Mean Values for Song Types			QDA Loadings	
		Center	High	Low	QDA 1	QDA 2
WT Song Types	Average Amplitude	-0.17 ± 0.02	-0.17 ± 0.02	-0.19 ± 0.02	**-0.64**	0.10
	Average Entropy	3.21 ± 0.06	3.32 ± 0.06	3.54 ± 0.06	**0.82**	-0.77
	Center Time (s)	1.01 ± 0.03	1.08 ± 0.03	1.06 ± 0.03	-0.01	-0.46
	IQR Bandwidth (Hz)	914.14 ± 63.02	1014.57 ± 76.50	1179.18 ± 66.84	< -0.01	< -0.01
	IQR Duration (s)	0.83 ± 0.03	0.82 ± 0.03	0.83 ± 0.02	0.19	**2.96**
	Low Frequency (Hz)	3117.00 ± 66.50	3172.02 ± 76.00	2802.50 ± 47.98	< -0.01	< 0.01
	Maximum Amplitude	4227.84 ± 835.96	2641.67 ± 530.90	3556.89 ± 663.02	< 0.01	< 0.01
	Maximum Entropy	4.38 ± 0.05	4.40 ± 0.05	4.64 ± 0.05	0.5	0.84
	Maximum Frequency (Hz)	4466.12 ± 94.45	4574.39 ± 110.00	4454.40 ± 90.32	< 0.01	< 0.01
	Minimum Entropy	1.79 ± 0.07	1.90 ± 0.06	2.04 ± 0.06	0.15	-0.03
	Peak Time (s)	0.97 ± 0.06	1.10 ± 0.06	1.04 ± 0.06	-0.06	-0.35
	PFC Average Slope (Hz/ms)	-0.16 ± 0.07	-0.15 ± 0.07	-0.09 ± 0.07	0.23	-0.05
	PFC Maximum Slope (Hz/ms)	614.89 ± 23.46	623.77 ± 25.25	675.51 ± 24.57	< 0.01	< 0.01
	1st Quartile Frequency (Hz)	4061.79 ± 63.26	4159.11 ± 78.75	3854.22 ± 51.16	< -0.01	< -0.01
	Time 5% (s)	0.23 ± 0.02	0.25 ± 0.02	0.23 ± 0.02	0.12	**1.69**
	Time 95% (s)	1.82 ± 0.03	1.89 ± 0.04	1.87 ± 0.03	**-0.41**	**-2.55**
ZT Song Types	Average Amplitude	-0.19 ± 0.06	-0.21 ± 0.09	-0.29 ± 0.08	< 0.01	**-1.37**
	Average Entropy	3.50 ± 0.14	3.30 ± 0.14	3.48 ± 0.14	-0.24	0.23
	Center Time (s)	1.02 ± 0.07	1.16 ± 0.10	1.15 ± 0.10	**3.27**	**2.92**
	IQR Bandwidth (Hz)	1023.03 ± 149.99	987.65 ± 147.23	1080.82 ± 180.39	< 0.01	< -0.01
	IQR Duration (s)	0.80 ± 0.06	0.74 ± 0.06	0.84 ± 0.07	0.31	**2.36**
	Low Frequency (Hz)	2940.82 ± 93.58	3295.96 ± 136.96	2922.47 ± 174.87	< -0.01	< -0.01
	Maximum Amplitude	4135.70 ± 1701.22	2949.33 ± 1536.21	3261.14 ± 1528.52	< -0.01	< 0.01
	Maximum Entropy	4.60 ± 0.13	4.29 ± 0.15	4.59 ± 0.17	**-1.05**	0.04
	Maximum Frequency (Hz)	4355.26 ± 198.49	4723.84 ± 307.62	4871.77 ± 283.13	< 0.01	< 0.01
	Minimum Entropy	2.07 ± 0.17	1.96 ± 0.16	2.25 ± 0.29	-0.31	0.61
	Peak Time (s)	1.17 ± 0.11	1.24 ± 0.15	1.12 ± 0.18	-0.54	**-1.37**
	PFC Average Slope (Hz/ms)	-0.13 ± 0.29	-0.06 ± 0.15	-0.08 ± 0.27	-0.03	-0.11
	PFC Maximum Slope (Hz/ms)	728.94 ± 64.89	625.93 ± 74.43	794.76 ± 70.56	< -0.01	< 0.01
	1st Quartile Frequency (Hz)	4019.74 ± 110.99	4257.99 ± 249.46	4209.05 ± 155.09	< 0.01	< -0.01
	Time 5% (s)	0.24 ± 0.03	0.30 ± 0.04	0.29 ± 0.06	**0.98**	**1.81**
	Time 95% (s)	1.83 ± 0.08	1.84 ± 0.11	1.93 ± 0.10	**-1.12**	**-1.23**

Table 2. 7: Means and discriminate function loadings for Field Sparrow songs recorded at Fort Riley, KS, and Jefferson Proving Grounds, IN (2011-2013).

Measurements	Song Type Mean Values				DFA Loadings		
	Complex Song Type	Simple Song Type					
		Down	Flat	Up	QD1	QD2	QD3
Average Amplitude	-0.15 ± 0.04	-0.18 ± 0.004	-0.18 ± 0.01	-0.19 ± 0.01	0.06	2.27	-5.75
Average Entropy (dB)	2.78 ± 0.20	2.79 ± 0.03	2.53 ± 0.04	2.53 ± 0.11	-0.14	1.89	-0.77
IQR Bandwidth (Hz)	401.78 ± 92.70	597.32 ± 16.61	351.32 ± 12.24	323.79 ± 36.25	< -0.01	**< -0.01**	**< -0.01**
IQR Duration (s)	1.42 ± 0.17	0.96 ± 0.03	0.95 ± 0.03	1.01 ± 0.06	0.42	**1.07**	0.02
Length	617.33 ± 79.08	525.16 ± 7.01	440.00 ± 7.20	474.74 ± 24.30	< -0.01	< 0.01	< -0.01
Minimum Entropy (dB)	1.42 ± 0.14	1.51 ± 0.03	1.28 ± 0.04	1.46 ± 0.09	**-0.26**	**-0.40**	1.57
Peak Time (s)	1.81 ± 0.48	1.49 ± 0.06	1.25 ± 0.06	1.30 ± 0.13	0.11	0.14	-0.19
PFC Average Slope (Hz/ms)	0.01 ± 0.09	0.03 ± 0.02	-0.002 ± 0.01	-0.02 ± 0.04	-0.40	-0.18	-0.76
1st Quartile Frequency (Hz)	3591.08 ± 59.38	3502.34 ± 15.99	3574.28 ± 33.83	3656.25 ± 101.99	< 0.01	< -0.01	< 0.01
RMS Amplitude	319.04 ± 86.93	391.45 ± 21.63	232.59 ± 12.44	298.95 ± 34.37	< -0.01	< 0.01	-0.45
Time 5% (s)	0.39 ± 0.12	0.45 ± 0.02	0.27 ± 0.02	0.32 ± 0.04	-0.81	-0.45	0.77

Table 2. 8: Means and discriminate function loading for Grasshopper Sparrow song types recorded at Fort Riley, KS (2011 & 2013).

Measurement	Song Type Mean values		QDA Loading
	Excited	Regular	
Low Frequency (Hz)	5942.16 ± 93.44	6134.64 ± 44.07	< -0.01
High Frequency (Hz)	9714.67 ± 90.14	9946.40 ± 29.03	< -0.01
3rd Quartile Frequency (Hz)	8132.47 ± 82.79	8504.69 ± 18.83	< 0.01
Aggregate Entropy	4.95 ± 0.06	4.78 ± 0.02	-2.59
Average Amplitude	-0.18 ± 0.03	-0.19 ± 0.01	-0.33
Average Entropy	4.09 ± 0.07	4.16 ± 0.01	2.92
Center Time (s)	0.94 ± 0.10	0.85 ± 0.02	-1.10
IQR Bandwidth (Hz)	1389.46 ± 109.00	1136.68 ± 26.59	< -0.01
IQR Duration (s)	0.66 ± 0.30	0.57 ± 0.01	-0.30
Peak Frequency (Hz)	7518.19 ± 197.39	7747.08 ± 53.24	< -0.01
Minimum Entropy	1.97 ± 0.10	2.07 ± 0.03	< 0.01
PFC Average Slope (Hz/ms)	-0.29 ± 0.29	-0.004 ± 0.06	0.12
Peak Time (s)	0.92 ± 0.13	0.79 ± 0.03	0.25
RMS Amplitude	364.84 ± 62.00	805.38 ± 51.80	< 0.01
Time 5% (s)	0.21 ± 0.02	0.30 ± 0.02	0.19

Table 2. 9: Means and discriminate function loadings for Prairie Warbler songs recorded at Fort Bragg, NC and Jefferson Proving Grounds, IN (2011-2013).

Measurements	Song Type Mean Values					Descriminant Function Loadings			
	Type A	Type B	Type B (Sing-Song)	Type B (FISP)	Other	QDA 1	QDA 2	QDA 3	QDA 4
Low Frequency (Hz)	4481.94 ± 27.14	46415.57 ± 39.46	4450.43 ± 17.48	4640.48 ± 90.32	4107.54 ± 109.77	< -0.01	< -0.01	< -0.01	< -0.01
Delta Frequency (Hz)	2459.43 ± 38.65	1930.91 ± 37.35	1240.87 ± 52.87	1890.14 ± 90.32	2077.23 ± 163.91	< -0.01	< 0.01	< -0.01	< -0.01
1st Quartile Frequency (Hz)	4965.35 ± 20.99	4982.01 ± 24.32	4798.52 ± 27.54	5169.86 ± 84.76	4557.05 ± 127.08	< 0.01	< 0.01	< 0.01	< -0.01
3rd Quartile Frequency (Hz)	5648.61 ± 33.39	5664.79 ± 23.51	5186.68 ± 41.61	5665.83 ± 87.89	5429.34 ± 92.19	< -0.01	< -0.01	< -0.01	< 0.01
Aggrogate Entroppy	3.91 ± 0.05	3.76 ± 0.05	2.95 ± 0.05	3.64 ± 0.12	4.18 ± 0.11	**-0.47**	**-0.99**	**1.05**	**0.88**
Average Amplitude	-0.19 ± 0.02	-0.20 ± 0.01	-0.17 ± 0.01	-0.17 ± 003	-0.20 ± 0.03	0.10	**0.97**	0.17	-0.04
Duration 90% (s)	1.37 ± 0.03	1.58 ± 0.04	1.63 ± 0.05	1.09 ± 0.07	1.67 ± 0.12	**1.29**	**-2.06**	**-0.97**	**-1.33**
Peak Frequency (Hz)	5183.39 ± 39.17	5290.62 ± 44.81	5054.28 ± 51.95	5434.48 ± 118.13	5095.11 ± 261.12	< 0.01	< -0.01	< -0.01	< 0.01
Minimum Entropy	1.65 ± 0.05	1.63 ± 0.06	1.54 ± 0.08	1.38 ± 0.10	1.72 ± 0.16	-0.32	0.55	-0.49	-0.29
PFC Average Slope (Hz/ms)	0.07 ± 0.04	0.05 ± 0.04	0.01 ± 0.04	-0.01 ± 0.11	0.04 ± 0.18	0.11	-0.04	-0.25	-0.07
PFC Max Slope (Hz/ms)	391.00 ± 9.51	351.60 ± 7.20	267.89 ± 12.46	272.74 ± 17.77	326.34 ± 32.16	< 0.01	< -0.01	< -0.01	< -0.01
Peak Time (s)	0.90 ± 0.04	1.20 ± 0.07	1.28 ± 0.10	0.80 ± 0.10	1.31 ± 0.23	-0.15	-0.15	0.27	0.30
RMS Amplitude	316.15 ± 16.79	225.83 ± 15.14	157.57 ± 53.30	189.36 ± 33.70	131.95 ± 27.61	< -0.01	< -0.01	< -0.01	< -0.01
Time 5% (s)	0.24 ± 0.02	0.30 ± 0.03	0.31 ± 0.04	0.21 ± 0.06	0.24 ± 0.08	**1.59**	**-1.40**	**-0.79**	**-0.71**

Chapter 3: Semi-supervised analysis methods affecting the outcome of automatic detection of bird songs of five sympatric grassland songbird species from passive continuous recordings

Abstract

Extraction and classification of bird songs from continuous recordings are challenging because song variation, overlapping noise, and low signal-to-noise ratios obscure targeted signals. There are proven approaches for extracting bird songs from continuous recordings under local, species-specific conditions, but no consistent approach that works for all species. Our goal was to develop a comprehensive, analytic approach that could be used for extraction and classification of any passerine song recorded from typical passive continuous recordings. We first extracted target songs from continuous recordings using a template matching method, and then investigated which post-extraction methods worked best for classifying extracted target songs from noise. Species with more tonal song elements had the greatest overall sensitivity values, and templates from population-wide sources (i.e. Xeno-Canto) worked better for song extraction for these species. Less tonal songs showed fewer consistent results; classification accuracy improved when the amplitude of the target song was much greater than the surrounding soundscape. There were no consistent results based on different semi-supervised methods. This study demonstrated the importance of understanding the nature of the target vocalizations when choosing an analytic approach for extraction and classification of songs from continuous acoustic recordings.

