RE-EVALUATING THE REPRODUCTIVE ECOLOGY OF THE ENDANGERED

HOUSTON TOAD (*BUFO* [=*ANAXYRUS*] *HOUSTONENSIS*) USING AUTOMATED

AUDIO MONITORING TECHNIQUES

by

Andrew Rance MacLaren, B.S., M.S.

A dissertation submitted to the Graduate Council of
Texas State University in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
with a Major in Aquatic Resources and Integrative Biology
August 2019

Committee Members:

    Michael R.J. Forstner, Chair

    Ben M. Bolker

    Shawn F. McCracken

    J. Andrew Royle

    Floyd W. Weckerly

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

Page

CHAPTER

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The Houston Toad (*Bufo* [=*Anaxyrus*] *houstonensis*) has been a species of conservation concern for approximately a half-century, yet many aspects of its life history are not well known. Technological innovations in the form of automated recording devices and automated detection have enabled all aspects of this dissertation. Monitoring for species that are difficult to detect, or become detectable within a narrow temporal window, such as the Houston Toad, has historically been done by human observers. This dissertation seeks to illustrate that through application of technological innovations, remotely monitoring populations of Houston Toads has given us the opportunity to re-evaluate the species' life history entirely, using data that are collected more frequently, by more reliable observers: automatons.

Populations of Houston Toads have been brought to the brink of extirpation in recent years, then returned to above average local abundances through captive propagation in select localities. These changes in demography are the result of anthropogenic disturbance, above all else, and changes to the landscape continue to occur within this species' remaining critical habitat. As a part of this study I sought to include data collected throughout these shifts in demography, across landscapes ranging in their relative quality as optimal habitat for the Houston Toad, and from portions of their range in which they are rarely found.

My research has resulted in a successful long-term study of the environmental abiotic correlates to male Houston Toad chorusing behavior, providing crucial

information to researchers, biological consultants, and governing agencies about the

seasonal and diurnal behavior of this endangered species. It has re-defined the approach

human surveyors must take when attempting to detect this species' vocalization using

auditory surveys, and for the first time offers researchers an understanding of how the

length, and frequency, of these auditory surveys impacts the probability of detection. I

evaluated the sound dampening qualities of the varying microhabitats surrounding

Houston Toad breeding locations, and the influence these qualities have on probability of

detection for a variety of animal vocalizations, including the Houston Toad. Finally, I

review the Houston Toad's historical range and its relationship with its nearest

congeneric relative the Dwarf American Toad (*Bufo* [=*Anaxyrus*] *americanus*

charlesmithi) by reviewing past literature, museum vouchers, and comparing the general

morphology and vocal repertoire of each species. These data fill gaps in the present

knowledge we hold for this species, and provide a foundation for future research to base

its hypotheses upon.

I. EVALUATION OF THE PHENOLOGY AND PROXIMAL ENVIRONMENTAL CUES OF CHORUSING BEHAVIOR IN THE ENDANGERED HOUSTON TOAD USING AUTOMATED RECORDING DEVICES.

Introduction

Amphibians and reptiles are declining globally (Gibbons et al. 2000; Beebee & Griffiths, 2005; Reading et al. 2010) as a result of habitat loss or anthropogenic disturbance (Welsh 1990; Waldron et al. 2006; Knutson et al. 1999; Becker et al. 2007; Templeton et al. 2001), interactions with invasive species (Riley et al. 2003; Allen et al. 2004), changes in weather and climate (Daszak et al. 2005; McMenamin et al. 2008), infection and widespread disease (Laurance et al. 1996; Briggs et al. 2005). For these reasons interest in improving amphibian monitoring programs, such as the Amphibian Research and Monitoring Initiative (ARMI; Adams et al. 2013) and the North American Amphibian Monitoring Program (NAAMP; Weir et al. 2009), has become a popular topic in conservation and management research. Research on amphibian activity patterns is important for maximizing detection (Steelman & Dorcas 2010; Cook et al. 2011), and thus data integrity (e.g., inferences about trends based on surveys). Towards these goals, the long-term monitoring of anuran populations is often carried out using automated recording devices (ARDs, hereafter; Oseen & Wassersug 2002; Saenz et al. 2006; Aide et al. 2013). This technology remotely collects audio of male anurans vocalizing to attract females during their reproductive season (Bridges & Dorcas 2000). ARDs allow researchers to apply immense survey effort, despite environments that may be inhospitable and difficult to access, while avoiding bias from observer disturbance to

vocalizing animals (Alldredge et al. 2007; Hutto & Stutzman 2009). As an alternative to manual human performed call surveys, ARDs may be more affordable, depending on the quality and number of capabilities researchers desire (Charif & Pitzrick 2008; Rempel et al. 2013). In addition to these advantages ARDs also avoid heterogeneity among observers, as is often found among researchers of varying experience (Miller et al. 2012).

The Houston Toad (*Bufo* [=*Anaxyrus*] *houstonensis*) is an explosive breeder (Price & Yantis 1990; 1992; 1993; Price 2003), meaning that choruses form for several nights and then disappear for intervals of several days to a week or more (Wells 1977). Chorusing is reported to initiate as early as January (Hillis et al. 1984; Jacobson 1989), or in some cases as late as March (Brown 1971) and cease as early as April (Price 2003) or May (Hillis et al. 1984; Jacobson 1989), although chorusing has been reported to end as late as June (Kennedy 1962). Precipitation is often associated with breeding activity among amphibians (Salthe & Mecham 1974), and it has been suggested that this is true among Houston Toads (Kennedy 1962). Subsequent studies reject this hypothesis, and report that chorusing behavior is most influenced by air temperature. For example, Hillis et al. (1984) and Jacobson (1989) state that minimum air temperature does not fall below 14 °C for the 24-hour period preceding chorusing, but also noted that this relationship did not account for asynchronous chorusing periods among simultaneously monitored sites. This value, treated as 57 °F, is incorporated into the United States Fish and Wildlife Service (USFWS, hereafter) guidelines for performing surveys aimed at determining absence of Houston Toad from a given location (USFWS 2007). Dixon et al. (1990) reports on a single year of Houston Toad activity at two ponds within Bastrop State Park, a historical stronghold for the Houston Toad, that were intensely surveyed (n=60 surveys

each). Their notable findings were that the 24-hour minimum air temperature wherein chorusing was observed was as low as 7 °C, which was later confirmed by Brown et al. (2013), and that toads prefer that the moon not be more than 50% illuminated during chorusing events. Dixon et al. (1990) also suggests that maximum air temperature in a 24-hour period may have an inhibitory effect late in the season, but the magnitude of this threshold is not evaluated or defined. Additionally, Price (2003) reports the chorusing requirements for Houston Toads as: overnight temperatures preceding chorusing > 55 °F, high relative humidity ($\geq$70%), and that cloud cover obscures the moon when illumination is approaching full. More recently, Brown et al. (2013) determined through multi-year monitoring that Houston Toad calling activity was influenced by daily average wind speed, daily average absolute humidity, and whether Houston Toads were heard calling within the previous 24 hours. These findings contradict those of Dixon et al. (1990) and Price (2003) regarding moon illumination. It is important to point out that each of these previous studies was carried out using human performed call surveys, and that each are testing for the detectability of Houston Toads rather than actual call occurrence. For example, the findings that increased wind speeds reduce the probability of chorusing may be a consequence of the effect of wind speed on the ability of human surveyors to hear animals vocalizing, rather than increased risk of desiccation as has been proposed in the past (Dixon et al. 1990; Brown et al. 2013).

Despite an apparent and maintained interest in detailing male Houston Toad chorusing behavior, and its causes, no previous study spans the entirety of this species' active season above ground. Rather, most focus on the perceived peak of breeding (i.e. chorusing) activity (Brown et al. 2013; Jackson et al. 2006). When, how, and why human

observers choose to monitor for Houston Toad chorusing has varied considerably through time, and the conflicting results of past studies exemplifies this. Many of the assumptions regarding the natural history of the Houston Toad are founded in documents that provide little or no statistical evidence for their claims. This is often true of gray literature (e.g., reports to state or federal agency), which lacks any peer-review.

We determine the range of environmental variables that could possibly influence calling behavior in Houston Toads, and discuss whether these conditions alone elicit vocalization or whether there is a calendar window in which the optimal environmental conditions must be met. That is to say, conditions seemingly favorable may take place days, weeks, or months prior to and after chorusing behavior takes place, yet the animal exhibits no response as a consequence of the day of year (i.e. circadian rhythm), as has been documented in other species of anuran (Canavero et al. 2008). We used eight years of call survey data collected through passive automated methods to build statistical models for Houston Toad chorusing activity across two Texas counties. Our findings will be useful for optimization of future Houston Toad monitoring efforts, highlighting factors in need of continued research and improving our general understanding of the reproductive ecology and behavior of this endangered species.

Methods

Study Area

This study includes only sites in which Houston Toads were detected, although depending on the year, we monitored a total number of sites ranging between 10 and 80.

Houston Toads were found within Bastrop and Robertson County, Texas, USA. This study includes 22 sites within Bastrop County, and 6 within Robertson County. Sites within Bastrop County include mature forests dominated by loblolly pine (*Pinus taeda*), post oak (*Quercus stellate*), blackjack oak (*Quercus marilandica*), and eastern red cedar (*Juniperus virginiana*), whereas Robertson County may also include stands dominated by bald cypress (*Taxodium distichum*), each of these forests with a dense understory consisting of yaupon holly (*Ilex vomitoria*), American beautyberry (*Callicarpa americana*), and farkleberry (*Vaccinium arboretum*). Much of the area monitored within Bastrop County suffered a high-intensity wildfire during 2011 (Brown et al. 2014).

Audio Recording

This study includes data collected between 2010 and 2018. No data were collected in 2012 as a result of the catastrophic Bastrop County Complex Fire which occurred in the fall of 2011, affecting much of our study area. At the beginning of each season we deployed SongMeter audio recording devices (ARDs; models SM2+, SM3, SM4; Wildlife Acoustics, Maynard MA, USA) to potential breeding locations for the Houston Toad. In an effort to ensure that the entirety of the active period for this species was recorded, deployment took place before or within the first week of January. However, during the 2013 spring breeding season ARDs were deployed on March 1, due to alternative research priorities or obligations. ARDs were retrieved from the field within July of each year. Each ARD was attached to a structure object within 10 meters of the edge of each water body. Recording schedules for each ARD varied across years, but consisted of four generally applied approaches; 1) record 10 minutes at the top of

each hour beginning at 1700 ending 0610 (14 per night); 2) record 10 minutes at the top of each hour beginning at 1800 ending 0510 (12 per night); 3) record continuously for 12 hours beginning at 1800 ending 0600; 4) record continuously, without pause. ARDs were powered using either rechargeable D-cell batteries (60 Wh; Tenergy, Fremont, California), conventional D-cell batteries (72Wh; Duracell, Bethel, CT, USA), or rechargeable sealed lead acid batteries (Power Sonic PS-6360 NB, 6V, 36.0 AH). Sites equipped with ARDs scheduled to record continuously were left in the field throughout all of 2017 to ensure chorusing was not taking place during the fall and winter months. Each ARD was stocked with two or more SD cards for storage of digital media. We did not deviate from the manufacturer's default settings for audio recordings with the exception of sample rate, which we lowered to 16,000 Hertz. This decreases the maximum frequency recorded, eliminating only ultrasonic frequencies not needed for analysis of most North American anuran vocalizations (Narins et al. 2004). This helped reduce overall file size, and thus extended the length we could allow ARDs to record without required field visits to each site. ARDs were checked for battery life approximately every 40 days. During these visits, removable digital media (i.e. SD cards) were retrieved and replaced. SongMeters do not require that they be fitted with previously formatted or erased cards because they feature on board software capable of completing this task.

Automated Detection

To analyze the large quantity of audio files collected, we trained an audio classifier using the software Kaleidoscope (Wildlife Acoustics). We followed the steps

outlined by the manufacturer for completing this process (Wildlife Acoustics 2017) and utilized the audio training data provided by MacLaren et al. (2018A) for the call of the Houston Toad. We chose to simply train towards two "clusters": Houston Toad vocalizations, and anything that is not a Houston Toad vocalization. This was efficient, and we achieved 100% detection of positive training data into the Houston Toad cluster within a single round of training, indicating that this recognizer returns near-perfect detection. Initially, all detections made by the software > 3 seconds in length were manually verified by A.R. MacLaren. We observed that detections below this length were mostly false positives. To ensure near perfect detection, we then sorted detections by date, taking care to search all detections on dates in which Houston Toads had not yet been detected. Houston Toad vocalizations can often overlap and may be classified in a single detection or multiple detections within Kaleidoscope (MacLaren et al. 2018A). Therefore, we refer only to 'detections' hereafter to differentiate between true vocalizations and vocalizations estimated via Kaleidoscope. Each detection is timestamped by the software, which enabled us to bin detections by hour in which they occurred. This resulted in a count of the number of detections per each hour at each site.

Abiotic Variables

Environmental variables for this study were retrieved the National Oceanic and Atmospheric Administration's (NOAA) "Quality Controlled Local Climatic Dataset" (QCLCD), measured at Giddings, Texas, USA (WBAN 53979), and Hearne, Texas, USA (WBAN 53973), for Bastrop and Robertson Counties, respectively. The QCLCD was decommissioned by NOAA at the end of 2017. Consequently, we used NOAA's

"hourlies" collated for each previously mentioned weather station, which are the origin

for the QCLCD, but have simply not been filtered by NOAA. However, we did not find

any egregious errors or missingness among the "hourlies" collected for 2018. This

provided us with data for temperature (°C), relative humidity (%), wind speed (kmph),

barometric pressure (mmHg at sea level), and daily total precipitation (cm). From these

data we calculated the difference, or change, in barometric pressure over 12 and 24 hours

previous. Absolute humidity is suggested as a preferred metric for air moisture in studies

addressing amphibian activity, due to relative humidity's dependence on air temperature

(Hillman et al. 2009). Thus, we calculated absolute humidity (g/m^3) from the temperature

and relative humidity provided, and compared these as predictors of chorusing. We

retrieved data for moon illumination (i.e. fraction of the moon illuminated) measured by

the U.S. naval observatory (USNO) for Central Time Zones. In addition to these

atmospheric conditions we also collated Julian day, hour (ranging 1 - 24), year

(categorical), and site name (categorical), corresponding to each unique hour/site/year

group of detections.

Statistical Models

     We conducted graphical data exploration to check for normality, homogeneity,

and collinearity of explanatory variables prior to performing any statistical analyses

(Bolker et al. 2009; Zuur et al. 2010). We performed all analyses within program R (R

core team 2018) primarily using the package "glmmTMB" (Brooks et al. 2017). Package

"glmmTMB" was used to fit generalized linear mixed models (GLMMs, hereafter) with

the Laplace approximation. GLMMs are ideal for ecological studies involving

presence/absence or count data (i.e. non-normal data) because they allow researchers to test for the appropriate error distribution, while implementing random effects (Bolker et al. 2009). Incorporating random effects into the models accounts for the potential non-independence of subsampled data points in a nested design; and GLMMs fit with a Poisson distribution and individual-level random effect or negative binomial distribution without an individual-level random effect allow for overdispersion (Bolker et al. 2009). The R package described by Brooks et al. (2017) also allows for offsets to be incorporated, allowing for greater flexibility concerning unbalanced sampling design. ARD's operate for differing lengths of time within each day. For example, many are programmed to record 10 minutes of each hour, and others record for the entirety of the hour. We calculated the proportion of each hour recorded and used that as an offset in our model, assuming that counts will be expressed proportionally to the amount of time that a recording device is active. Among all years and sites, the first date Houston Toads were detected was Jan 20 (i.e. Julian day 20) and the latest date was June 6 (i.e. Julian day 157). The seasonal timing of emerging from and returning to fossorial behavior is not well known for this species. Thus, in an effort to exclude periods where adult Houston Toads may cease to be active above ground, we restricted our analysis to exclude days prior to Julian day 20, and beyond Julian day 157.

We used generalized linear mixed models (GLMMs) to test for the influence of seasonal, diel, and environmental abiotic variables on abundance of detected Houston Toad vocalizations, using a Poisson and Negative Binomial error distribution, with a log link function. These distribution families, and their variants, are best suited for data organized by count (Ver Hoef & Boveng 2007; Lindén & Mäntyniemi 2011). Prior to

fitting any models, we standardized (e.g. scaled and centered) all continuous variables by subtracting the mean and dividing by the standard deviation. The R package "glmmTMB" provides the ability to fit Poisson and negative binomial error distributions with or without zero-inflation. First, we fit a fully saturated model (i.e. containing all fixed predictors and random effects) and then performed model reduction using the Poisson, zero-inflated Poisson, type 1 negative binomial (e.g. linear mean-variance relationship), zero-inflated NB1, type 2 negative binomial (e.g. quadratic mean-variance relationship), and zero-inflated NB2, error distributions. We included the following seasonal, diel, and environmental variables as fixed effects: Julian day, the quadratic effect of Julian day (i.e. day^2), hour, the quadratic effect of hour (i.e. $hour^2$), temperature, the quadratic effect of temperature (i.e. $temperature^2$), humidity (relative or absolute), wind speed, moon illumination, barometric pressure, change in pressure over 12 hours previous, change in pressure over 24 hours previous, and daily total precipitation.

Houston Toads breed in large aggregations (Gaston et al. 2010) and may continue to chorus despite unfavorable atmospheric conditions or time of day, due to the influence of competition on their behavior. To account for this effect, we included the fixed predictor of count of detections in previous one hour (Lag Call, hereafter). Not all sites were monitored in all years, and effort at each site was not constant across years (i.e. date of deployment or retrieval of ARDs varied). For these reasons, we treated year as a random effect in each model we fit. Unlike previous studies of Houston Toad chorusing behavior, we are not primarily interested in assessing the probability of occupancy at sites where Houston Toads have not been detected (Brown et al. 2013; Jackson et al. 2006). Alternatively, we seek to refine knowledge concerning the atmospheric and seasonal

requirements for chorusing behavior within this species with greater detail than has been achieved in the past. For these reasons, this study includes only sites where Houston Toads were detected. We assume that sites which received monitoring via ARD, but produced no Houston Toad detections were not unoccupied due to seasonal or atmospheric effects, rather they occur in unsuitable habitat or are experiencing a local extirpation of the species. As stated previously not all sites are represented in each year, which may reflect a lack of monitoring effort as well as these local extirpation events. Because we cannot differentiate between these influences, and to address the potential non-independence (i.e. spatial correlation) of breeding locations sampled in each year, sites were treated as a random effect in each model fit.

We began the analysis with a full model containing all fixed and random effects. We first identified the best-fit error distribution of the full model using the Akaike Information Criterion (AIC, hereafter) before proceeding (Burnham and Anderson 2002). We carried out model reduction by first removing non-significant main fixed effects using the AIC to determine best-fit models (Burnham & Anderson 2002). Only models within ΔAIC≤2 were considered similar to best fit models for support of the data and reported herein (Burnham & Anderson 2002). Significance tests for fixed effects and their interactions were done using Wald *Z*-tests (Bolker et al. 2009). During model reduction, we first allowed the package "glmmTMB" to match predictors of structural zeros (i.e. zero-inflation) to those in the conditional model (i.e. counts as response variable), then attempted to improve global model fit by removing predictors from the zero-inflated and conditional models independently.

Results

Across all years and sites we collected 144,790 hourly counts that were ultimately utilized in this study. Of these, 2,269 (1.57%) contained Houston Toad detections. Only a single site possessed Houston Toads in all years wherein monitoring took place, and the maximum number of sites to contain Houston Toads within a single year was 14, in the year 2018. The greatest number of years between detection events at a single site was 7 years, but it must be noted again that 2012 received no monitoring effort via ARDs (Table 1). The greatest number of detections made within a single hour, at a single site, was 316 (mean = 19.26). We observed chorusing at temperatures as low as 4.3 °C, but none when temperature exceeded 31.3 °C. Three quarters of all hours containing detections occurred when relative humidity exceeded 73%, and half occur when humidity exceeded 85%. Although rarely, we did observe chorusing during wind gusts reaching up to 29 km/h. Houston Toads called during all phases of the moon, without discrimination (min = 0, mean = 0.5, max = 1.0).

Preliminarily fitting fully saturated models under different error distributions revealed that our data were strongly overdispersed (i.e. deviance exceeds residual degrees of freedom). Ultimately, the type 2 negative binomial error distribution, with estimates of zero-inflation, best fit our dataset. The next best fit distribution, determined via AIC, was the type 1 negative binomial, with zero-inflation, however this was ranked 130.1 AIC units greater than the alternative. For all references to models hereafter, the use of the type 2 negative binomial error distribution is assumed. We found that the fully saturated model featuring relative humidity was a better fit than the same model using absolute humidity instead ($\chi^2 = 438.24$, p < 0.01). Temperature was more strongly correlated with

absolute humidity (ρ=0.714) than with relative humidity (ρ=0.433), and this may be cause for the improved model fit.

Model selection procedures resulted in five competing models (i.e. Table 2). The top four models differed only in their zero-inflated model component, and feature identical conditional model components. All fixed predictors were significant in the conditional model common among the top four best fit global models. Fixed predictors common among these top four models were as follows: the quadratic influence of Julian day, temperature, and hour; the linear influence of Julian day relative humidity, moon illumination, barometric pressure, precipitation, and the influence of calling detected within the one hour preceding each count (i.e. Lag Call). The most parsimonious (i.e. fewest predictors) model differs from the top model by 0.2 AIC units (Table 2). Within the zero-inflated model component of the most parsimonious model all fixed effects were significant (Table 3). The two competing models with smaller AIC values than the most parsimonious model contained one additional predictor each, precipitation, and moon illumination, respectively. However, in each of these models the additional predictor was not significant.

