AUTOMATED DETECTION OF RARE AND ENDANGERED ANURANS USING

ROBUST AND RELIABLE DETECTION SOFTWARE

by

Andrew Rance MacLaren, BS

A thesis submitted to the Graduate Council of
Texas State University in partial fulfillment
of the requirements for the degree of
Master of Science
with a Major in Biology
August 2015

Committee Members:

Michael R.J. Forstner

Shawn F. McCracken

Floyd W. Weckerly

FAIR USE AND AUTHOR'S PERMISSION STATEMENT

ACKNOWLEDGEMENTS

I would like to express my thanks to Dr. Michael R. J. Forstner for his willingness to take me on as his responsibility, a task for which many would call him brave. Without his knowledge and experience I would not have been able to follow through with this process. From the onset of our professional relationship I was privy to the limits of Dr. Forstner's patience and expectations, and with this work I am confident that I have exceeded them both. Yet through it all, I never sensed doubt from him, for that I am very thankful.

I must thank Dr. Shawn F. McCracken who gave me the confidence to achieve an education beyond my baccalaureate. He has been the most critical sounding board throughout my research, and is simply the best professor I have ever had. Thank you for seeing something in me that I was unable to see in myself.

I am also thankful for the patience and guidance of Dr. Floyd W. Weckerly, who agreed to participate as a committee member. His input has elevated the quality of my work by orders of magnitude. The Forstner Lab members deserve thanks, namely Zachary Adcock, who has consistently been a fountain of great ideas and good advice.

Finally, I must thank all the friends and family that have shown support for my ambitions throughout this journey. Also, a very special thanks to my Grandparents, Lois and Ron Duquette, for your continued love and support throughout my adventures, whatever they may be.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Amphibian populations are experiencing rapid rates of decline, the causes of which are sometimes controversial. The vocalization of the male anuran is used as an indication of a potential breeding event. Researchers have been relying on these vocalizations to monitor the health, reproductive status, and diversity of anuran populations for centuries. As technology advances so does our ability to innovate and improve the way anuran populations are monitored. One such innovation comes in the form of portable commercially available audio recording devices (ARD). These tools enable researchers to capture the sounds produced by populations of any vocalizing animal species and analyze them using machine-learning techniques of pattern recognition. The application of these techniques is understudied and not well documented for anurans. I conducted rigorous testing of these techniques to improve methods of monitoring populations of the endangered Houston Toad (*Bufo houstonensis*). The desired result of these tests would be a reliable and robust tool for recognizing the call of the Houston Toad. This would allow researchers to search vast quantities of digital audio files for the unique sound of this animal. I also compared the efficacy of this machine-learning technique to a highly trained professional listening for the call. Researchers often doubt the reliability of automated techniques, thus my recognizer must perform capably. Additionally, I employed these automated machine-learning techniques to document the presence or absence of the Houston Toad in two counties of Texas, and then coupled those data with

highly resolute details of the environmental conditions to examine calling activity of the Houston Toad and graphically visualize this behavior across a complete chorusing season.

CHAPTER I

DEVELOPMENT OF RECOGNITION TOOLS FOR INCORPORATION WITHIN SONG SCOPE (©WILDLIFE ACOUSTICS) AUDIO ANALYTICAL SOFTWARE

Introduction

Inquiring minds have been perplexed and intrigued by the vocalization of anurans dating back to the early Greek philosophers (Capranica 1965). Anurans represent the first clade to have evolved laryngeal vocalization in the vertebrate phylum, and controversy over when and why these animals produce such sounds persists to this day (Capranica 1965, Bridges and Dorcas 2000, Jackson et al. 2006).

Presently amphibian populations are experiencing dramatic and alarming rates of decline, most notable in anuran populations (Phillips 1990, Stuart et al. 2004, Gibbs et al. 2005). Long-term monitoring of anuran populations is required to gain an understanding of population dynamics and causes for decline (Pechmann et al. 1991). During the breeding season, when male anurans vocalize to attract females, researchers commonly conduct auditory surveys to determine presence or absence of anuran species (Bridges and Dorcas 2000, Crouch and Paton 2002, Schmidt 2003, Pierce and Gutzweiller 2004, Jackson et al. 2006). Data from these surveys can be used for estimates of the relative abundance of calling male anurans (Zimmerman 1994) or for monitoring the occurrence of anuran populations (Weir et al. 2005).

Methods of how best to access the calling activity of anurans have evolved alongside technological innovations. As our technological capability to record and analyze audio has developed, questions of how these innovations may serve our scientific interests have been raised. Sound recording equipment that seems primitive by today's standards, such as tape recorders (Parris et al. 1999), brought state of the art technology to surveying methods allowing researchers to bring the sounds from the field back to their laboratories. With the digital revolution tape recording devices became obsolete. The last two decades have brought dramatic improvements in both the portability and the recording quality of audio devices. There are now commercially available devices designed specifically to record and store the sounds of nature for the sole purpose of biological monitoring. These devices record high-quality digital audio files directly onto removable digital media, such as an SD card (©Wildlife Acoustics, Maynard, Massachusetts). Audio recording devices (ARD) like these have been used to monitor the activity of anurans acoustically (Bridges and Dorcas 2000). ARD's give researchers the opportunity to exploit advantages that are not available to monitoring programs utilizing purely traditional survey techniques, such as the ability to monitor multiple sites simultaneously including places and times that might be difficult for researchers to physically access and survey (Hsu et al. 2005). With the availability of efficient, cost-effective ARD's a single researcher is easily capable of obtaining many hours of audio recordings of frog calls. However, this requires a non-trivial investment in person-hours required for listening and interpretation of audio data.

As stated before, technological advances bring methodological advances to biologists. Machine-learning methods for identifying the unique call of vocalizing fauna have come to be critical tools for the monitoring of avian species (e.g. Raven Pro 2014, Cornell Ornithology Lab), as well as chiropteran mammals (e.g. Anabat Titley Scientific, SM2BAT+ ©Wildlife Acoustics). Commercially available recognition tools for many species of bird and bat are available for researchers, but no such tools for anurans have come to market. Moreover, the efficacy of these "canned" recognition tools has been criticized in the literature (Barclay 1999, Towley 2012). Machine-learning methods for identification of anuran vocalizations have been developed (Taylor et al. 1996, Brandes et. Al. 2006), but only very recently have commercially available platforms for designing these tools become common.

Recently, Song Scope (©Wildlife Acoustics) bioacoustics monitoring software has been examined for its usefulness as a platform for generating recognition tools for use in monitoring anuran populations (Eldridge 2011, Waddle et al. 2009). One critical aspect of recognizer development that is overlooked in both the previously mentioned studies is the subjectivity of these recognition tools. The methods used to produce recognition tools can vary between software packages and study designs, and thus the reliability of the resulting recognizers varies just the same.

The goal of this study is to design a reliable methodology for generating recognition tools using Song Scope Bioacoustics Software (©Wildlife Acoustics). The endangered Houston Toad (*Bufo houstonensis)* will be the focal species for this study. In regard to the Houston Toad, permitted observers rely primarily on auditory cues for species detection (*Bufo houstonensis*; Forstner and Swannack 2004, Jackson et al. 2006). Therefore the need for reliable and robust recognition tools for locating the unique vocalization of the Houston Toad is of critical concern for researchers. Though automated methods have been proposed for the Houston Toad, high costs and low quality of machine-learning techniques for automated detection have been cited as obstacles preventing widespread implementation (Jackson et al. 2006).

Houston Toads, *Bufo* (=*Anaxyrus*) *houstonensis* (Sanders 1953) are a rare species of anuran endemic to southeastern Central Texas, both federally listed under the Endangered Species Act (U.S. Endangered Species Act 1973) and internationally protected under the IUCN red list (Hammerson 2004). Continuous habitat loss and fragmentation throughout its range are the major drivers of population declines, with Bastrop County historically considered to hold the largest populations of Houston Toads across its range (Brown 1971; USFWS 1984). Houston Toads were historically present in twelve Texas counties, but have been extirpated in three of those counties since the 1960's. In September of 2011 the Bastrop County Complex Fire burned over 34,000 acres of habitat including 96% of Bastrop State Park, which was believed to be the species last stronghold. This catastrophic natural disaster and the ongoing severe drought present the additional stressors on already declining populations. Therefore, continuous species monitoring across its remaining range is essential for implementing proper management and conservation regimes, as well as continuing head-start efforts. However, due to time, personnel, and financial constraints, thorough long term, range-wide monitoring is often times difficult to maintain.

Materials and Methods

Audio data collection was carried out in the first half of 2014 (January 3-July 12). Thirty-five commercially available ARD's (Song Meter model SM1, SM2, and SM3) were deployed at potential breeding locations for the Houston Toad in two counties of east central Texas. 11 of these ARD's were placed in Bastrop County; Four ponds on the

1967-ha Griffith League Ranch, one pond from the adjacent Welsh ranch, five ponds or drainages bordering Highway 290, and one pond located on the Bluebonnet Electric Company headquarters each received an ARD (Figure 1.1). The remaining 24 ARD's were deployed in Robertson County, Texas for use in biological monitoring throughout the installation of an oil pipeline bisecting a patch of occupied Houston Toad habitat with adjacent potential breeding ponds. In total, 24 potential breeding locations were identified along the nearly 12 miles of Right of Way being monitored and each received an ARD (Figure 1.2).

ARD's were secured to trees or structure objects <10m from pond, drainage, or water body edge and oriented such that the device would be "facing" the water. Each ARD was programmed to record 10 minutes each hour, on the hour, from 18:00 to 05:00 the following morning. This resulted in a total of 12 - 10 minute segments (120 min) of audio per logger, per night. To reduce file size the WAC data format was selected, and sample rates were lowered to 16000 Hz. This effectively lowered the "ceiling" on the spectrum of audio being recorded, eliminating only ultrasonic frequencies not needed for my analyses. I equipped each ARD with four Tenergy (Tenergy, Fremont, California) brand rechargeable D-cell batteries rated at 60Wh as opposed to the conventional 72Wh in other non-rechargeable types. Finally each ARD was stocked with a 32GB SD card for primary data collection, as well as a 16 GB or 8GB SD card to be used as a backup in case of failure. Under these settings my ARD's required battery changes approximately every 40 days. During these visits the existing SD cards containing digital audio files were swapped for blank replacements.

For comparison, traditional nocturnal anuran calling surveys were carried out roughly following the protocol for Houston Toad given by the United States Fish and Wildlife Service (USFWS) (Figures 1.1 and 1.2). Surveys were conducted on nights that met the environmental conditions prescribed by the USFWS, as well as on nights that would not be considered under the current protocol, to insure that no chorusing events would be missed. Traditional driving surveys with preselected listening stations across this habitat patch required, at minimum, five researchers to complete, one group of surveyors for each county (Bastrop and Robertson Counties, Texas). Fifty-eight listening posts in Bastrop County were monitored by dividing them among three routes, each route assigned to one researcher. Twenty-two listening posts in Robertson County were divided into two routes, requiring two researchers.

Song Scope spectrographic visualization software was used to review collated audio data files. It is one of the few commercially available spectrographic visualization software that offer the ability to build and customize your own recognizers. Seven audio files containing isolated and well-defined Houston Toad vocalizations were identified by cross referencing positive detections during traditional survey outings. The vocalizations contained within these files were used as training data for my recognizer. The calls contained within these seven files ranged in number from 13 to 61 vocalizations made by a single male Houston Toad. Each file was opened in Song Scope and annotated. Annotating a bit of audio in Song Scope is a simple “click and drag” type highlighting process used to define the physical bounds of the focal species vocalization within the viewable spectrograph (Figure 1.3). Practically, an annotation is the physical definition of the moments in time when a vocalization of interest starts and ends, and which frequencies this vocalization occupies, for the purpose of incorporating the sound segment into a recognizer.