Introduction

Bird songs contain a wide variety of information that can be used to learn about passerine behavior (Borror 1961, Kroodsma 1981, Acevedo et al. 2009). However, the variability in bird songs within and among species also creates challenges for using passive acoustic recording approaches to monitor avian populations. Advances in recording technology and data storage have greatly facilitated the collection of acoustic data, but the ability to process the volumes of data efficiently and accurately is still problematic for avian species (Anderson et al. 1996, Bardeli et al. 2010). New analytical techniques with standardized approaches must be created to manage and analyze the use of these large datasets to fully utilize the potential of acoustic monitoring. Study into what works best for different types of songs is warranted, as it is unlikely that one approach will work well for all species or situations.

Automatic detection of animal vocalizations has proven to be a useful tool in monitoring populations of many species including birds. For example, bats are commonly recorded using acoustic equipment and standardized monitoring methods exist in North America and Europe which show reliable results across a number of species (Whitby et al. 2014). Their short, precise calls in a relatively quiet acoustic landscape make it easy to record and classify vocal cues. However, the minute differences between species often make generalizing to genus necessary in identification Arrays of acoustic monitoring stations have been used to passively to monitor communities of beluga whales (*Delphinapterus leucas*) and songbirds during migrations at night (Roy et al. 2010, Harlow et al. 2013, Stepanian et al. 2016).

Underwater acoustic surveys of whales have been utilized for many declining species and can greatly increase accuracy of population estimation when combined with visual detection methods (Barlow and Taylor 2005, von Benda-Beckmann et al. 2010). These methods work in part because they are tailored to each species/vocalization type and fit specific research questions.

Avian species are monitored by sight and sound in systematic, well-documented ways including point counts, line transects, and spot mapping (Bibby et al. 2000). Passive acoustic recording systems can replace traditional methods and collect more data when time, accessibility, or human expertise are in short supply (Hutto and Stutzman 2009). However, extensive training must be done prior to application of acoustic monitoring to understand how to implement an acoustic monitoring approach in the field. Additional training is needed to analyze the data to produce consistent results (Brumm et al. 2017). The foundation of the analysis of acoustic monitoring data is the use of various analytic approaches to first extract and then classify bird songs (Potamitis et al. 2014). Study-specific monitoring objectives may warrant different approaches in how acoustic signals are processed for each step. Currently there is no standardized computer program, classification method, or training/sample size that could generate comparable results across studies (Knight et al. 2017). Many programs and custom algorithms have been used to detect a particular species' songs, but these approaches are often species and location specific and typically are generated using small populations of interest or training datasets. In addition, variability in bird song may influence the analytic approach as well as the interpretation of the output (Anderson et al. 1996, Acevedo et al. 2009, Raghuram et al. 2016). The use of passive recording

systems to monitor wildlife would become more wide-spread and generally accessible if there was a standardized method or analytical pathway for the user to follow.

Methods for song extraction can either be based on waveform (amplitude) or spectrogram views. Tonal sounds that are loud compared to their background and have little overlap with other sounds work well with waveform detection, as it focuses on finding the loudest events in a recording (Evans and Mellinger 1999, Brandes 2008). Many studies focused on flight calls of migratory birds at night or other short and relatively loud signals in a quiet environment have had success with this type of detection (Farnsworth 2005, Zwart et al. 2014, Knight et al. 2017).

Spectrogram-based extraction can either be based on band-limiting energy detectors (those that look for high levels of energy within specific, user-defined frequency bands) or more frequently correlation template matching. Like waveform detection methods, the band-limited energy detector approach works for short, precise, and identical acoustic signals, but isn't robust enough to maintain good results when vocalizations are variable in either frequency or time (Keen et al. 2014). The eXtensible BioAcoustic Tool (XBAT) developed by Cornell's Lab of Ornithology (CLO) is one commonly used template correlation-based extraction tool that is run in MATLAB software (MathWorks Inc., Natick, MA). XBAT compares user-defined song templates against continuous recordings and with sliding window analysis identifies acoustic signals that exceed a user-specified correlation threshold. This template method effectively combines extraction and classification into one step. Based on this type of pattern matching approach, reported positive predictive values (the percentage of true positives out of all positive results, PPV; 0.39 up to 1.0) and sensitivities (true positives /

false negatives + true positives; 0.15 to 0.70) were highly variable in different studies (Charif and Pitzrick 2008, Fristrup and Clark 2009, Goh 2011, Goyette et al. 2011). The number of detectors used in these studies varied from as low as one to more than three hundred individual human-extracted templates. Template matching approaches in general work best when there is low variation between individuals and when the background noise is similar between the template and in the continuous recording. These conditions are seldom met when analyzing passive acoustic monitoring recordings. Nevertheless, template matching may still be useful as a signal extraction method for continuous recordings when ideal conditions can't be met.

More complex variations on template matching are frequently used and are incorporated into some commercially available software, like Kaleidoscope, Song Scope (Wildlife Acoustics, Inc., Maynard, MA) and Raven Pro (Cornell Lab of Ornithology, Ithaca NY). Hidden Markov models (HMMs) flag songs in continuous recordings that maximize probability of matching a template based on the arrangement of song syllables and their metrics (Kogan and Margoliash 1998). Kaleidoscope uses HMMs and also k-means clustering, which can incorporate templates of many different species at once. Some studies have used dynamic-time-warping (DTW) successfully to accommodate unpredictable variation within bird songs (Kaewtip et al. 2013, Ruse et al. 2016). Instead of matching templates exactly in both frequency and time dimensions, DTW allows for plasticity in the time dimension of a potential target call. However, DTW can require many templates and be hard to implement computationally, especially in continuous recordings (Buck and Tyack 1993).

Many of the extraction methods mentioned above combine classification into the extraction process but separating the extraction and classification processes may lead to more flexibility in approach and greater overall success. Classification methods start with either user or computer extracted song clips and typically utilize multivariate statistical techniques to separate target sounds from ‘noise’. Neural networks (Ashiya and Nakagawa 1993, McIlraith and Card 1997, Nickerson et al. 2006) decision trees (Acevedo et al. 2009, Digby et al. 2013), hidden markov models (Trifa et al. 2008), and discriminant analysis (Anderson et al. 1996) have all been used with various levels of success to classify bird songs. Spectral peak or frequency tracking has also been proven to work on both birds with clean/tonal songs and other vocalizing species (Forrest 1994, Chen and Maher 2006, Brandes 2008, Kershenbaum and Garland 2015). These statistically-based methods have been successful in telling species apart, but are only accurate if the measurements used to describe bird songs capture the appropriate variability needed to separate them (see Chapter 2)

There is currently no one best approach for classification of bird songs (Acevedo et al. 2009, Knight et al. 2017). Choice of target vocalizations, recording method, sample size, and extraction type all may affect correct classification rates and the consistency of the results. Acevedo (2009) found that support vector machines were best at discriminating between songs of three different avian species compared to discriminant function analysis and neural networks, whereas Knight (2017) had the best classification success with convolutional neural networks, and Digby (2013) produced good results with decision trees. results from studies comparing classification methods are often hard to interpret, because they frequently classify songs into groups by

species, and do not classify songs into target/non-target bins. Labeling discrete songs, even with inherent variation, are easier to handle statistically than separating all forms of non-target noise from songs of interest.

Ease of use of the classification methods is also an issue. Researchers have addressed the limitations in commercially available automatic detection software programs by creating custom algorithms in programming software to improve performance for specific research species or questions. However, the programs required to run more advanced techniques can be difficult to use and may not be widely available. Statistically-based approaches have worked under limited circumstances or with simple vocalizations, where they end up over-fitting data, but have had limited success in detecting bird songs from typical continuous recordings (Potamitis et al. 2014). Even when understanding the source and type of variation in bird songs, there are unavoidable issues. Passive recording systems often record wind, insects, overlapping target or non-target songs, and anthropocentric noises, all which make accurate automatic detection much more difficult and render some types of analysis (i.e. harmonic or waveform based) unusable (Lasseck 2013, Keen et al. 2014). Some newer studies have abandoned the detection of individual songs and instead successfully focused on the complexity of a soundscape as a measure of species diversity (Kasten et al. 2010, Gasc et al. 2015). However, there is still a great value in detecting individual songs in a recording as a means to monitoring individual species of interest. More work needs to be done to both better isolate target bird songs in typical field-based recordings as well as create general recommendations for recognizer development and deployment (Acevedo et al. 2009, Blumstein et al. 2011, Knight et al. 2017)

Our goal in this study was to develop a better understanding of the process of extraction and classification of bird songs. Based on the knowledge gained, we developed an analytic pathway that can be used for a diversity of species songs and other vocalizations using acoustic recordings from typical passive monitoring systems. We chose to analyze song extraction and classification as separate processes to enable flexibility in approach and to allow for more specific evaluation of performance. We evaluated the performance tradeoffs in different approaches based on three specific analytical issues in the pathway: how to choose a template source for extraction, how to set a correlation threshold for extraction, and how to select a semi-supervised learning technique for classification (Figure 3.1). We only used moving window correlation analysis to initially extract song selections, as it was a simple, effective way to extract target signals from typical field recordings where masking songs of conspecifics or other species were abundant. The subsequent questions for the study were as follows:

1. Do templates from a local or global (generic) population perform better?
2. What is the best system of choosing a correlation threshold for extraction?
3. Do different semi-supervised learning methods perform better when separating results between target and non-target sounds?
4. How much does the signal to noise ratio affect classification performance?
5. To what degree do these options differ by species/song types?

Methods

Acoustic monitoring database

Passive acoustic monitoring was conducted during May-July2011-2012 at Jefferson Proving Grounds, Indiana, Fort Bragg, North Carolina, and Fort Riley, Kansas. Five sympatric grassland/shrubland songbird species were selected for analysis that had a variety of song types:Bachman's Sparrow (*Peucaea aestivalis*), Field Sparrow (*Spizella pusilla*), Grasshopper Sparrow (*Ammodramus savannarum*), Henslow's Sparrow (*Ammodramus henslowii*), and Prairie Warbler (*Setophaga discolor*). Typical songs from these species comprise a wide range of acoustic characteristics including strong tonal notes (Bachman's Sparrow, Field Sparrow, and Prairie Warbler), fast trills (Bachman's Sparrow and Grasshopper Sparrow) and fast chirps (Henslow's Sparrow). The song variation at the individual, population, and regional scale varied by species because of different song learning methods and communication strategies (Chapter 2).

Bird songs were recorded using SM2 Song Meters (Wildlife Acoustics, Inc., Maynard, MA) with 48dB internal gain and a 24 kHz sampling rate to record target species across the three study sites. Song Meters were placed in the middle of a target individual's territory from early May to mid-July. Nine different focal males of each species were targeted for recording at each study site per year. Although recorders were placed within the territory of a focal male, songs from 1-4 neighboring males were present on the recordings and were included in the analysis. Thus, the acoustic database included songs of >25 individual males per species. Thirty 5-min passively

recorded song clips were randomly chosen for each species across all study sites, with each clip selected to contain ≥25 target songs. The acoustic database for each species, therefore, included >750 individual songs for extraction and classification. Chips, excited notes, and calls were not used in this study to focus correlation analysis on primary song types. Stereo recordings with two omnidirectional microphones pointed in opposite horizontal directions were made from sunrise to four hours past sunrise. The channel with the least noise or interference was retained for use in the analysis. Trained technicians used RavenPro 1.5 to manually listen to and view recorded data and to tag each target bird song. I then reviewed tagged files to ensure accuracy and consistency. One third of the 5-minute clips for each species were randomly chosen and set aside as a training dataset.

Templates were either chosen from local recordings made at the study locations or from clips available from Xeno-Canto (http://www.xeno-canto.org). I selected templates from both sources that included all species' song types with minimal background noise. Local templates were obtained from individuals that were independent of those recorded in the 5-minute clips. Xeno-Canto is a website dedicated to sharing user-submitted bird songs from around the world, and currently holds recordings of 9000+ species, or 90% of all known birds (August et al. 2015). I chose Xeno-Canto templates from all available clips for each species to represent the greatest possible geographic range.

Acoustic song metrics useful in accurately discriminating between inter- and intra-species songs (Chapter 2) were calculated on 200+ potential templates for each species from each template source (local or Xeno-Canto). I used correlation analysis

on each set of templates using the ‘mclust’ package in Program R to screen out templates that were the most highly correlated with others, in effect grouping templates based on similar song measurement characteristics. The remaining templates captured the majority of variation in each species’ songs. The number of templates retained for analysis ranged from 30 (Bachman’s Sparrow) to 8 (Henslow’s Sparrow), reflective of the variability in song for a given species.