Among temporal (i.e. season and diel) predictors the quadratic influence of hour had the greatest impact (β=0.437, Z=9.759, p<0.001). Although we observed Houston Toads chorusing during daytime hours, the greatest number of detections is distributed around midnight (Figure 1). Counts increased linearly with Julian Day (β=0.125, Z=2.089, p=0.037), but we also found that the quadratic influence of Julian Day was borderline significant (β=-0.118, Z=-1.922, p=0.055), which indicates that some decrease in counts must occur as Julian Day increases (Figure 2). Temperature (quadratic) had the

greatest influence on Houston Toad chorusing (β=-0.451, Z=-6.238, p<0.001), followed

by humidity (β=0.153, Z=3.104, p=0.002). The remaining environmental predictors,

moon illumination (β=0.078, Z=2.243, p=0.025), barometric pressure (β=-0.085, Z= -

2.045, p=0.041), and precipitation (β=-0.092, Z= -3.431, p<0.001), were each significant,

but had a smaller magnitude (Table 3). The single most significant predictor within our

conditional model was the one-hour lag applied to counts (Lag Call; β=0.118, Z= 17.692,

p<0.001; Table 3). This was expected, and reinforces that conditions may deviate from

favorable and Houston Toads may continue to chorus despite this, due to strong

competitive cues. Estimates for our random effects within our conditional model

component were greater for the variation among years (variance = 0.297, standard

deviation = 0.545) than sites (variance = 0.070, standard deviation = 0.265).

Discussion

General calling information is imperative to designing successful monitoring

programs for anurans and is particularly pertinent given the potential for ongoing habitat

disturbance, and climate change to affect breeding phenology and reproductive success.

Our study is the first to examine the chorusing behavior of the Houston Toad that does

not rely solely on information collected by human observers, and is thus biased by human

observer availability as well as preconceptions concerning the Houston Toads' behavior.

For example, Houston Toad call survey data utilized by Brown et al. (2013) was collected

under monitoring guidelines outlined by Jackson et al. (2006), wherein surveys are not

randomly selected, but rather were conducted only if weather conditions were seemingly

preferable (USFWS 2007). In other words, surveys only occurred when researchers had

reason to believe Houston Toads will be chorusing. By utilizing ARDs to monitor a fraction, or the entirety, of each day, we likely captured Houston Toad chorusing that has been overlooked by previous researchers due to these preconceptions. Because ARDs passively monitor toads we can be certain that no effect from observer disturbance has influenced our estimates, or impacted their behavior. Further, our study is the first, for this species, to monitor the same location multiple times in a single evening, again highlighting the advantage ARDs present to researchers. The use of daily average metrics may serve as one potential cause for differences between our findings, which utilizes hourly metrics, and previous research (Brown et al. 2013; Dixon et al. 1990).

Our study is also the first to consider modeling the detection of Houston Toad chorusing within a single unified model at all stages. Previously, due to low sample size, lack of model fit, or limitations within statistical software, multiple seasons (i.e. years) have been treated, in part, using independent models (Brown et al. 2013; Jackson et al. 2006). By treating years as random effects within our analysis, we account for and estimate variation among years without the need to make post-hoc comparisons of multiple model estimates. We also treated sites as random factors, due to the potential lack of independence within a season. That is to say that Houston Toads may move from one pond to the other within a single season, as has been described previously (Brown et al. 2013; Dixon et al. 1990). Although we observed greater variation among years than sites, one can argue that both of these serve as a proxy for local abundance, which inherently varies across space and time. This is especially true of populations situated in a nadir of range-wide abundance in the recent past (Gaston et al. 2010; Duarte et al. 2011). Site specific abundances have been seemingly impacted, albeit positively, by population

supplementation as well in the final years included in this study, which also contribute to the estimated variation in both year, and site, random intercepts.

We found that Houston Toads are active from January to June, but that this may vary among years (Figure 2), most likely due to stochastic shifts in weather patterns, and local abundance of toads. It is clear from the data gathered for this study that peak breeding does occur between February and April, upholding much of the previous qualitative and quantitative research performed for this species. (Hillis et al.1984; Jacobson 1989; Brown et al. 2013). Our study is the first to examine whether Houston Toads are active throughout the day within these irregular intervals of reproductive behavior that occur during peak breeding season. We found Houston Toads are more active at night than during the day, but certainly call during daytime hours when multiple individuals are active on the landscape surrounding ponds or wetlands. Anecdotally, when researchers visit breeding locations during daytime hours, Houston Toads have not been observed chorusing from pond edges, rather they call from uplands, within burrows, or under cover objects. Past research has indicated that chorusing and movement of adults is initially motivated by weather, but that once active, Houston Toads may remain active for several days despite poor weather (Brown et al. 2013). In our study the strong influence of the Lag Call predictor, that is, the count of detection in the preceding one hour, indicates that this is true also of chorusing behavior within a single overnight period. This effect is not surprising given the nature of aggregative breeding that is documented within the species (Gaston et al. 2010).

We found that Houston Toad chorusing activity reached a maximum in the middle of the range of temperatures, or rather that the quadratic influence of this variable was the

strongest among all weather-related predictors. This type of relationship is not unique among anurans (Willacy et al. 2015). The lower bound of 4.3 °C that we observed is lower than that of previous studies (Brown et al. 2013; Dixon et al. 1990), and upper bound of 31.3 °C is also lower than previous reports (Dixon et al. 1990). Humidity, as has repeatedly been suggested (Price 2003; Brown et al. 2013), was influential on the chorusing behavior of the Houston Toad, and our findings are on concordance with this hypothesis. Although we did discover Houston Toads calling during periods of very low relative humidity, as we stated previously, the overwhelming majority occurred during periods of water vapor rich atmosphere. Wind speed was not retained as a reliable predictor within the conditional model component of the most parsimonious model. However, we did find a slight positive relationship between the occurrence of structural zero's and elevated wind speeds, indicating that winds contribute to events of non-detection more so than detection of chorusing.

Despite being explosive breeders, like other members of the genus *Bufo* within this region (Saenz et al. 2006), Houston Toad activity was not positively correlated with rainfall events. Rather, there was a significant, but slight, negative relationship with daily rainfall metrics. This is a markedly different result than previous studies which found no relationship, negative or otherwise, between Houston Toad behavior and rainfall events (Brown et al. 2013; Hillis et al. 1984; Jacobson 1989). Houston Toads exhibit irregular intervals of explosive breeding (Brown et al. 2013; Dixon et al. 1990), and are thus characterized as ephemeral breeders (Oseen & Wassersug 2002) which may be less responsive to weather. However, Houston Toads primarily occur on a landscape with many large permanent water bodies, and are observed utilizing these for reproduction

annually. This behavior not only enabled our study to be carried out, but may also contribute to a lack of relationship among chorusing behavior and rainfall within this species.

The marginal influence of moon illumination and barometric pressure are novel findings for this species. The influence of the moon has been all but ruled out in previous studies (Brown et al. 2013), yet it remains an integral part of the monitoring protocol for Houston Toads (USFWS 2007). The USFWS (2007) state that the moon should not be well lit, and that if it is that cloud cover must obscure this, in order for chorusing to be most likely. We did not find support for this, but rather the opposite. Our results may be a consequence of our large sample size, or statistical power. As discussed previously, summary statistics for moon illumination indicate that Houston Toads chorused under all phases, yet our model yields a slightly positive relationship, confounding what we observe within the raw data. However, no explicit measure of cloud cover was made during our study, and this is suggested by USFWS (2007) to obscure the moons illumination sufficiently enough to influence Houston Toad behavior. Similarly to our findings regarding moon illumination, precipitation is estimated to have a slight negative association with counts of detections, which is not observed within the raw data, and certainly is not in line with general amphibian biology (Oseen & Wassersug, 2002).

In conclusion, this study expands and refines the body of statistically based support for the abiotic factors influencing chorusing of the endangered Houston Toad. Brown et al. (2013) called for continued long-term monitoring of Houston Toad populations, and I believe that was achieved in this study. Further, the authors suggest studies investigate Houston Toads across their entire range. While our study also

achieves this, we were still limited in our ability to include multiple sites across multiple years from outside the most densely populated areas within their range (i.e. Bastrop, TX, USA). Although, preliminary research (MacLaren et al. 2018B), and overall trends observed throughout this study indicate that chorusing phenology, at least, is consistent throughout the remaining range for the species. Future research into the variation in Houston Toads, and their behavior, across their geographic range will be contingent upon research access on private lands.

Table 1. Matrix of sites and years wherein Houston Toads were detected. Color code indicates recording schedule type. Recording schedules consisted of: orange, 14-10 min. recordings per night; yellow, 12-10 min. recordings per night; green, 12-1 hour recordings per night; blue, continuous recording. Dashes in grey indicate no loggers were used at a site in a given year. An "X" has been placed at each site that was fully monitored but no Houston Toads were detected.

Site	Year								
	2010	2011	2012	2013	2014	2015	2016	2017	2018
BASM02	BASM02	X	--	BASM02	BASM02	X	X	X	X
BASM03	BASM03	X	--	X	BASM03	X	X	X	X
BASM05	BASM05	X	--	BASM05	BASM05	BASM05	BASM05	BASM05	BASM05
BASM06	BASM06	X	--	X	X	X	X	X	BASM06
BASM07	BASM07	BASM07	--	BASM07	BASM07	BASM07	BASM07	BASM07	BASM07
BASM08	BASM08	X	--	X	X	X	X	X	X
BASM10	X	X	--	BASM10	BASM10	X	BASM10	X	BASM10
BSP02	--	--	--	--	--	X	X	BSP02	BSP02
BSP10	--	--	--	--	--	BSP10	X	X	X
BSP18	--	--	--	--	--	X	X	BSP18	X
BSP23	--	--	--	--	--	X	X	BSP23	X
BSP27	--	--	--	--	--	X	X	BSP27	X
BSP30	--	--	--	--	--	X	X	BSP30	BSP30
GLR01	--	--	--	--	--	X	X	GLR01	GLR01
GLR04	--	--	--	--	--	X	X	GLR04	GLR04
GLR05	--	--	--	--	--	X	GLR05	GLR05	GLR05
GLR06	--	--	--	--	--	X	X	GLR06	GLR06
GLR07	--	--	--	--	--	X	X	GLR07	GLR07
GLR09	--	--	--	--	--	X	X	X	GLR09
GLR10	--	--	--	--	--	X	X	X	GLR10
GLR13	--	--	--	--	--	X	X	GLR13	X
GLR17	--	--	--	--	--	X	X	X	GLR17
RBSM01	--	--	--	--	RBSM01	--	--	--	--
RBSM02	--	--	--	--	RBSM02	--	--	--	--
RBSM03	--	--	--	--	RBSM03	--	--	--	--
RBSM04	--	--	--	--	RBSM04	--	--	--	--
RBSM05	--	--	--	--	RBSM05	--	--	--	--
RBSM11	--	--	--	--	RBSM11	--	--	--	--

Table 2. Best supported models with AIC scores less than that of the fully saturated model (model 8). All models were fit using a type 2 negative binomial error distribution, as well as random intercepts for categories of Year (n=8) and Site (n=28). Formulas in grey indicate they refer to the zero-inflated model, each occurs below its corresponding conditional model.

Model	Formula	ΔAIC	DF
1).	$day + day^2 + hour^2 + temp^2 + hum + moon + pressure + precip + Lag\ Call$	0.0	28
	$day + day^2 + hour + hour^2 + temp + temp^2 + hum + moon + wind + pressure + \mathbf{\Delta}24\ pressure + Lag\ Call$		
2).	$day + day^2 + hour^2 + temp^2 + hum + moon + pressure + precip + Lag\ Call$	0.1	29
	$day + day^2 + hour + hour^2 + temp + temp^2 + hum + moon + wind + pressure + \mathbf{\Delta}24\ pressure + precip + Lag\ Call$		
3).	$day + day^2 + hour^2 + temp^2 + hum + moon + pressure + precip + Lag\ Call$	0.2	27
	$day + day^2 + hour + hour^2 + temp + temp^2 + hum + wind + pressure + \mathbf{\Delta}24\ pressure + Lag\ Call$		
4).	$day + day^2 + hour^2 + temp^2 + hum + moon + pressure + precip + Lag\ Call$	0.3	28
	$day + day^2 + hour + hour^2 + temp + temp^2 + hum + wind + pressure + \mathbf{\Delta}24\ pressure + precip + Lag\ Call$		
5).	$day + day^2 + hour^2 + temp^2 + hum + moon + wind + pressure + precip + Lag\ Call$	1.3	30
	$day + day^2 + hour + hour^2 + temp + temp^2 + hum + moon + wind + pressure + \mathbf{\Delta}24\ pressure + precip + Lag\ Call$		
6).	$day + day^2 + hour^2 + temp^2 + hum + moon + wind + pressure + precip + Lag\ Call$	3.0	31
	$day + day^2 + hour + hour^2 + temp + temp^2 + hum + moon + wind + pressure + \mathbf{\Delta}24\ pressure + precip + Lag\ Call$		
7).	$day + day^2 + hour^2 + temp + temp^2 + hum + moon + wind + pressure + \mathbf{\Delta}24\ pressure + precip + Lag\ Call$	5.0	32
	$day + day^2 + hour + hour^2 + temp + temp^2 + hum + moon + wind + pressure + \mathbf{\Delta}24\ pressure + precip + Lag\ Call$		
8).	$day + day^2 + hour + hour^2 + temp + temp^2 + hum + moon + wind + pressure + \mathbf{\Delta}12\ pressure + \mathbf{\Delta}24\ pressure + precip + Lag\ Call$	7.2	35
	$day + day^2 + hour + hour^2 + temp + temp^2 + hum + moon + wind + pressure + \mathbf{\Delta}12\ pressure + \mathbf{\Delta}24\ pressure + precip + Lag\ Call$		

Table 3. Summary of most parsimonious among best supported models, with 95% confidence interval for estimated coefficients, including the random intercept coefficients for Year (n=8) and Site (n=28).

Conditional Model					
	2.50%	Estimate	97.50%	Z	p
Intercept	0.168	0.543	0.917	2.838	0.005
day	0.008	0.125	0.243	2.089	0.037
day^2	-0.238	-0.118	0.002	-1.922	0.055
$hour^2$	0.349	0.437	0.525	9.759	<0.001
$temperature^2$	-0.593	-0.451	-0.309	-6.238	<0.001
humidity	0.056	0.153	0.250	3.104	0.002
moon illumination	0.010	0.078	0.145	2.243	0.025
pressure	-0.167	-0.085	-0.004	-2.045	0.041
precipitation	-0.145	-0.092	-0.040	-3.431	0.001
Lag Call	0.105	0.118	0.131	17.692	<0.001
Random Effects					
Year	0.145	0.265	0.483		
Site	0.329	0.545	0.903		

Zero-inflated model					
	2.50%	Estimate	97.50%	Z	p
Intercept	3.816	4.624	5.431	11.224	<0.001
day	0.758	0.889	1.020	13.309	<0.001
day^2	0.696	0.825	0.955	12.475	<0.001
hour	-0.155	-0.093	-0.032	-2.966	0.003
$hour^2$	-0.585	-0.504	-0.424	-12.255	<0.001
temp	-1.516	-1.336	-1.156	-14.547	<0.001
$temp^2$	0.292	0.459	0.627	5.391	<0.001
hum	-0.803	-0.688	-0.573	-11.717	<0.001
wind	0.040	0.133	0.226	2.798	0.005
pressure	0.019	0.127	0.236	2.301	0.021
delta24P	-0.274	-0.174	-0.075	-3.429	<0.001
Lag Call	-21.351	-19.116	-16.880	-16.757	<0.001
Random Effects					
Year	0.511	0.869	1.479		
Site	0.891	1.208	1.638		
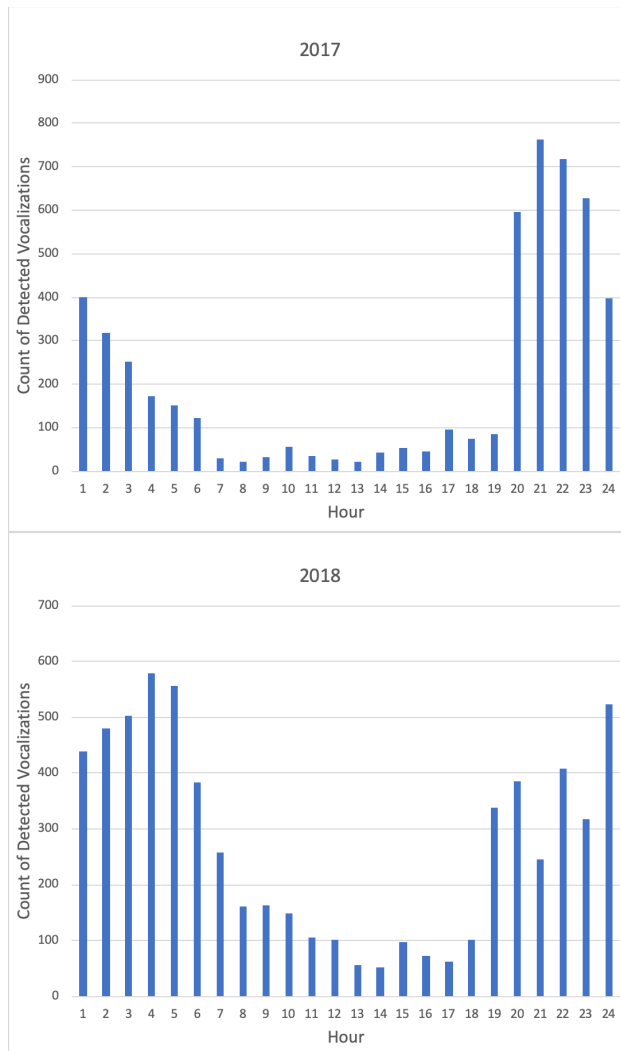
Figure 1. The hourly count of Houston Toad vocalizations made by software Kaleidoscope. These data include the years 2017 and 2018, summed across two sites (BASM05 and BASM07) within Bastrop County, Texas. Sites with labels BASM05 and BASM07 recorded continuously throughout the day within these two years, whereas other sites did not.
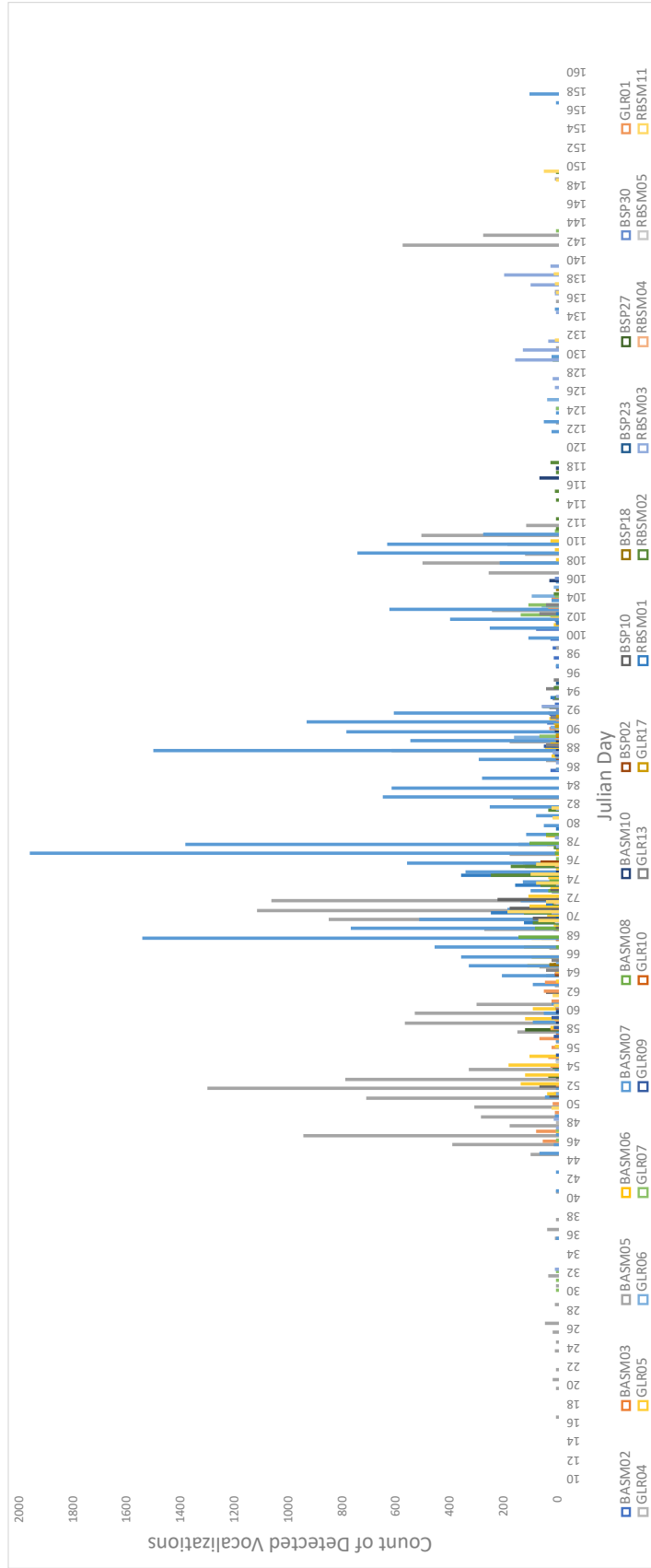
Figure 2. The complete count of Houston Toad vocalizations made by software Kaleidoscope among all years (2010-2018) and all sites (n=28) across two Texas counties, by Julian day (10-160). Sites with labels BASM05 and BASM07 recorded continuously, contributing to their large peaks.