In order to track the performance of my recognizer as it was being built, I simulated one full survey night of positive detections. I used the same cross-referencing technique to generate a population of 105 files that contained the call of the Houston Toad. I used program R (R-project: Gentleman 2009) to randomly select 12 files from this population. These 120 minutes (12, 10 minute audio files) were then manually searched to determine the number of calls contained within (n=186).

Song Scope offers two proprietary filters used for removing unwanted results from appearing in your output. Quality is one filter, and it is a measurement of the signal characteristics. Quality can range from 0.00 to 9.99; 5.00 indicating signal characteristics of an identified sound are average to the characteristics in the recognizers training data. However, the results table created from batch processing in Song Scope shows Quality as ranging from 0.00 to 99.99; Figure 1.4 follows these terms. The second filter is Score, which can range from 0 to 100%, and measures the statistical fit of a vocalization to the model estimated by the recognizer. As a benchmark, vocalizations of interest should fall within the range of the cross-training score given for the recognizer (Song Scope User Manual 2014). For this experiment each recognizer was used to scan the sample 120 minute audio batch with Quality and Score set at zero, in order to determine the lower threshold for true positive detections. The purpose of this step is to insure a zero tolerance for false negative identifications (Type II error), given that my focal species is rare and elusive and known to call sporadically (Jackson et. al. 2006).

I began the process of building a recognizer for the Houston Toad by incorporating the first of seven files of annotations (25 vocalizations) and adjusting the parameters offered within the software such that they summarize the variation present within those 25 individual calls. A detailed list of those parameters, and how they change as I improved my recognizer, can be found in Tables 1.2 and 1.3. In order to confirm that the parameters have been appropriately set I then recursively scanned the single file in which the annotations used as training data were taken from using the final recognizer. This allowed me to view the recognizers “opinion” of itself. Parameters were adjusted until the recognizer was able to accurately identify all of the vocalizations that were used to build it. Results from these self-tests were manually reviewed to insure that true detection of each vocalization was made. Also, the number of identifications made were counted and compared to the number of true vocalizations they represent, and shown relative to the total number of identifications (false or otherwise) for a batch run with filters fixed at zero and with filters adjusted to lowest true positive.

Once the recognizer was constructed, parameterized, and self-tested, I used it to scan the randomly selected 120 minute audio sample with both Quality and Score fixed at zero (Table 1.2). The output of these scans were reviewed manually to confirm detections and insure no false negatives (i.e. true vocalizations not identified by the recognizer) were committed. The batch was then scanned a second time with Quality and Score adjusted to match the lowest true positive detection (Table 1.2).

These steps were carried out for each new set of annotations incorporated into the recognizer. Once all seven files worth of training data were present, unwanted annotations were removed that had potentially negative effects (i.e. overlapping vocalizations, short bursts, weak signals, or poor quality). Overall eight steps were carried out to complete the process of developing a recognizer for *Bufo houstonensis.*

Results

The 35 ARDs individually recorded between 1465 and 2272 audio files. A total of 657,350 minutes of audio were collated during the breeding season of 2014. I experienced data loss at several ARD's due to inconsistent battery life, which is to be expected when managing a rather large collection of rechargeable cells used in demanding circumstances.

Results from the final step of my recognizer's self-tests (Table 1.1) revealed that true positive identifications (those used as training data) make up 68% of the results given when settings are adjusted to exclude results with a Quality and Score lower than that of the lowest true positive detection. This adjustment of Quality and Score removes 27.8% of the unwanted false positives identified. As training data was added Quality and Score lowered with every step, with the exception of the final step in which non-perfect annotations were removed from the recognizer. The number of vocalizations present in test data was underestimated in steps 4, 6, 7, and 8 (Tables 1.1 and 1.2). That is, the number of identifications used to identify 100% of the true positives was less than the number of positives present. This is due to a single identification accounting for more than one vocalization. The number of vocalizations were overestimated in steps 1, 2, 3, and 5, such that more than one identification was made per true positive present in the training data (Tables 1.1 and 1.2). That is, a single call from a single male at one time was recognized by the algorithm multiple times, resulting in multiple positives recorded for the single call event. Upon the final step of development, self-tests show that the finalized *Bufo houstonensis* recognizer underestimates the number of vocalizations present by 2.7% (n=191), in low deviation from the actual number present (n=186).

Recognizer performance when tested using the 120 minutes of simulated anuran activity are summarized in Tables 1.2 and 1.3. Between initial creation and completion the total number of false positives were reduced by 97%. Of the final 219 identifications made by the recognizer, only 28 results measured shorter than one second in Duration, indicating they represent erroneous false positives (verified as false positives manually).

Parameters that underwent the greatest change throughout this process included cross and total training percentages, model states, state usage, and mean duration. Cross training and total training both refer to the "fit" of model being built, represented as the average and standard deviation. Cross training refers only to annotations excluded from the recognizer, whereas total includes every annotation. Model states describes the size of the model as the number of states it contains. The number of states increases as call complexity increases, thus a small number of states indicates a non-complex call. State usage represents the average and standard deviation of the number of different states traversed by each vocalization. Mean duration represents the average and standard deviation of the length of each vocalization in seconds.

These parameters, more than any others, had a strong influence on the performance of the recognizer. Training percentages dropped as variation from annotations increased, however, these percentages ranged from 70.97 to 83.3 percent. Such high estimates shouldn't be considered low in this context, and it is safe to assume that the majority of variation within Houston Toad vocalizations is represented by the recognizer. Mean duration ranged from 5.92 to 10.45 seconds. This parameter limits the Quality given for lowest true positive considerably. Thus, the 8th iteration aimed towards increasing the mean duration to influence this metric.

False negatives were located in two of the seven training files and in one of the files used to create the simulated 120 minute audio sample. Only one false negative per each of these three files were found, and they occur at the origin of the file (Figure 1.3). False positives within the recognizer output consisted primarily of identifications shorter than one second. These false positives were triggered by the sound of wind, rain, automobile traffic, birds, and other anurans, namely *Hyla versicolor* or *Pseudacris crucifer*.

Discussion

My approach for the development and optimization of this recognizer followed a strict criterion of zero tolerance for false negatives. While this may not be necessary for recognizers focused on all species, for a rare and elusive anuran such as the Houston Toad this method is necessary and proved to be quite effective. Studies have shown that there is a tradeoff between false positive and false negatives (type I and type II errors) (Eldridge 2011, Waddle et al. 2009.) However, given the sporadic and unpredictability of the Houston Toad's vocalizations, false negatives pose a much greater concern than false positives for researchers. The results in Table 1.2 illustrate that with little to no instance of false negatives I was able to achieve nearly 90% true positive results within my control audio.

Though the zero tolerance approach was overall a success, there was found to be a unique situation in which a false negative will occur. In the event a vocalization of interest is taking place when a recording begins, that portion of sound that is captured will go unidentified by the software (Figure 1.3). This phenomenon was seen a total of only three times throughout my study, once within the self-test phase and twice more within the control data. I believe this to be the only occasion, when using a zero tolerance approach, in which a false negative should occur. I have one suggestion for how this error could be corrected albeit purely hypothetical. Song Scope produces recognitions in two passes. The first pass makes large identifications, the second focuses on smaller less ideal identifications. Both these passes move in the same direction through the file (chronologically). If the direction of either pass was to run in the opposite direction, I believe this error would be solved. This hypothesis is supported by several instances where Song Scope correctly identifies vocalizations of interest that are taking place as a recording ends, leaving only a portion of sound captured. If either pass were to view the beginning of a recording as the "end" then this vocalization should be identified appropriately. It is not clear at this time how this can be enabled, but an approach that digitally reverses the initial recording of each file is one blunt force approach to a solution without requiring changes to the software itself. Another hypothetical solution is to physically add five seconds of silence to the beginning of every recording, which may allow enough time or variation present to capture the initial vocalization of interest.

One difficulty I faced in developing this tool was that throughout the process I was unable to move the minimum Quality ranking of my true positive identifications above 2.4. This problem is explained inherently in the definition given for Quality by Wildlife Acoustics in the Song Scope users manual. Quality is the ranking compared to the "average characteristic of the recognizer." In Figure 1.4 you will see that Quality ranking and Duration of identification share a normal distribution. That is to say that identifications closer to the mean duration of the annotations making up a recognizer, the higher the Quality. Therefore, the greater the difference in Duration of the identification in either a positive or negative direction (longer or shorter) the lower the Quality ranking for that vocalization. I do not believe this phenomenon to be limited to only effects based on Duration of vocalization, but it is the only parameter having an effect on the efficacy of my recognizer to identify the call of the Houston Toad, that I am able to detect. Given that the Houston Toad's call has a constant frequency that is rather narrow-banded, most of the variation present within the call is in regard to its length. Because Quality is defined as the mean of all characteristics, it is fair to consider the outliers that violate the normal distribution presented in Figure 1.4, as those vocalizations that exhibit un-characteristic variation in a metric other than Duration.

In total the process of preparing and optimizing this recognizer required a comprehensive time investment of approximately 24 hours once the methodological approach was confirmed. This is a shorter build time, start to finish, than comparable studies (Eldridge 2011, Waddle et. al. 2009). Of course, if not for the advantage of gleaning my training data from known positive dates there would be a definite increase in time investment. One other advantage to my methods is the small yet effective amount of control data. 120 minutes of audio requires <2.0 minutes to process. Processing times are dependent on a multitude of factors, for example the computational ability and processing power of the machine running the software, and thus they may vary greatly between researchers.

What I am able to gather from other studies that attempted to access the efficacy of automated or machine-learning techniques of detection, when compared to my own study, is that subjectivity is key. Critical to the efficacy of a recognition tool is the attention to detail given from the researcher designing said tool. The results reported in studies similar to my own show dramatic errors associated with automated techniques (Eldridge 2011, Waddle et. al. 2009), that I believe can be eliminated with more robust and carefully assembled tools.

There is an egregious lag between the discovery of errors and acceptance of proposed solutions in regards to current government protocols for monitoring populations of threatened and endangered anurans (Jackson et al. 2006). These errors are inherent to traditional call survey methods, stemming from observer bias, temporal variation, ease of access, right of entry, hazardous roadways, presence of observer effects, etc. Many, but not all, of these errors can be corrected for via the implementation of audio recording devices. While ARD's suffer from their own suite of errors (i.e. data loss, battery life, theft, and requiring ROE for many breeding locations) the advantages they offer to researchers far outweigh these shortcomings. Although remote audio recording devices are becoming more commonly implemented, I am unaware of any ongoing long term monitoring program for anurans that uses automated detection in practice. This represents a growing body of anuran datasets with limited availability of useful tools to analyze said data. Given recent advancements, software now offer simple user friendly foundations for complete development of robust and reliable automated pattern recognition tools. As indicated by my research involving the endangered Houston Toad, a proper methodological approach to creating these tools would enable researchers to better understand the sounds they capture.