Correlation Analysis

I used spectrogram cross correlation in the R package ‘monitoR’ as the foundation for the song extraction analysis (Hafner and Katz 2017). MonitoR is a template-matching correlation algorithm that is highly flexible and easily accessible at no charge to the user. All spectrograms were created using a Fast Fourier Transform (FFT) algorithm with a 512-pt Hann window and 50% overlap. Each template was run as a ‘moving window’ over the target spectrogram at a frame by frame basis. The running correlation result between the template and target spectrogram records a ‘hit’ when the correlation value crosses as user-defined threshold. The individual template with the greatest correlation value was recorded when multiple templates flagged the same song within a recording. Local and Xeno-Canto templates for each species were used in analysis as independent groups of templates on the same 5-minute clips. Thresholds for separating detections from ‘noise’ were chosen either based on correctly identifying 95% of all true signals in 5-minute clips or by using receiver operating characteristic (ROC) curves to maximize detection of true positive signals while

minimizing false positive classifications. Correlation threshold values were chosen by pooling all test files for a species together to minimize variation between test files.

Semi-supervised methods of categorizing

I used three commonly used methods of semi-supervised learning to classify signals from all the monitoR output selections. First, the 11 to 19 species-specific measurements useful in discrimination of species and song types (Chapter 2) were calculated using Raven Pro 1.5 software for the signals that had been extracted via template matching. After recording the efficiency of the extraction process, only the false positive and true positive correlation selection outputs were retained after initial analysis. Real-world song files would not have the target song clips labeled, and as such only unlabeled, positive results would be retained when using the monitor package. The resulting tables were analyzed with different but commonly used methods of semi-supervised learning to classify true and false signals. I ran neural network analysis on the training datasets using the pattern recognition and classification application within the Matlab Neural Network package using scaled function backpropagation. The ideal number of hidden nodes was determined for each combination of template source (local vs Xeno-Canto) and correlation threshold score (95% vs. ROC), while minimizing overfitting of the testing data. The neural network algorithms were then applied back to the evaluation dataset for each species to determine correct classification rates. I conducted decision tree analysis to classify the selection clips in the training datasets using the ‘rpart’ and ‘rpart.plot’ packages in

Program R (Therneau et al. 2017). Trees were created until new splits were no longer statistically significant ($P > 0.05$), and then pruned to minimize the cross-validated error and over-fitting. The final trees were then applied to the evaluation dataset for each species to determine correct classification rates for each species. Finally, I used the 'MASS' package in R to conduct quadratic discrimination analysis (QDA) on the training datasets. This method, as compared to linear discrimination analysis, allowed for unequal covariances between the selection measurements that better fit the underlying input data. All measurements for QDA were scaled to standardize the mean and variance to avoid bias during calculations.

I compared the results from each analytical approach using confusion matrices for all steps in the pathway. Initial metrics comparing program monitoR results included total number of selections per file averaged over species, prevalence of true positives over all selections, sensitivity (the percent of all true occurrences that were correctly classified), and the false discovery rate (percent of all predicted positives that are false negatives). Global sensitivity was also calculated for ROC files, which included all true positives retained after using the ROC score cutoff in addition to any rejected true positive clips under the cutoff. The primary metrics for assessing accuracy of the final semi-supervised analysis methods included sensitivity, specificity (the percent of all false occurrences that were correctly classified) and the positive predictive value (PPV: the percentage of all predicted true occurrences that were actually true positives).

Average power (dB) for each program monitoR selection was recorded in Raven Pro 1.5 with other song metrics. The average power within the same frequency bounds for each five-minute clip was also recorded, and the ratio of signal to average noise was

calculated for all true positive results of each species. I used this ratio to calculate the probability of accurately predicting a true signal with logistic regression in program JMP (SAS Institute Inc., Cary, NC).

Results

Henslow's Sparrows generally had four times more detections than all other species per file for all template/score cutoff combinations (Table 3.1). The greatest value for sensitivity, the proportion of properly identified positive results (true positives / false negatives + true positives), was 0.99 ± 0.01 for Henslow's Sparrows with local templates. Sensitivity was also high for Prairie Warblers across the board and for Bachman's Sparrows (0.97 ± 0.01) and Field Sparrows (0.95 ± 0.02) using Xeno-Canto templates and the 95% correlation threshold. Sensitivity for Grasshopper Sparrows varied the most among all template and correlation threshold combinations. False discovery rate was the greatest in Grasshopper Sparrows (0.71 ± 0.04) and Henslow's Sparrows (0.73 ± 0.03) and least for all species overall when using the ROC correlation threshold.

Thirty local Bachman's Sparrow templates and twenty-one templates from Xeno-Canto were retained after correlation for analysis. Neural network hidden nodes ranged from eight to nine and an average of nine terminal leaves were used in the final pruned decision trees (Figure 3.2). Greatest sensitivity rates occurred with Xeno-Canto templates and ROC correlation thresholds for all three final analysis methods, the greatest being 0.82 ± 0.04 with decision trees (Table 3.2). Specificity was greatest for

all analysis methods using the 95% correlation threshold and the greatest value was again generated with decision trees (0.84 ± 0.04). Positive predictive values were relatively high for local templates and ROC correlation thresholds, and greatest for neural networks (0.70 ± 0.09).

Analysis of Field Sparrow data used twenty local templates, eighteen Xeno-Canto templates, 7-8 hidden nodes in neural networks, and an average of 12 terminal leaves in the pruned decision trees (Figure 3.3). Sensitivity was greatest across the board for decision trees, and reached 1.00 ± 0.02 for neural networks using Xeno-Canto templates and the ROC correlation threshold (Table 3.3). Specificity had the greatest range for Xeno-Canto templates and the ROC correlation threshold: the greatest overall specificity was attained with QDA (0.87 ± 0.12) and the least specificity was attained with neural networks (0.11 ± 0.12). PPV was again greatest for all analysis types with Xeno-Canto templates and the ROC correlation threshold.

Grasshopper sparrow analysis was completed with seven local and eight Xeno-Canto templates. Eight or nine hidden nodes were used to maximize neural network outputs, and an average of nine leaves were left after pruning decision trees (Figure 3.4). There were not enough individual selections left after using the higher ROC correlation threshold, so all files needed to be pooled for those combinations. Sensitivity was greatest for all analysis types using local templates and the ROC correlation threshold, but sensitivity ranged from 0.51 ± 0.10 with QDA to 0.71 ± 0.09 for neural networks (Table 3.4). Specificity was maximized with local templates and ROC correlation threshold but minimized across all analysis types.

Analysis for Henslow's Sparrows was run with 8 local and 8 Xeno-Canto templates, 8-10 hidden neural network nodes, and an average of seven pruned decision tree leaves (Figure 3.5). Sensitivity for Henslow's Sparrows was maximized (0.89 ± 0.07 with neural networks and ROC correlation threshold) and minimized (0.17 ± 0.05 with decision trees and 95% correlation threshold) when using local templates (Table 3.5). Specificity was very similar across all analysis combinations except for local templates with ROC correlation thresholds, which were much lower. Positive predictive values were all around 0.7 to 0.8 for all analysis combinations.

Prairie Warbler results were consistent across templates, with 16 from local sources and 18 from Xeno-Canto (Table 3.6, Figure 3.6). The correlation threshold method had a large effect on classification metrics. Using the ROC correlation threshold increased sensitivity over all final analysis methods, slightly better positive predictive values, and decreased specificity over all analysis methods. Sensitivity was greatest for QDA with local templates and ROC correlation threshold (0.80 ± 0.06) and specificity was greatest with neural networks, local templates, and the 95% correlation threshold (0.80 ± 0.06).

All species had an average signal to noise ratio of roughly 1.0 (Table 3.7). Logistic regression for combinations of species and analysis types were all significant ($P<0.05$). True positive songs of Grasshopper and Henslow's Sparrows were more likely to be correctly classified when the ratio between songs and the background was the greatest. (1.30 ± 0.05 and 1.39 ± 0.01 respectively). The average inflection point for the ratio at which true Field Sparrow songs were most likely to be correctly classified was the least at 0.87 ± 0.05.

Discussion

The potential for using passive acoustic recording systems to monitor songbird populations has been limited because of the challenges in analyzing the massive volumes of data that can be easily acquired (Acevedo et al. 2009). In this study we sought to answer specific questions about how to optimize the analytic pathway for automated extraction and classification of bird songs for five sympatric grassland song species recorded though passive acoustic monitoring. Through examination of the results relevant to the specific questions, we sought to define a generalized approach that other researchers could use in the analysis of passive continuous recordings for monitoring songbird populations.

We split the extraction and classification processes into separate analyses and used a simple template-matching approach widely available in Program R to conduct the extraction process. We structured the simple question of how to develop the template set for extraction into two likely possibilities- either use templates recorded locally or use templates that are available across a wide geographic range (i.e., Xeno-Canto). Bachman's Sparrows and Prairie Warblers had the most consistent results when comparing local and Xeno-Canto-based templates. The similarity between results from both template types suggests that both methods were successful in capturing the variety in songs necessary to be effective in the extraction process. Prairie Warblers have an overall ascending, tonal/buzzy song notes, with less documented variety than the other two tonal-type songbirds in this study (Houlihan 2000, Byers et al. 2013). Bachman's Sparrows and Field Sparrows have more variety in their song types, and the

type of variation for each species could have had an effect in the results (Thorpe and Lade 1961, Searcy 1992). The songs Bachman's Sparrows sing are highly variable between individual (Borror 1971), rendering the difference between regional or global populations moot. However, results between the two template sources showed similar relationships and values between all analysis results. The greatest sensitivity values were found in the Xeno-Canto template results, suggesting that while both sources were successful at capturing variability in Bachman's Sparrow songs, the Xeno-Canto templates apparently captured additional variation that was useful in this correlation analysis. Field Sparrows tend to learn one song type early on from their neighbors and use it throughout their lives (Searcy 1992). This song behavior can lead to development of regional dialects such that use of local templates may perform better in the correlation analysis. Contrary to this expectation, however, Xeno-Canto templates were actually better than local templates at extracting songs leading to correct classification. Conversely, some of the worst specificity rates resulted from use of Xeno-Canto templates. These global templates were good at extracting target songs of all species if they matched the templates, but also introduced more false positives.

The selection of a correlation threshold for template matching also affected the classification results and varied by template (local or global) and by species. For Grasshopper Sparrows, the ROC correlation threshold generated some of the best classification outcomes across all semi-supervised learning methods with local templates, but the ROC threshold eliminated too many true positives with Xeno-Canto templates to produce useful results. The relatively low amplitude trill of Grasshopper Sparrows, as well as the frequency overlap with common insects apparently made it a

difficult song type to detect in the best (signal/noise) circumstances. The ROC correlation threshold with local templates in Prairie Warblers produced the best sensitivity rates (0.73 to 0.88), whereas the 95% correlation threshold with local templates had the worst sensitivity overall (0.49 to 0.57). For this tonal species, which performed the most consistently well across template and correlation threshold alternatives, the ROC threshold was very useful in minimizing false negatives while maintaining the ability to correctly label true positives.

More variation in classification outcomes came from the options during the extraction process (template source and correlation threshold) than from the selection of the semi-supervised analysis method. Field Sparrow songs were best classified by decision trees, with the best combination of high sensitivity and specificity across all template and correlation threshold options. Classification of Grasshopper Sparrow songs, in contrast, produced inconsistent results by classification method. Quadratic discrimination analysis produced the worst sensitivities for Grasshopper Sparrow songs, decision trees were intermediate, and neural networks produced the best sensitivities based on local templates. However, other performance metrics (i.e, specificity, PPV, and FPR) did not follow this same pattern demonstrating that one classification method may perform better at identifying true positives but may perform more poorly in eliminating false positives. This result emphasizes the importance of making informed choices for both extraction and classification analyses rather using a standardized approach for species with different song characteristics.

These semi-supervised classification methods all used the same input measurements and generally produced similar fairly similar results, with no one method

out-performing the others for all five species. More differentiation between the classification methods could be exhibited by including more templates from the start to ensure greater sensitivity rates going into analysis, or by using more sophisticated approaches to the semi-supervised classification. For example, I used a fairly simplistic shallow learning neural network analysis. There are more complex, deep learning options available which are typically used for image classification and speech recognition and may produce better results, but would be harder to implement for the average user (Koops et al. 2015). Similarly, there are many more types of decision trees available that may be better suited for our data (i.e., random forests) that may produce greater classification rates. Alternatively, the similarity of results between classification methods may simply reflect the nature of the differences inherent in the input data. Use of more complex analysis may not show further differentiation in the classification results.