References

Adams, M.J., Miller, D.A., Muths, E., Corn, P.S., Grant, E.H.C., Bailey, L.L., Fellers, G.M., Fisher, R.N., Sadinski, W.J., Waddle, H. and Walls, S.C., 2013. Trends in amphibian occupancy in the United States. *PloS one*, *8*(5), p.e64347.

Aide, T.M., Corrada-Bravo, C., Campos-Cerqueira, M., Milan, C., Vega, G. and Alvarez, R., 2013. Real-time bioacoustics monitoring and automated species identification. *PeerJ*, *1*, p.e103.

Alldredge, M.W., Pollock, K.H., Simons, T.R., Collazo, J.A. and Shriner, S.A., 2007. Time-of-detection method for estimating abundance from point-count surveys. *The Auk*, *124*(2), pp.653-664.

Allen, C.R., Epperson, D.M. and Garmestani, A.S., 2004. Red imported fire ant impacts on wildlife: a decade of research. *The American Midland Naturalist*, *152*(1), pp.88-104.

Becker, C.G., Fonseca, C.R., Haddad, C.F.B., Batista, R.F. and Prado, P.I., 2007. Habitat split and the global decline of amphibians. *Science*, *318*(5857), pp.1775-1777.

Beebee, T.J. and Griffiths, R.A., 2005. The amphibian decline crisis: a watershed for conservation biology?. *Biological conservation*, *125*(3), pp.271-285.

Bolker, B.M., Brooks, M.E., Clark, C.J., Geange, S.W., Poulsen, J.R., Stevens, M.H.H. and White, J.S.S., 2009. Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in ecology & evolution*, *24*(3), pp.127-135.

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

Briggs, C.J., Vredenburg, V.T., Knapp, R.A. and Rachowicz, L.J., 2005. Investigating the population-level effects of chytridiomycosis: an emerging infectious disease of amphibians. *Ecology*, *86*(12), pp.3149-3159.

Brooks, M.E., Kristensen, K., van Benthem, K.J., Magnusson, A., Berg, C.W., Nielsen, A., Skaug, H.J., Machler, M. and Bolker, B.M., 2017. glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R journal*, *9*(2), pp.378-400.

Brown, D.J., Duarte, A., Mali, I., Jones, M.C. and Forstner, M.R., 2014. Potential impacts of a high severity wildfire on abundance, movement, and diversity of herpetofauna in the Lost Pines ecoregion of Texas. *Herpetological Conservation and Biology*, *9*(1), pp.192-205.

Brown, D.J., Swannack, T.M. and Forstner, M.R., 2013. Predictive models for calling and movement activity of the endangered Houston toad. *The American Midland Naturalist*, pp.303-321.

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

Burnham, K.P., and Anderson, D.R., 2002. Model Selection and multimodel Inference: a practical information-theoretic approach. New York: Springer-Verlag. 488 p.

Canavero, A., Arim, M., Naya, D.E., Camargo, A., Da Rosa, I. and Maneyro, R., 2008. Calling activity patterns in an anuran assemblage: the role of seasonal trends and weather determinants. *North-Western Journal of Zoology*, *4*(1).

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

Cook, R.P., Tupper, T.A., Paton, P.W. and Timm, B.C., 2011. Effects of temperature and temporal factors on anuran detection probabilities at Cape Cod National Seashore, Massachusetts, USA: implications for long-term monitoring. *Herpetological Conservation and Biology*, *6*(2), pp.25-39.

Daszak, P., Scott, D.E., Kilpatrick, A.M., Faggioni, C., Gibbons, J.W. and Porter, D., 2005. Amphibian population declines at Savannah River site are linked to climate, not chytridiomycosis. *Ecology*, *86*(12), pp.3232-3237.

Dixon, J.R., Dronen, N.O., Godwin, J.C. and Simmons, M.A., 1990. The amphibians, reptiles, and mammals of Bastrop and Buescher State Parks: with emphasis on the Houston toad (*Bufo houstonensis*) and the short-tailed shrew (*Blarina sp*.). *Texas Parks and Wildlife Department*.

Duarte, A., Brown, D.J. and Forstner, M.R., 2011. Estimating abundance of the endangered Houston toad on a primary recovery site. *Journal of Fish and Wildlife Management*, *2*(2), pp.207-215.

Gaston, M.A., Fuji, A., Weckerly, F.W. and Forstner, M.R., 2010. Potential component allee effects and their impact on wetland management in the conservation of endangered anurans. *PLoS One*, *5*(4), p.e10102.

Gibbons, J.W., Scott, D.E., Ryan, T.J., Buhlmann, K.A., Tuberville, T.D., Metts, B.S., Greene, J.L., Mills, T., Leiden, Y., Poppy, S. and Winne, C.T., 2000. The Global Decline of Reptiles, Déjà Vu Amphibians: Reptile species are declining on a global scale. Six significant threats to reptile populations are habitat loss and degradation, introduced invasive species, environmental pollution, disease, unsustainable use, and global climate change. *BioScience*, *50*(8), pp.653-666.

Hillis, D.M., Hillis, A.M. and Martin, R.F., 1984. Reproductive ecology and hybridization of the endangered Houston toad (*Bufo houstonensis*). *Journal of Herpetology*, pp.56-72.

Hillman, S. S., Withers, P. C., Drewes, R. C., and Hillyard, S. D., 2009. Ecological and environmental physiology of amphibians. Oxford University Press, New York. 467 p.

Hutto, R.L. and Stutzman, R.J., 2009. Humans versus autonomous recording units: A comparison of point-count results. *Journal of Field Ornithology*, *80*(4), pp.387-398.

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

Jacobson, N.L., 1989. Breeding dynamics of the Houston toad. *The Southwestern Naturalist*, pp.374-380.

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

Knutson, M.G., Sauer, J.R., Olsen, D.A., Mossman, M.J., Hemesath, L.M. and Lannoo,
M.J., 1999. Effects of landscape composition and wetland fragmentation on frog
and toad abundance and species richness in Iowa and Wisconsin,
USA. *Conservation Biology*, *13*(6), pp.1437-1446.

Laurance, W.F., McDonald, K.R. and Speare, R., 1996. Epidemic disease and the
catastrophic decline of Australian rain forest frogs. *Conservation Biology*, *10*(2),
pp.406-413.

Lindén, A. and Mäntyniemi, S., 2011. Using the negative binomial distribution to model
overdispersion in ecological count data. *Ecology*, *92*(7), pp.1414-1421.

MacLaren, A.R., McCracken, S.F. and Forstner, M.R., 2018A. Development and
Validation of Automated Detection Tools for Vocalizations of Rare and
Endangered Anurans. *Journal of Fish and Wildlife Management*, *9*(1), pp.144-
154.

MacLaren, A.R., McCracken, S.F. and Forstner, M.R., 2018B. Automated Monitoring
Techniques Reveal New Proximate Cues of Houston Toad Chorusing
Behavior. *Herpetological Review*, *49*(4), pp.622-626.

McMenamin, S.K., Hadly, E.A. and Wright, C.K., 2008. Climatic change and wetland
desiccation cause amphibian decline in Yellowstone National Park. *Proceedings
of the national Academy of Sciences*, *105*(44), pp.16988-16993.

Miller, D.A., Weir, L.A., McClintock, B.T., Grant, E.H.C., Bailey, L.L. and Simons,
T.R., 2012. Experimental investigation of false positive errors in auditory species
occurrence surveys. *Ecological Applications*, *22*(5), pp.1665-1674.

Narins, P.M., Feng, A.S., Lin, W., Schnitzler, H.U., Denzinger, A., Suthers, R.A. and Xu, C., 2004. Old World frog and bird vocalizations contain prominent ultrasonic harmonics. *The Journal of the Acoustical Society of America*, *115*(2), pp.910-913.

Oseen, K.L. and Wassersug, R.J., 2002. Environmental factors influencing calling in sympatric anurans. *Oecologia*, *133*(4), pp.616-625.

Price, A.H. 2003. The Houston Toad in Bastrop State Park 1990 - 2002: a narrative. Open- file report 03-0401, Texas Parks & Wildlife Department.

Price, A.H. and Yantis, J.H., 1990. Houston Toad status survey. Sec. 6 Perf. Rep., No. E-1-2(8):46pp.

Price, A.H. and Yantis, J.H., 1992. Houston Toad (*Bufo houstonensis*) status survey. Sec. 6 Perf. Rep., No. E-1-3(8):38pp.

Price, A.H. and Yantis, J.H., 1993. Houston Toad (*Bufo houstonensis*) status survey. Sec. 6 Perf. Rep., No. E-1-4(8):30pp.

R Development Core Team, 2018. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Reading, C.J., Luiselli, L.M., Akani, G.C., Bonnet, X., Amori, G., Ballouard, J.M., Filippi, E., Naulleau, G., Pearson, D. and Rugiero, L., 2010. Are snake populations in widespread decline?. *Biology letters*, *6*(6), pp.777-780.

Rempel, R.S., Francis, C.M., Robinson, J.N. and Campbell, M., 2013. Comparison of audio recording system performance for detecting and monitoring songbirds. *Journal of Field Ornithology*, *84*(1), pp.86-97.

Riley, S.P., Shaffer, H.B., Voss, S.R. and Fitzpatrick, B.M., 2003. Hybridization between a rare, native tiger salamander (*Ambystoma californiense*) and its introduced congener. *Ecological Applications*, *13*(5), pp.1263-1275.

Saenz, D., Fitzgerald, L.A., Baum, K.A. and Conner, R.N., 2006. Abiotic correlates of anuran calling phenology: the importance of rain, temperature, and season. *Herpetological Monographs*, *20*(1), pp.64-82.

Salthe, S.N. and Mecham, J.S., 1974. Reproductive and courtship patterns. *Physiology of the Amphibia*, *2*, pp.309-521.

Steelman, C.K. and Dorcas, M.E., 2010. Anuran calling survey optimization: developing and testing predictive models of anuran calling activity. *Journal of Herpetology*, *44*(1), pp.61-69.

Templeton, A.R., Robertson, R.J., Brisson, J. and Strasburg, J., 2001. Disrupting evolutionary processes: the effect of habitat fragmentation on collared lizards in the Missouri Ozarks. *Proceedings of the National Academy of Sciences*, *98*(10), pp.5426-5432.

[USFWS] US Fish and Wildlife Service, 2007. Section 10 (a)(1)(A) Scientific permit requirements for conducting Houston Toad presence/absence surveys. *Austin, TX: US Fish and Wildlife Service*.

Ver Hoef, J. M., and Boveng, P. L., 2007. Quasi-Poisson vs. negative binomial regression: How should we model over-dispersed count data? *Ecology* 88:2766–2772.

Waldron, J.L., Bennett, S.H., Welch, S.M., Dorcas, M.E., Lanham, J.D. and Kalinowsky, W., 2006. Habitat specificity and home-range size as attributes of species vulnerability to extinction: a case study using sympatric rattlesnakes. *Animal conservation*, *9*(4), pp.414-420.

Weir, L., Fiske, I.J. and Royle, J.A., 2009. Trends in anuran occupancy from northeastern states of the North American Amphibian Monitoring Program. *Herpetological Conservation and Biology*, *4*(3), pp.389-402.

Wells, K.D., 1977. The social behaviour of anuran amphibians. *Animal Behaviour*, *25*, pp.666-693.

Welsh Jr, H.H., 1990. Relictual amphibians and old-growth forests. *Conservation Biology*, *4*(3), pp.309-319.

Wildlife Acoustics. 2017. Kaleidoscope 4.3.1 user manual. Maynard, Massachusetts: Wildlife Acoustics.

Zuur, A.F., Ieno, E.N. and Elphick, C.S., 2010. A protocol for data exploration to avoid common statistical problems. *Methods in ecology and evolution*, *1*(1), pp.3-14.

II. OPTIMIZING THE POWER OF HUMAN PERFORMED AUDIO SURVEYS FOR MONITORING THE ENDANGERED HOUSTON TOAD USING AUTOMATED RECORDING DEVICES

Introduction

Long-term monitoring of anuran populations is required to gain an understanding of population dynamics (Pechmann et al. 1991). Researchers commonly conduct auditory surveys to determine presence or absence of anuran species (Bridges & Dorcas 2000; Crouch & Paton 2002; Schmidt 2003; Pierce & 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).

There are presently 14 threatened and endangered anurans in the United States with the inclusion of Puerto Rico (*Anaxyrus californicus, Anaxyrus canorus, Bufo hemiophrys baxterii, Bufo houstonensis, Eleutherodactylus cooki, Eleutherodactylus Jasper, Eleutherodactylus juanariveroi, Peltophryne lemur, Rana chiricahua, Rana draytonii, Rana muscosa, Rana pretiosa, Rana sierra,* and *Rana sevosa*). Guidelines for conducting presence/absence surveys authored by the United States Fish and Wildlife Service (USFWS, hereafter) exist for only four of these 14 (*B. houstonensis, A. californicus, R. chiricahua,* and *R. draytonii*; USFWS 1999; USFWS 2005; USFWS 2006; USFWS 2007). Two species (*R. pretiosa* and *R. sierra*) have general guideline overviews of currently applied survey methods, authored by the United States Forest Service, but these have not been established as legitimized federal rules. Federal

guidelines for conducting surveys of an additional two species (*A. canorus,* and *R. muscosa*) are reported to be pending approval. This leaves six species without formal documentation dictating the method to confirm species presence and subsequent site management protocol, leaving this distinction open ended. These guidelines are intended to ensure that independent researchers are performing surveys when environmental conditions are optimal for male chorusing (i.e. potential breeding), and thus optimal for detecting the species when present. Among the documents cited above the guidelines regarding Houston Toads (*B. houstonensis*) are some of the most specific (USFWS 2007).

The efficacy of automated methods of call detection for this species has been examined (MacLaren et al. 2018A), but these methods have not been vetted against the USFWS guidelines currently in place. These guidelines dictate that at minimum six surveys, per year, are required at each listening post (i.e. potential breeding location); surveys must be conducted for three consecutive years; surveys should be spread out between February through April; temperatures must be at or above 57°F; surveys do not begin until about 30 minutes after sundown and cease if a drop in temperature occurs (presumably below 57°F, however, this is unclear); wind speeds must not exceed 15 miles per hour (USFWS 2007). These guidelines also include a number of much more vague recommendations intended to increase the likelihood that Houston Toads are chorusing on nights that are chosen to survey. These include nights in which humidity is greater than 70%; cloud cover is present or the moon is not full; rainfall occurring or recent rainfall has occurred. MacLaren et al. (2018B) summarizes the putative sources for these environmental correlates, given that the federal rule does not provide citations for

these claims. The efficacy of these guidelines has been studied previously, as a result the federal rule now recommends increasing the number of surveys necessary within each season from six to twelve (Jackson et al. 2006; USFWS 2007).

Many studies have revealed the advantages of automated audio recording systems in determining the influential exogenous environmental factors associated with vocalizing behavior when compared to manual call surveys (Bridges & Dorcas 2000; Oseen & Wassersug 2002; Hsu et al. 2005; Acevado & Villanueva-Rivera 2006; Dorcas et al. 2009; Willacy et al. 2015; MacLaren et al. 2018B). However, I am aware of no studies that model the potential outcomes of a manual survey protocol using data acquired from an automated recording system. The purpose of this study is to determine the probability of detecting Houston Toads through repeat audio surveys, varying in survey selection method and survey effort. We hypothesize that as number of surveys conducted, and length of surveys conducted, both increase that power to detect toads will increase accordingly.

Methods

Study Site

We carried out this study utilizing data gathered from the Griffith League Ranch (GLR), located in Bastrop County, Texas, USA. The GLR is a private property owned and operated by the Boy Scouts of America. This property is commonly thought of as the primary recovery site for the Houston Toad (Duarte et al. 2014), and received population supplementation through captive propagation efforts for this endangered species both during and prior to when this study was conducted.

We used Song Meter SM3 (Wildlife Acoustics, Maynard, MA) audio recording devices to monitor for the call of male Houston Toads at two breeding locations on the GLR. Song Meters were programmed to record continuously, beginning in January and ending in July, for four years (2015-2018). We powered the Song Meters using rechargeable sealed lead acid batteries (Power Sonic PS-6360 NB, 6V, 36.0 AH). We stored the external batteries in plastic cases, secured to a structure object adjacent to each Song Meter. We equipped each Song Meter with four 64GB SD cards for media storage. The additional costs and data storage requirements associated with continuously monitoring limited us to only two locations. We selected monitoring locations based on a history of maintaining a large number of chorusing male Houston Toads relative to other ponds within the GLR, likely because these sites received repeated annual supplementation events.

To analyze the large quantity of audio files collected, we trained an audio classifier using the software Kaleidoscope (Wildlife Acoustics). We followed the steps outlined by the manufacturer for completing this process (Wildlife Acoustics 2017) and used the audio training data provided by MacLaren et al. (2018A) for the call of the Houston Toad. We chose to simply train towards two "clusters", Houston Toad vocalizations, and anything that is not a Houston Toad vocalization. This was efficient, and we achieved 100% detection of positive training data into the Houston Toad cluster within a single round of training, indicating that this recognizer returns near-perfect detection. Initially, all detections made by the software > 3 seconds in length were

manually verified by ARM. We observed that detections below this length were mostly false positives. To ensure near perfect detection, we then sorted detections by date, taking care to search all detections on dates in which Houston Toads had not yet been detected. We binned detections into 5-minute intervals, and summarized them as binomial, where 1 and 0 indicate detection and non-detection of Houston Toads, respectively.

Simulation

We simulated Houston Toad audio surveys under three sampling protocols. First, by randomly selecting survey data. Second, by restricting available survey data according to environmental conditions presented in the United States Fish and Wildlife Services protocol for conducting surveys for this species (USFWS 2007). And finally, we sought to identify which environmental conditions, if any, contain boundaries that yield increased detection probability. For environmental variables we used the National Oceanic and Atmospheric Administration's quality controlled local climatic dataset, measured at Giddings, Texas, USA, ca. 25 km East (WBAN 53979). We utilized moon illumination measured by the U.S. naval observatory (USNO) for Central Time Zones.

We removed all instances in which the Song Meters did not record, then pooled all Houston Toad occurrence data across the four years and the two sites (N=92,652). We randomly selected one 5-minute interval for every date within the pooled dataset and repeated this 1,000 times. This was done to eliminate the possibility of randomly selecting multiple surveys within a given date, which more correctly reflects how surveys are conducted in practice (i.e. human performed surveys). We calculated detection probability by raw proportion of detections within each permutation, resulting in 1,000

estimates for detection probability. We conducted this simulation for surveys of length 5

minutes, through 60 minutes, by 5-minute intervals. We used the formula provided by

Pellet and Schmidt (2005) for calculating the minimum number of surveys required to be

95% confident in Houston Toad absence, $N_{min} = \frac{\log{(0.05)}}{\log{(1-p)}}$. We calculated this metric for

the mean, upper and lower 95% confidence interval boundaries for each distribution of

probabilities (12 per sampling protocol).

The decision to conduct Houston Toad audio surveys is often made in advance of

appropriate environmental conditions occurring, based largely on weather forecasts. To

reflect the uncertainty implicit in this practice we chose to assume that if environmental

thresholds were met at any point within a calendar date, that all data for this date may be

surveyed. This is reflected in the results as "dates surveyable" under each protocol. Each

of our three survey protocols selected only for intervals occurring in the months

February, March, and April, before 0600 and after 1800 hours of each date. This not only

reflects roughly what is currently required (USFWS 2007) it also coincides with peak

activity periods for the Houston Toad. To implement a random survey selection protocol,

all 5-minute intervals within this time frame were considered. To replicate the restrictions

within USFWS (2007) we removed dates wherein environmental variables failed to meet

the following thresholds: temperature > 14 °C, relative humidity > 70%, wind speeds <

24 kmph, and percent moon illumination < 0.5.

For our final protocol we searched for improved environmental thresholds by

which researchers may choose to conduct a human performed audio survey for the

Houston Toad. We calculated summary statistics for the following environmental

variables, when Houston Toad vocalizations were detected: temperature (°C), relative

humidity (%), wind speed (kmph), moon illumination (%), hourly precipitation (mm), cumulative precipitation over the previous 24 hours (mm), barometric pressure (mmHg at sea level), difference in barometric pressure across 24 hours (mmHg at sea level). We then examined which, if any, of these variables offered thresholds useful towards restricting the number of dates containing non-detections. We calculated the ratio of detections/non-detections for all combinations of thresholds both above and below all values of temperature, the change in barometric pressure over 24 hours, and cumulative precipitation over the previous 24 hours. This allowed us to identify which thresholds excluded large periods of inactivity within the breeding season. We then applied these thresholds in the same way as described above and carried out the simulation under these new restrictions.

Results

We collected detection/non-detection data on 433 dates across all four years. We observed a total of 3,975 5-minute intervals (intervals, hereafter) found to possess one or more Houston Toad vocalizations. Environmental variables measured for the dates which include these intervals are given in Table 4. Of these intervals, only 122 occur outside of Feb-April. All results henceforth refer to sampling within this peak period only. When we selected surveys randomly, ignoring environment, detection probability ranged from 0.038 – 0.100 (mean = 0.063) for surveys 5 minutes in duration, and 0.088 – 0.171 (mean = 0.121) for surveys 60 minutes in duration (Figure 3, Table 5). These probabilities result in requiring on average 47 (range = 29-79) surveys to be 95% confident in determining

absence of the species, when conducting 5-minute-long surveys, and on average 24 (range = 16-33) one-hour-long surveys (Figure 3, Table 5).