Table 1.1 Performance of Houston Toad (*Bufo houstonensis*) recognizer during self identification tests. The following table summarizes the self-identification performance of the *Bufo houstonensis* recognizer for each step of development ("steps" columns 1 – 8). That is, the recognizers ability to identify the calls it is built from. Parameters include the minimum "Quality" (min Q) and "Score" (min S) assigned to a true positive detection, the number of results produced when filters "Quality" and "Score" are fixed at zero (highlighted blue) (no. results), the number of results with those filters adjusted to represent the numbers presented in the first two rows (highlighted orange) (adj. results), the number of results that represent positive *B. houstonensis* detections (pos. det), and the true number of *B. houstonensis* vocalizations used for each step (no. calls).

	step							
parameter	1	2	3	4	5	6	7	8
min Q	11.8	3.5	2.3	6.1	8.2	6.6	1.3	3.5
min S	78.28	67.63	43.5	47.93	47.18	46.69	40.43	44.89
no. results	492	892	603	221	294	338	377	338
adj. results	26	80	122	126	177	216	333	244
pos. det	26	78	90	111	134	163	185	166
no. calls	25	58	77	117	130	191	219	170

=Quality and Score fixed at zero (blue)

=Quality and Score set to lowest true positive (orange)

	step							
parameter	1	2	3	4	5	6	7	8
cross training (%)	0 ± 0	75.93 ± 18.88	76.77 ± 28.76	73.72 ± 12.02	73.81 ± 12.23	72.95 ± 10.92	74.2 ± 10.62	73.05 ± 11.62
total training (%)	83.3 ± 5.38	76.39 ± 13.57	77.88 ± 31.37	75.67 ± 11.53	75.08 ± 14.63	70.97 ± 12.13	73.91 ± 12.24	73.91 ± 13.72
model states	38	30	40	45	44	40	40	46
state usage	33 ± 7	12 ± 8	16 ± 10	21 ± 7	21 ± 9	20 ± 8	19 ± 7	24 ± 9
feature vector	8	8	13	13	13	13	13	13
mean symbols	675 ± 271	432 ± 377	471 ± 457	594 ± 411	514 ± 413	525 ± 392	514 ± 371	578 ± 399
syllable types	8	12	7	10	11	7	12	6
mean duration (s)	8.75 ± 3.51	5.92 ± 4.9	6.06 ± 5.33	8.99 ± 4.66	9.42 ± 5.1	9.51 ± 4.82	9.48 ± 4.62	10.45 ± 4.47
max complexity	48	48	48	48	48	48	48	48
max resolution	8	8	13	13	13	13	13	13
freq. min (Hz)	30	30	26	26	26	26	26	26
freq. range (Hz)	5	6	11	11	11	11	11	11
max syllable (ms)	200	176	328	1504	1504	1504	1504	1504
max syl. Gap (ms)	96	64	256	1048	1048	1048	1048	1048
max song (ms)	16960	16960	16544	16544	16544	16544	16544	16544
annotations	25	58	77	117	130	191	219	170
Total ID's	6829.00	9990.00	4400.00	797.00	797.00	797.00	797.00	797.00
ID's > 1s	234.00	262	692	642	642	642	642	642
Total ID's	6829.00	8377.00	360.00	206.00	210.00	204.00	208.00	219.00
ID's >1s	234.00	244.00	206	187	189	188	189	191
Quality	0	1.3	1.3	1.3	1.3	1.9	1.3	2.7
Score	0	19	46	57	57	55	56	54

=Quality and Score fixed at zero
=Quality and Score set to lowest true positive

Table 1.2 Changes in parameterization of recognizer for Houston Toad (*Bufo houstonensis*). The following table summarizes the performance of the *Bufo houstonensis* recognizer for each step of development (“steps” columns 1 – 8) when used to scan 120 minutes of control data containing known vocalizations (n=186) of the Houston Toad. Each parameter estimated by the software Song Scope is given for each iterative step of development (“steps” 1 – 8). Rows highlighted in blue represent the total number of results (Total ID’s) and the number of results that measure longer than one second (ID’s >1s) when filters “Quality” and “Score” are set to zero. Rows highlighted in orange represent the those same measurements when “Quality” and “Score” are adjusted to represent the minimum boundary of positive Houston Toad detection, those minimum boundaries are given in the last two rows.

Table 1.3 Parameters held constant during recognizer for Houston Toad (*Bufo houstonensis)* development. In the table below the parameters held constant throughout all eight steps of recognizer development for automated detection of Houston Toad vocalizations within Song Scope software are given.

sample rate (Hz)	16000	amplitude gain (Db)	0
playback speed	normal	background filter	1s
max sample delay	64	dynamic range (Db)	20
FFT size	256	algorithm	2
FFT overlap	1/2		

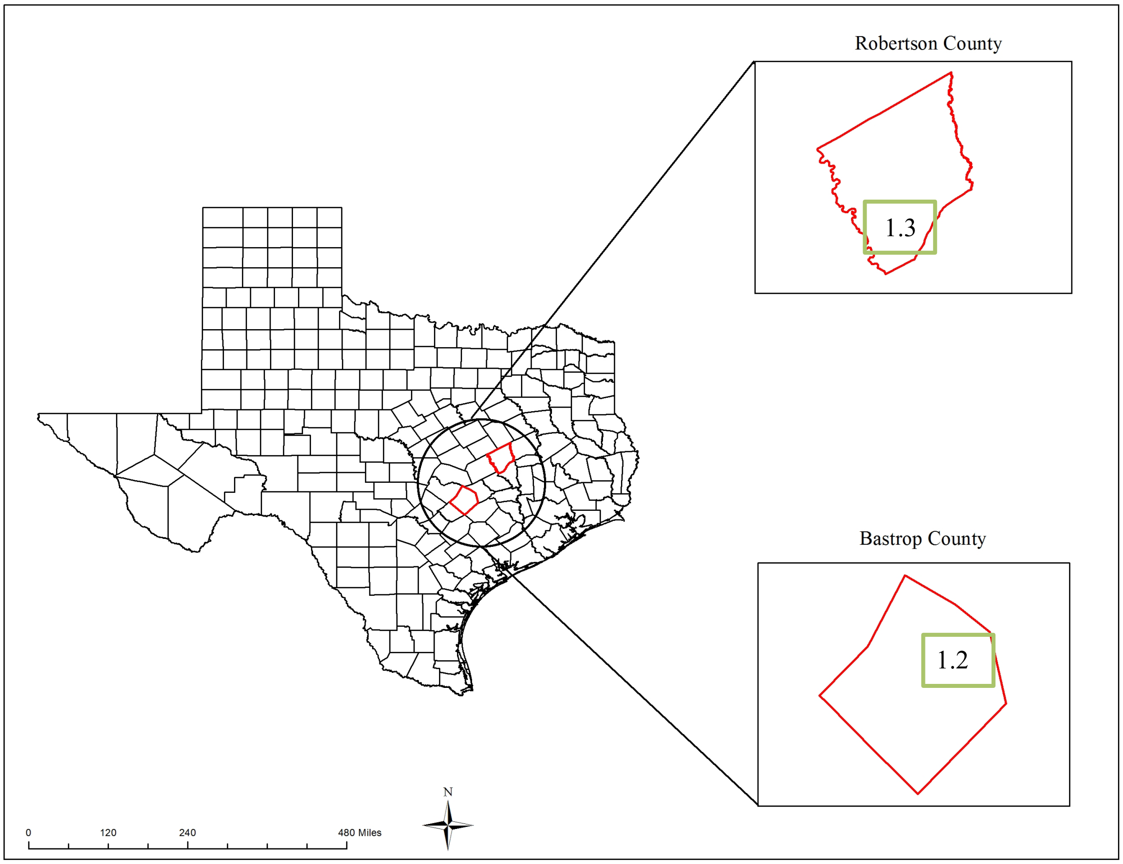

Figure 1.1 Map of Texas showing Bastrop and Robertson counties, Texas, (in context). Figure Legend for Robertson and Bastrop Counties outlined in red to showing location within the state of Texas. Green boxes represent areas described by Figures 1.2 and 1.3 in their respective counties.

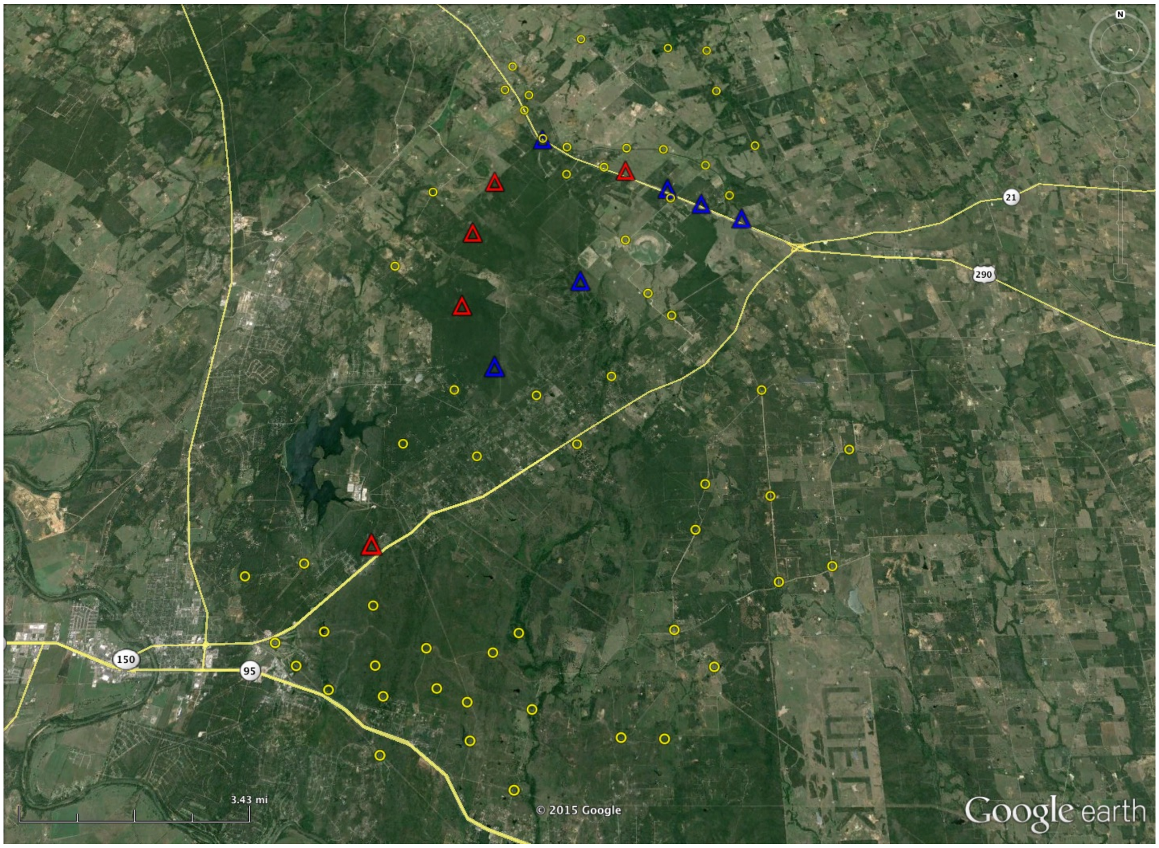

Figure 1.2 Map of area monitored for Houston Toad (*Bufo houstonensis*) within Bastrop County, Texas. Audio Recording Devices indicated as triangles, red indicates presence of *Bufo houstonensis***,** blue indicates absence of *Bufo houstonensis.* Open yellow circles indicate traditional survey listening posts. For context within the county refer to Figure 1.1.