Finally, I addressed the question about how the nature of a given species song compared to the background acoustic environment (noise) affected the classification outcomes by regressing the probability of predicting a true positive signal with the signal/noise ratio. All five species showed significant relationships between the signal/noise ratio and the probability of predicting a true positive, suggesting that the distinctness of a given song affected the ability to correctly classify it. Grasshopper and Henslow's sparrow song classification was more confounded by noise than the other species, and required a greater signal/noise ratio to accurately classify their songs These two types of songs were the least tonal, and overlapped the most with insect noise and call notes from several other species. The average signal/noise ratio for

Grasshopper Sparrows also accounted for the lack of true positives in the dataset after the extraction process. Although we targeted this species in the same fashion as the other species during the recording sessions and song rates for this species were similar to the others (Prevost 2016), their songs often blended into the background and were not extractable in the first place. Exploratory recordings and analysis prior to implementing an acoustic monitoring plan would be helpful for these types of bird songs to allow for screening of files with sufficient signal/noise ratios to ensure successful extraction and classification results.

Different objectives may be met by optimizing the sensitivity rate, the specificity rate, or balancing the performance of the two metrics during extraction and classification analyses. Optimizing the sensitivity rate would be desirable for detecting a rare species (e.g., Ivory-billed Woodpecker [*Campephilus principalis*]), where detection of true positives was much more important than excluding false positives. Balancing the sensitivity and specificity rates may be appropriate when acoustic monitoring is being used to track relative abundance of more common species. As such, selection of the extraction and classification methods may be dependent on project objectives in the first place.

When successful, automated detection of bird songs from continuous acoustic recordings can reduce manual analysis time by >80% (Digby et al. 2013, Ross and Allen 2014). However, there is little consistency in approach and documentation between studies, which complicates the ability to implement an automated detection analysis (Knight et al. 2017). Our results highlight the potential to achieve very high correct classification rates (e.g., >80%) for a suite of songbird species with varied song

types with the proper extraction and classification approach. We documented considerable variability in the optimal approach, however, depending upon the species being studied and the nature of their songs. This key result suggests, then, that a pilot study on extraction and classification is warranted prior to developing an acoustic monitoring plan to ensure that the analytic approach will successfully meet monitoring objectives.

Appendix

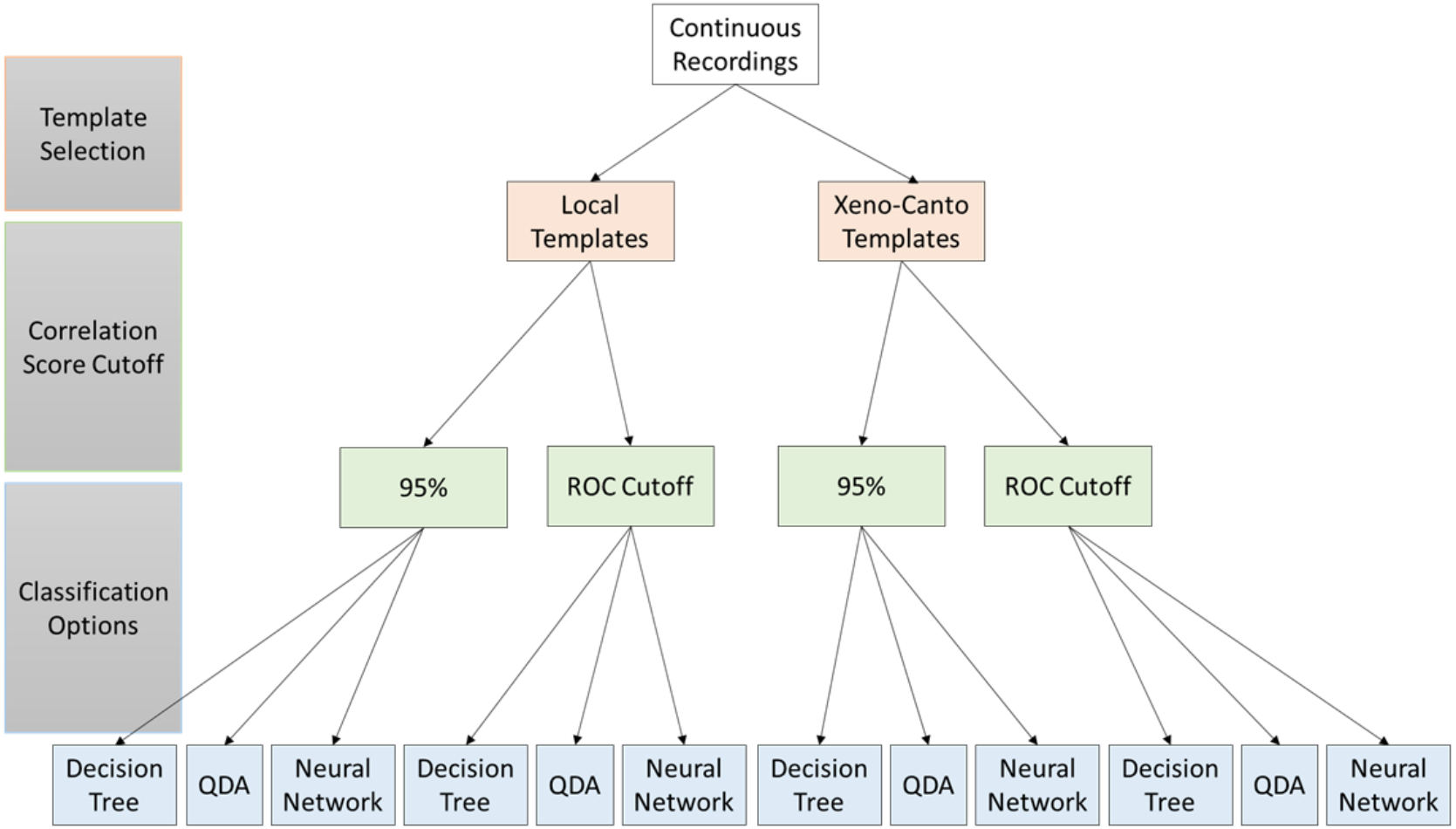

Figure 3. 1: Proposed workflow for bird song analysis of continuous recordings.

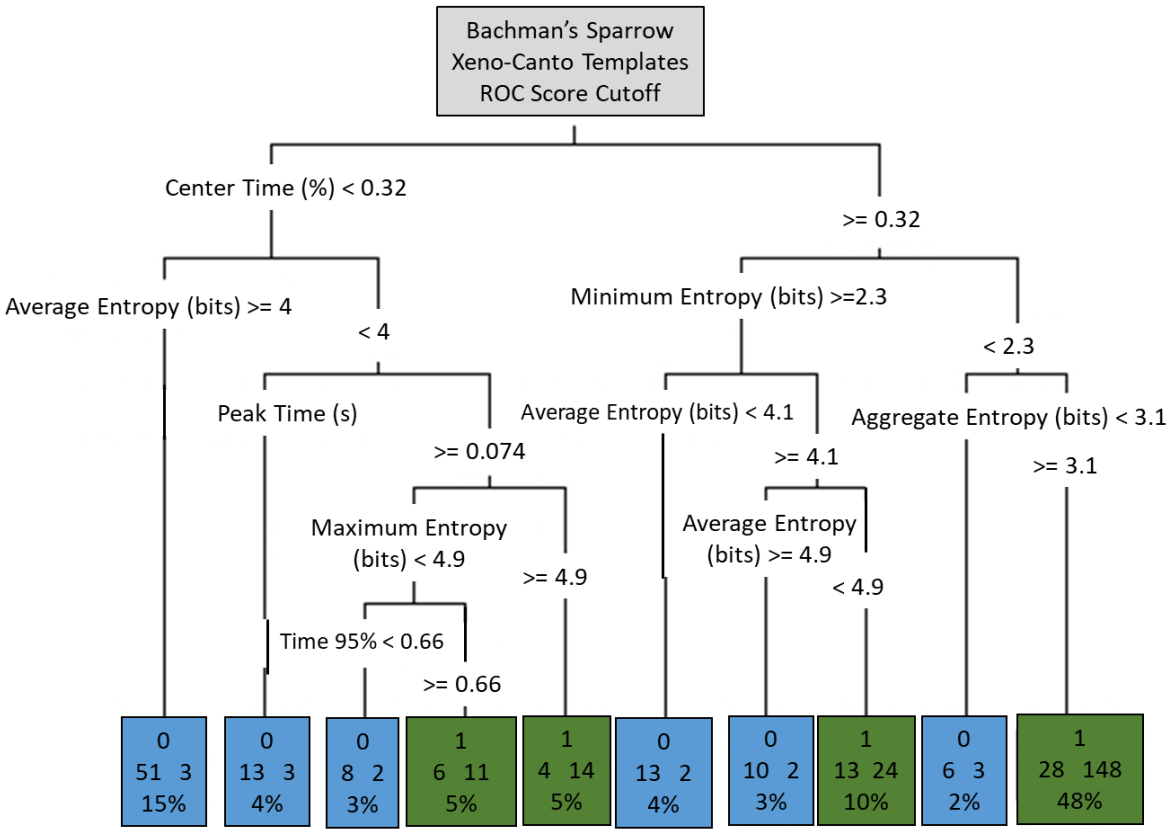

Figure 3. 2: Pruned decision tree for Bachman's Sparrow with the greatest specificity rate (0.82 ± 0.04).

Terminal leaves show the final outcome (0 = False, 1 = True), total number of songs in the leaf that was originally a False + and then True +, and the percentage of total songs that ended up in that leaf.

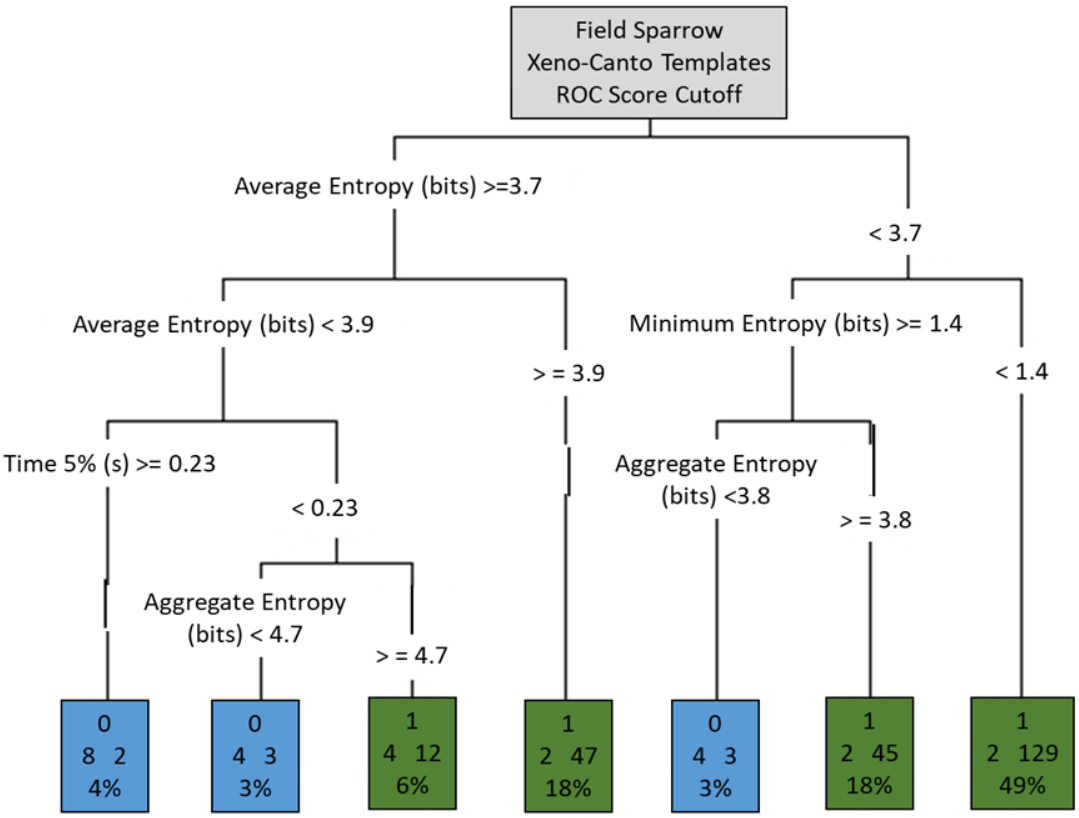

Figure 3. 3: Pruned decision tree for Field Sparrow with the greatest specificity rate (0.95 ± 0.02).

Terminal leaves show the final outcome (0 = False, 1 = True), total number of songs in the leaf that was originally a False + and then True +, and the percentage of total songs that ended up in that leaf.