Restricting survey appropriate dates to only those that reflect the environmental thresholds given by USFWS (2007) reduced the number of available survey dates to 118. Under this paradigm only 1,958 intervals containing detections are available to be sampled, leaving 1,895 intervals unobservable to surveyors. Detection probability ranged from 0.042 – 0.119 (mean = 0.080) for surveys 5 minutes in duration, and 0.093 – 0.186 (mean = 0.142) for surveys 60 minutes in duration (Figure 3, Table 5). These probabilities result in requiring on average 36 (range = 24-69) surveys to be 95% confident in determining absence of the species, when conducting 5-minute-long surveys, and on average 17 (range = 15-30) one-hour-long surveys (Figure 3, Table 5).

Calculating the proportion of detections to non-detections over a vast range of environmental thresholds revealed that a unique combination of temperature (> 16 °C), precipitation (> 0 mm/day), and change in barometric pressure (< -0.07 mmHg) provided the greatest advantage. These thresholds allow 133 dates to be surveyable, comparable to USFWS (2007), yet provide 2,569 intervals, a 31% relative increase. Under this paradigm detection probability ranged from 0.066 – 0.146 (mean = 0.105) for surveys 5 minutes in duration, and 0.133 – 0.229 (mean = 0.179) for surveys 60 minutes in duration (Figure 3, Table 5). These probabilities result in requiring on average 27 (range = 19-44) surveys to be 95% confident in determining absence of the species, when conducting 5-minute-long surveys, and on average 16 (range = 12-21) one-hour-long surveys (Figure 3, Table 5).

Discussion

Heterogeneity in probability of detecting vocalizing anurans, including the Houston Toad, during acoustic surveys has been shown to be a consequence of daily variation in environmental parameters (Brown et al. 2013; Crouch & Paton 2002; Oseen & Wassersug 2002; Saenz et al. 2006; Willacy et al. 2015). Our simulation sought to re-evaluate (Jackson et al. 2006) how to reliably detect the Houston Toad by manual, human performed, audio survey given this heterogeneity. This re-evaluation has been necessitated by shifts in the demography of the Houston Toad (Gaston et al. 2010; Duarte et al. 2014), which are a consequence of both positive (e.g. captive propagation and population supplementation) and negative (e.g. catastrophic wildfire) impacts (Brown et al. 2014). One benefit of our study is that we focus on both the number and duration of surveys required to reliably detect Houston Toads, whereas previous studies often focus on one aspect of this tradeoff alone (Bridges & Dorcas 2000, Crouch & Paton2002, Pierce & Gutzweiller 2004).

Jackson et al. (2006) hypothesized that the use of technological innovations would positively impact the detection of animals during auditory surveys, but also state that these technologies may be too expensive for researchers to reasonably employ. Today, automated recording technology ranges in cost and performance capabilities, and can, in some cases, be more affordable than human performed manual call surveys (Charif & Pitzrick 2008; Rempel et al. 2013). There are many reasons that such technology are preferred. For example, a large number of surveys may be required, or occupied habitats may be difficult to access (Alldredge et al. 2007; Hutto & Stutzman 2009). For these

41

reasons, anuran monitoring is routinely performed using automated recording devices (Saenz et al. 2006; Aide et al. 2013; Willacy et al. 2015).

This study updates and expands upon the findings of Jackson et al. (2006). For perspective the previous study (Jackson et al. 2006) utilized at most 14 5-minute surveys (70 minutes) within a given year, whereas within a single year our ARDs provided us with approximately 5000 minutes of audio. Using these vast and detailed data we found that detection probabilities, for surveys of any length, and under any sampling protocol, were lower than what has been previously estimated for this species (Jackson et al. 2006). By suggesting more accurate environmental thresholds under which surveys should be conducted, and evaluating surveys of varying duration, we have provided researchers with the greatest probability of detecting Houston Toads possible. Our approach to simulating survey effort, if adopted by the USFWS, will allow researchers to choose the length or survey they find most convenient, and maintain 95% confidence in determinations of absence by conducting the number of surveys we have recommended for that given duration. Like Jackson et al. (2006) our results suggest that the USFWS should change the mandatory survey guidelines to require more surveys in each season than is currently specified. However, unlike the previous study, our findings reinforce that selecting surveys based on environmental factors increases the probability of survey success.

We found that previously suggested environmental correlates to chorusing among Houston Toads offered improved detection probabilities over randomly selected surveys. However, we found that not all suggested weather criteria within USFWS (2007) were useful, specifically, moon illumination, humidity, and wind speed. This is either because

these variables share no true relationship with chorusing within Houston Toads, as is the case for moon illumination or because they do not serve as a hard boundary, as is the case for relative humidity. For example, relative humidity may range between 10% and 90% within a given single date in response to natural diel cycle. We identified definitive thresholds among temperature, precipitation, and shifts in barometric pressure that improve the probability of detection for Houston Toads beyond what USFWS currently suggests.

Table 4. Summary statistics of environmental variables during dates in which Houston Toads were detected.

	MIN.	1ST QU.	MEDIAN	MEAN	3RD QU.	MAX.
TEMPERATURE (°C)	-1.15	16.63	19.20	19.24	22.00	31.40
RELATIVE HUMIDITY (%)	14.33	72.33	88.00	82.63	95.67	100.00
WIND SPEED (KMPH)	0.00	3.67	7.00	7.19	9.67	28.33
BAROMETRIC PRESSURE (MMHG)	29.06	29.38	29.47	29.47	29.57	29.99
PRESSURE CHANGE (MMHG)	-0.35	-0.11	-0.04	-0.03	0.04	0.38
PRECIPITATION (MM)	0.00	0.03	0.07	0.24	0.22	4.50
MOON ILLUMINATION (%)	0.00	11.00	40.00	48.76	91.00	100.00

Table 5. Mean and 95% confidence bounds for the probability of detection and number of surveys required to be 95% confident in absence of Houston Toads during acoustic surveys, for three sampling paradigms, as duration of survey increases from 5 to 60 minutes.

Duration (mins)	Detection Probability			Number of Surveys		
	2.50%	Mean	97.50%	2.50%	Mean	97.50%
	Randomly Selected					
5	0.038	0.063	0.100	78.4	46.1	28.4
10	0.046	0.074	0.113	63.9	38.9	25.1
15	0.050	0.082	0.121	58.4	34.9	23.3
20	0.058	0.089	0.129	49.8	32.3	21.7
25	0.063	0.094	0.138	46.4	30.3	20.3
30	0.067	0.099	0.142	43.4	28.6	19.6
35	0.071	0.104	0.146	40.8	27.4	19.0
40	0.075	0.107	0.150	38.4	26.3	18.4
45	0.075	0.111	0.154	38.4	25.4	17.9
50	0.079	0.115	0.158	36.3	24.6	17.4
55	0.083	0.118	0.167	34.4	23.9	16.4
60	0.088	0.121	0.171	32.7	23.2	16.0
	USFWS Protocol					
5	0.042	0.080	0.119	69.2	36.1	23.7
10	0.051	0.094	0.136	57.4	30.4	20.6
15	0.059	0.103	0.153	49.0	27.5	18.1
20	0.068	0.110	0.153	42.7	25.7	18.1
25	0.076	0.116	0.161	37.8	24.3	17.1
30	0.076	0.121	0.169	37.8	23.2	16.1
35	0.085	0.126	0.169	33.8	22.3	16.1
40	0.085	0.129	0.178	33.8	21.6	15.3
45	0.085	0.133	0.178	33.8	21.0	15.3
50	0.093	0.136	0.178	30.6	20.4	15.3
55	0.093	0.140	0.186	30.6	19.9	14.5
60	0.093	0.143	0.186	30.6	19.4	14.5
	Optimized Protocol					
5	0.060	0.101	0.143	48.3	28.0	19.4
10	0.075	0.118	0.165	38.3	23.9	16.6
15	0.083	0.128	0.173	34.7	21.8	15.8
20	0.090	0.136	0.180	31.7	20.4	15.1
25	0.098	0.143	0.188	29.1	19.5	14.4
30	0.105	0.149	0.195	26.9	18.6	13.8
35	0.113	0.154	0.195	25.0	17.9	13.8
40	0.113	0.159	0.203	25.0	17.3	13.2
45	0.120	0.163	0.203	23.4	16.9	13.2
50	0.120	0.167	0.211	23.4	16.4	12.7
55	0.128	0.170	0.211	21.9	16.1	12.7
60	0.128	0.173	0.218	21.9	15.7	12.2
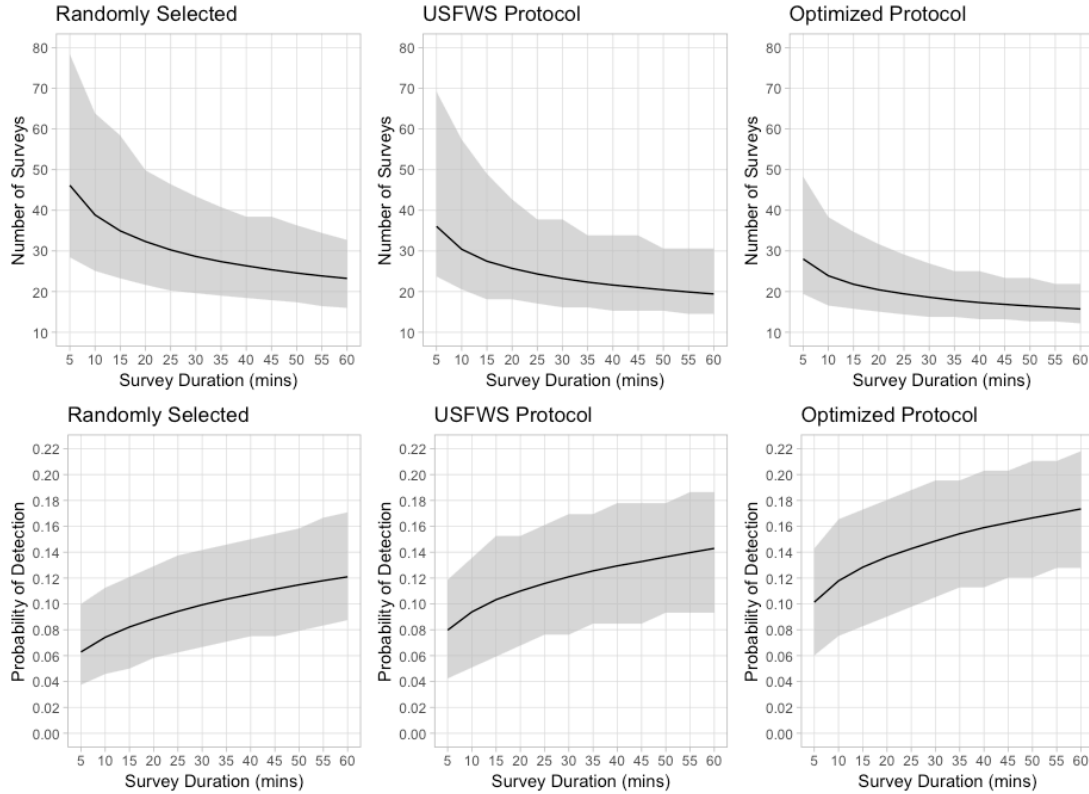
Figure 3. Results of simulation to assess mean probability of detection of Houston Toads (bottom panel), and the mean number of surveys necessary for a given probability of detection (top panel), and their 95% confidence bounds, as the length in minutes of each auditory survey increases along the x-axis, for three approaches to survey selection: Random selection (left), following USFWS 2007 (middle), and under our proposed optimization for survey selection (right).

References

Acevedo, M.A. and Villanueva-Rivera, L.J., 2006. Using Automated Digital Recording Systems as Effective Tools for the Monitoring of Birds and Amphibians. *Wildlife Society Bulletin*, pp.211-214.

Aide, T.M., Corrada-Bravo, C., Campos-Cerqueira, M., Milan, C., Vega, G. and Alvarez, R., 2013. Real-time bioacoustics monitoring and automated species identification. *PeerJ*, *1*, p.e103.

Alldredge, M.W., Pollock, K.H., Simons, T.R., Collazo, J.A. and Shriner, S.A., 2007. Time-of-detection method for estimating abundance from point-count surveys. *The Auk*, *124*(2), pp.653-664.

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

Brown, D.J., Duarte, A., Mali, I., Jones, M.C. and Forstner, M.R., 2014. Potential impacts of a high severity wildfire on abundance, movement, and diversity of herpetofauna in the Lost Pines ecoregion of Texas. *Herpetological Conservation and Biology*, *9*(1), pp.192-205.

Brown, D.J., Swannack, T.M. and Forstner, M.R., 2013. Predictive models for calling and movement activity of the endangered Houston toad. *The American Midland Naturalist*, pp.303-321.

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

Crouch, W.B. and Paton, P.W., 2002. Assessing the use of call surveys to monitor breeding anurans in Rhode Island. *Journal of Herpetology*, *36*(2), pp.185-193.

Dorcas, M.E., Price, S.J., Walls, S.C. and Barichivich, W.J., 2009. Auditory monitoring of anuran populations. *Amphibian ecology and conservation: a hand book of techniques. Oxford University Press, Oxford*, pp.281-298.

Duarte, A., Brown, D.J. and Forstner, M.R., 2014. Documenting extinction in real time: decline of the Houston toad on a primary recovery site. *Journal of Fish and Wildlife Management*, *5*(2), pp.363-371.

Gaston, M.A., Fuji, A., Weckerly, F.W. and Forstner, M.R., 2010. Potential component allee effects and their impact on wetland management in the conservation of endangered anurans. *PLoS One*, *5*(4), p.e10102.

Hsu, M.Y., Kam, Y.C. and Fellers, G.M., 2005. Effectiveness of amphibian monitoring techniques in a Taiwanese subtropical forest. *The Herpetological Journal*, *15*(2), pp.73-79.

Hutto, R.L. and Stutzman, R.J., 2009. Humans versus autonomous recording units: A comparison of point-count results. *Journal of Field Ornithology*, *80*(4), pp.387-398.

Jackson, J.T., Weckerly, F.W., Swannack, T.M. and Forstner, M.R.J., 2006. Imperfect detection and number of auditory surveys for Houston toads. *J. Wildlife Management*, *70*, pp.1461-1463.

MacLaren, A.R., McCracken, S.F. and Forstner, M.R., 2018A. Development and

Validation of Automated Detection Tools for Vocalizations of Rare and

Endangered Anurans. *Journal of Fish and Wildlife Management*, *9*(1), pp.144-

154.

MacLaren, A.R., McCracken, S.F. and Forstner, M.R., 2018B. Automated Monitoring

Techniques Reveal New Proximate Cues of Houston Toad Chorusing

Behavior. *Herpetological Review*, *49*(4), pp.622-626.

Oseen, K.L. and Wassersug, R.J., 2002. Environmental factors influencing calling in

sympatric anurans. *Oecologia*, 133(4), pp.616-625.

Pechmann, J.H., Scott, D.E., Semlitsch, R.D., Caldwell, J.P., Vitt, L.J. and Gibbons,

J.W., 1991. Declining amphibian populations: the problem of separating human

impacts from natural fluctuations. *Science*, *253*(5022), pp.892-895.

 Pellet, J. and Schmidt, B.R., 2005. Monitoring distributions using call surveys:

estimating site occupancy, detection probabilities and inferring

absence. *Biological Conservation*, 123(1), pp.27-35.

Pierce, B. and Gutzweiller, K., 2004. Auditory sampling of frogs: detection efficiency in

relation to survey duration. *Journal of Herpetology* 38(4), pp.495-500.

R Development Core Team, 2018. R: A language and environment for statistical

computing. Vienna, Austria: R Foundation for Statistical Computing.

Rempel, R.S., Francis, C.M., Robinson, J.N. and Campbell, M., 2013. Comparison of

audio recording system performance for detecting and monitoring

songbirds. *Journal of Field Ornithology*, 84(1), pp.86-97.

Saenz, D., Fitzgerald, L.A., Baum, K.A. and Conner, R.N., 2006. Abiotic correlates of
anuran calling phenology: the importance of rain, temperature, and
season. *Herpetological Monographs*, *20*(1), pp.64-82.

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

[USFWS] U.S. Fish and Wildlife Service, U.S.F.W.S. 1999. Survey Protocol for the
Arroyo Toad. U.S. Fish and Wildlife Service, Carlsbad & Ventura, California.

[USFWS] U.S. Fish and Wildlife Service, U.S.F.W.S. 2005. Revised Guidance on Site
Assessments and Field Surveys for the California Red-legged Frog. U.S. Fish and
Wildlife Service, Sacramento, California.

[USFWS] U.S. Fish and Wildlife Service, U.S.F.W.S. 2006. Chiricahua Leopard Frog
(*Rana chiricahua*) Draft Recovery Plan with appendices. Appendix E. E-1 – E-15
pp.

[USFWS] 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.

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

Weir, L., Fiske, I.J. and Royle, J.A., 2009. Trends in anuran occupancy from northeastern
states of the North American Amphibian Monitoring Program. *Herpetological
Conservation and Biology*, *4*(3), pp.389-402.

Wildlife Acoustics, 2017. Kaleidoscope 4.3.1 user manual. Maynard, Massachusetts:
Wildlife Acoustics.

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

III. OBSERVER-FREE EXPERIMENTAL EVALUATION OF HABITAT AND DISTANCE EFFECTS ON THE DETECTION OF ANURAN AND BIRD VOCALIZATIONS[1]

Abstract

Acoustic surveys of vocalizing animals are conducted to determine density, distribution, and diversity. Acoustic surveys are traditionally performed by human listeners, but automated recording devices (ARD) are becoming increasingly popular. Signal strength decays, or attenuates, with increasing distance between source and receiver and some habitat types may differentially increase attenuation beyond the effects of distance alone. These combined effects are rarely accounted for in acoustic monitoring programs. We evaluated the performance of three playback devices and three ARD models using the calls of six anurans, six birds, and four pure tones. Based on these evaluations we determined the optimal playback and recording devices. Using these optimal devices we broadcast and recorded vocalizations in five habitat types along 1000m transects. We used generalized linear models to test for effects of habitat, distance, species, environmental, and landscape variables. We predicted detection probabilities for each vocalization, in each habitat type, from 0 to 1000 m. Among playback devices, only a remote predator caller simulated vocalizations consistently. Differences of ~10dB were observed among ARDs. For all species, we found differences in detectability between open and closed canopy habitats. We observed large differences

in predicted detection probability among species in each habitat type, as well as along 1000 m transects. Increases in temperature, barometric pressure, and wind speed significantly decreased detection probability. However, aside from differences among species, habitat, and distance, topography impeding a line of sight between sound source and receiver had the greatest negative influence on detections. Our results suggest researchers should model the effects of habitat, distance, and frequency on detection probability when performing acoustic surveys. To optimize survey design we recommend pilot measurements among varying habitats.

Key Words: observer free; audio recording device; detection probability; vocalization; sound attenuation; generalized linear models.

Introduction

Automated recording devices (ARD) are utilized to document a large variety of vocalizing animals. Ecologists use these systems to monitor the behavior of birds (Digby et al. 2013), bats (Bader et al. 2015), anurans (Oseen and Wassersug 2002; Aide et al. 2013), insects (Romer & Lewald 1992; Lehmann et al. 2014), and both terrestrial (Mielke & Zuberbühler 2013) and marine mammals (Wiggins & Hildebrand 2007; Selby et al. 2016). ARDs are commonly applied to determine species richness (Wimmer et al. 2013; Hsu et al. 2005) or construct general biodiversity indices (Zimmerman 1994; Sueur et al. 2014). Sound recordings are also used for more specific functions, such as assessing coral reef health (Piercy et al. 2014) or documenting changes in forest noise via soundscapes

53

(Pijanowski et al. 2011). ARDs may offer advantages over human performed surveys when a large number of sites must be visited multiple times to achieve adequate sampling effort, when monitoring inaccessible and inhospitable environments, or to avoid bias from subject disturbance (Alldredge et al. 2007a; Hutto & Stutzman 2009). The commercial availability of ARDs has grown, offering a variety of costs and performance capabilities, making them, in some cases, less expensive than human performed surveys (Charif 2008; Rempel et al. 2013; Yip et al. 2017a).

Effectively employing acoustic surveys for ecological research or biological monitoring requires an understanding of how the components of survey design and implementation affect the probability of detecting animal vocalizations. As all species are detected imperfectly, estimates of detection probability are essential for determining the true presence or absence of animal populations in occupancy studies (MacKenzie et al. 2002) and also in localization of sources for density estimation (Marques et al. 2014) and acoustic telemetry (Kessel et al. 2014). Whether surveyors detect a vocalization depends on two major factors. Vocalizations must occur during the survey period and the signal strength of vocalizations needs to be sufficient to be detected at the listening post. Heterogeneity in detection probability due to the surrounding environment is widely recognized in aquatic systems where "range-testing" is routinely carried out to aid in localization of sources in acoustic telemetry studies (Kessel et al. 2014; Selby et al. 2016). Range-testing evaluates the influence of habitat and environmental factors on the probability of detecting an animal's acoustic signal, either a vocalization or sound emitting tag.   Recent studies illustrate that localization can be effective in terrestrial systems as well (Dawson & Efford 2009; Borchers et al. 2015; Measey et al. 2017). Thus,

54

understanding spatial effects on sound attenuation, and thus detection probability, are crucial at two levels:. The "among-site" level, which is routinely used to estimate occupancy from bioacoustic data, and the "within-site" level required for localization of sources.

The probability of detecting vocalizations depends, in part, on the physical processes affecting sound traveling through a medium. For example, a relationship exists between the signal strength of a vocalization and the distance from the sound source. This is due to spherical spreading of propagating sound pressure waves (Embleton 1996), scattering and reflection of sound waves by structural objects in the intervening habitat (Selby et al. 2016; Yip et al. 2017b), atmospheric absorption, and refraction from water vapor and air temperature (Lawrence & Simmons 1982; Öhlund & Larsson 2015). Studies using human observers have shown that habitat-specific reverberations influence sound attenuation (Morton 1975; Richards & Wiley 1980; Bibby & Buckland 1987). Detection distance may vary between open and closed environments (Fricke 1984), sound source and receiver heights (Kime et al. 2000; Mathevon et al. 2005), and among species with different sound pressure level and frequency components (Nelson 2003; Llusia et al. 2011). Ambient background noise may also vary among habitats, potentially masking vocalizations, and reducing detection probability (Bormpoudakis et al. 2013).