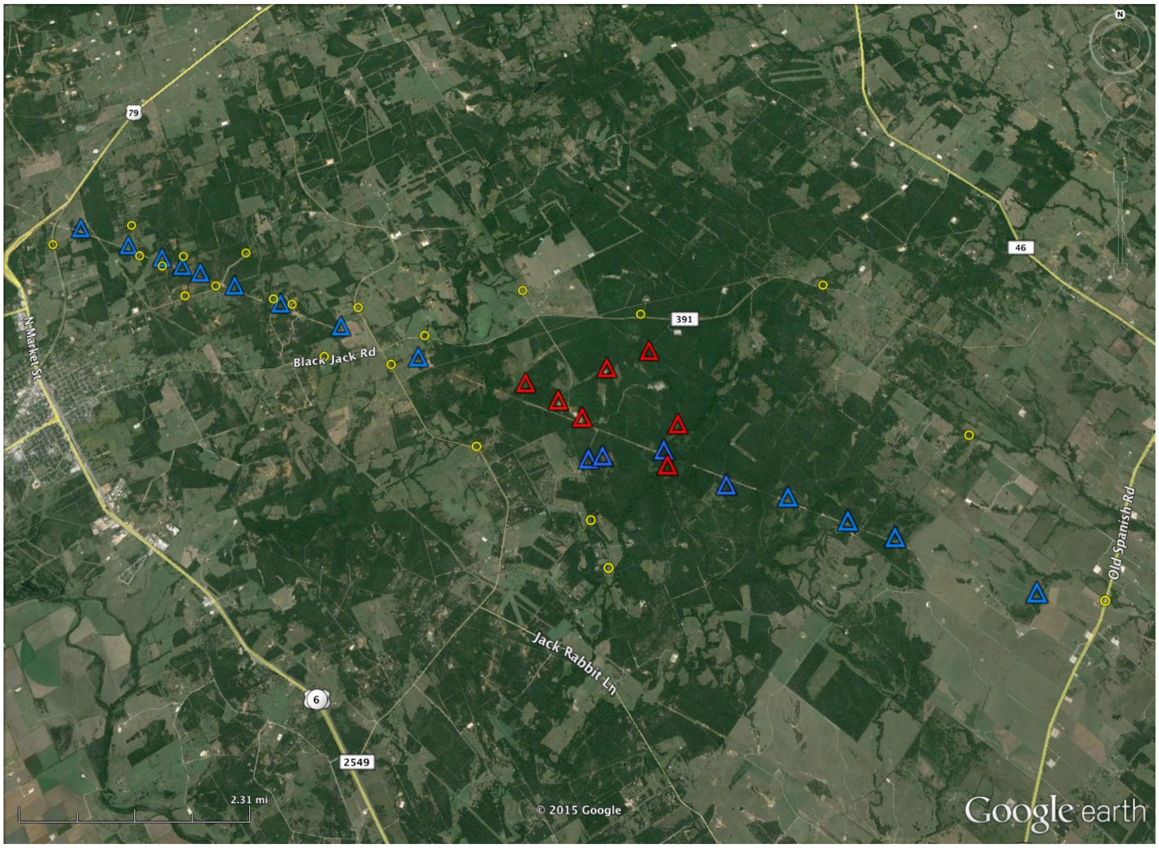

Figure 1.3 Map of area monitored for Houston Toad (*Bufo houstonensis*) within Robertson County, Texas. Audio Recording Devices indicated as triangles, red indicates presence of *Bufo houstonensis*, blue indicates absence of *Bufo houstonensis*. Open yellow circles indicate traditional survey listening posts. For context within the county refer to Figure 1.1.

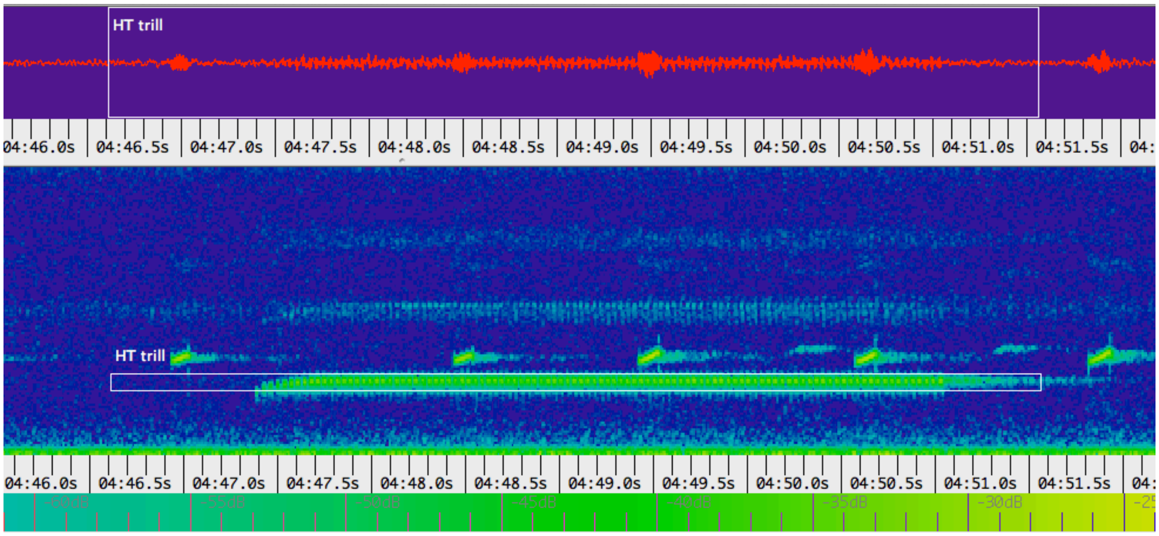

Figure 1.4 Spectrographic annotation of a Houston Toad (*Bufo houstonensis*) vocalization in Song Scope Software (©Wildlife Acoustics). Screen capture of Song Scope in which a Houston Toad vocalization is annotated (selected for analysis), as indicated by the white box surrounding the vocalization of interest.

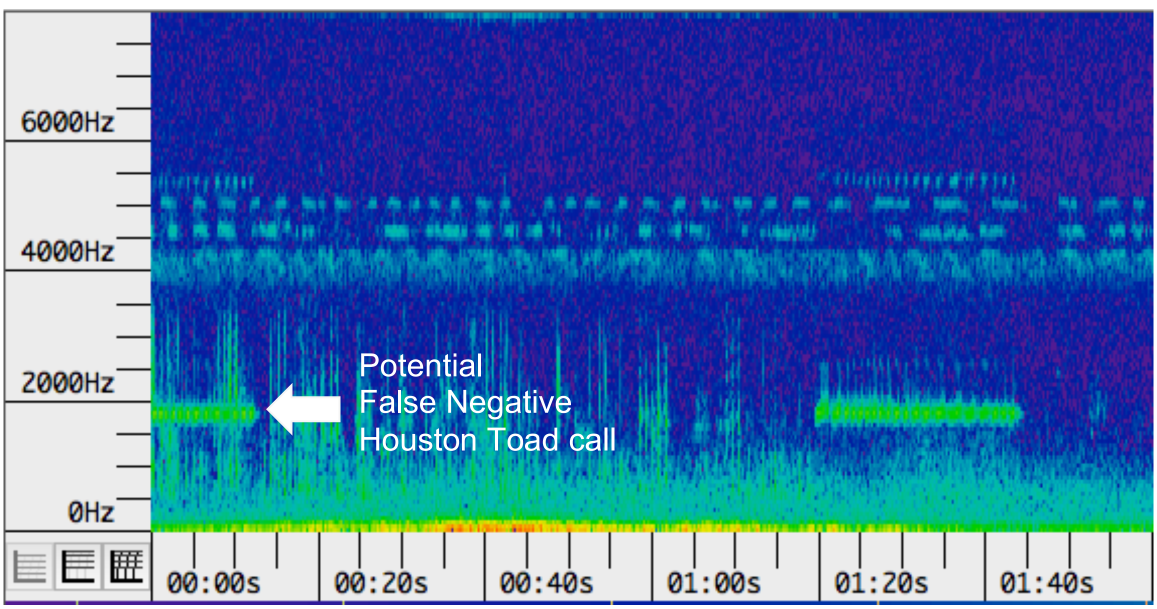

Figure 1.5 Spectrographic of potential false negative of a Houston Toad (*Bufo houstonensis*) vocalization in Song Scope Software (©Wildlife Acoustics).. Screen capture of Song Scope in which a Houston Toad vocalization is taking place at the origin of the recording, leading to a false negative or non-detection, as a result of an inherent error within the software.

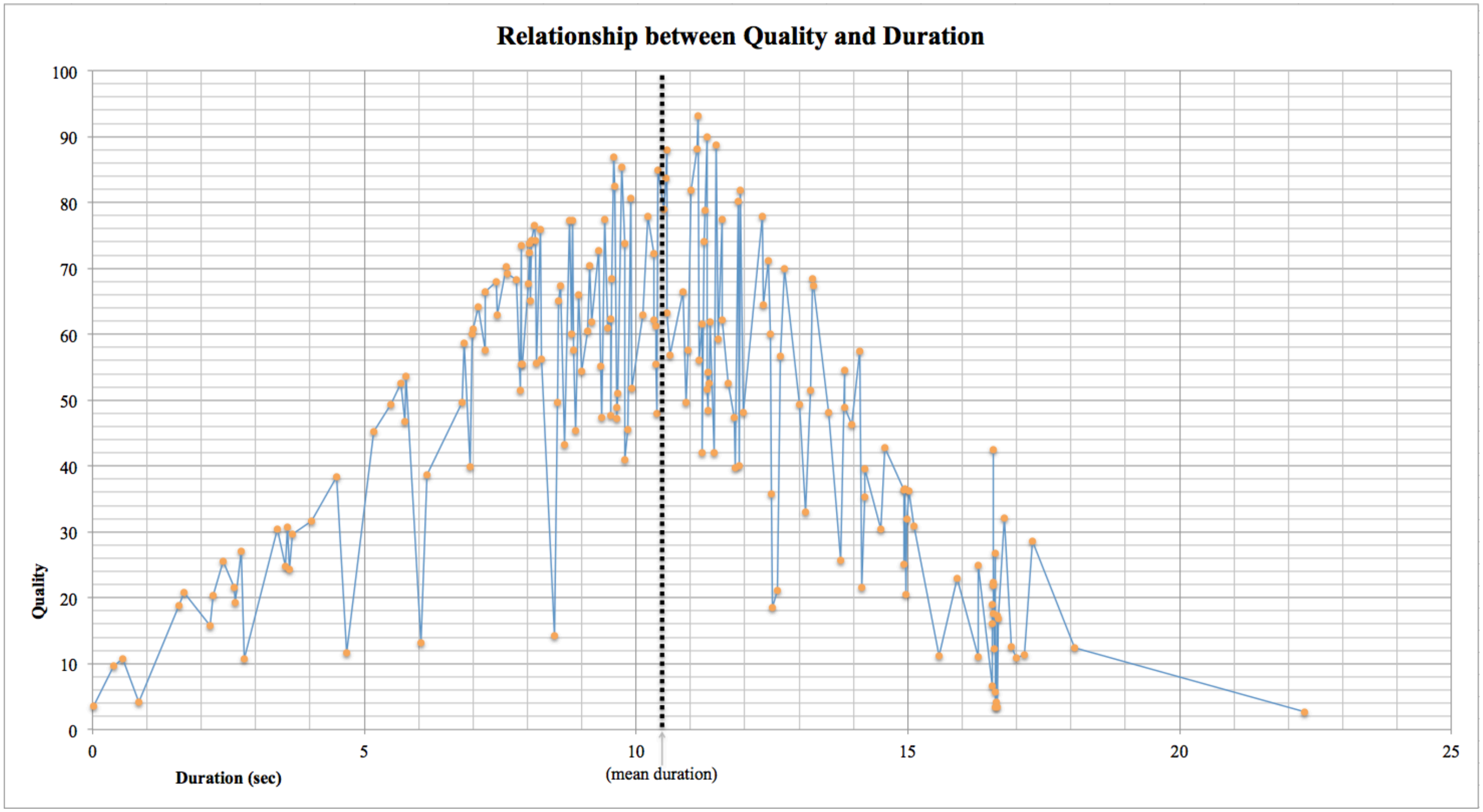

Figure 1.6 Graph of the "Quality" by "Duration" for detections of Houston Toad (*Bufo houstonensis*) vocalizations. This graph illustrates that vocalizations identified as closer to the mean Duration are assigned greater "Qualities" by the algorithm used in Song Scope. Duration is the length of the vocalization produced by a Houston Toad. Quality is a proprietary estimate used by Song Scope software. Quality can range from 0.00 to 99.9; 50.0 indicating signal characteristics of an identified sound are average to the characteristics in

CHAPTER II

COMPARING EFFICACY OF AUTOMATED DETECTION SOFTWARE VERSUS A TRAINED RESEARCHER

Introduction

As discussed in Chapter I, researchers have relied on audio recording techniques for as long as portable recording technologies have been in existence. If practices of automated detection and machine-learning techniques are to be widely accepted by researchers it is important that we understand how these methods compare to current survey methodologies.