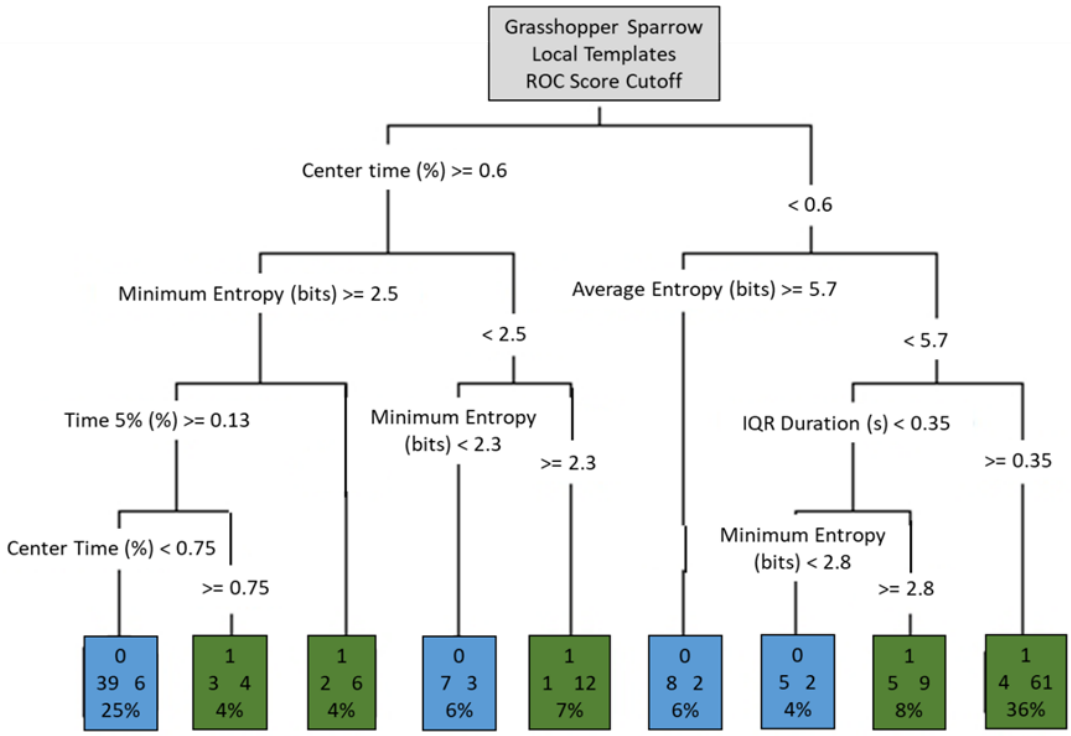

Figure 3. 4: Pruned decision tree for Grasshopper Sparrow with the greatest specificity rate (0.61 ± 0.05).

Terminal leaves show the final outcome (0 = False, 1 = True), total number of songs in the leaf that were originally a False + and then True +, and the percentage of total songs that ended up in that leaf.

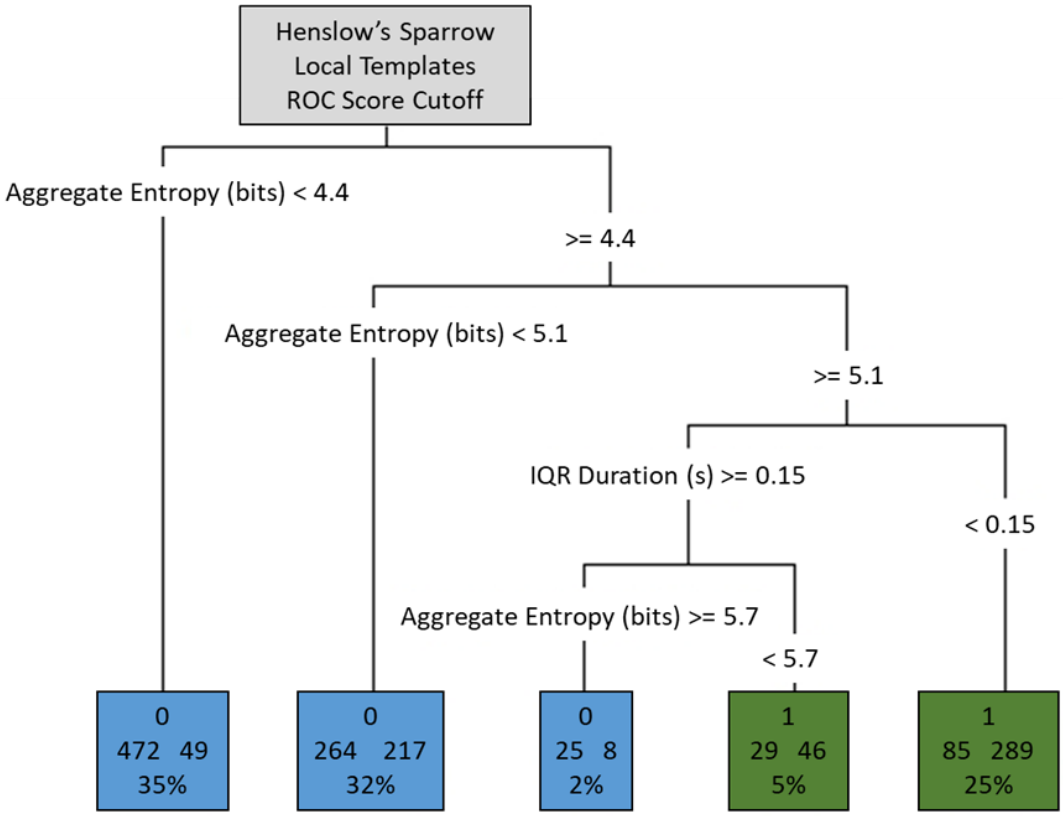

Figure 3. 5: Pruned decision tree for Henslow's Sparrow with the greatest specificity rate (0.86 ± 0.06).

Terminal leaves show the final outcome (0 = False, 1 = True), total number of songs in the leaf that were originally a False + and then True +, and the percentage of total songs that ended up in that leaf.

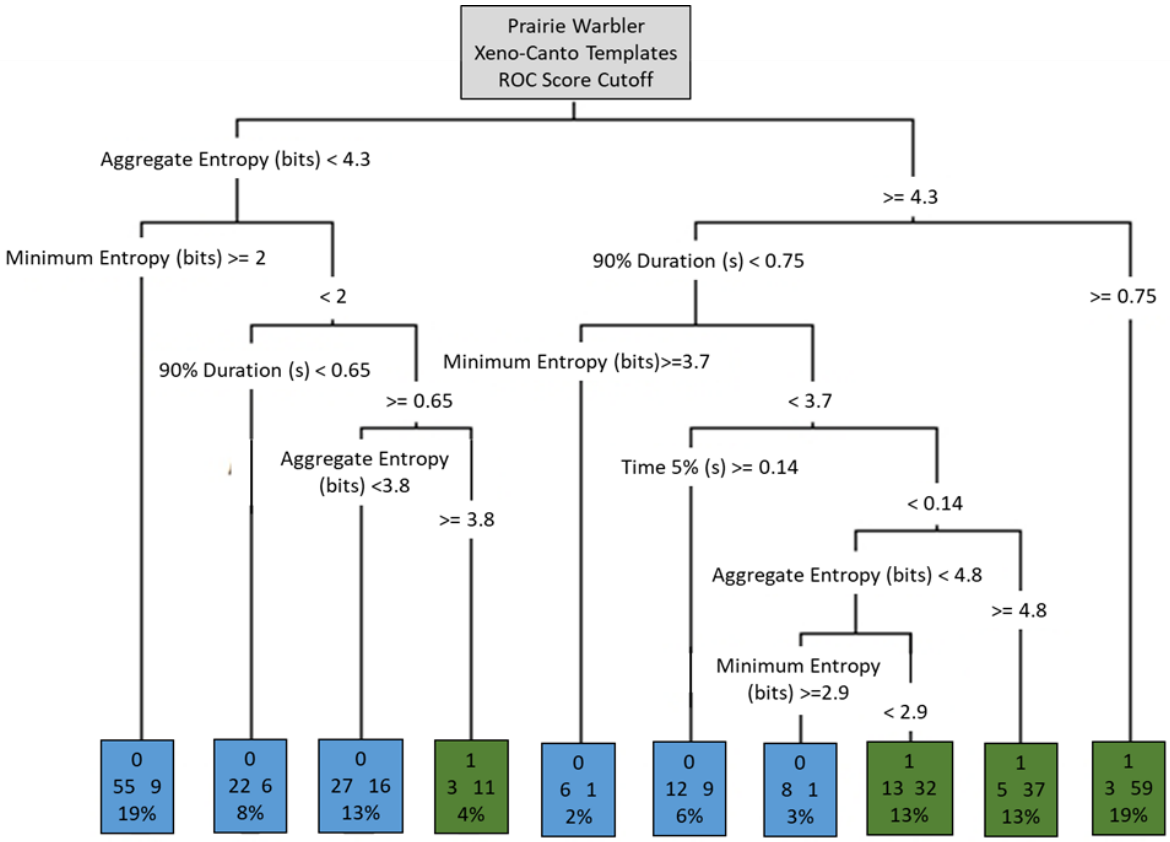

Figure 3. 6: Pruned decision tree for Prairie Warbler with the greatest specificity rate (0.79 ± 0.05).

Terminal leaves show the final outcome (0 = False, 1 = True), total number of songs in the leaf that was originally a False + and then True +, and the percentage of total songs that ended up in that leaf.

Table 3. 1: Selection table results from correlation analysis in program R using two sets of templates and five grassland - shrubland species.

Template Detection	Score Cutoff	Metric	Bachman's Sparrow	Field Sparrow	Grasshopper Sparrow	Henslow's Sparrow	Prairie Warbler
Local Templates	95% True + Capture	Total Selections	101.37 ± 7.50	105.41 ± 2.91	100.85 ± 6.62	408.83 ± 30.17	116.83 ± 8.18
		Prevalence	0.31 ± 0.04	0.55 ± 0.04	0.35 ± 0.04	0.27 ± 0.04	0.38 ± 0.04
		Sensitivity	0.77 ± 0.04	0.90 ± 0.02	0.93 ± 0.03	0.99 ± 0.00	0.97 ± 0.01
		False Discovery Rate	0.65 ± 0.05	0.42 ± 0.04	0.64 ± 0.04	0.73 ± 0.03	0.61 ± 0.04
	ROC Curve	Total Selections	38.67 ± 5.45	43.93 ± 7.97	37.00 ± 6.45	182.76 ± 45.51	55.17 ± 7.57
		Prevalence	0.46 ± 0.08	0.63 ± 0.07	0.57 ± 0.08	0.55 ± 0.09	0.63 ± 0.06
		Sensitivity	0.63 ± 0.07	0.78 ± 0.06	0.86 ± 0.06	0.99 ± 0.01	0.96 ± 0.03
		Global Sensitivity	0.45 ± 0.07	0.40 ± 0.06	0.54 ± 0.09	0.58 ± 0.08	0.73 ± 0.06
		False Discovery Rate	0.37 ± 0.09	0.25 ± 0.08	0.37 ± 0.08	0.45 ± 0.09	0.35 ± 0.07
Xeno-Canto Templates	95% True + Capture	Total Selections	115.37 ± 6.35	104.38 ± 3.34	116.28 ± 2.95	408.83 ± 51.72	109.14 ± 5.67
		Prevalence	0.47 ± 0.05	0.61 ± 0.05	0.27 ± 0.03	0.39 ± 0.06	0.38 ± 0.04
		Sensitivity	0.97 ± 0.01	0.95 ± 0.02	0.81 ± 0.03	0.96 ± 0.02	0.96 ± 0.02
		False Discovery Rate	0.53 ± 0.05	0.37 ± 0.05	0.71 ± 0.04	0.60 ± 0.06	0.62 ± 0.04
	ROC Curve	Total Selections	55.27 ± 9.12	47.73 ± 8.03	47.00 ± 1.93	185.55 ± 44.65	56.14 ± 9.33
		Prevalence	0.53 ± 0.07	0.82 ± 0.08	0.26 ± 0.04	0.53 ± 0.09	0.59 ± 0.06
		Sensitivity	0.95 ± 0.02	0.91 ± 0.05	0.61 ± 0.05	0.91 ± 0.05	0.94 ± 0.03
		Global Sensitivity	0.50 ± 0.07	0.56 ± 0.07	0.31 ± 0.09	0.52 ± 0.09	0.71 ± 0.06
		False Discovery Rate	0.46 ± 0.08	0.13 ± 0.08	0.69 ± 0.05	0.41 ± 0.11	0.39 ± 0.07

Sensitivity = ability to correctly detect true positives out of the total true positives (true positives / true positives + false negatives)

Specificity = ability to correctly reject true negatives out of total true negatives (true negatives / (true negatives + false positives)

PPV = positive predictive value (true positives / true positives + false positives)

FPR = false positive rate (true negatives / false negatives + true negatives)

Table 3. 2: Final output results for Bachman's Sparrow automatic detecting and song classification using two different template sources and three semi-supervised classification methods.