When distance is not known, as it may be through experimentation (McClintock et al. 2010) or localization (Measey et al. 2017), researchers have attempted to control for the confound of distance by estimating rate of decay using the half-normal detection function (Buckland et al. 2001; Solymos et al. 2013), fixed-radius survey methods (Hutto 2016), or by treating sound intensity (i.e. amplitude) as a function of distance (Pieretti et

al. 2011). Many species may modulate the amplitude of their acoustic signal (Rose and Brenowitz 1991; Stewart and Bishop 1994), further confounding the use of this metric as a meaningful measure of true distance from sound receiver. Additionally, researchers often estimate detection distance using non-empirical methods, such as binning estimated distance into categories based arbitrarily on observer experience (Vold et al. 2017). Nevertheless, these approaches routinely do not account for how distance might interact with landscape or species specific effects. Without explicitly examining landscape effects on the detection process, specifically habitat structure, researchers may conclude differences in animal occurrence or abundance among differing habitats erroneously, thus causing heterogeneity in attenuation to affect patterns in animal occurrence (Gasc et al. 2018). Similarly, the effect of call frequency (Hz) and structure is routinely overlooked in assessing species richness, or biodiversity, where detectability is assumed to be equal among multiple species at the same distance (e.g. fixed-radius surveys; Sadoti et al. 2018).

Experimentation is a useful mechanism for understanding how we observe populations and environments. For example, Simons et al. (2007) employed an experimental system that simulates bird songs to study sources of heterogeneity in detection and misidentification among human observers. Through similar call reproduction experiments, birds and anurans have been shown to respond to playback vocalizations (Kearns et al. 1998; Mannan et al. 2014; James et al. 2015). Mandated survey protocols sometimes require researchers to perform playbacks when monitoring for endangered species (USFWS 2007). Nonetheless, call broadcasting devices vary widely among studies, making comparison or replication of studies challenging, and

introducing a potential source of bias. Given recent development and usage of bioacoustic technology, experimental manipulation of both sound sources and receivers is required to evaluate effectiveness in field studies.

Here, we conduct a thorough terrestrial range-testing experiment to evaluate the effects of varying habitats on the detection probability of acoustic signals of two major groups of vocalizing organisms, birds and anurans. Our objectives were to determine how distance and habitat affect the probability of detecting a variety of bird and anuran vocalizations. *A priori*, we predicted that differences among habitat types would introduce heterogeneity in detection probability across distance, and among different species. These sources of error are not routinely incorporated into acoustic monitoring programs and could lead to biased about species occurrence and other population or ecological quantities. Additionally, we evaluate the performance of three playback devices, and compare sensitivity among three commercially available ARDs.

Materials and Methods

Study site.- We performed the field work for this study on the Griffith League Ranch (GLR), a 1,948 ha property owned by the Boy Scouts of America, located in Bastrop County, Texas (Fig. 4). Presently, the GLR contains mature forests dominated by loblolly pine (*Pinus taeda*), post oak (*Quercus stellate*), blackjack oak (*Quercus marilandica*), and eastern red cedar (*Juniperus virginiana*), an area recovering from a high-intensity wildlife fire that occurred in 2011 (Brown et al. 2014), and a small central prairie. Mechanical understory thinning of yaupon holly (*Ilex vomitoria*), American

beautyberry (*Callicarpa americana*), and farkleberry (*Vaccinium arboretum*) has occurred within sections of the mature forests to create firebreaks and reduce fuel.

Playback audio.- The playback file consisted of four pure tones at 1, 3, 5, and 7 kHz, six anuran calls, and six bird calls (Supporting Audio S1; Supporting Table S2). The anuran calls used were the wood frog (*Rana* [*Lithobates*] *sylvaticus*), California red-legged frog (*Rana* [*Lithobates*] *draytonii*), Houston toad (*Bufo* [*Anaxyrus*] *houstonensis*), Arroyo toad (*Bufo* [*Anaxyrus*] *californicus*), American bullfrog (*Rana* [*Lithobates*] *catesbeiana*), and the spring peeper (*Pseudacris crucifer*). The bird calls used were the golden-cheeked warbler (*Dendroica chrysoparia*), black-capped vireo (*Vireo atricapilla*), red-cockaded woodpecker (*Picoides borealis*), black rail (*Laterallus jamaicensis*), spotted owl (*Strix occidentalis*), and the painted bunting (*Passerina ciris*). These species were selected because they are rare, endangered, the subject of audio monitoring to determine site occupancy, or widely used in acoustics research. This collection of vocalizations includes wide variation in call structure (e.g., pulses, trills, number of syllables), duration, and frequency. We assembled, edited, and volume-balanced playback audio using GarageBand (Apple Inc., Cupertino, CA, USA).

Playback devices.- We selected three playback devices representing the range of equipment used for biological monitoring. We used a smartphone (iPhone 6s, Apple Inc., Cupertino, CA, USA), a Bluetooth speaker (Swimmer, Polk Audio, Baltimore, MD, USA), and a remote predator caller (Inferno, FoxPro, Lewiston, PA, USA). We broadcast playback audio 10 times from each device toward a sound level meter set to "A" weighting (R8050, Reed Instruments, Wilmington, NC, USA) from 1 and 5 meters away. The sound level meter was calibrated with a 1 kHz tone at 94 dB (re 20 µPa) prior to use

(R8090, Reed Instruments). We measured the maximum amplitude (dB) of each vocalization broadcast and compared these values among playback devices. To judge device utility we generated frequency response curves by plotting amplitude measurements for each vocalization from each device.

Automated recording devices.- There are many recording devices for biologists to select from, including commercially available and custom created units. We tested three generations of SongMeter acoustic recorders (SM2+, SM3, and SM4, Wildlife Acoustics, Maynard MA, USA) because they are commonly used for monitoring or research (Digby et al. 2013; Yip et al. 2017a; Yip et al. 2017b). We chose the manufacturer's default settings to record, with the exception of sample rate, which we set to 22 kHz. To calibrate SongMeters we recorded a 1 kHz tone at 94 dB played directly into all microphones, then used Raven (version 1.5.0) to measure the amplitude of the 1 kHz tone recorded. The difference in amplitude from 94 dB within the recording represents the individual sensitivity of each microphone on each device. We account for this difference by adding or subtracting equivalent dB to the amplitude measurements of audio recordings in subsequent experiments. For example, the SM4 microphones produced recordings of the calibration tone (1 kHz at 94 dB) that measured 110.3 dB (both left and right channels) within Raven. Thus, we subtracted 16.3 dB from that estimate of amplitude in all subsequent experiments. This is equivalent to removing the effect of the default internal amplifier (+16 dB gain) without sacrificing signal to noise ratio. We secured each ARD to a structural object, then broadcast playback audio from the remote predator caller mounted on a tripod 1 m above the ground 10 times every 50 m on a 1 km unpaved road

(n=200 playbacks). We then used Raven to measure the amplitude of the recorded calls for comparison among species and recorders.

Effect of habitat on detection.- We performed the range-testing experiment between 0800 and 2200 hrs, October 25th to November 27th 2017. We avoided performing this study during the spring breeding season to reduce likelihood that resident target organisms might be calling coincidentally and be mistaken for broadcast signals. We established 30 1 km transects using a geographic information system (GIS) within five different habitat types. We performed six 1 km transects each in 1) mature forest with a mechanically thinned understory, 2) mature forest with an unmanaged understory, 3) post-wildfire recovering forest, 4) prairie, and 5) unpaved roads. Our previous experiments indicated the SM4 and remote predator caller were the most sensitive and least biased devices for playback and detection, and thus we utilized only these devices throughout the range-testing experiment. We attached a SongMeter (SM4) to a structural object at the start of each transect and broadcast vocalizations twice every 100 m (n=20 playbacks per transect). Playbacks were broadcast using the remote predator caller mounted on a tripod 1 m above the ground. We used a Garmin 64st handheld GPS (Garmin International, Inc., Olathe, KA, USA) to maintain the appropriate line transect path and ensure that the remote predator caller was oriented towards the ARD. In order to model the effects of environmental variables, we recorded wind speed (kph), air temperature (C), and relative humidity (%) using a Kestrel 3000 (Kestrel Instruments, Boothwyn, PA, USA). We obtained barometric pressure from the nearest weather station (Giddings Airport, 25 km east of the GLR). To account for an effect of the rolling topography of the GLR we determined if there was an unobstructed line-of-sight (LOS)

between the playback device and recorder. We projected each transect in a GIS onto a digital elevation model (National Elevation Dataset 2013, U.S. Geological Survey) and extracted the transect elevation. We calculated a linear regression slope from the SM4 to each transect point and added 1 m to account for source and receiver height, providing an estimate of the elevation that cannot be exceeded to maintain LOS. This method approximates a hypothetical scenario in which the topography between our source and receiver is perfectly flat. If any elevation between source and receiver exceeded this estimate, we considered LOS to be obstructed. We estimated this for all 10 stops along each transect and scored LOS as 0 = obstructed or 1 = unobstructed.

   Extraction of acoustic measurements.- All acoustic measurements and spectrograms were produced using Raven. To determine whether the call was detected, we isolated the playback audio from each distance along each transect. We visually examined spectrograms and, if needed, listened to recordings to determine the presence or absence of each vocalization. We manipulated all possible spectrogram settings (e.g., color, brightness, contrast, etc.) needed to be confident of our decision. Visual inspections were carried out by ARM and PSC, and detections were only recorded if both authors agreed they were visible within the spectrogram window. This resulted in 9,600 detection/non-detection events (30 transects, 10 distances, 16 calls, played twice). To estimate masking effects of background noise (dB), we measured one second of the recording immediately before the start of the playback sequence, with frequency ranges adjusted to match each vocalization.

   Statistical models.- To estimate detection probability we used generalized linear models (GLMs) with binomial response distribution complementary log link function

(Baddeley et al. 2010), implemented in R.3.4.2 (R Core Team 2018). We treated all categorical variables (LOS, habitat, and species), as well as scaled and centered continuous variables (distance, air temperature, relative humidity, barometric pressure, wind speed, and background noise) as fixed covariates. *A priori* we developed a list of candidate models to examine in an information-theoretic framework using AIC (Burnham and Anderson 2002). We built a full model containing all measured variables, including an interaction between habitat, distance, and species. We then built another 21 models of reduced complexity, including an intercept-only model, to test hypotheses about the importance of background noise, distance, habitat, environmental, and landscape variables. We ranked models using Akaike's Information Criteria (AIC) using the package AICcmodavg (Mazerolle 2013) and considered models competitive if they were ≤ 2 AIC points of the top ranked model. Using estimates from the top ranked model, we predicted detection probabilities for all 16 sounds among the five habitats between 0 and 1000 m, with 95% confidence intervals.

Results

Among the three playback devices we tested, only the remote predator caller reproduced animal vocalizations at the volume required (94dB), without excessive variance among frequencies. The Bluetooth speaker reached appropriate volumes, unlike the smartphone which was approximately 20dB too quiet for our purposes. Yet both the Bluetooth speaker and the smartphone were found to possess an inherently biased frequency response (Fig. 5).

Our evaluation of differences among three generations of SongMeter revealed that the

SM2+ measured on average 10 dB lower than the SM3 and SM4. However, we observed

a large overlap in standard deviation among all three generations of SongMeter (Fig. 6).

We utilized the SM4 for our range-testing experiment due to its acceptable performance

and more convenient size.

We performed a total of 9,536 playbacks, of which 4,036 were detectable.

Missing data (n=64 playbacks) occurred due to SongMeters dividing continuous audio

into multiple files, or because property boundaries differed from those used when

designing transects. Model comparison indicated two competitive top models (i.e. ΔAIC

< 2.0; Table 6). The full model (AIC = 5423.19; Table 6) including all measured

variables and an interaction between habitat, distance, and species, was found to be less

parsimonious than the same model minus humidity (ΔAIC = 1.60; AIC = 5421.59; Table

6). Humidity was not a significant predictor of detection within the full model. It was

expected that an interaction between habitat and distance would be important, and the top

seven models within our comparison contain this predictor. The model with distance as

the only predictor fit the data better than the model with habitat as the only predictor, and

a model with just species out-performed both. However, the model including an

interaction of these three predictors alone performed better than any single predictor

alone (Table 6).

A high degree of variability in predicted detection probability was found among

species and habitats (Fig. 7). In general, we found that species may be arranged by

probability of detection from least to greatest according to the dominant frequency within

their call, with high frequency calls being least detectable and low frequency calls being

the most detectable (Fig. 7). However, this general trend is not without exception, as is exemplified by the interaction of species, habitat, and distance in our top model (Tables 6 and 7). The greatest predicted detection probability at 1000 m distance was 0.79 (95% CI: 0.56-0.95) for the Houston Toad within prairies, and then burned forests (0.62, 95% CI: 0.41-0.83; Supporting Figure S3). For comparison, we predicted equal detection probability (0.52, 95%CI: 0.32-0.75) for the call of the Arroyo Toad within both prairies and burned forests, illustrating that the influence of habitat type is not constant across all species and sounds (Supporting Figure S3). With the exception of these two calls, unpaved roads were found to attenuate acoustic signals the least (Figure 7; Supporting Figure S3). That is, unpaved roads allowed vocalizations to travel the farthest distance before predicted detection probability reached zero. Within each habitat type the species with the highest predicted probability of detection is highly contingent upon distance. With few exceptions, no single species remains the most easily detected within a single habitat type for the entirety of a 1 km transect (Figure 7; Supporting Figure S3).

Coefficients for our five habitat types decrease in the following order: road>prairie>burned>thinned>unthinned (Table 7, Supporting Figure S3). Prairie and road treatments did not differ from the reference category, the burned treatment, indicating open canopy habitat types attenuate sound similarly (Table 7). However, the unmanaged mature forest treatment differed significantly from all other treatments, indicating heterogeneity between our two closed canopy treatments as well. Further, as was hypothesized, the influence of distance upon these habitat types follows the same pattern as above (Table 7). Aside from categories of species and habitat, distance was the most influential variable estimated in the top model (Table 7). Detection probability

decreased when topography obstructed a LOS between the remote predator caller and ARD (Table 7).

Discussion

Detections of acoustic signals are influenced by the environment between the sound source and receiver (Darras et al. 2016; Selby et al. 2016). Our results indicate that the probability of detecting an acoustic signal by an ARD is highly variable among different habitat types. Predicted detection probability was reduced in closed habitats (i.e. thinned or unmanaged forest). Evidently, these habitat types have higher densities of physical structures that may impede or scatter sound, relative to open habitats (i.e. burned, prairies, or roads), which lack disruptive structures at heights greater than 1 m. In general, this is consistent with the findings of previous studies of the influence of habitat type on acoustic signals (Yip et al. 2017b; Fricke 1984). Our results are unique from previous studies in that they illustrate a clear interaction between habitat, distance, and species. In closed habitats, where disruptive structures occur, as distance increases between source and receiver we observed sound attenuation increase beyond the effects attributed to distance alone (Pacifici et al. 2008). This is seemingly due to an accumulation of disruptive structures within these habitats, as the sound source becomes further away from the receiver. Attenuation due to interactions between habitat and distance were strongest in the unmanaged mature forest treatment, and weakest along unpaved roads. The implications of these findings are broadly applicable for biological monitoring programs that utilize acoustic monitoring technology. Primarily, there is habitat-induced heterogeneity in detection probability, which is relevant in any study that

employs occupancy models or spatial capture-recapture models to estimate source density. Therefore, it is imperative to model these habitat-effects explicitly, further emphasizing the importance of methods that allow for the explicit modeling of detection probability. Without estimating the effect of habitat on detection probability, researchers run the risk of concluding reduced occurrence of vocalizing animals among habitats dense with structures disruptive to traveling sounds. While the emergence of acoustic sampling using ARDs offers observer-free monitoring of bird and anuran communities, they are not a panacea, as illustrated by the clear influence of habitat and distance on their performance.

For this study we chose to evaluate the vocalization of species that fall into two broad categories: species of conservation concern (i.e., federally endangered, or a candidate for listing), or species of ubiquity whose call is well studied (e.g., American bullfrog). For rare or endangered species that can be surveyed using acoustic methods, studies evaluating the efficacy of such approaches are imperative to species conservation and recovery. Predicted detection probability varied widely among species within and among habitats, when measured at the same distance. Nonetheless, all species illustrated a clear and similar pattern of relative sound attenuation with respect to habitat. In general, the trend within our results is that high frequencies (i.e. 7 kHz tone) decay most rapidly, traveling the shortest distance within all habitat types, and that low frequencies (i.e. 1 kHz tone) travel further. Exceptions to this general trend occur for the very lowest frequency sounds we broadcast (i.e. California red-legged frog and bullfrog), which showed reduced detection probability at distances <500 m relative to sounds with slightly higher dominant frequency (Figure 7). We initially hypothesized that this deviation may

be caused by sound masking due to increased ambient noise occurring at lower frequencies (Bee & Swanson 2007), however this was not observed within our data. Rather, ambient noise showed no apparent pattern with respect to frequency. These findings are particularly pertinent to indices of biodiversity that measure the abundance multiple species. Without accounting for variation in detection probability among species, researchers may conclude reduced abundances or more restricted distributions for animals that may simply be difficult to detect, relative to species that are easy to detect.

The pattern we observed among species of varying dominant frequencies in different habitats may support the hypothesis that animals might evolve to vocalize at frequencies that are favored by the surrounding habitat. This has been referred to as the "sound window" hypothesis (Morton 1975; Marten & Marler 1977). While our study site is home to many species of bird and anuran, it is primarily utilized for researching the Houston toad, whose vocalization carried further, in all three open habitat types, than all other species considered, providing additional anecdotal support for this hypothesis (Ey & Fischer 2009). One potential confound that complicates our study is that not all calls broadcast are the same length. Based on our results, one could argue that animals with the longest calls are more likely to be detected, but to our knowledge, this hypothesis is yet to be tested. However, some female anurans select for males with longer calls, perhaps because they are most easily detected (Cocroft & Ryan 1995).

One complication with most studies involving auditory surveys of vocalizing fauna is observer bias, or an analogous example from automated methods; bias among ARDs (Miller et al. 2012; Yip et al. 2017a). We found differences between the three

generations of SongMeters to be negligible among the SM3 and SM4, and that the SM2+ is on average 10 decibels less sensitive. The manufacturers of these devices state that improvements in signal to noise ratio have occurred with the introduction of each new model, from >62dB in the SM2+, >68dB in the SM3, to ~80dB in the SM4. Despite these improvements we observed a large amount of overlap in sensitivity along both distance and frequency gradients. Although we failed to distinguish an optimal unit among the recording devices, we feel researchers should always evaluate the sensitivity and performance of their devices through range-testing, prior to choosing a recording platform (i.e. model of recorder). Previous studies have illustrated that variation in sensitivity exists among microphones of varying use and age, within a single model of ARD, further necessitating the need for calibration prior to deployment (Turgeon et al. 2017). Adoption of this approach should provide researchers with improved replicability and the ability to quantify error in their estimates of abundance or biodiversity that might be caused by ARD choice alone. By comparing popular devices used for broadcasting animal vocalizations we discovered remarkable differences in the devices' ability to reliably reproduce vocalizations across a frequency gradient at a constant volume. Researchers utilizing this method vary widely in their selection of playback device, and rarely, if ever, provide readers with precise information about the frequency response or volume capabilities of their respective device.

Within our study, we found temperature, wind speed, and barometric pressure to be significant predictors of detection probability. For these factors variation is caused by both prevailing atmospheric conditions as well as habitat type. For example, during high winds all surveys in open habitats will be impacted, whereas closed habitats will not

suffer additional attenuation due to increased winds by virtue of their inherent sound interference. That is to say, the same disruptive structures that obscure sound will also obscure wind. With respect to variation due to temperature and pressure, the broad scale atmospheric impacts on sound dampening are well understood (Lawrence & Simmons 1982; Öhlund & Larsson 2015). Within our study, temperature shares an apparent relationship with transect, where each transect experiences a unique and independent series of temperatures, and temperatures may not overlap at all among transects. We are unable to examine this relationship in detail because we did not repeat surveys within the same transect across multiple days, or across a gradient of atmospheric conditions.

We chose to control for effects related to rolling topography on the GLR by calculating LOS for each distance between sound source and receiver, across all transects. In nature, animals have been documented to overcome these topographical obstacles by seeking perches (Kime et al. 2000; Mathevon et al. 2005). Although our measurements reflect the variation that practitioners might consider among species and habitats, we caution against considering our findings calibration or correction factors that could be applied to future studies for these reasons. Additionally, the monitoring of living animals, rather than technological homologs such as playback devices, includes stochasticitystochasticity that cannot be controlled for in most cases, such as the direction, volume, and structure of real animal vocalizations.

To maximize detection probability, researchers should minimize the distance between sound source and receiver, especially within habitats featuring extensive disruptive structures. When monitoring anuran populations, ARDs are typically placed adjacent to water bodies used for congregations of breeding adults. At small wetlands and

ponds, the minimized distance between sound source and receiver should overcome any problems associated heterogeneity in detection probability caused by habitat or species. However, when monitoring large wetlands that may sample a variable amount of the available anuran habitat, this may not be true. Furthermore, avian monitoring is usually focused on habitat patches where individuals are less clustered. The effect of habitat type, species, and the distance between sound source and receiver on detection probability is likely to be more complicated in these monitoring scenarios, and caution should be used when evaluating the underlying assumptions of equal detectability.