More recently studies have examined whether source of error in human listeners and automated methods are similar or dissimilar (Eldridge 2011). The authors studied whether listeners with between 3 and 12 hours of training would be able to identify the unique call of eight different anurans, then compared their estimates to that of the author's recognizers for each species. While there is useful information to be gleaned from understanding what sources of variation exist in errors committed by human listeners that will not be the focus of this chapter.

In an effort to further test the efficacy of the recognizer constructed in the first chapter I will examine the abilities of a certified and permitted surveyor of Houston Toads and compare their findings with those of my recognizer. I hypothesize that given the iterative and rigorous parameterization in Chapter I that my recognizer will perform to the standards expected of an expert certified by the USFWS.

Methods

I used the entire 2014 season of audio taken from a single location to test the efficacy of the recognition tools created in Chapter I and compared its abilities to that of a trained professional researcher. I chose the location of interest that represented the highest chance of containing male Houston Toad choruses, "Pond 12" of the Griffith League Ranch (Bastrop, Texas, USA), based on historical data from Dr. Michael R. J. Forstner (Figure 2.1, 2.2, 2.3). Digital audio files were captured using ARD's following the methods outlined in Chapter I.

Digital audio files were reviewed by myself, and verified by Dr. Shawn F. McCracken. Both individuals are federally permitted to conduct Houston Toad audio surveys and represent expert level surveyor effort. Reviewing files manually was done mostly spectrographically with some verification requiring acoustic confirmation. Song Scope was used for spectrographic visualization and listening of digital audio files. The number of Houston Toad vocalizations contained within each file were quantified. These numbers are best estimates based on the discretion of the researcher, given that overlapping vocalizations can be confounding and difficult to confidently verify.

The files were then scanned using the recognizer for *Bufo houstonensis* built in Chapter I. I counted the number of times the recognizer selected the call of the Houston Toad accurately, as well as the number of true vocalizations that were missed by the recognizer. If false negatives occurred, files containing such errors were reanalyzed with filters (Quality and Score) set to zero to investigate the source of error within the recognizer.

Results

In total 1,945 files were searched for the unique vocalization of the Houston Toad. This equates to 19,450 minutes of audio (ten minute recordings). Manual methods of review and detection estimated 393 Houston Toad vocalizations. Houston Toads were present in 53 audio files. Of those 53 positive recordings the number of vocalizations per file ranged from 1 to 30, averaging 7.42 vocalizations per file.

The recognizer made 437 true positive detections in 51 audio files. Of those 51 positive recordings the number of vocalizations per file ranged from 1 to 26, averaging 8.57 identifications per file. There were 11 incidents of false negatives. Six of these incidents were a consequence of a vocalization taking place at the origin of the recording, as outlined in Figure 1.3. The five other incidents of false negatives included faint or weak calls, and only a single instance of not detecting a vocalization that is characteristically "normal" that went overlooked by my recognizer (see below). Taking all vocalizations into account my recognizer correctly identified 97.2% of the true vocalizations present within the studied audio.

By dropping the Quality and Score setting to zero and reanalyzing those files that contained false negatives (noted above), I found that my recognizer did in fact identify them. The "Scores" assigned for these faint, weak, or uncharacteristic vocalizations were below the bounds determined in Chapter I for my recognizer. Lowering this filter such that it includes these five vocalizations increased the number of false positives by 224% (n_1=1399 to n_2=3133).

Manual methods of audio review required approximately 32 hours to complete, that is roughly one minute to review each digital audio file. For files that contained no Houston Toad vocalizations, this process was fast and simple. However, for those files that were found to contain vocalizations of interest, quantification and interpretation required greater time investment. Automated methods of detection required < 6 hours to complete batch processing. It required an additional hour to quantify and interpret the results of the automated scanning process. In summation, automated methods of detection required nearly one-fifth the amount of time manual methods required to complete.

Discussion

Former studies (Eldridge 2011, Waddle et al. 2009) have remarked on the trade-off that exists between false positive and false negative detections when using automated techniques. My study exhibits a similar relationship, however I experienced dramatically reduced rates of error when compared to these other studies. Or rather, the results of my study are skewed such that false negatives are forced to their extreme low, driving false positives to rates similar, yet still reduced when compared to findings of others.

There is one glaring dissimilarity between this study and the predecessors which I have referred to throughout (Eldridge 2011, Waddle et. al. 2009). The former studies used relatively commonplace anurans to test the efficacy of automated tools, whereas this study focuses on a very rare endangered anuran. The very nature of my focal species is that it vocalizes sporadically, and periods of activity are quite difficult to predict (Jackson et al. 2006; Brown et al. 2015). As I have touched on in the previous chapter, this places an added value on true detections. Reduction in false negatives must be preferred over reducing false positives, in this instance.

There is a strong emphasis on high-fidelity methods of detection for the Houston Toad due to the nature of policies aimed at protecting the species. Presence/Absence data drives Houston Toad conservation policy, and thus any missed detections that result in false conclusions of absence are of great concern for researchers. The protocol for surveying prescribed by the USFWS has shown to be insufficient in its ability to insure that there are no false negative conclusions (Jackson et al. 2006). For these reasons, false negatives committed by the recognizer could not be tolerated.

As a result of this strong emphasis on reducing false negative detections researchers studying the Houston Toad train for multiple seasons, and must be certified by the USFWS through a current permit holder. It is my opinion, that it is, in part, a consequence of this lengthy and arduous training period that I have found dramatically reduced errors associated with manual review of digital audio files, as opposed to those in comparable studies (Eldridge 2011).

Although my recognizer for *Bufo houstonensis* was found to commit false negatives, these few instances did not cause any single date of positive occurrence for the species to go undetected. That is to say, false negatives occurred in files that contained other detected vocalizations of the Houston Toad, or that files that contained only one vocalization that went undetected (i.e. false negative) occurred on a night (of a date) in which another survey recording was positive for Houston Toad vocalizations. Given these circumstances I am compelled to stand by the findings in Chapter I. Batch processing carried out using filters Quality equal to 2 and Score equal to 54 yielded a low number of false negatives, and was found to have not concluded absence of *Bufo houstonensis* such that it would impact information relevant to the policies of the USFWS regarding this species.

The discrepancy between the number of vocalizations estimated by a trained professional and the estimate provided by my recognizer could have a number of causes. One such cause is that Song Scope often treats overlapping calls in a manner that over estimates the number of vocalizations present in a given file (Figure 2.4). A simple overlap of two Houston Toad calls can result in three individual hits by the recognizer. The recognizer will identify the singular vocalization, make a second identification for the portion of vocalization that is overlapping, and a third identification will take place once one of the two toads ceases to call. Houston Toads are known to call in large groups, making this type of overlap in vocalizations common among the digital audio files collated for this study.

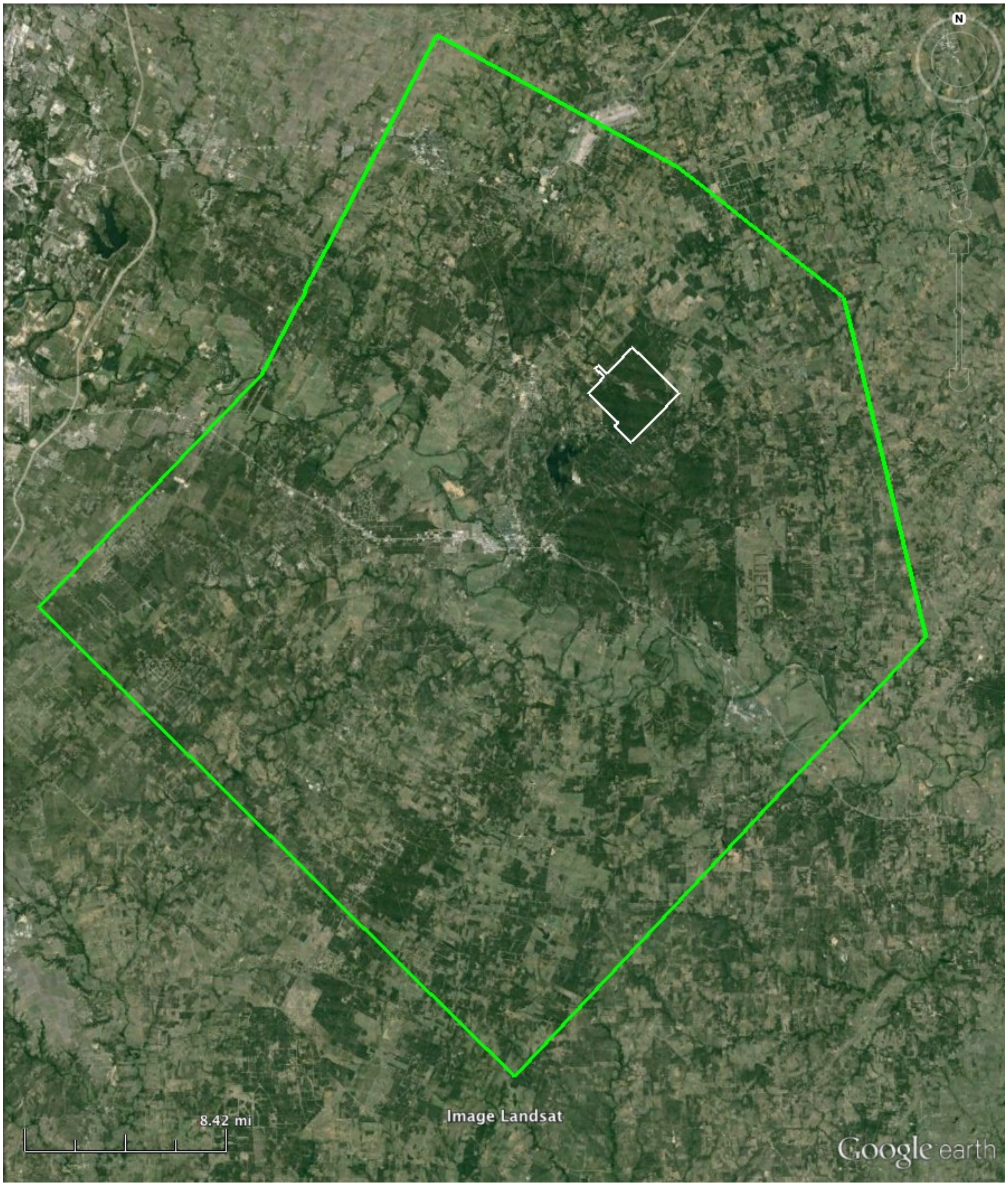

Figure 2.1 Map of Bastrop County, Texas, highlighting location of the Griffith League Ranch. The Griffith League Ranch bordered in white, in the context of Bastrop County, bordered in green.