Analysis Method	Metric	Local Templates		Xeno-Canto Templates	
		95% True + Score Cutoff	ROC Score cutoff	95% True + Score Cutoff	ROC Score cutoff
Decision Trees	Sensitivity	0.41 ± 0.07	0.69 ± 0.08	0.62 ± 0.06	0.82 ± 0.04
	Specificity	0.84 ± 0.04	0.38 ± 0.12	0.69 ± 0.09	0.48 ± 0.07
	PPV	0.55 ± 0.10	0.64 ± 0.10	0.66 ± 0.08	0.61 ± 0.09
	FPR	0.45 ± 0.10	036 ± 0.10	0.34 ± 0.08	0.39 ± 0.09
Quadratic Discriminant Analysis	Sensitivity	0.55 ± 0.08	0.73 ± 0.08	0.61 ± 0.07	0.77 ± 0.06
	Specificity	0.82 ± 0.09	0.53 ± 0.11	0.74 ± 0.06	0.66 ± 0.10
	PPV	0.65 ± 0.10	0.72 ± 0.09	0.66 ± 0.07	0.69 ± 0.10
	FPR	0.35 ± 0.10	0.28 ± 0.10	0.34 ± 0.07	0.31 ± 0.10
Neural Network	Sensitivity	0.51 ± 0.10	0.76 ± 0.09	0.61 ± 0.10	0.77 ± 0.07
	Specificity	0.78 ± 0.10	0.42 ± 0.12	0.71 ± 0.09	0.62 ± 0.12
	PPV	0.62 ± 0.12	0.70 ± 0.09	0.64 ± 0.08	0.66 ± 0.10
	FPR	0.38 ± 0.12	0.30 ± 0.09	0.36 ± 0.08	0.34 ± 0.10

Sensitivity = ability to correctly detect true positives out of the total true positives (true positives / true positives + false negatives)

Specificity = ability to correctly reject true negatives out of total true negatives (true negatives / (true negatives + false positives)

PPV = positive predictive value (true positives / true positives + false positives)

FPR = false positive rate (true negatives / false negatives + true negatives)

Table 3. 3: Final output results for Field Sparrow automatic detecting and song classification using two different template sources and three semi-supervised classification methods.

Analysis Method	Metric	Local Templates		Xeno-Canto Templates	
		95% True + Score Cutoff	ROC Score cutoff	95% True + Score Cutoff	ROC Score cutoff
Decision Trees	Sensitivity	0.80 ± 0.04	0.81 ± 0.07	0.78 ± 0.07	0.95 ± 0.02
	Specificity	0.42 ± 0.04	0.43 ± 0.13	0.51 ± 0.06	0.17 ± 0.09
	PPV	0.64 ± 0.05	0.77 ± 0.09	0.74 ± 0.04	0.92 ± 0.06
	FPR	0.36 ± 0.05	0.23 ± 0.09	0.26 ± 0.04	0.08 ± 0.06
Quadradic Discriminant Analysis	Sensitivity	0.59 ± 0.07	0.72 ± 0.11	0.65 ± 0.06	0.50 ± 0.09
	Specificity	0.73 ± 0.07	0.54 ± 0.16	0.64 ± 0.07	0.87 ± 0.12
	PPV	0.76 ± 0.05	0.84 ± 0.08	0.77 ± 0.04	0.98 ± 0.01
	FPR	0.24 ± 0.05	0.16 ± 0.08	0.23 ± 0.04	0.02 ± 0.01
Neural Network	Sensitivity	0.70 ± 0.05	0.79 ± 0.08	0.93 ± 0.09	1.00 ± 0.02
	Specificity	0.59 ± 0.08	0.53 ± 0.15	0.18 ± 0.09	0.11 ± 0.12
	PPV	0.70 ± 0.06	0.70 ± 0.07	0.64 ± 0.07	0.92 ± 0.06
	FPR	0.30 ± 0.06	0.16 ± 0.07	0.36 ± 0.07	0.08 ± 0.06

Sensitivity = ability to correctly detect true positives out of the total true positives (true positives / true positives + false negatives)

Specificity = ability to correctly reject true negatives out of total true negatives (true negatives / (true negatives + false positives)

PPV = positive predictive value (true positives / true positives + false positives)

FPR = false positive rate (true negatives / false negatives + true negatives)

Table 3. 4: Final output results for Grasshopper Sparrow automatic detecting and song classification using two different template sources and three semi-supervised classification methods.

Analysis Method	Metric	Local Templates		Xeno-Canto Templates	
		95% True + Score Cutoff	ROC Score cutoff	95% True + Score Cutoff	ROC Score cutoff
Decision Trees	Sensitivity	0.24 ± 0.06	0.61 ± 0.05	0.23 ± 0.08	0.29
	Specificity	0.84 ± 0.03	0.31 ± 0.08	0.79 ± 0.10	0.81
	PPV	0.46 ± 0.07	0.62 ± 0.09	0.40 ± 0.06	0.31
	FPR	0.54 ± 0.07	0.38 ± 0.09	0.60 ± 0.06	0.69
Quadratic Discriminant Analysis	Sensitivity	0.20 ± 0.06	0.51 ± 0.10	0.20 ± 0.06	0.05
	Specificity	0.85 ± 0.08	0.68 ± 0.12	0.87 ± 0.05	0.91
	PPV	0.56 ± 0.14	0.77 ± 0.08	0.44 ± 0.09	0.13
	FPR	0.44 ± 0.14	0.23 ± 0.08	0.56 ± 0.09	0.87
Neural Network	Sensitivity	0.22 ± 0.06	0.71 ± 0.09	0.24 ± 0.07	0.05
	Specificity	0.91 ± 0.03	0.45 ± 0.14	0.84 ± 0.07	0.94
	PPV	0.57 ± 0.10	0.71 ± 0.08	0.48 ± 0.09	0.20
	FPR	0.43 ± 0.10	0.29 ± 0.08	0.52 ± 0.09	0.80

Sensitivity = ability to correctly detect true positives out of the total true positives (true positives / true positives + false negatives)

Specificity = ability to correctly reject true negatives out of total true negatives (true negatives / (true negatives + false positives)

PPV = positive predictive value (true positives / true positives + false positives)

FPR = false positive rate (true negatives / false negatives + true negatives)

Table 3. 5: Final output results for Henslow's Sparrow automatic detecting and song classification using two different template sources and three semi-supervised classification methods.

Analysis Method	Metric	Local Templates		Xeno-Canto Templates	
		95% True + Score Cutoff	ROC Score cutoff	95% True + Score Cutoff	ROC Score cutoff
Decision Trees	Sensitivity	0.17 ± 0.05	0.86 ± 0.06	0.28 ± 0.06	0.46 ± 0.10
	Specificity	0.98 ± 0.01	0.47 ± 0.10	0.92 ± 0.02	0.87 ± 0.04
	PPV	0.80 ± 0.07	0.75 ± 0.08	0.70 ± 0.08	0.81 ± 0.08
	FPR	0.20 ± 0.07	0.25 ± 0.08	0.30 ± 0.08	0.19 ± 0.08
Quadratic Discriminant Analysis	Sensitivity	0.24 ± 0.07	0.77 ± 0.10	0.29 ± 0.07	0.36 ± 0.08
	Specificity	0.96 ± 0.01	0.60 ± 0.11	0.93 ± 0.01	0.92 ± 0.03
	PPV	0.76 ± 0.08	0.79 ± 0.08	0.71 ± 0.08	0.81 ± 0.09
	FPR	0.24 ± 0.08	0.21 ± 0.08	0.29 ± 0.08	0.19 ± 0.09
Neural Network	Sensitivity	0.27 ± 0.05	0.89 ± 0.07	0.38 ± 0.07	0.49 ± 0.10
	Specificity	0.94 ± 0.02	0.31 ± 0.13	0.83 ± 0.06	0.86 ± 0.06
	PPV	0.68 ± 0.09	0.73 ± 0.09	0.69 ± 0.09	0.81 ± 0.08
	FPR	0.32 ± 0.09	0.27 ± 0.09	0.31 ± 0.09	0.19 ± 0.08

Sensitivity = ability to correctly detect true positives out of the total true positives (true positives / true positives + false negatives)

Specificity = ability to correctly reject true negatives out of total true negatives (true negatives / (true negatives + false positives)

PPV = positive predictive value (true positives / true positives + false positives)

FPR = false positive rate (true negatives / false negatives + true negatives)

Table 3. 6: Final output results for Prairie Warbler automatic detecting and song classification using two different template sources and three semi-supervised classification methods.

Analysis Method	Metric	Local Templates		Xeno-Canto Templates	
		95% True + Score Cutoff	ROC Score cutoff	95% True + Score Cutoff	ROC Score cutoff
Decision Trees	Sensitivity	0.56 ± 0.06	0.73 ± 0.07	0.51 ± 0.07	0.79 ± 0.05
	Specificity	0.81 ± 0.04	0.59 ± 0.08	0.82 ± 0.04	0.61 ± 0.09
	PPV	0.66 ± 0.07	0.78 ± 0.06	0.65 ± 0.06	0.76 ± 0.07
	FPR	0.34 ± 0.07	0.22 ± 0.06	0.35 ± 0.06	0.24 ± 0.07
Quadratic Discriminant Analysis	Sensitivity	0.57 ± 0.07	0.83 ± 0.05	0.59 ± 0.07	0.86 ± 0.04
	Specificity	0.75 ± 0.06	0.52 ± 0.10	0.73 ± 0.08	0.49 ± 0.13
	PPV	0.62 ± 0.07	0.80 ± 0.06	0.62 ± 0.07	0.73 ± 0.08
	FPR	0.38 ± 0.07	0.20 ± 0.06	0.38 ± 0.07	0.27 ± 0.08
Neural Network	Sensitivity	0.49 ± 0.08	0.88 ± 0.06	0.42 ± 0.06	0.80 ± 0.06
	Specificity	0.80 ± 0.06	0.49 ± 0.11	0.77 ± 0.07	0.46 ± 0.12
	PPV	0.63 ± 0.08	0.77 ± 0.06	0.61 ± 0.08	0.70 ± 0.08
	FPR	0.37 ± 0.08	0.23 ± 0.07	0.39 ± 0.08	0.30 ± 0.08

Sensitivity = ability to correctly detect true positives out of the total true positives (true positives / true positives + false negatives)

Specificity = ability to correctly reject true negatives out of total true negatives (true negatives / (true negatives + false positives)

PPV = positive predictive value (true positives / true positives + false positives)

FPR = false positive rate (true negatives / false negatives + true negatives)

Table 3. 7: Average song power to background power ratio (dB) for all species and inflection point for predicting correct positive classification.

Species	Average Power (dB)	Ratio at Inflection Point (dB)
Bachman's Sparrow	1.02 ± 0.003	0.92 ± 0.15
Field Sparrow	0.99 ± 0.003	0.87 ± 0.05
Grasshopper Sparrow	0.98 ± 0.004	1.30 ± 0.04
Henslow's Sparrow	0.96 ± 0.002	1.39 ± 0.01
Prairie Warbler	0.97 ± 0.003	1.03 ± 0.03

Conclusion

This study introduced a new method of recording bird songs from a floating platform and walked through analysis of automatically detecting bird songs from passive recordings, like those recorded by the AAARS or other ARUs. We demonstrated the usefulness of the AAARS in situations where time, accessibility, or manpower isn't enough to effectively monitor a species of interest. The power of the AAARS comes from its ability to create high-quality recordings as well as its flexibility; with further modifications it can be fit to monitor ultrasonic signals, fly nearly unlimited distances, or monitor any number of species. The AAARS excels in inaccessible areas, especially when recording frequently vocalizing and territorial species. For less vocal species the AAARS would need multiple flights over the same area, higher flight (if possible), and days with ideal weather.

Use of machines like the AAARS and other ARUs can generate hours upon hours of real-world recordings. A typical manager or researcher may not have the time or resources to complete lengthy or computationally complicated analysis. We aimed to utilize easily available software to find detailed measurements that could differentiate between species in similar frequency bands. Overlapping signals can be an issue when trying to classify songs in real-world recordings, like bird songs during the dawn chorus. Using manually extracted song clips we were able to show that a relatively simplistic method of classifying songs based on a variety of metrics is more than adequate. However, the results are less straightforward when combining this classification with song clips extracted automatically by a program and not by hand.

More species-specific inconsistencies emerge when combining the extraction and classification together in one analysis.

All of our methods suggested that a species-specific approach is necessary every step of the way, from data recording to analysis. Unlike species with less variable vocalizations, bird songs, especially from passerines, can be highly plastic in nature and are hard to accommodate in acoustic analysis without a pilot study. We used species-specific approaches for the footprint analysis, cue count data, template selection, and final analysis, discussing the differences between species along the way. The issues related to species and location-specific song and song attenuation can't be ignored If users want to generate accurate density estimations.

Analysis pathway

We worked through the recoding and analyzing bird songs from continuous recordings through this entire manuscript. Below is the synopsis of steps necessary to complete a similar analysis from scratch in a realistic setting (Figure 4.1). We recommend the following workflow as a necessary pilot study approach for any analysis, with the understanding that some steps would be more or less complex depending on the species.

The first step, and imperative for density estimation, is to determine the recording area and exposure available to the passive recorder being used. For the AAARS we generated his empirically using realistic representations of target species and generating exposure footprints. This could be done for stationary recoding units by

pairing human observers with ARUs to estimate both detection distance by species and also cue count rates (see Prevost, 2016).