To estimate and account for biases in detection probability due to distance and habitat, researchers may borrow techniques used in marine environments (Selby et al. 2016). Detection of marked individuals is achieved through implanted or attached acoustic transponders, and calibration of acoustic signals with respect to environmental conditions is achieved using fixed-location sentinel tags (Kessel et al. 2014). Researchers could achieve similar rigor in terrestrial environments using call or tone broadcasts from fixed locations at regular intervals, as described in this study. Alternatively, estimation of habitat effects can be achieved in situ using methods of spatial capture-recapture methods (Dawson & Efford 2009; Borchers et al. 2015, Measey et al. 2017). A subset of prior research has shown that human performed avian surveys are more effective than ARD surveys (Hutto & Stutzman 2009; Yip et al. 2017). However, human observers have been shown to produce biased estimates of density, as well as detection distance, during avian point counts (Alldredge et al. 2007b; Simons et al. 2007). It has been shown that the variability we observed in detection probabilities among different habitats holds true for human detection of acoustic signals as well, with the exception that detection radius can

be greater among human observers than in ARDs (Pacifici et al. 2008; Yip et al. 2017a). Nevertheless, the assumption that detection probability may be equal for different habitat types should likely be examined closely by future researchers.

Our results indicate pilot studies aimed directly at quantifying habitat and species-specific detection probabilities are valuable when attempting to achieve quality monitoring of avian or anuran populations, as we have shown these effects can create large variation in detection probability. While habitats, in general, may be open or closed, fine scale differences within each of these groups (e.g. roads and prairies), and potentially their constituent subcategories are responsible for a large amount of variation in detection probability. When utilizing well-established survey protocols (e.g., North American Amphibian Monitoring Program [NAAMP] or the North American Breeding Bird Survey [BBS]) careful examination and appreciation for the implicit assumptions about these sources of heterogeneity in probability of detection may be required, or researchers are at risk of failing to detect targeted taxa within seemingly uniform habitats.
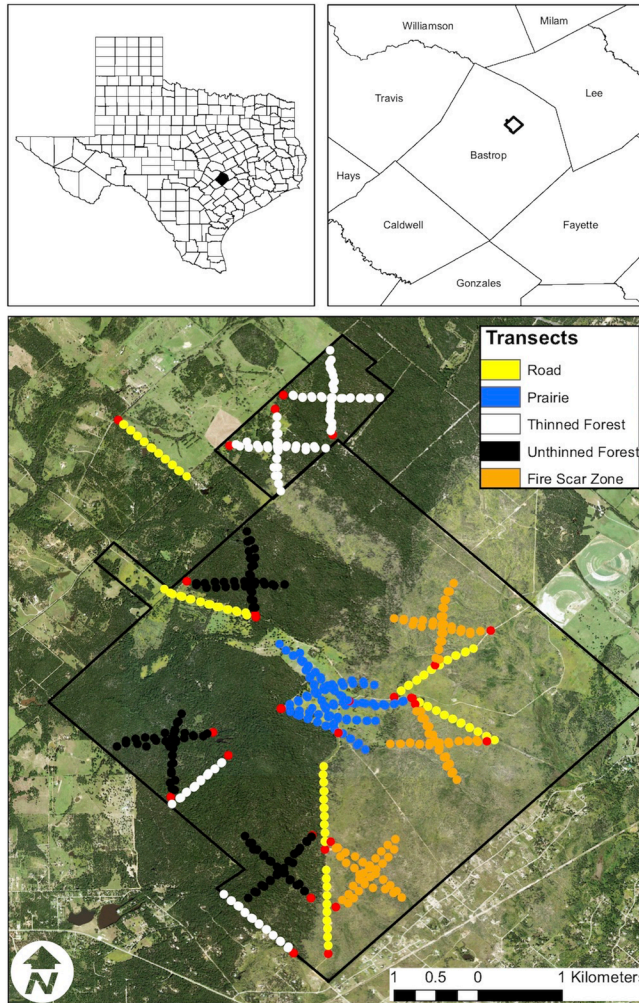
Figure 4. Aerial map of the Griffith League Ranch, Bastrop County, TX, USA. Top left: Texas, showing Bastrop County in solid black. Top right: Bastrop County, TX, with the Griffith League Ranch boundary in bold black line. Bottom: Satellite image of the Griffith League Ranch (outlined in black), with transect points in colors, representing habitat types, and red dots indicating automated recording device locations.
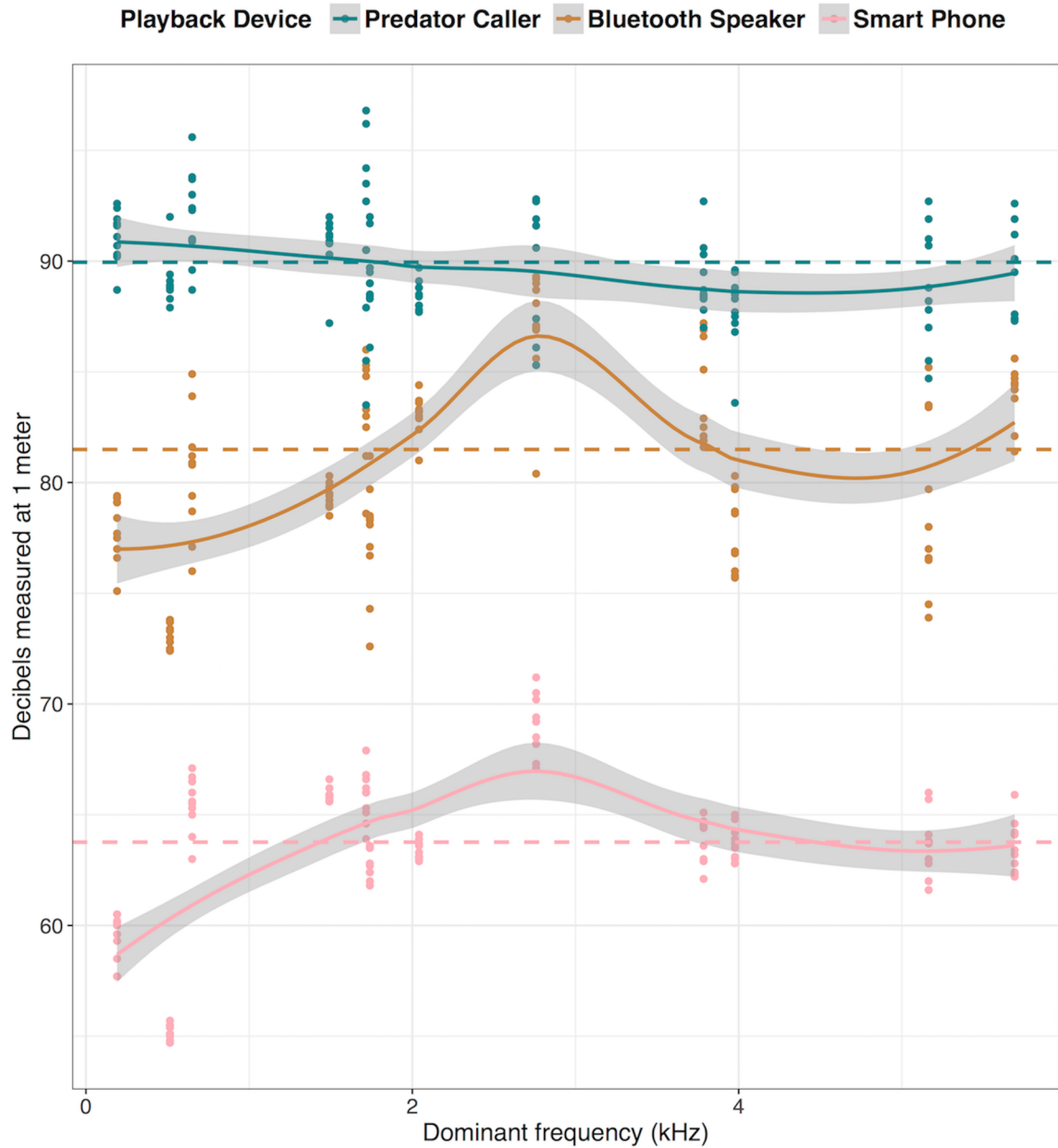
Figure 5. Playback device frequency response curves. Locally weighted smoothed scatterplot of the decibel level produced by each of the three playback devices tested, plotted for the 6 anuran and 6 bird species used within this study. Sounds are arranged along the x-axis according to dominant frequency from least to greatest. Colored dashed lines represent mean decibels produced by each playback device across all frequencies. Solid colored lines with grey envelope represent smoothed regression line and standard error.
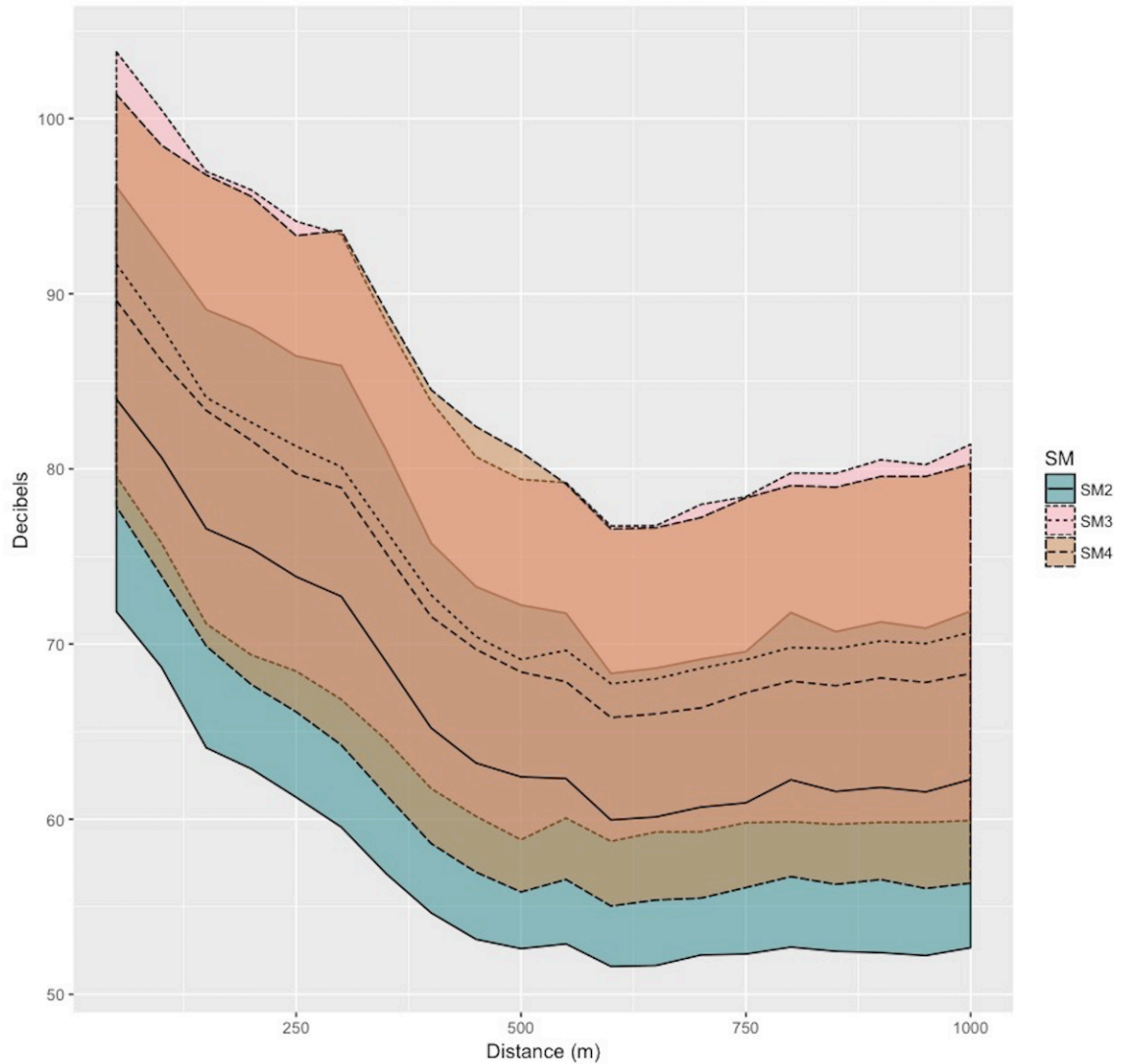
Figure 6. The received sound levels of three audio recording devices for detecting 6 birds, 6 anurans, and 4 pure tones as distance increases. Sounds were played using a FoxPro Inferno remote predator caller. Black lines represent mean decibel detected, line type represents recorder type, and shaded regions represent standard deviation for each device respectively.
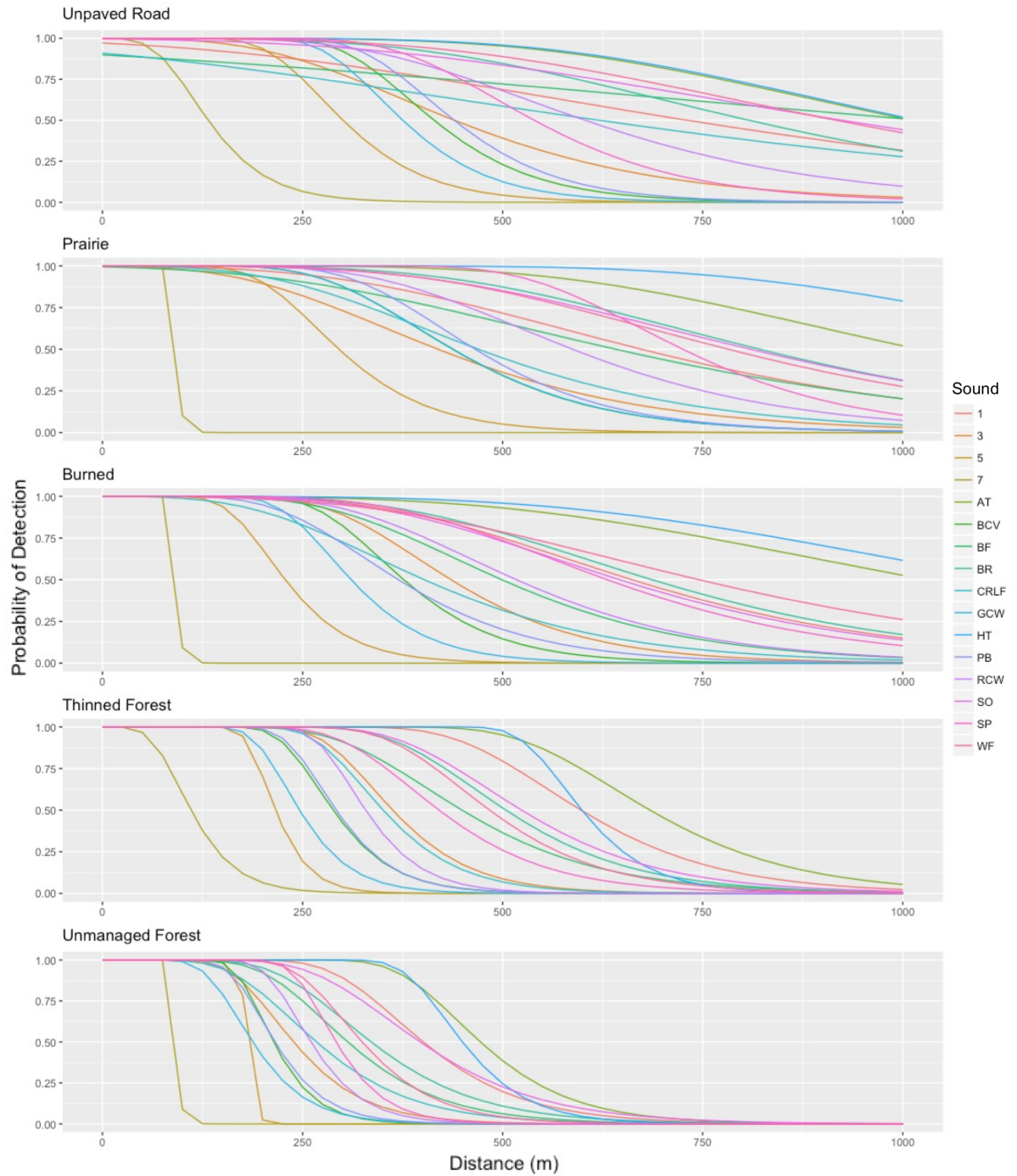
Figure 7. Predicted probability of detection for 16 sounds, among 5 habitat types, along a distance of 1000 m. Sounds include four pure sine waves, the vocalizations of six anurans, and six birds. Each sound was broadcast on a private ranch in Bastrop Country, Texas, USA, using a FoxPro Inferno predator calling device and recorded using a SongMeter SM4. Values of detection probability were predicted using the estimates from our top generalized linear model with an interaction between habitat, distance, and species.

Table 6. Model output and rankings. Generalized linear models (GLM) tested with the number of parameters in each model (Pars), difference in model AIC (dAIC), AIC weight (Wt.), and log-likelihood (LL); fixed categorical variables include line of sight (1 or 0), habitat (burned, prairie, road, thinned, unthinned), and species. Scaled and centered continuous fixed variables include distance = distance from recorder and source (m), temp. = air temperature (°C), hum. = relative humidity (%), pressure = barometric pressure (mmHg), wind = wind speed (kph), and noise = background noise (decibels, dB) measured 1 second prior to recording.

No.	Model statement	Pars.	ΔAIC	Wt.	LL
3	y ~ species * habitat * distance + temp. + wind + pressure + LOS + noise	165	0.00	0.69	-2545.8
1	y ~ species * habitat * distance + temp. + hum. + wind + pressure + LOS + noise	166	1.60	0.31	-2545.6
7	y ~ species * habitat * distance + temp. + hum. + wind + pressure + LOS	165	31.93	0.00	-2561.76
5	y ~ species * habitat * distance + temp. + hum. + wind + LOS + noise	165	44.13	0.00	-2567.86
4	y ~ species * habitat * distance + temp. + hum. + pressure + LOS + noise	165	108.99	0.00	-2600.29
6	y ~ species * habitat * distance + temp. + hum. + wind + pressure + noise	165	159.57	0.00	-2625.58
2	y ~ species * habitat * distance + hum. + wind + pressure + LOS + noise	165	314.76	0.00	-2703.17
10	y ~ species + habitat * distance + temp. + hum. + wind + pressure + LOS + noise	162	396.38	0.00	-2746.99
8	y ~ species * habitat * distance + LOS + noise	160	507.83	0.00	-2804.71
9	y ~ species * habitat * distance	31	539.01	0.00	-2949.3
11	y ~ species + habitat * distance + LOS + noise	27	931.64	0.00	-3149.61
13	y ~ species * habitat + distance + temp. + hum. + wind + pressure + LOS + noise	87	984.28	0.00	-3115.93
17	y ~ species + habitat + distance + temp. + hum. + wind + pressure + LOS + noise	27	1016.63	0.00	-3192.11
14	y ~ species * habitat + distance + LOS + noise	83	1433.97	0.00	-3344.78
18	y ~ species + habitat + distance + LOS + noise	23	1440.81	0.00	-3408.2
22	y ~ 1	1	9079.56	0.00	-7249.58
21	y ~ species	16	8162.25	0.00	-6775.92
20	y ~ distance	2	3512.68	0.00	-4465.13
19	y ~ habitat	5	8828.17	0.00	-7119.88
16	y ~ habitat : distance + species : distance + distance	21	3233.93	0.00	-4306.76
15	y ~ species * habitat	80	7937.71	0.00	-6599.65
12	y ~ habitat * distance	10	2809.28	0.00	-4105.44

Table 7. Summary of the selected top model. Table showing the estimate, standard error, z-value, and P-value for the fixed factors species, habitat, distance, temperature, wind, pressure, line of sight, and background noise. 1kHz, burned, and 0 (impeded) were used as reference categories for species, habitat, and line of sight, respectively. Values for interaction terms are given in Supporting Material S4.

Fixed Effects	Estimate	SE	z value	Pr(>|z|)
(Intercept)	-0.139	0.159	-0.870	0.384
Species Effects				
3kHz	-1.053	0.298	-3.529	<0.001
5kHz	-4.941	1.068	-4.627	<0.001
7kHz	-61.015	1902.742	-0.032	0.974
Arroyo toad	0.576	0.215	2.682	0.007
black-capped vireo	-1.828	0.440	-4.152	<0.001
bull frog	-0.616	0.257	-2.393	0.017
black rail	0.086	0.223	0.386	0.700
California red-legged frog	-1.207	0.300	-4.021	<0.001
golden-cheeked warbler	-3.028	0.668	-4.530	<0.001
Houston toad	0.754	0.219	3.448	0.001
painted bunting	-1.619	0.365	-4.435	<0.001
red-cockaded woodpecker	-0.503	0.254	-1.977	0.048
spotted owl	-0.053	0.226	-0.232	0.817
spring peeper	-0.010	0.229	-0.043	0.965
wood frog	0.062	0.217	0.288	0.773
Habitat effects				
prairie	-1.519	0.334	-4.553	<0.001
road	0.316	0.280	1.128	0.259
thinned	-0.131	0.219	-0.596	0.551
unthinned	-0.253	0.206	-1.224	0.221
distance	-1.233	0.219	-5.632	<0.001
temperature	-0.484	0.026	-18.502	<0.001
wind	-0.262	0.025	-10.653	<0.001
pressure	-0.224	0.027	-8.434	<0.001
line of sight	0.664	0.053	12.596	<0.001
noise	-0.169	0.028	-5.930	<0.001

References

Aide, T., C. Corrada-Bravo, M. Campos-Cerqueira, C. Milan, G.Vega, and R. Alvarez. 2013. Real-time bioacoustics monitoring and automated species identification. *PeerJ* 1:e103.