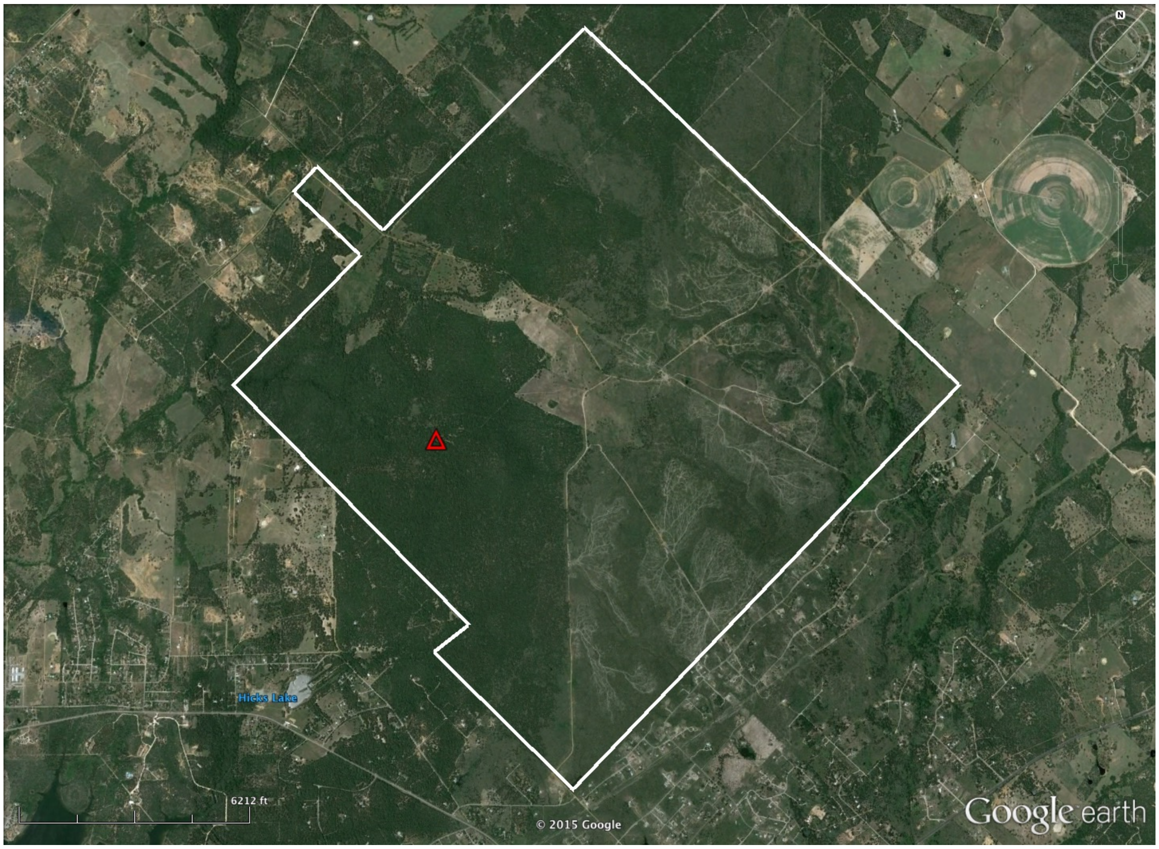

Figure 2.2 "Pond 12" of the Griffith League Ranch in Bastrop County, Texas (in context). Illustrating the location of the pond in the portion of the ranch that remains dominated by loblolly pine (*Pinus taeda*) in context with the main ranch road. Red triangle indicates location of the pond. Property boundary of the Griffith League Ranch bordered in white.

Figure 2.3 "Pond 12" of the Griffith League Ranch in Bastrop County, Texas (aerial). Aerial photograph of Pond 12 of the Griffith League Ranch. Red triangle indicates location of the Song Meter used to monitor breeding activity of *Bufo houstonensis.*

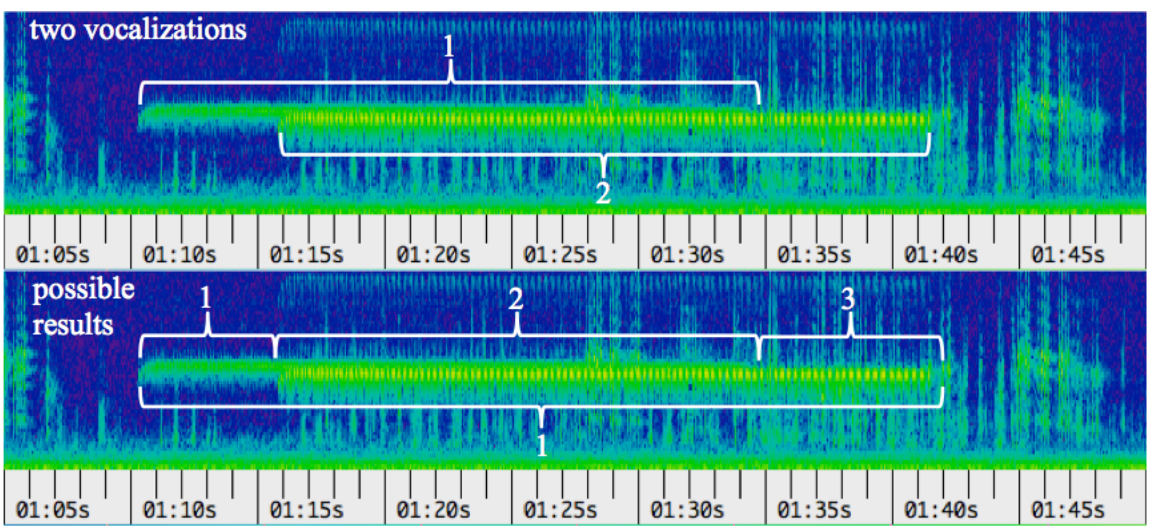

Figure 2.4 Spectrograph of two overlapping Houston Toad (*Bufo houstonensis)* calls. Spectrographic view of two overlapped Houston Toad vocalizations shown on top. Below are illustrations of how Song Scope may misrepresent the number of vocalizations by inaccurately identifying overlapping calls.

CHAPTER III

CHARACTERIZING THE CALLING ACTIVITY OF THE HOUSTON TOAD GRAPHICALLY

Introduction

One of the many factors that make the Houston Toad a challenge to research is that the male Houston Toads mating call has never been described or summarized in a single document. The USFWS' protocol for nocturnal audio surveys monitoring the breeding activity of Houston Toads includes information dating back to the 1960's (Kennedy 1962). Over 50 years of ingenuity and invention have provided researchers with advanced tools and techniques that could help reexamine the conditions in which *Bufo houstonensis* perform their mating call, as evident in Chapters I and II.

Audio recording devices (ARD) and portable remote logger devices used for monitoring environmental conditions have only become commercially available to researchers within the last decade. With the advent of these technologies site specific microclimate changes can be tracked along with calling patterns for anurans, shedding new light on what cues trigger species to vocalize. The monitoring of local weather conditions has also changed over time. By way of the internet researchers can now receive up-to-the-minute updates on weather conditions. The resolution and accuracy of these measurements did not exist to the previous generation of researchers, and thus offers an advantage to the current generation of researchers. Furthermore, it warrants investigating former conclusions about environmental conditions that influence the behavior of anurans.

In this chapter I will illustrate the observed calling activity of the Houston Toad (*Bufo houstonensis*) using information gathered from ARD's, temperature and humidity loggers, and online databases, using modern techniques of visualization. To investigate which environmental parameters influence Houston Toad breeding events I will employ multivariate Principal Components Analysis.

Methods

For this study I used 20 of the 35 locations monitored for anuran activity during 2014 in Bastrop and Robertson Counties, Texas. Eleven locations in Bastrop County were used, this includes all locations monitored via ARD's in 2014. Eleven locations in Robertson County were considered for this study, however two locations were dropped due to extended periods of data loss due to battery failure. All of the ARD's from Robertson County used in this study were located on a single tract of land that was found to contain a small population of Houston Toads. Digital audio files from these 20 ARDs were scanned using the *Bufo houstonensis* recognizer developed in Chapter I of this thesis. Temperature and relative humidity were collected using iButton Hygrochrons (Maxim Integrated, San Jose, California). A hygrochron was attached to the bottom of each ARD using adhesive backed Velcro to allow for ease of access and removal during data downloads. Hygrochrons were set to log environmental conditions in one-hour increments. Hourly data for wind speed, barometric pressure at sea level, precipitation, and moon illumination were obtained from the nearest weather station available from Weather Underground (Weather Underground, San Francisco, CA).

Program R (R-project: Gentleman 2009) was used to conduct a Principal

Components Analysis of the environmental covariates that may influence the calling period of the Houston Toad using data from the total 20 ARD's from both Robertson and Bastrop Counties. Parameters estimated in this analyses included date (1 to 32), hour of night (18:00 to 05:00; overnight), temperature (°C), percent relative humidity, percent moon illumination, cumulative precipitation (in.) (previous 24 hours), sea level pressure (in.), and wind speed (mph). Presence/absence was used categorically to illustrate the influence of the former parameters (Figure 3.3).

Results

The mating call of the Houston Toad was detected at a maximum of four of the total 11 locations monitored in Bastrop County. At least one location tested positive for Houston Toad vocalizations on 16 of 31 dates in March of 2014. The greatest number of positive surveys conducted at a single location in Bastrop County was 52 of the 372 carried out by each ARD. Figure 3.1 illustrates that in Bastrop County the Houston Toad breeding events clearly coincide with cyclical periods of low barometric pressure.

Houston Toads were detected at seven of nine locations monitored in Robertson County. At least one location tested positive for Houston Toad vocalizations on 16 of 31 nights in March of 2014. These dates differ from those in Bastrop County on two occasions; the March 23rd and 28th tested positive for Robertson, but not Bastrop, whereas March 25th and 26th tested positive for Bastrop County, but not Robertson. The greatest number of positive surveys conducted at a single location in Robertson County was 45 of the 372 carried out. Figure 3.2 illustrates the same coincident relationship between low barometric pressure and calling activity in Robertson County as found in Bastrop County.

Output of the principal components analysis indicates relationships between Houston Toad detectability (i.e. vocalizations) and high temperatures as well as low barometric pressure (Table 3.1). Temperature had a loading of 0.6267 for the first principal component; meaning as temperatures increase so does detections. Because temperature had the greatest loading (absolute) relative to other covariates one can infer that temperature had the greatest influence on detectability of Houston Toads. Barometric pressure had a loading of -0.4965, the second greatest (absolute), and thus one can infer that barometric pressure is also influencing detectability of Houston Toads. 25.23% and 16.19% of the variation present in the dataset are accounted for by the first two principal components, respectively (Table 3.1). Figure 3.3 illustrates that of the 4,464 surveys collated for this study, the 264 that tested positive for Houston Toad vocalizations are defined by a set of environmental conditions that are unique, given that portions of the confidence ellipses for Detection and Non-Detection surveys do not entirely overlap.