Next, generating a training data set is necessary to test out the entire pathway and determine accuracy for each species. We relied on a template matching analysis to extract target songs, and as such the results are highly susceptible to local issues with weather and overlapping conspecific signals. It may be worth noting that we did not employ any filtering in this experiment aside from the and pass filter inherent in template matching. Median filtering is a common method that aims to separate the desired signal from the background noise (Briggs et al. 2012). However, the target species may not be the loudest in the recording. Depending on the environment, any given birdsong could be a desired signal, background noise, or both. Utilizing template matching extraction techniques attempts to find correlation within continuous recordings regardless of the overlapping or supplemental background noise.

Knight (2017) argues that there are very few established rules for reporting accuracy of automatic detection of birdsongs from continuous recordings. They found that only eleven published papers reported the correlation score cutoff used minimize false positives selected by template correlation extraction. However, we showed that our two species with less tonal songs (i.e. Grasshopper and Henslow's Sparrows) need to be louder than the surrounding frequency-specific background noise to increase the probability of being classified correctly at the later stages of the proposed pathway. It is important, then, to manually create a training dataset for a subset of recordings and investigate this relationship before extraction and classification. If necessary, the

available recording area and ROC score cutoffs can be adjusted to accommodate these types of species.

Once a training dataset is generated, it is relatively easy to extract metrics from all clips and generate data tables. After correlation analysis the collinearity is minimized, and only unique metrics are left to use in classification after extraction. The success of our original extraction and classification pathway relied heavily on the template selection and score cutoff options than classification methods. We suggest using Xeno-Canto templates for those tonal species with the highest amount of variation both between and within individuals (Borror 1961), and possibly also increasing the number of templates used for extraction to better capture these more verbose repertoires.

We were not able to suggest any one semi-supervised classification technique to fit all species. There was minimal variation between the results. However, the use of more complicated, deep-learning techniques may produce better results and is worth investigating further in these real-world scenarios.

Results should be examined at the end of this proposed pilot study pathway to determine the efficiency of the outcomes. Options should be tweaked iteratively to maximize outcomes. Unfortunately, not all species may be extracted and classified adequately with these methods. More hands-on methods may be warranted to generate similar results for species with less obviously tonal songs. Raven Pro 1.5 software, for example, allows the user to visually inspect each flagged clip in a recording quickly in a grid format. Songs for less tonal species could be reviewed in this manner in leu of computer-based classification methods to increase accuracy.

Additionally, we would suggest investigating the use of citizen science or species-richness models to generate better results for more complicated species or soundscapes if this pathway fails.

With the inherent variability in bird songs in a natural setting, there is not going to be one catch-all solution that works for all song types. We believe that we have demonstrated a solid, relatively simplistic pathway that is available and useful to managers and researchers to record and analyze passive recordings now, without relying on complicated computer methods.

Appendix

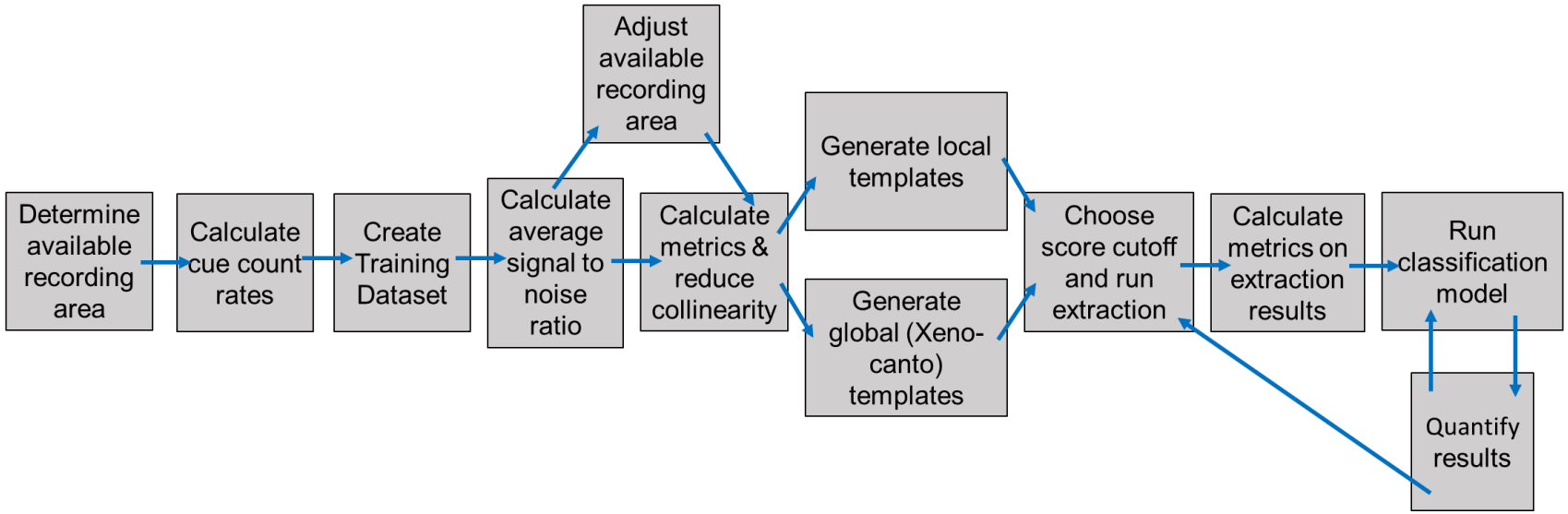

Figure 4. 1: Simplified version of the full proposed analysis pathway

Literature Cited

Acevedo, M. A., C. J. Corrada-Bravo, H. Corrada-Bravo, L. J. Villanueva-Rivera, and T. M. Aide. 2009. Automated classification of bird and amphibian calls using machine learning: A comparison of methods. Ecological Informatics 4:206-214.

Althoff, D. P., P. S. Gipson, J. S. Pontius, and P. B. Woodford. 2006. Plant community and bare ground trends on Fort Riley, Kansas: implications for monitoring of a highly disturbed landscape. Transactions of the Kansas Academy of Science 109:101-119.

Anderson, S. E., A. S. Dave, and D. Margoliash. 1996. Template-based automatic recognition of birdsong syllables from continuous recordings. The Journal of the Acoustical Society of America 100:1209-1219.

Ashiya, T., and M. Nakagawa. 1993. A proposal of a recognition system for the specices of birds receiving birdcalls--an application of recognition systems for environmental sound. IEICE transactions on fundamentals of electronics, communications and computer sciences 76:1858-1860.

August, T., M. Harvey, P. Lightfoot, D. Kilbey, T. Papadopoulos, and P. Jepson. 2015. Emerging technologies for biological recording. Biological Journal of the Linnean Society 115:731-749.

Aylor, D. 1972. Noise reduction by vegetation and ground. The Journal of the Acoustical Society of America 51:197-205.

Baker, M. C., T. K. Bjerke, H. U. Lampe, and Y. O. Espmark. 1987. Sexual response of female yellowhammers to differences in regional song dialects and repertoire sizes. Animal Behaviour 35:395-401.

Baker, M. C., and M. A. Cunningham. 1985. The biology of bird-song dialects. Behavioral and Brain Sciences 8:85-100.

Bardeli, R., D. Wolff, F. Kurth, M. Koch, K.-H. Tauchert, and K.-H. Frommolt. 2010. Detecting bird sounds in a complex acoustic environment and application to bioacoustic monitoring. Pattern Recognition Letters 31:1524-1534.

Barlow, J., and B. L. Taylor. 2005. Estimates of sperm whale abundance in the northeastern temperate Pacific from a combined acoustic and visual survey. Marine Mammal Science 21:429-445.

Bent, A. C., and O. L. Austin. 1968. Life histories of North American cardinals, grosbeaks, buntings, towhees, finches, sparrows, and allies: order Passeriformes: family Fringillidae. Smithsonian Institution Press, Washington, D.C.

Bibby, C. J., N. D. Burgess, D. A. Hill, and S. H. Mustoe. 2000. Bird census techniques, 2nd Edition. Academic Press, London, UK.

Blumstein, D. T., D. J. Mennill, P. Clemins, L. Girod, K. Yao, G. Patricelli, J. L. Deppe, A. H. Krakauer, C. Clark, and K. A. Cortopassi. 2011. Acoustic monitoring in terrestrial environments using microphone arrays: applications, technological considerations and prospectus. Journal of Applied Ecology 48:758-767.

Borker, A. L., M. W. McKown, J. T. Ackerman, C. Eagles-Smith, B. R. Tershy, and D. A. Croll. 2014. Vocal activity as a low cost and scalable index of seabird colony size. Conservation Biology 28: 1100-1108.

Borror, D. J. 1961. Intraspecific variation in passerine bird songs. The Wilson Bulletin:57-78.

_____. 1971. Songs of Aimophila sparrows occurring in the United States. The Wilson Bulletin:132-151.

Borror, D. J., and W. W. Gunn. 1965. Variation in White-throated Sparrow songs. The Auk:26-47.

Brandes, S. T. 2008. Automated sound recording and analysis techniques for bird surveys and conservation. Bird Conservation International 18:S163-S173.

Briggs, F., B. Lakshminarayanan, L. Neal, X. Z. Fern, R. Raich, S. J. Hadley, A. S. Hadley, and M. G. Betts. 2012. Acoustic classification of multiple simultaneous bird species: A multi-instance multi-label approach. The Journal of the Acoustical Society of America 131:4640-4650.

Brumm, H., S. A. Zollinger, P. T. Niemelä, and P. Sprau. 2017. Measurement artefacts lead to false positives in the study of birdsong in noise. Methods in Ecology and Evolution 8:1617-1625.

Buck, J. R., and P. L. Tyack. 1993. A quantitative measure of similarity for *Tursiops truncatus* signature whistles. The Journal of the Acoustical Society of America 94:2497-2506.

Byers, B. E., B. A. Kramer, M. E. Akresh, and D. I. King. 2013. Interspecific song imitation by a Prairie Warbler. Journal of Field Ornithology 84:181-186.

Catchpole, C. K., and P. J. Slater. 2003. Bird song: biological themes and variations. Cambridge University Press, Cambridge, England.

Chabot, D., and D. M. Bird. 2015. Wildlife research and management methods in the 21st century: Where do unmanned aircraft fit in? Journal of Unmanned Vehicle Systems 3:137-155.

Charif, R. A., and M. Pitzrick. 2008. Automated detection of Cerulean Warbler songs using XBAT data template detector software. Preliminary Report. Bioacoustics Research Program Cornell Laboratory of Ornithology Technical Report:08-02.

Chen, Z., and R. C. Maher. 2006. Semi-automatic classification of bird vocalizations using spectral peak tracks. The Journal of the Acoustical Society of America 120:2974-2984.

Cully, J. F., and H. L. Michaels. 2000. Henslow's Sparrow habitat associations on Kansas tallgrass prairie. The Wilson Bulletin 112:115-123.

Darras, K. F. A., P. Batary, B. Furnas, I. Fitriawan, Y. Mulyani, and T. Tscharntke. 2017. Autonomous bird sound recording outperforms direct human observation: Synthesis and new evidence. bioRxiv.

Digby, A., M. Towsey, B. D. Bell, and P. D. Teal. 2013. A practical comparison of manual and autonomous methods for acoustic monitoring. Methods in Ecology and Evolution 4:675-683.

Eberly, C., and J. Keating. 2006. Birds and bombs: How bird conservation planning and the military mission work together. Federal Facilities Environmental Journal 17:51-65.

Evans, W. R., and D. K. Mellinger. 1999. Monitoring grassland birds in nocturnal migration. Studies in Avian Biology 19:219-229.

Farnsworth, A. 2005. Flight calls and their value for future ornithological studies and conservation research. The Auk 122:733-746.

Forrest, T. G. 1994. From sender to receiver: Propagation and environmental effects on acoustic signals. American Zoologist 34:644-654.

Fristrup, K. M., and C. W. Clark. 2009. Acoustic monitoring of threatened and endangered species in inaccessible areas. Cornell University, Ithaca, NY.

Frommolt, K.-H., and K.-H. Tauchert. 2014. Applying bioacoustic methods for long-term monitoring of a nocturnal wetland bird. Ecological Informatics 21:4-12.

Gasc, A., S. Pavoine, L. Lellouch, P. Grandcolas, and J. Sueur. 2015. Acoustic indices for biodiversity assessments: analyses of bias based on simulated bird assemblages and recommendations for field surveys. Biological Conservation 191:306-312.

Goebel, M. E., W. L. Perryman, J. T. Hinke, D. J. Krause, N. A. Hann, S. Gardner, and D. J. LeRoi. 2015. A small unmanned aerial system for estimating abundance and size of Antarctic predators. Polar Biology 38:619-630.

Goh, M. 2011. Developing an Automated acoustic monitoring system to estimate abundance of Cory's Shearwaters in the Azores. Thesis, Imperial College, London, Englan

Goyette, J. L., R. W. Howe, A. T. Wolf, and W. D. Robinson. 2011. Detecting tropical nocturnal birds using automated audio recordings. Journal of Field Ornithology 82:279-287.