Alldredge, M., K. Pollock, T. Simons, J.Collazo, S. Shriner, and D. Johnson. 2007. Time-of-detection method for estimating abundance from point-count surveys. *The Auk* 124(2):653–664.[a]

Alldredge, M., T. Simons, and K. Pollock. 2007. A field evaluation of distance measurement error in auditory avian point count surveys. *Journal of Wildlife Management* 71(8):2759-66.[b]

Baddeley, A., M. Berman, N. Fisher, A. Hardegen, R. Milne, D. Schuhmacher, R. Shah, & R. Turner. 2010. Spatial logistic regression and change-of-support in Poisson point processes. *Electronic Journal of Statistics* 4:1151–1201.

Bader, E., K. Jung, E. Kalko, R. Page, R. Rodriguez, and T. Sattler. 2015. Mobility explains the response of aerial insectivorous bats to anthropogenic habitat change in the Neotropics. *Biological Conservation* 186:97-106.

Bee, M., and E. Swanson. 2007. Auditory masking of anuran advertisement calls by road traffic noise. *Animal Behaviour* 74(6):1765-76.

Bibby, C., and S. Buckland. 1987. Bias of bird census results due to detectability varying with habitat. *Acta Oecologica* 8(2):103-12.

Borchers, D., B. Stevenson, D. Kidney, L. Thomas, and T. Marques. 2015. A unifying model for capture–recapture and distance sampling surveys of wildlife populations. *Journal of the American Statistical Association* 110(509):195-204.

Bormpoudakis, D., J. Sueur, and J. Pantis. 2013. Spatial heterogeneity of ambient sound
at the habitat type level: ecological implications and applications. *Landscape
Ecology* 28(3):495-506.

Brown, D., W. Nowlin, E. Ozel, I. Mali, D. Episcopo, M. Jones, and M. Forstner. 2014.
Comparison of short term low, moderate, and high severity fire impacts to aquatic
and terrestrial ecosystem components of a southern USA mixed pine/hardwood
forest. *Forest ecology and management* 312:179-92.

Buckland, S.T., D. Anderson, K. Burnham, J. Laake, D. Borchers, & L. Thomas. 2001.
*Introduction to Distance Sampling: Estimating Abundance of Biological
Populations.* Oxford University Press, Oxford, UK.

Charif, R., 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 08-2.

Cocroft, R., and M. Ryan. 1995. Patterns of advertisement call evolution in toads and
chorus frogs. *Animal Behaviour* 49(2):283-303.

Darras, K., P. Pütz, K. Rembold, and T. Tscharntke. 2016. Measuring sound detection
spaces for acoustic animal sampling and monitoring. *Biological Conservation*,
201, 29-37.

Dawson, D., and M. Efford. 2009. Bird population density estimated from acoustic
signals. *Journal of Applied Ecology* 46(6):1201-9.

Digby, A., M. Towsey, B. Bell, and P. Teal. 2013. A practical comparison of manual and
autonomous methods for acoustic monitoring. *Methods in Ecology and Evolution*
4(7):675-83.

Embleton, T. 1996. Tutorial on sound propagation outdoors. *The Journal of the Acoustical Society of America* 100(1):31-48.

Ey, E. and J. Fischer. 2009. The "acoustic adaptation hypothesis"—a review of the evidence from birds, anurans and mammals. *Bioacoustics* 19(1-2):21-48.

Fricke, F. 1984. Sound attenuation in forests. *Journal of Sound and Vibration* 92(1):149-158.

Gasc, A., B. Gottesman, D. Francomano, J. Jung, M. Durham, J. Mateljak, and B. Pijanowski. 2018. Soundscapes reveal disturbance impacts: biophonic response to wildfire in the Sonoran Desert Sky Islands. *Landscape Ecology* 33(8):1399-1415.

Hsu, M., Y. Kam, and G. Fellers. 2005. Effectiveness of amphibian monitoring techniques in a Taiwanese subtropical forest. *The Herpetological Journal* 15:73-79.

Hutto, R. 2016. Should scientists be required to use a model-based solution to adjust for possible distance-based detectability bias? *Ecological Applications* 26(5):1287-1294.

Hutto, R., and R. Stutzman. 2009. Humans versus autonomous recording units: a comparison of point-count results. *Journal of Field Ornithology* 80(4):387-98.

James, M., M. Stockwell, J. Clulow, S. Clulow, and M. Mahony. 2015. Investigating behaviour for conservation goals: conspecific call playback can be used to alter amphibian distributions within ponds. *Biological Conservation* 192:287-93.

Kearns, G., N. Kwartin, D. Brinker, and G. Haramis. 1998. Digital Playback and Improved Trap Design Enhances Capture of Migrant Soras and Virginia Rails. *Journal of Field Ornithology* 69(3):466-73.

Kessel, S., S. Cooke, M. Heupel, N. Hussey, C. Simpfendorfer, S. Vagle, and A. Fisk. 2014. A review of detection range testing in aquatic passive acoustic telemetry studies. *Reviews in Fish Biology and Fisheries* 24(1):199-218.

Kime, N., W. Turner, and M. Ryan. 2000. The transmission of advertisement calls in Central American frogs. *Behavioral Ecology* 11(1):71-83.

Lawrence, B., and J. Simmons. 1982. Measurements of atmospheric attenuation at ultrasonic frequencies and the significance for echolocation by bats. *The Journal of the Acoustical Society of America* 71(3):585-90.

Lefcheck, J. 2016. piecewiseSEM: Piecewise structural equation modelling in r for ecology, evolution, and systematics. *Methods in Ecology and Evolution* 7(5):573-9.

Lehmann, G., K. Frommolt, A. Lehmann, and K. Riede. 2014. Baseline data for automated acoustic monitoring of Orthoptera in a Mediterranean landscape, the Hymettos, Greece. *Journal of insect conservation* 18(5):909-25.

Llusia, D., R. Márquez, and R. Bowker. 2011. Terrestrial sound monitoring systems, a methodology for quantitative calibration. *Bioacoustics* 20(3):277-86.

MacKenzie, D., J. Nichols, G. Lachman, S. Droege, J. Royle, and C. Langtimm. 2002. Estimating site occupancy rates when detection probabilities are less than one. *Ecology* 83(8):2248-2255.

Mannan, R., G. Perry, D. Andersen, and C. Boal. 2014. Call broadcasting and automated recorders as tools for anuran surveys in a subarctic tundra landscape. *The Journal of North American Herpetology* 1:47-52.

Marten, K., and P. Marler. 1977. Sound transmission and its significance for animal

   vocalization. *Behavioral ecology and sociobiology* 2(3):271-90.

Mathevon, N., T. Dabelsteen, and S. Blumenrath. 2005. Are high perches in the blackcap

   Sylvia atricapilla song or listening posts? A sound transmission study. *The

   Journal of the Acoustical Society of America* 117(1):442-9.

Mazerolle, M. 2013. AICcmodavg: model selection and multimodel inference based on

   (Q) AIC (c). R package version, 1, 35.

McClintock, B., L. Bailey, K. Pollock, and T. Simons. 2010. Experimental investigation

   of observation error in anuran call surveys. *The Journal of Wildlife

   Management* 74(8):1882-1893.

Measey, G., B. Stevenson, T. Scott, R. Altwegg, and D. Borchers. 2017. Counting chirps:

   acoustic monitoring of cryptic frogs. *Journal of Applied Ecology* 54(3):894-902.

Mielke, A., and K. Zuberbühler. 2013. A method for automated individual, species and

   call type recognition in free-ranging animals. *Animal behaviour* 86(2):475-82.

Miller, D., L. Weir, B. McClintock, E. Grant, L. Bailey, and T. Simons. 2012.

   Experimental investigation of false positive errors in auditory species occurrence

   surveys. *Ecological Applications* 22:1665-1674

Morton, E. 1975. Ecological sources of selection on avian sounds. *The American

   Naturalist* 109(965):17-34.

Nelson, B. 2003. Reliability of sound attenuation in Florida scrub habitat and behavioral

   implications. *The Journal of the Acoustical Society of America* 113(5):2901-11.

Öhlund, O., and C. Larsson. 2015. Meteorological effects on wind turbine sound

   propagation. *Applied Acoustics* 89:34-41.

Oseen, K., and R. Wassersug. 2002. Environmental factors influencing calling in

   sympatric anurans. *Oecologia* 133(4):616-25.

Pacifici, K., T. Simons, and K. Pollock. 2008. Effects of vegetation and background

   noise on the detection process in auditory avian point count surveys. *The Auk*

   125:600-607.

Piercy, J., E. Codling, A. Hill, D. Smith, and S. Simpson. 2014. Habitat quality affects

   sound production and likely distance of detection on coral reefs. *Marine Ecology*

   Progress Series 516:35-47.

Pieretti, N., A. Farina, and D. Morri. 2011. A new methodology to infer the singing

   activity of an avian community: The Acoustic Complexity Index

   (ACI). *Ecological Indicators*, 11(3):868-873.

Pijanowski, B., L. Villanueva-Rivera, S. Dumyahn, A. Farina, B. Krause, B. Napoletano,

   S. Gage, and N. Pieretti. 2011. Soundscape ecology: the science of sound in the

   landscape. *BioScience* 61(3):203-16.

R Core Team. 2018. R: A language and environment for statistical computing. R

   Foundation for Statistical Computing, Vienna, Austria.

Rempel, R., C. Francis, J. Robinson, and M. Campbell. 2013. Comparison of audio

   recording system performance for detecting and monitoring songbirds. Journal of

   Field Ornithology 84(1):86-97.

Richards, D., and R. Wiley. 1980. Reverberations and amplitude fluctuations in the

   propagation of sound in a forest: implications for animal communication. *The*

   American Naturalist 115(3):381-99.

Romer, H., and J. Lewald. 1992. High-frequency sound transmission in natural habitats: implications for the evolution of insect acoustic communication. *Behavioral Ecology and Sociobiology* 29(6):437-44.

Rose, G. and E. Brenowitz. 1991. Aggressive thresholds of male Pacific treefrogs for advertisement calls vary with amplitude of neighbors' calls. *Ethology* 89(3):244-252.

Sadoti, G., K. Johnson, J. Smith, and N. Petersen. 2018. Influences of spatial variation in vegetation on avian richness and abundance vary by season in the Chihuahuan Desert. *Journal of Arid Environments* 151:49-57.

Selby, T., K. Hart, I. Fujisaki, B. Smith, C. Pollock, Z. Hillis-Starr, I. Lundgren, and M. Oli. 2016. Can you hear me now? Range-testing a submerged passive acoustic receiver array in a Caribbean coral reef habitat. *Ecology and evolution* 6(14):4823-35.

Simons, T., M. Alldredge, K. Pollock, and J. Wettroth. 2007. Experimental analysis of the auditory detection process on avian point counts. *The Auk* 124(3):986-99.

Sólymos, P., S. Matsuoka, E. Bayne, S. Lele, P. Fontaine, S. Cumming, D. Stralberg, F. Schmiegelow, and S. Song. 2013. Calibrating indices of avian density from non-standardized survey data: making the most of a messy situation. *Methods in Ecology and Evolution* 4(11):1047-1058.

Stewart, M. and P. Bishop. 1994. Effects of increased sound level of advertisement calls on calling male frogs, *Eleutherodactylus coqui*. *Journal of Herpetology* 28(1):46-53.

Sueur, J., A. Farina, A. Gasc, N. Pieretti, and S. Pavoine. 2014. Acoustic indices for
biodiversity assessment and landscape investigation. *Acta Acustica united with
Acustica* 100(4):772-81.

Turgeon, P., S. Van Wilgenburg, and K. Drake. 2017. Microphone variability and
degradation: implications for monitoring programs employing autonomous
recording units. *Avian Conservation and Ecology* 12(1):9.

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

Vold, S., C. Handel, and L. McNew. 2017. Comparison of acoustic recorders and field
observers for monitoring tundra bird communities. *Wildlife Society
Bulletin* 41(3):566-576.

Wiggins, S., and J. Hildebrand. 2007. High-frequency Acoustic Recording Package
(HARP) for broad-band, long-term marine mammal monitoring. In *Underwater
Technology and Workshop on Scientific Use of Submarine Cables and Related
Technologies Symposium*, 551-557.

Wimmer, J., M. Towsey, P. Roe, and I. Williamson. 2013. Sampling environmental
acoustic recordings to determine bird species richness. *Ecological Applications*
23(6):1419-28.

Yip, D., E. Bayne, P. Sólymos, J. Campbell, and D. Proppe. 2017. Sound attenuation in
forest and roadside environments: Implications for avian point-count surveys. *The
Condor* 119(1):73-84.[b]

Yip, D., L. Leston, E. Bayne, P. Sólymos, and A. Grover. 2017. Experimentally derived

    detection distances from audio recordings and human observers enable integrated

    analysis of point count data. *Avian Conservation and Ecology* 12(1):11.

Zimmerman, B. 1994. Audio Strip Transects. In R. Heyer, M. Donnelly, M. Foster, R.

    Mcdiarmid (Eds). *Measuring and monitoring biological diversity: standard*

    methods for amphibians. Smithsonian Institution.

IV. COMPARISON OF THE MORPHOLOGY AND VOCALIZATION OF THE
HOUSTON TOAD AND THE DWARF AMERICAN TOAD WITH IMPLICATIONS
ON THEIR HISTORIC RANGE IN TEXAS.

Introduction

Vertebrates are currently experiencing range contractions and population declines globally, contributing to our sixth mass extinction event (Ceballo et al. 2018). Among declining vertebrates, amphibians are one of the groups reported to be most at risk (Wake et al 2008; Bishop et al. 2012). In response to these steep declines, nearly 300 species have been proposed as candidates for the U.S. Endangered Species Act (U.S. Endangered Species Act [ESA] 1973, as amended). Despite an effort by researchers to place species on the endangered list, a recent study found that approximately 13% of candidate species possessed any kind of pre-listing conservation plan (BenDor et al. 2017). The basis of listing decisions for candidate species is under a directive to follow the best available science ([ESA] 1973, as amended), but for many species these data are not available; furthermore, agencies often demonstrate a preference for uncertain findings regarding candidate species (Murphy et al 2016).

Conservation biologists rely on historical data to document changes in the distribution and abundance of species (Skelly et al. 2003). Long-term monitoring is the ideal way to document these changes, but these data do not exist for most species of concern. In the absence of these data, existing information on the historical distribution of a species may be gleaned from natural history collections, which are the only verifiable source of species presence at a given time and place. For species of conservation concern

these data are routinely collated and compared to contemporary observations, describing

expansion or contraction in geographic distribution (Shaffer et al. 1998). Surprisingly, for

the United States' first federally protected amphibian, the Houston Toad (*Bufo*

[=*Anaxyrus*] *houstonensis*; Gottschalk 1970; Sanders 1953), these data have not been

examined.

The Houston Toad (*B. houstonensis*) is a diminutive member of the *Bufo*

americanus complex, occurring in East-central Texas, USA (Sanders 1953; Masta et al.

2002). The adults of the species show a preference for deep sandy soils where forest or

woodland cover occurs proximal to breeding ponds (Potter et al. 1984). In its description,

the species' range is hypothesized to extend North and East to Arkansas and Oklahoma,

following the occurrence of these features (Sanders 1953). However, the eastern limit for

this species' distribution has been speculated to occur at the Trinity River, citing the lack

of detections beyond this perceived barrier (Seal 1994). The historic range of *B.*

houstonensis includes occupied habitat, where toads are detected annually (e.g., Bastrop

& Robertson counties), and locations where only a single or few detections have occurred

(e.g., Freestone & Brazos counties). Additionally, extirpation of the species has been

reported within a portion of this range (e.g., Liberty, Harris, and Fort Bend counties;

Sanders 1953; Price & Yantis 1993). Long-term monitoring for the species has taken

place within occupied counties only, but monitoring occurs only opportunistically within

adjacent counties.

The Houston Toad is reported to be morphologically distinct from its congeners

(Sanders 1953). The nearest relative to *B. houstonensis* is *Bufo* [=*Anaxyrus*] *americanus*

charlesmithi. Historically, *B. a. charlesmithi* was thought to be excluded from Texas via

the Red River, occurring only in Oklahoma (Sanders 1953), but this has since been disproven (Dixon 2013). Putatively these species are allopatric within Texas. Morphologically they are differentiated on the basis of *B. houstonensis* possessing enlarged cranial features and warts on the hind limbs (Sanders 1953). However, it has been suggested that misuse of the largest possible specimen as the Holotype created this notion, when in natural populations these features are highly variable (Seal 1994).

Here we report the existence of additional, historically overlooked, museum records for *B. houstonensis* in the context of its putative historic range, and discuss errors associated with the curation of these specimens whose identity and nomenclature have been ephemeral. In order to overcome errors associated with misidentification or curation of museum records, we qualitatively review the specific epithet given to all specimens that might indicate Houston Toad presence outside of its accepted historical range. Additionally, we attempt to differentiate between formalin fixed, preserved, voucher specimens of *B. houstonensis* and *B. a. charlesmithi* using morphological characters, as well as advertisement call characteristics.

Methods

Unresolved Museum Records.— In order to locate museum vouchers for *B. houstonensis* and *B. a. charlesmithi* we accessed VertNet (vertnet.org), a collaborative data portal funded by the National Science Foundation that makes biodiversity information, including specimen collections records, freely and easily accessible to the public. We requested photos of the dorsal, ventral, and cranial features, for each vouchered specimen reported by VertNet to have been collected outside of the known

range for these species (Table 8, Figure 8). We qualitatively evaluated these photographs to determine if specimens may have been misidentified at the time of collection, or accession into museum collections, based on the presence or absence of diagnostic morphological characters using multiple dichotomous keys (Tipton et al. 2012; Powell et al. 2012; Dixon 2013). We qualified photos based on whether specimens possessed parotoid glands closest at midpoint, versus closest anteriorly; whether warts of the tibia were larger than those of the thigh, or uniform in size on both tibia and thigh; whether post-orbital cranial crests were enlarged relative to the supra-orbitals, or equal in size to the supra-orbitals; whether the chest was spotted, pale, or entirely black; and the number of dorsal warts encircled in dark pigment. Specimens that have been formalin fixed, and stored in ethanol can suffer shrinking, decalcification of bone, and darkening or fading of pigmentations (Simmons 1995). Thus, many of the characters used to differentiate among North American Bufonids may be absent or indistinguishable, obscuring our ability to reliably identify toads from photographs. Sex of these specimens was not able to be determined in the photograph, and was not provided by the collectors for most specimen evaluated. Thus, sex was not utilized as a predictor within our analyses

Morphometric Comparison.—We tested a suite of morphometric measurements for their ability to differentiate between these two species, once fixed in preservatives. We measured eight morphological features present among vouchers of *B. houstonensis* (n=51) and *B. a. charlesmithi* (n=50), which have been previously reported as capable of diagnosing species assignment as well as hybrid status for *B. houstonensis* and some of its congeners (Hillis et al. 1984). Additionally, we included 39 Coastal Plains Toads (*B.*

[=*Icilius*] *nebulifer*), which are known to be morphologically unique, to serve as a
control. The features we measured included head width (HW; at widest part of head),
distance between interocular crests (DBIC; at widest point), mean parotoid length (MPL;
horizontal length of each parotoid gland), mean parotoid width (MPW; maximum width
of parotoid glands), mean length of tibiofibula (MTFL; distance between knee and top of
foot), snout urostyle length (SUL; tip of snout to posterior end of urostyle), mean distance
between parotoid and postorbital ridge (MPPG; the gap between anterior of parotoid and
the postorbital cranial feature, at its narrowest point), and mean thickness of postorbital
ridge (MPT; measured at its widest point anteriorly to posteriorly). For characteristics
that could be measured on both the right and left side, a mean was taken. All
morphological characters were measured by A. R. MacLaren using Mitutoyo digital
calipers to a precision of 0.01 mm. To control for the effect of body size we performed
residuals analysis using ordinary least-squares regression resulting in morphological
characters adjusted for body size (i.e., snout-urostyle length) for each specimen
measured. We conducted linear discriminant function analysis (LDA) to visualize size-
adjusted morphological characters that may discriminate between species. We treated
species as *a priori* groups to test for the presence of differentiability between *B.
houstonensis* and *B. a. charlesmithi*, once in collection (i.e., after preservation). We
calculated pairwise Pearson's correlation coefficient for all morphological characters to
test for collinearity.

Advertisement Call Comparison.—We also examined advertisement call
structure of *B. houstonensis* and *B. a. charlesmithi*. Geographically dispersed populations

of Houston Toads were utilized, as well as a single population of *B. a. charlesmithi*.

Houston Toad vocalizations were collected from one site within Bastrop, and Robertson counties, Texas, USA. *B. a. charlesmithi* vocalizations were collected in Cleveland county, Oklahoma, USA. In Texas, we recorded vocalizations using a SongMeter SM3 (Wildlife Acoustics, Maynard MA, USA). Deployment of these devices follows the methods outlined in MacLaren et al. (2018A). In Oklahoma, *B. a. charlesmithi* calls were recorded using a hand held digital recording device (Tascam DR05, Montabella CA, USA) which was oriented in the direction of the chorusing individuals. As is the case with other remote acoustic monitoring efforts, it is difficult or even impossible to determine how many individuals can be heard in any given recording. Therefore, we are unable to control for variation at the individual level, as is the case in some studies of this nature (Moriarty and Cannatella 2004). Likewise, in Oklahoma, although the recording device is hand held, not all animals chorusing were visually observed, or counted in a systematic manner. Ultimately, we collated and described 50 Houston Toad individual calls, and 49 *B. a. charlesmithi* individual calls, from an unknown number of individuals at each location. We extracted characteristics from each vocalization in program Raven Pro (V.1.5, Cornell University; Bioacoustics Research Program 2014). These characteristics include call dominant frequency (DF; frequency in Hz at peak call amplitude), call length (CL; duration in seconds from first call pulse to last call pulse), pulse number (PN; the number of pulses in each call divided by the call length), and frequency range (FR; the number of Hz between lowest and highest frequency) to provide an estimate for whether each call is narrow or broad in its frequency breadth. We explored patterns in advertisement call data by plotting univariate box and whisker plots

of raw data or performing a linear discriminant function analysis with R function "lda" (package "MASS", Venables & Ripley 2002; R Development Core Team 2018).