Discussion

The vague understanding researchers currently have of the environmental conditions required to evoke the call of the Houston Toad has led to the necessity of advanced and more resolute methods of characterizing breeding events. With the aid of ARD's, compact remote logging devices, and the vast network of online resources I was able to produce concise yet exhaustively detailed graphical representations of the conditions in which male Houston Toads are driven to perform their mating call (Figures 3.1 and 3.2). By viewing the subtle shifts in environmental conditions in sync with the breeding activity of Houston Toads we are able to see relationships that have been overlooked or undiscovered in the past, such as the impact barometric pressure may have on the Houston Toads calling activity. By running principal components analysis on all survey data collated I was able to validate and rank the influence of each environmental covariate investigated.

In regards to the USWFS' protocol for Houston Toad call surveys, I offer a comparison between the conditions in which calls are prescribed and the conditions in which ARD's recorded Houston Toads calling (Table 3.2). Based on this comparison researchers should be concerned that the currently defined (USFWS 2007) protocol for Houston Toad call surveys excludes instances when calling could very well be taking place. My findings violate the assumptions of the given protocol for temperature, humidity, moon illumination, and precipitation (Table 3.2). Furthermore, the USFWS does not acknowledge barometric pressure as an influential covariate. It is time to reexamine the context in which Houston Toads exist currently given the effects of our changing climate, the degradation and fragmentation of critical habitat, and the advent of more advanced methods of monitoring anuran populations.

Table 3.1 Variable loadings for principal components analysis of Houston Toad (*Bufo houstonensis*) calling activity. Summary of the variable loadings and variance explained by the first two principal components from principal components analysis of environmental covariates influencing Houston Toad calling activity. Loadings for Temperature and Barometric pressure (in bold) were shown to have the greatest influence on calling activity of the Houston Toad.

	Principal components	
Parameter	Component 1	Component 2
Temperature	**0.6267**	0.0652
Wind speed	0.2423	0.3156
Moon illumination	0.1905	-0.1376
Hour	-0.0054	0.076
Cumulative precipitation	-0.113	0.6588
Date	-0.3354	0.0434
Relative humidity	-0.3748	0.5283
Barometric pressure	**-0.4965**	-0.3954
Standard deviation	1.4207	1.1381
Variance explained	0.2523	0.1619
Cumulative proportion	0.2523	0.4142

Table 3.2 Comparison of observed conditions of Houston Toad (*Bufo houstonensis*) chorusing to USFWS protocol. Summary of observed environmental conditions in which Houston Toads were detected as compared to the conditions required by the USFWS for performing call surveys.

Parameter	Observed Phenology			USFWS
	min	mean	max	
Temperature (°C)	11.73	18.06	25.95	>14.0
Relative Humidity (%)	37.75	88.09	102.1	>70.0
Moon Illumination (%)	0	74.16	100	"dark"
Cumulative Precipitation (in.)	0	0.036	0.27	"recent"
Barometric Pressure (in.)	29.61	29.85	30.26	---
Wind Speed (mph)	0	6.505	15	<15.0

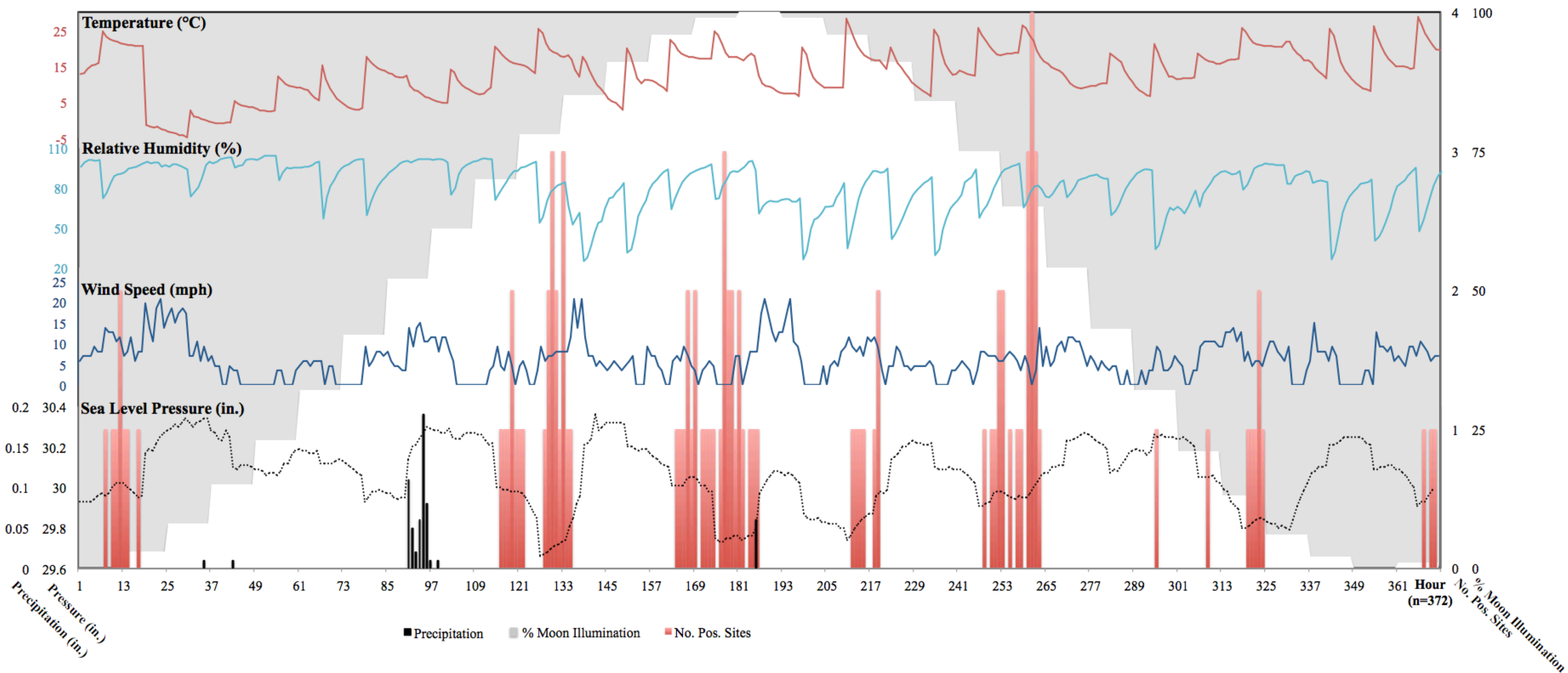

Figure 3.1 Graph of Houston Toad (*Bufo houstonensis*) detection and environmental covariates, March 2014, Bastrop County, Texas, USA. This graph summarizes the environmental covariates contributing to occurrences of breeding vocalizations 11 monitored breeding locations in Bastrop County, Texas. Covariates include temperature, relative humidity, sea level pressure, cumulative precipitation, and moon illumination. Also included in the figure are the number of sites where Houston Toads were detected at each given hour.

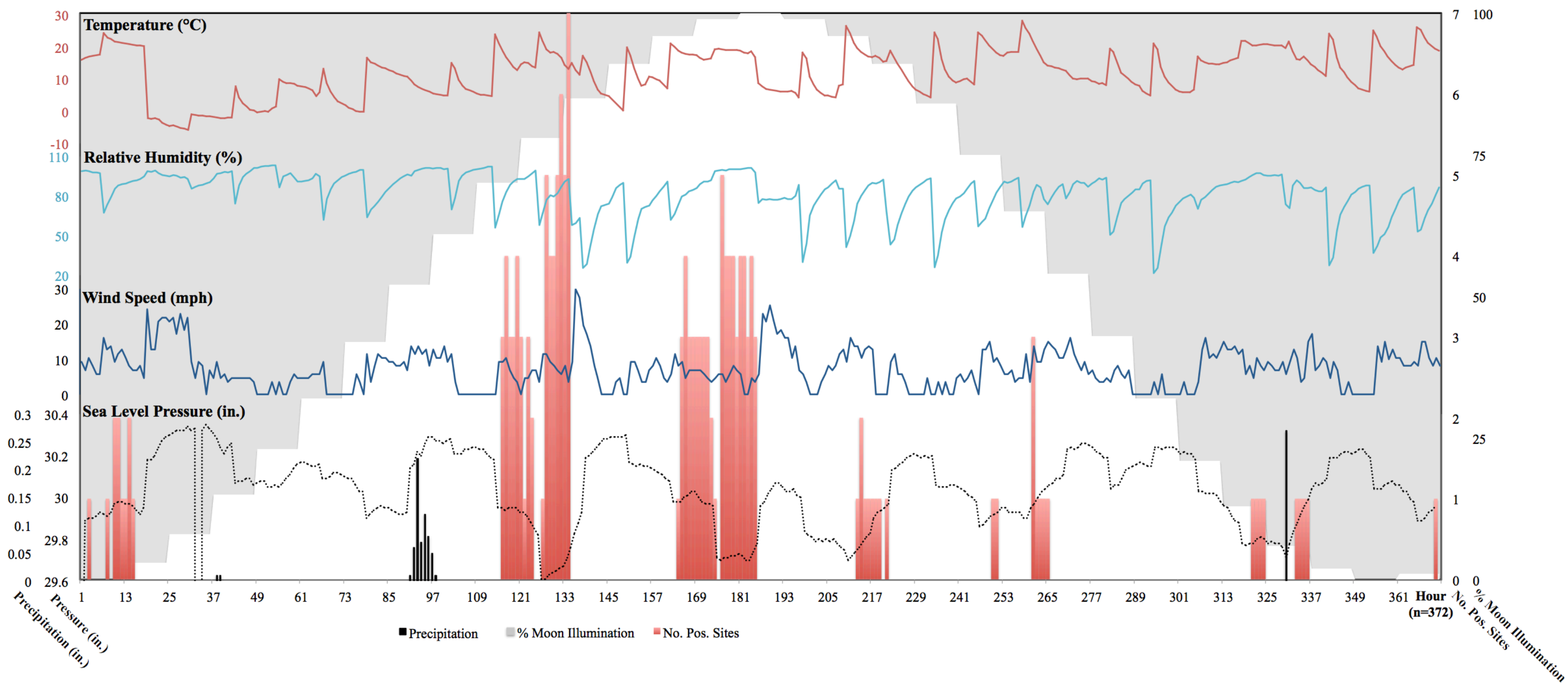

Figure 3.2 Graph of Houston Toad (*Bufo* houstonensis) detection and environmental covariates, March 2014, Robertson County, Texas, USA. This graph summarizes the environmental covariates contributing to occurrences of breeding vocalizations 9 monitored breeding locations in Bastrop County, Texas. Covariates include temperature, relative humidity, sea level pressure, cumulative precipitation, and moon illumination. Also included in the figure are the number of sites where Houston Toads were detected at each given hour.