Hafner, S. D., and J. Katz. 2017. Tools for automated acoustic monitoring within the R package monitoR. Bioacoustics, 25:197-210.

Harlow, Z., T. Collier, V. Burkholder, and C. E. Taylor. Acoustic 3D localization of a tropical songbird. Pages 220-224 *in* 2013 IEEE China Summet and Internaional Conference on Signal and Information Processing, China.

Houlihan, P. W. 2000. The singing behavior of Prairie Warblers (*Dendroica discolor*). University of Massachusetts Amherst, Ann Arbor, MI.

Hourdequin, M., and D. G. Havlick. 2011. Ecological restoration in context: Ethics and the naturalization of former military lands. Ethics, Policy and Environment 14:69-89.

Hunter, M. L., and J. R. Krebs. 1979. Geographical variation in the song of the great tit (parus major) in relation to ecological factors. Journal of Animal Ecology 48:759-785.

Hutto, R. L., and R. J. Stutzman. 2009. Humans versus autonomous recording units: a comparison of point-count results. Journal of Field Ornithology 80:387-398.

Isler, M. L., P. R. Isler, and B. M. Whitney. 1998. Use of vocalizations to establish species limits in antbirds (Passeriformes: Thamnophilidae). The Auk:577-590.

Ivošević, B., Y.-G. Han, Y. Cho, and O. Kwon. 2015. The use of conservation drones in ecology and wildlife research. Journal of Ecology and Environment 38:113-118.

Jančovič, P., and M. Köküer. 2011. Automatic detection and recognition of tonal bird sounds in noisy environments. EURASIP Journal on Advances in Signal Processing 2011:1-10.

Kaewtip, K., L. N. Tan, A. Alwan, and C. E. Taylor. A robust automatic bird phrase classifier using dynamic time-warping with prominent region identification. *in* 2013 IEEE China Summet and Internaional Conference on Signal and Information Processing, China.

Kasten, E. P., P. K. McKinley, and S. H. Gage. 2010. Ensemble extraction for classification and detection of bird species. Ecological Informatics 5:153-166.

Keen, S., J. C. Ross, E. T. Griffiths, M. Lanzone, and A. Farnsworth. 2014. A comparison of similarity-based approaches in the classification of flight calls of four species of North American wood-warblers (Parulidae). Ecological Informatics 21:25-33.

Kershenbaum, A., and E. C. Garland. 2015. Quantifying similarity in animal vocal sequences: which metric performs best? Methods in Ecology and Evolution 6:1452-1461.

Klapel, J. 2014. Acoustic Measurements with a Quadcopter: Embedded System Implementations for Recording Audio from Above.Thesis, Institute for Electronics and Telecommunication Engineers, India.

Knight, E. C., K. C. Hannah, G. J. Foley, C. D. Scott, R. M. Brigham, and E. Bayne. 2017. Recommendations for acoustic recognizer performance assessment with application to five common automated signal recognition programs. Avian Conservation and Ecology 12:4.

Kogan, J. A., and D. Margoliash. 1998. Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden Markov models: A comparative study. The Journal of the Acoustical Society of America 103:2185-2196.

Koops, H. V., J. Van Balen, and F. Wiering. Automatic segmentation and deep learning of bird sounds. Springer, 2015.

Kroodsma, D. E. 1981. Geographical variation and functions of song types in warblers (parulidae). The Auk 98:743-751.

Kroodsma, D. E., E. H. Miller, H. Ouellet, and E. H. Miller. 1982. Acoustic communication in birds. Volume 1. Academic Press, New York, NY.

Lack, D. 1964. Population studies of birds. Princeton University Press, Princeton, NJ.

Lambert, K. T., and P. G. McDonald. 2014. A low-cost, yet simple and highly repeatable system for acoustically surveying cryptic species. Austral Ecology 39:779-785.

Lasseck, M. Bird song classification in field recordings: winning solution for NIPS4B 2013 competition. 2013 MLSP Challenge.

Leger, D. W., D. J. Mountjoy, and D. A. Nelson. 2003. Geographic variation in song of the Bright-rumped Attila (Tyrannidae: *Attila spadiceus*): implications for species status. The Auk 120:69-74.

MacDougall-Shackleton, E. A., and S. A. MacDougall-Shackleton. 2001. Cultural and genetic evolution in Mountain White-crowned Sparrows: song dialects are assicated with population structure. Evolution 55:2568-2575.

Marler, P., and M. Tamura. 1962. Song "Dialects" in Three Populations of White-Crowned Sparrows. The Condor 64:368-377.

Marler, P., and M. Tamura. 1964. Culturally transmitted patterns of vocal behavior in sparrows. Science 146:1483-1486.

Marques, T. A., L. Thomas, S. W. Martin, D. K. Mellinger, J. A. Ward, D. J. Moretti, D. Harris, and P. L. Tyack. 2013. Estimating animal population density using passive acoustics. Biological Reviews 88:287-309.

Marten, K., and P. Marler. 1977. Sound transmission and its significance for animal vocalization. Behavioral Ecology and Sociobiology 2:271-290.

Martin, S. W., T. A. Marques, L. Thomas, R. P. Morrissey, S. Jarvis, N. DiMarzio, D. Moretti, and D. K. Mellinger. 2013. Estimating minke whale (Balaenoptera acutorostrata) boing sound density using passive acoustic sensors. Marine Mammal Science 29:142-158.

McIlraith, A. L., and H. C. Card. 1997. Birdsong recognition using backpropagation and multivariate statistics. IEEE Transactions on Signal Processing 45:2740-2748.

Meretsky, V. J., R. L. Fischman, J. R. Karr, D. M. Ashe, M. J. Scott, R. F. Noss, and R. L. Schroeder. 2006. New directions in conservation for the National Wildlife Refuge System. BioScience 56:135-143.

Nelson, D. A. 1992. Song overproduction and selective attrition lead to song sharing in the field sparrow (*Spizella pusilla*). Behavioral Ecology and Sociobiology 30:415-424.

Nicholls, J., J. Austin, C. Moritz, and A. Goldizen. 2006. Genetic population structure and call variation in a passerine bird, the satin bowerbird, Ptilonorhynchus violaceus. Evolution 60:1279-1290.

Nickerson, C. M., L. L. Bloomfield, M. R. Dawson, and C. B. Sturdy. 2006. Artificial neural network discrimination of black-capped chickadee (*Poecile atricapillus*) call notes. The Journal of the Acoustical Society of America 120:1111-1117.

Nolan, V. 1978. The ecology and behavior of the Prairie Warbler, Dendroica discolor. Ornithological Monographs 551-572.

Nolan, V. J., E. D. Ketterson, and C. A. Buerkle. 1999. Prairie Warbler (*Setophaga discolor*). The Birds of North America Online (A. Poole, Ed). <http://www.birdna.org>. Accessed 1 Jan 2016.

Nott, M. P., D. F. Desante, and N. Michel. 2003. Management strategies for reversing declines in landbirds of conservation concern on military installations: A landscape-scale analysis of MAPS data. Legacy Resources management Program, U.S. Department of Defense.

Nowacek, D. P., F. Christiansen, L. Bejder, J. A. Goldbogen, and A. S. Friedlaender. 2016. Studying cetacean behaviour: new technological approaches and conservation applications. Animal Behaviour 120:235-244.

Payne, R. B. 1981. Song learning and social interaction in indigo buntings. Animal Behaviour 29:688-697.

Podos, J., and P. S. Warren. 2007. The evolution of geographic variation in birdsong. Advances in the Study of Behavior 37:403-458.

Potamitis, I., S. Ntalampiras, O. Jahn, and K. Riede. 2014. Automatic bird sound detection in long real-field recordings: Applications and tools. Applied Acoustics 80:1-9.

Prevost, S. C. 2016. Estimating avian populations with passive acoustic technology and song behavior. Thesis, University of Tennessee, Knoxville, TN.

Raghuram, M., N. R. Chavan, R. Belur, and S. G. Koolagudi. 2016. Bird classification based on their sound patterns. International Journal of Speech Technology:1-14.

Ross, J. C., and P. E. Allen. 2014. Random forest for improved analysis efficiency in passive acoustic monitoring. Ecological Informatics 21:34-39.

Roth, K. L., and K. Islam. 2008. Habitat selection and reproductive success of Cerulean Warblers in Indiana. Wilson Journal of Ornithology 120:105-110.

Roy, N., Y. Simard, and C. Gervaise. 2010. 3D tracking of foraging belugas from their clicks: Experiment from a coastal hydrophone array. Applied Acoustics 71:1050-1056.

Ruse, M. G., D. Hasselquist, B. Hansson, M. Tarka, and M. Sandsten. 2016. Automated analysis of song structure in complex birdsongs. Animal Behaviour 112:39-51.

Searcy, W. A. 1992. Song repertoire and mate choice in birds. American Zoologist 32:71-80.

Shiovitz, K. A., and R. E. Lemon. 1980. Species identification of song by indigo buntings as determined by responses to computer generated sounds. Behaviour 74:167-198.

Shiovitz, K. A., and W. L. Thompson. 1970. Geographic variation in song composition of the indigo bunting, Passerina cyanea. Animal Behaviour 18, Part 1:151-158.

Smith, R. L. 1959a. The songs of the Grasshopper Sparrow. The Wilson Bulletin:141-152.

_____. 1959b. The Songs of the Grasshopper Sparrow. The Wilson Bulletin 71:141-152.

Sorrie, B. A., J. B. Gray, and P. J. Crutchfield. 2006. The Vascular Flora of the Longleaf Pine Ecosystem of Fort Bragg and Weymouth Woods, North Carolina. Castanea 71:129-161.

Stepanian, P. M., K. G. Horton, D. C. Hille, C. E. Wainwright, P. B. Chilson, and J. F. Kelly. 2016. Extending bioacoustic monitoring of birds aloft through flight call localization with a three-dimensional microphone array. Ecology and Evolution 6:7039-7046.

Therneau, T., B. Atkinson, and B. D. Ripley. 2017. rPart: Recursive partitioning for classfication, regression and survival tres. R package version 4.1-11.

Thome, D. M., and T. M. Thome. 2000. Radio-controlled model airplanes: inexpensive tools for low-level aerial photography. Wildlife Society Bulletin:343-346.

Thorpe, W., and B. Lade. 1961. The songs of some families of the Passeriforems. The songs of the Buntings (Emberizidae). Ibis 103:246-259.

Toews, D. P. L., and D. E. Irwin. 2008. Cryptic speciation in a Holarctic passerine revealed by genetic and bioacoustic analyses. Molecular Ecology 17:2691-2705.

Trifa, V. M., A. N. Kirschel, C. E. Taylor, and E. E. Vallejo. 2008. Automated species recognition of antbirds in a Mexican rainforest using hidden Markov models. The Journal of the Acoustical Society of America 123:2424-2431.

Venables, W. N., and B. D. Ripley. 2013. Modern applied statistics with S-PLUS. Springer Science & Business Media, Berlin/Heidelberg, Germany.

von Benda-Beckmann, A. M., F. P. A. Lam, D. J. Moretti, K. Fulkerson, M. A. Ainslie, S. P. van Ijsselmuide, J. Theriault, and S. P. Beerens. 2010. Detection of Blainville's beaked whales with towed arrays. Applied Acoustics 71:1027-1035.

Weisman, R., and L. Ratcliffe. 1989. Absolute and relative pitch processing in black-capped chickadees, Parus atricapillus. Animal Behaviour 38:685-692.

Whitby, M. D., T. C. Carter, E. R. Britzke, and S. M. Bergeson. 2014. Evaluation of mobile acoustic techniques for bat population monitoring. Acta Chiropterologica 16:223-230.

Wilson, A. M., J. M. Barr, and M. E. Zagorski. The feasibility of using drones to count songbirds, *in* Conference The feasibility of using drones to count songbirds. April 19, 2015.

Yu, K., R. Hudson, Y. D. Zhang, K. Yao, C. Taylor, and Z. WANG. 2013. Low-complexity 2d direction-of-arrival estimation for acoustic sensor arrays. IEEE Signal Processing Letters.

Zwart, M. C., A. Baker, P. J. McGowan, and M. J. Whittingham. 2014. The use of automated bioacoustic recorders to replace human wildlife surveys: an example using nightjars. PloS one 9:e102770.

Vita

Emily Hockman is from Brandon, Wisconsin. She attended the Rosendale-Brandon school district for her K-12 Education. After graduation she attended the University of Wisconsin Eau Claire and graduated with honors with a Bachelor's degree in Biology and minor in Natural Sciences. She spent four and a half years pursuing technician work on avian projects across the country and internationally before accepting a graduate position at the University of Tennessee Knoxville under Dr. David Buehler. She completed her Masters of Science degree in the Forestry, Wildlife, and Fisheries department in 2013 and stayed on in the same department for her Doctorate of Philosophy, Natural Sciences concentration.