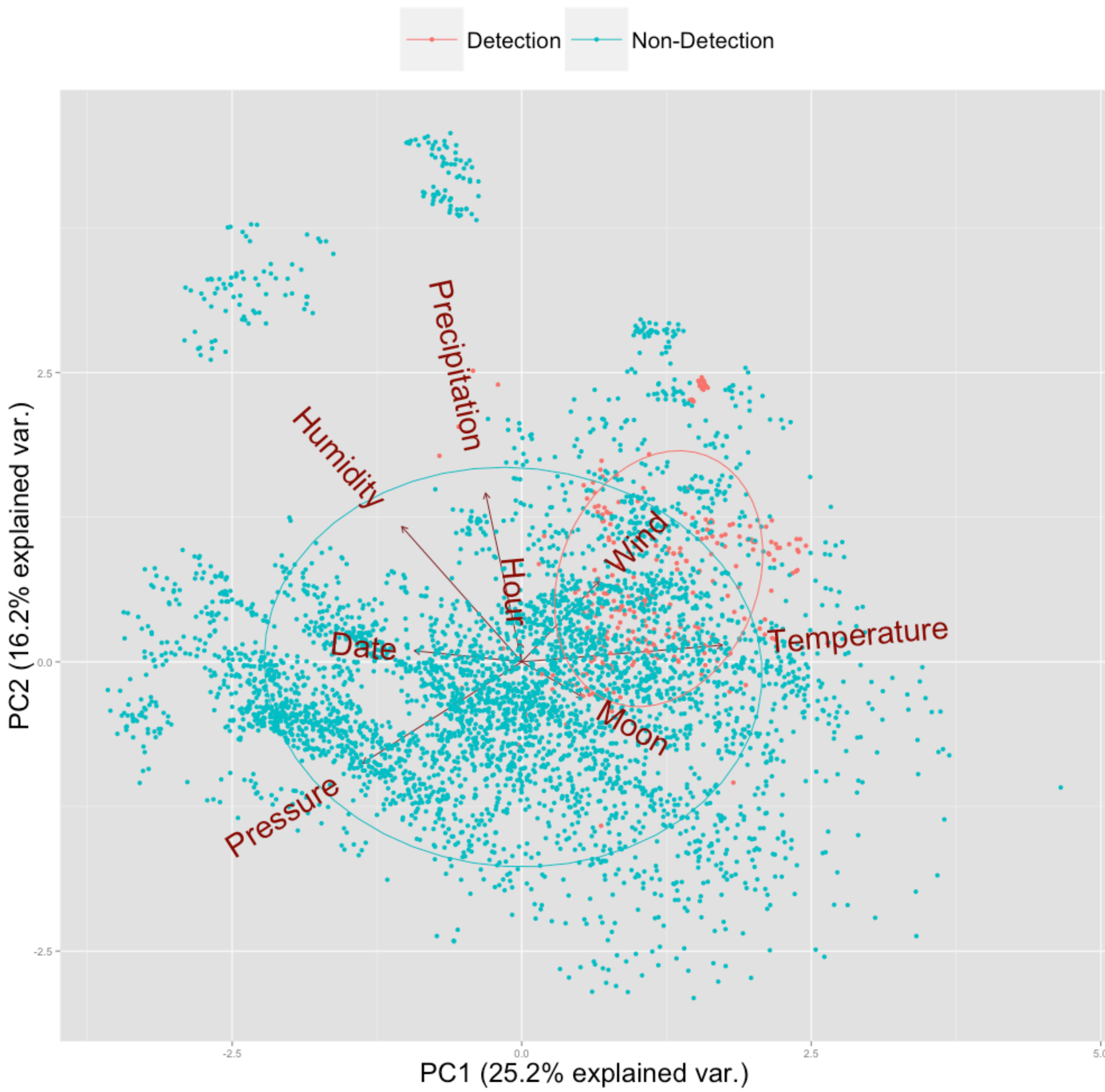

Figure 3.3 Biplot of environmental covariates and their effects on Houston Toad (*Bufo* houstonensis) detectability. The first two principal components, each explaining 25.23% and 16.19% of the variation present respectively, summarizing the influence of date, hour of night, temperature, relative humidity, moon illumination, cumulative precipitation, barometric pressure, and wind speed on the detection of Houston Toads. Factors that appear to have the greatest influence are temperature and barometric pressure.

CHAPTER IV

CONCLUSION

Machine-Learning techniques of pattern recognition have not been exhaustively studied for their efficacy or utility in identifying the unique vocalizations of male anurans. This approach has true potential to enable data to better understand and potentially address anuran populations that are experiencing declines. Wide application could even provide insights to the underlying commonality among species of anuran now recognized to part of global amphibian declines. This thesis outlines the advancements that have been made in applying these methods to anuran species, and also illustrates my personal efforts to utilize and analyze these methods for their use in the detection of the threatened and endangered Houston Toad. Exhaustive efforts were taken to produce a recognition tool capable of identifying the call of the Houston Toad. Through this process an iterative methodology for development and optimization of recognition tools using software Song Scope was defined. To insure that this tools performance was adequate its ability to locate the call of the Houston Toad were compared to a trained professional. To investigate the potential application of the recognition tool, it was also used in part to build robust and cogent visualizations of the calling activity of the Houston Toad.

In order to better serve the scientific community and continue to make significant contributions to our collective breadth of knowledge researchers must work towards innovation. The fate of current threatened and endangered anurans will be determined by the limitations researchers face when trying to conserve species. Machine-learning methods of pattern recognition for the purpose of identifying anuran vocalizations have the potential to aid researchers in an impactful way. Researchers have only recently discovered the advantages that these techniques can provide. These devices give researchers the ability to capture audio from virtually any place, from the deep ocean to the tropical upper canopies. Application software suites like Song Scope offer researchers ways of analyzing the forms of life occupying these audio recordings at a rate that is incomparable to traditional manual methods of detection.

If we are to conserve what species are left, we must embrace these innovations and perfect them. Future research must show that these methods of detection are capable of providing insights beyond those realized from decades of traditional manual methods of detection. As researchers we must improve these digitized techniques such that we can extract information on population health, diversity, abundance, and more, from these digital audio surveys. Tools such as this are the future of acoustic monitoring efforts for all forms of life, and improving upon them only serves to better our understanding of the life that surrounds us, as well as enabling us to better protect it.

LITERATURE CITED

Barclay, R. M. (1999). Bats are not birds—a cautionary note on using echolocation calls to identify bats: a comment. Journal of Mammalogy, 80(1), 290-296.

Bioacoustics Research Program. (2014). Raven Pro: Interactive Sound Analysis Software (Version 1.5) [Computer software]. Ithaca, NY: The Cornell Lab of Ornithology. Available from http://www.birds.cornell.edu/raven.

Brandes, T. S., Naskrecki, P., & Figueroa, H. K. (2006). Using image processing to detect and classify narrow-band cricket and frog calls. The Journal of the Acoustical Society of America, 120(5), 2950-2957.

Bridges, A. S., & Dorcas, M. E. (2009). Temporal variation in anuran calling behavior: implications for surveys and monitoring programs. Copeia, 2000(2), 587-592.

Brown, D. J., Swannack, T. M., & Forstner, M. R. J. (2015) Using calling activity to predict calling activity: A case study with the endangered Houston Toad (*Bufo [Anaxyrus] houstonensis*). The Journal of North American Herpetology, Volume 2015(1): 12-16.

Brown, L. E. (1971). Natural hybridization and trend toward extinction in some relict Texas toad populations. The Southwestern Naturalist, 16(2), 185-199.

Capranica, R. R. (1965). The evoked vocal response of the bullfrog; A study of communication by sound. Massachusetts Inst. of Tech. Cambridge Research Lab of Electronics, (No. STR-12).

Corben, C., & O'Farrell M. J. (1999). Anabat system user's guide. Anabat system manual, 2nd ed.

Crouch, W. B., and P. W. Paton. (2002). Assessing the use of call surveys to monitor breeding anurans in Rhode Island. Journal of Herpetology 36:185–192.

Eldridge, J. D. (2011). A comparison of current anuran monitoring methods with emphasis on the accuracy of automatic vocalization detection software. M.S. Thesis. Western Kentucky University, Bowling Green, Kentucky, USA. 122 pp..

Forstner, M. R J. and T. M. Swannack. (2004). The Houston toad in context. Final project report submitted to TPWD/USFWS. U. S. Fish and Wildlife Service. Austin, Texas, USA.

Gentleman, R., Ihaka, R., & Bates, D. (2009). The R project for statistical computing. URL: http://www. r-project. org/254.

Gibbs, J.P., K.K. Whiteleather, and F.W. Schueler. (2005). Changes in frog and toad populations over 30 years in New York state. Ecological Applications 15(4), 1148– 1157.

Hammerson, G. and D. Shepard (2004). *Anaxyrus houstonensis*. The IUCN red list of threatened species. Version 2014.3. <www.iucnredlist.org>.

Hsu, M.Y ., Y .C. Kam, and G.M. Fellers. (2005). Effectiveness of amphibian monitoring techniques in a Tiawanese subtropical forest. Herpetological Journal 15, 73–79.

Jackson, J. T., Weckerly, F. W., Swannack, T. M., & Forstner, M. R. (2006). Inferring absence of Houston toads given imperfect detection probabilities. Journal of Wildlife Management, 70(5), 1461-1463.

Kennedy, J. P. (1962). Spawning season and experimental hybridization of the Houston toad, *Bufo houstonensis*. Herpetologica, 17(4), 239-245.

Parris, K. M., Norton, T. W., & Cunningham, R. B. (1999). A comparison of techniques for sampling amphibians in the forests of south-east Queensland, Australia. Herpetologica, 55(2), 271-283.

Pechmann, J. H. K., D. E. Scott, R. D. Semlitsch, J. P. Caldwell, L. J. Vitt, And J. W. Gibbons. (1991). Declining amphibian populations: the problem of separating human impacts from natural fluctuations. Science 253, 892–895.

Phillips, K. (1990). Where have all the frogs and toads gone? Bioscience 40(6), 422–424.

Pierce, B., and K. Gutzweiller. (2004). Auditory sampling of frogs: detection efficiency in relation to survey duration. Journal of Herpetology 38(4), 495–500.

Sanders, O. (1953). A new species of toad, with a discussion of morphology of the bufonid skull. Herpetologica 9(1), 25-47.

Schmidt, B. R. (2003). Count data, detection probabilities, and the demography, dynamics, distribution, and decline of amphibians. Comptes Rendus Biologies 326(1), 119–124.

Stuart, S. N., J. S. Chanson, N. A. Cox, B. E. Young, A. S. L. Rodrigues, D. L. Fischman, And R. W. Waller. 2004. Status and trends of amphibian declines and extinctions worldwide. Science 306:1783– 1786.

Taylor, A., G. Watson, G. Grigg, and H. McCallum. (1996). Monitoring frog communities: an application of machine learning. Proceedings of the 8th Innovative Applications of Artificial Intelligence Conference, Portland, Oregon, USA, 1564–1569.

Towsey, M., Planitz, B., Nantes, A., Wimmer, J., & Roe, P. (2012). A toolbox for animal call recognition. Bioacoustics, 21(2), 107-125.

U.S. Fish and Wildlife Service, U.S.F.W.S. (1984). Houston Toad recovery plan. U.S. Fish and Wildlife Service, Albuquerque, N.M.

U.S. Fish and Wildlife Service, U.S.F.W.S. (2007). Section 10(a)(1)(A) Scientific permit requirements for conducting Houston Toad presence/absence surveys. U.S. Fish and Wildlife Service, Austin, Texas.

Waddle, J. H., Thigpen, T. F., & Glorioso, B. M. (2009). Efficacy of automatic vocalization recognition software for anuran monitoring. Herpetological Conservation and Biology, 4(3), 384-388.

Weir, L.A., and M.J. Mossman. (2005). North American Amphibian Monitoring Program (NAAMP). In Amphibian Declines: The Conservation Status of United States Species. University of California Press, Berkeley, California, USA. 307– 313.

Zimmerman, B.L. (1994). Audio strip transects. In measuring and monitoring biological diversity: standard methods for amphibians. Smithsonian Institution Press, Washington, D.C., USA. 92–97.