Results

Unresolved museum records were discovered for each species of toad both within and outside of their known historic range within the state of Texas. Dixon (2013) identifies 16 Texas counties where *B. a. charlesmithi* is reported to occur presently, or in the past. Of these 16, we were able to locate museum vouchers for 14 counties. Museum vouchers were discovered for an additional 12 counties, unreported by Dixon (2013). For *B. houstonensis* Dixon (2013) identifies 12 Texas counties where they have been previously found. Of these, we recovered museum records for 10, leaving two counties questionable. We recovered museum vouchers of *B. houstonensis* for an additional seven Texas counties.

Qualitative assessment of visible morphological features showed that museum vouchers putatively identified as *B. a. charlesmithi* were found to be mostly *B. woodhousei* (Woodhouses Toad), a more common and widely dispersed toad in Texas (Table 8; Powell et al. 2012; Dixon 2013), although records collected from Sabine, Panola, and Smith counties have seemingly been correctly identified as *B. a. charlesmithi* (Table 8). Collected in the 1930's, 60's and 2000's respectively, these records indicate that the range of *B. a. charlesmithi* might presently extend, or previously have extended, further south into Texas following the Sabine River (Figure 8).

Similarly, museum vouchers for *B. houstonensis* were comprised mostly of what must be misidentifications by the collectors in the field, or at the time of accession into

93

museum collections. Thirteen of the total 18 specimen photographs reviewed possessed qualitative traits of *B. woodhousei*. Specimens recovered for counties already known to currently, or previously, contain populations of *B. houstonensis* were correctly identified as such (Table 8, Figure 8). We discovered a single meaningful voucher that could be discerned as a preserved Houston Toad originating from outside its previously known range, which expands the historic distribution to include Brazos County, Texas (MacLaren and Forstner 2017).

Multiple specimens we sought to verify returned as photographs of juvenile animals, lacking many of the morphological features used to differentiate among these creatures (Table 8). A toad collected from Orange county, Texas, in 1968 was reportedly a voucher of *B. americanus*, but has since been cleared and stained, a preparation that also prevented us from qualifying the species identity due to lack of visible morphology (e.g., wart pigmentation).

Our attempt to replicate and expand the morphological comparison provided by Hillis et al. (1984) was generally inconclusive. The previous authors did not account for body size, with the exception of removing it from the analysis entirely through stepwise procedures. After conducting regression analysis, the remaining seven morphological characters were all highly correlated to one another. This is evident within the matrix of Pearson's correlation coefficients for these variables (Figure 9; Table 10). This may indicate several things. First, *B. houstonensis* and *B. a. charlesmithi* may truly be morphologically identical. Second, the morphological characters measured may simply not capture any signal of morphological differentiation present among the specimens considered here. Third, the use of a univariate measure of body size may be inappropriate

94

and perhaps should be expanded to include a multivariate consideration of snout-urostyle length and head-width (McCoy et al. 2006).

Advertisement call data indicate that *B. houstonensis* differs from *B. a. charlesmithi* with respect to dominant frequency. We found the mean dominant frequency for *B. houstonensis* to be 2033.95 ± 62.87, and 1891.7 ± 153.97 for *B. a. charlesmithi*. Other advertisement call features measured varied little among vocalizations collected for these two species (Table 9, Figure 10). Results from linear discriminant function analysis further reinforce that differences are present between the calls of these species when all four call characteristics are summarized (Figure 10).

Discussion

B. houstonensis was described prior to the establishment of modern conventions for the description of a new species (Sanders 1953; Winston 1999). This species uniqueness among contemporary congeners was determined based on a qualitative assessment of external morphological and skeletal features. The approach utilized by Sanders (1953) is not consistent with contemporary species descriptions. For example, primary evidence presented by Sanders (1953) hinges upon qualitative skeletal features. In fact, only two whole specimens were held as paratypes, while nine were skeletonized for this purpose. While Sanders' study provides much of the first general information about the species, it does not provide us with an assessment of the Houston Toad's morphology or call relative to its nearest relatives. Bragg (1954) described the Dwarf American Toad just one year later. Yet again this description is without a proper treatment of quantified differentiation relative to nearby congeners. No mention of *B.*

houstonensis is made in Bragg's report of the existence of *B. terrestris* [=*americanus*] *charlesmithi* (Bragg 1954).

The use of measuring anuran advertisement calls to infer phylogenetic relationships, or hybrid status, had not become commonplace until after this species was given independent status (Zweifel 1968; Cocroft and Ryan 1995). Researchers have previously examined the relationship among the calls of *B. houstonensis* and *B. americanus.* Brown (1973) reports that differences in advertisement call are dramatic, primarily pulse number (32.2 and 48.3, respectively), and concludes that for this reason *B. houstonensis* cannot be a subspecies as had been suggested by Blair (1957) despite utilizing a sample of only 11 calls. Likely due to attention following the listing as endangered Thomas and Dessauer (1982) re-examined this question, stating that the findings of Brown (1973) were intentionally biased through the use of *B. americanus* calls originating from New Jersey, rather than from a more nearby population. Thus, the calls of *B. houstonensis* (n=2) were compared to *B. a. charlesmithi* from Oklahoma (n=2), resulting in the conclusion that no differences occur within the advertisement call of these two congeners, refuting the previous author's findings. Our findings reveal that some differences are present, mainly that the dominant frequency is lower for *B. a. charlesmithi*. It has been shown that pulse number and dominant frequency decrease at lower body temperatures (Zweifel 1968). Within our study, body temperature was not known, and average air temperature varied by less than 3 °C at the time recordings were taken. Additionally, we have routinely anecdotally observed Houston Toad choruses which contain increased variation in dominant frequency due to male-male competition

96

(Figure 11) that causes us to suspect that such variation does not carry phylogenetic signal, as has been suggested in the past (Zweifel 1968; Cocroft and Ryan 1995).

The identity and phylogeny of the toads of Texas has been questioned previously, specifically the status of the East Texas Toad (*Bufo* [=*Anaxyrus*] *velatus*; Bragg and Sander 1951; Dixon 2013; Masta et al 2002; Fontenot et al. 2011). To the best of our knowledge, no study has examined or compared the morphology of sympatric toads within Texas. Thus, the features used in discriminating among species (e.g. within dichotomous keys) are at best qualitative, and perhaps erroneous. Through qualitative review of specimen photographs we found that morphological structures proposed as unique to *B. houstonensis* (i.e., enlarged postorbital crests) are found in toads occurring outside their putative range, and conversely we found toads within this range that possess features unique to *B. a. charlesmithi* (i.e., enlarged tibial warts). Nearly all toads vouchered as either *B. houstonensis* or *B. a. charlesmithi* possessed parotoid glands nearest together at their midpoint in contradiction with contemporary dichotomous keys (Tipton et al 2012; Powell et al 2012; Dixon 2013). In life, *B. houstonensis* have an overall darker appearance than its sympatric congeners; however, due to variability in the date of collection and procedures or reagents used to preserve the specimens examined the overall pattern or color was uninformative to this study.

We discovered museum records catalogued as Houston Toads collected outside the putative range described for this species (Figure 8). Interestingly, we found three counties where toads have been collected under both names (i.e., Harris, Harrison, & Robertson counties; Figure 8), exemplifying the influence of the date of collection, taxonomic subjectivity of the collector or curator, and the potential for confounding

morphology. Additionally, we discovered multiple errors in which specimen tags had been misread, or contained misleading information. For example, toads collected from Houston, Texas were cataloged as toads collected from Houston County, Texas (Table 8). One toad was reportedly collected from "Sam Houston State Park" which can be interpreted as an erroneous concatenation of the two terms "Sam Houston National Forest" with "Huntsville State Park," because the two co-occur within Walker County, Texas. Alternatively, a park in western Louisiana was also formerly known by this name, although the collector of this particular toad could not recall conducting any work outside of Texas during this time period (T. Matthews, pers. comm.). Dixon (2013) provides species distribution maps by county, based on the existence of museum vouchers, photographs, written accounts, and personal observations. As a consequence, Washington County is provided among those occupied by *B. houstonensis*, which is likely a personal observation or simply an error in the map of revised volume because no apparent documentation has been found supporting this record and no mention of this "new" county is made in text. At the same time, Liberty and Fort Bend counties have been left blank, despite the presence of populations previously occurring there, yet both of these counties appear to be without any museum specimen vouchers remaining (Yantis and Price 1993). One challenge we repeatedly faced when categorizing physical specimen, or photographs of vouchers, is the prevalence of juveniles or poorly preserved animals. As juveniles Bufonids are difficult, if not impossible, to differentiate to species. The features that are reported to distinguish Houston Toads from their allies (i.e. wart size and crests among the head) are not fully formed until adulthood. Further, changes in taxonomic status are often disputed for long periods of time, causing researchers and curators to

hesitate, or fail to altogether, in changing or updating voucher specimen designations (Pauly et al. 2009). Innovations such as VertNet, as illustrated in our study, aid in identifying these shortcomings so that they can begin to be addressed.

Hillis et al. (1984) reported morphological differences among Houston Toads and several of their sympatric congeners, using stepwise discriminant function analysis. Here we attempt to replicate and expand this approach to include *B. a. charlesmithi*. However, in the previous study no attempt was made to account for body size among morphological characters. When we account for this effect in a univariate manner (i.e. using SUL), we found collinearity among nearly all remaining size-adjusted morphological characters (Figure 9; Table 10). Ultimately, this indicates that few, if any, significant differences are present within the morphology of these two species. These findings further support our claims that the morphological features traditionally used to differentiate among Bufonids of the *americanus* complex are not reliable, especially among vouchered specimen. The physical frame of a generalized Bufonid can easily be quantified using digital calipers, and quickly compared among its allies as we have shown here. However, variation exists among the coloration and patterning within this group yet is seldom accounted for. In many cases these differences are used as differentiating features (Dixon 2013).

Within the counties we examined during this study there is reportedly a total of eight species of Bufonid (*B. americanus, B. debilis, B. houstonensis, B. nebulifer, B. punctatus, B. speciosus, B. velatus, B. woodhousii*; Dixon 2013). Research has shown that hybridization among sympatric Bufonids is common (Strecker 1915; Blair 1963; Ballinger 1966; Brown 1971; Hillis et al. 1984; Masta et al. 2002; Chivers 2016). This can result in shared morphological characters that confuse or undermine identification at

the species level. Many contradictory opinions exist within the primary literature with respect to the morphology of these animals. For example, in its initial description Sanders (1953) describes the parotoid glands of *B. houstonensis* as "diverging posteriorly", and Bragg (1954) describes those of *B. a. charlesmithi* as "nearest together in the middle". Nevertheless, contemporary dichotomous keys (Tipton et al. 2012; Dixon 2013) report that both possess parotoid glands that are nearest together anteriorly. Ultimately, *B. houstonensis* are reported to be differentiated from *B. a. charlesmithi* by the presence of enlarged warts which occur on the tibia of the latter, and enlarged post-orbital crests among the former. Our findings indicate that the statement (Sanders 1953) with respect to size of post-orbital crests may be false, supporting the remarks of Seal et al. (1994). The relative wart size of these animals has not been examined in any quantifiable manner that we are aware of at this time. The efficacy of the existing dichotomous keys available for North American bufonids is ultimately contingent upon the strengths of the findings within the primary literature from which they are drawn. Small mistakes in foundational literature, such as those proposed by Seal et al. (1994), may confound or undermine efforts by researchers, biological consultants, conservation groups, and governing agencies, whose focus may primarily be on enacting protections for endangered species.

Table 8. List of toads photographed and qualitatively assessed. *A priori* indicates the epithet under which each specimen is catalogued, parotoid indicates the point at which the two swollen glands of the toad are nearest to one another, crest size indicates whether the supraorbital and postorbital are uniform in thickness or the postorbital are enlarged, chest pattern is the degree to which the venter of each toad is colored, warts per spot indicates the number of large dorsal warts encircled by black pigment. Using these characteristics, we provide a qualitative assessment of the likely identity of these specimen, as well as the county, year, museum, and specimen number taken from each catalogued record.

#specimen verified as *B. houstonensis* by A. Bragg
*specimen tag reads 'Houston, Texas' indicating the locality of Houston County is in error
^specimen tag reads 'Texas, Sam Houston State Park' creating confusion over the exact location of collection

A priori	Parotoid	Wart Size	Crest Size	Chest Pattern	Warts per Spot	Qualitatively	County	Year	Museum	Specimen no.
americanus	middle	n/a	uniform	black throat	n/a	*woodhousei / velatus*	Chambers	1955	BYU	38773
americanus	--	--	--	--	--	juvenile	Dallas	1928	USNM	75355-56
americanus	--	--	--	--	--	juvenile	Grimes"	--	OMNH	21728
americanus	middle	enlarged	enlarged	spotted	1	*houstonensis*	Harris	1952	SDNHM	42045
americanus	middle	uniform	uniform	pale	2	*woodhousei*	Harris	1966	UAMZ	A1535
americanus	middle	enlarged	enlarged	pale	7	*woodhousei*	Hunt	1957	UTEP	14438
americanus	anteriorly	enlarged	uniform	flecks	6	*woodhousei*	Nacogdoches	1968	TNHC	79899
americanus	--	--	--	--	--	cleared and stained	Orange	1968	YPM	6842
americanus	middle	uniform	uniform	spotted	2	*houstonensis / americanus*	Sabine	1934	USNM	99771
americanus	anteriorly	uniform	uniform	spotted	1	*houstonensis / americanus*	Panola	1961	KU	70013
americanus	anteriorly	uniform	uniform	black throat	4	*woodhousei / velatus*	Smith	2000	KU	289496
americanus	anteriorly	enlarged	uniform	black throat	0	*woodhousei / velatus*	Smith	2000	KU	289469
americanus	anteriorly	enlarged	uniform	spotted	1	*americanus*	Smith	2001	KU	289499
houstonensis	middle	uniform	uniform	spotted	1	*houstonensis*	Brazos	1958	MSUM	n/a
houstonensis	middle	uniform	uniform	black throat	3	*woodhousei / velatus*	Erath	1967	ASNHC	14667
houstonensis	middle	enlarged	enlarged	black throat	0	*woodhousei / velatus*	Houston	1967-68	ASNHC	14657
houstonensis	middle	n/a	uniform	black throat	1	*woodhousei / velatus*	Houston	1967-68	ASNHC	14668
houstonensis	middle	enlarged	uniform	black throat	3	*woodhousei / velatus*	Houston	1967-68	ASNHC	14671
houstonensis	middle	enlarged	uniform	pale	2	*woodhousei / velatus*	Houston	1967-68	ASNHC	14672
houstonensis	middle	uniform	enlarged	spotted	1	*houstonensis*	Houston*	1959	LSUMZ	9309
houstonensis	middle	uniform	uniform	spotted	2	*houstonensis*	La Calcasieu^	1969	LSUMZ	47849
houstonensis	middle	uniform	uniform	flecks	5	*woodhousei*	Travis	--	UTA	42188
houstonensis	middle	uniform	uniform	flecks	5	*woodhousei*	Travis	--	UTA	41629
houstonensis	middle	n/a	uniform	n/a	3	*woodhousei*	Trinity	--	UTA	42438
houstonensis	middle	uniform	uniform	n/a	2	*woodhousei*	Walker	--	UTA	40636
houstonensis	middle	n/a	uniform	n/a	0	*woodhousei*	Walker	--	UTA	41633
houstonensis	middle	n/a	enlarged	n/a	1	*woodhousei*	Walker	--	UTA	41634
houstonensis	middle	uniform	enlarged	n/a	n/a	*woodhousei*	Walker	--	UTA	40638
houstonensis	middle	n/a	uniform	n/a	6	*woodhousei*	Walker	--	UTA	40637
houstonensis	--	--	--	--	--	juvenile	Harrison	1972	OSUM	n/a
terrestris	middle	uniform	uniform	black throat	1	*houstonensis*	Leon	1945	FMNH	46795

Table 9. Summary statistics of call characteristics for the calls of the Houston Toad (*Bufo houstonensis*) and Dwarf American Toad (*B. a. charlesmithi*). CL = Call length, FR = Frequency Range, PN = Pulses per second, DF = Dominant Frequency.

Dwarf American Toad				
	Min	Mean	Median	Max
CL (s)	2.66	12.65	12.97	26.325
FR (Hz)	363.8	708.85	691.2	1247.6
PN	21.5	26.64	26.57	34
DF (Hz)	1050	1891.7	1894.9	2239.5
Houston Toad				
	Min	Mean	Median	Max
CL (s)	2.61	11.14	10.782	18.86
FR (Hz)	269.8	564.97	582.8	1088.4
PN	22	27.07	27.25	31.5
DF (Hz)	1875	2033.95	2046.85	2125

Table 10. Matrix of pairwise Pearson's correlation coefficient for size-adjusted morphological characters among Houston Toads (*Bufo houstonensis*) and Dwarf American Toads (*Bufo americanus charlesmithi*). HW= head width, DBIC= Distance between interocular crests, MPL/W= Mean parotoid length/width, MTFL= mean tibiofibular length, MPPG = mean parotoid to postorbital gap, MPT = mean postorbital thickness.

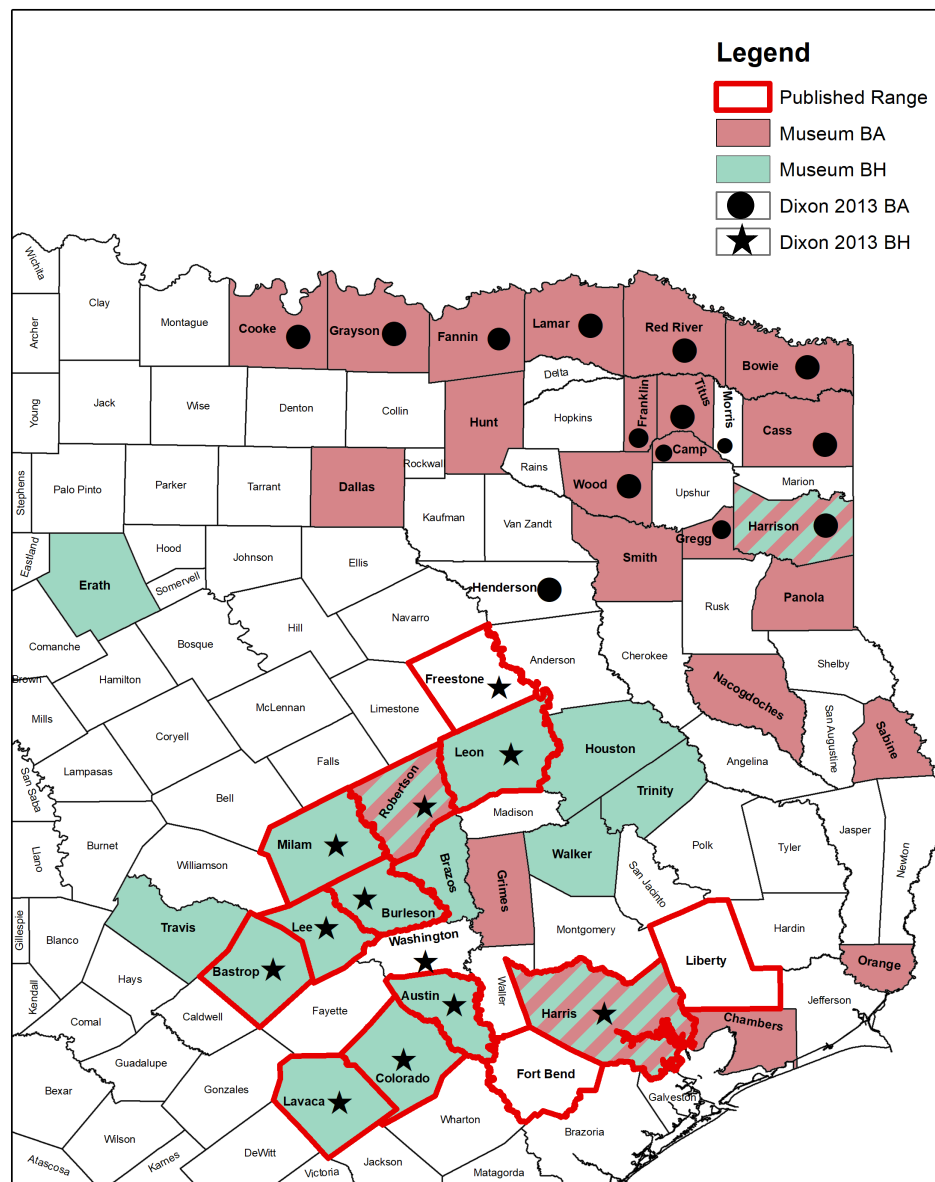|	HW	DBIC	MPL	MPW	MTFL	MPPG	MPT
HW	1	0.142	0.628	0.342	0.476	0.476	0.476
DBIC	0.142	1	0.859	0.978	0.938	0.938	0.938
MPL	0.628	0.859	1	0.946	0.983	0.983	0.983
MPW	0.342	0.978	0.946	1	0.989	0.989	0.989
MTFL	0.476	0.938	0.983	0.989	1	1	1
MPPG	0.476	0.938	0.983	0.989	1	1	1
MPT	0.476	0.938	0.983	0.989	1	1	1

Figure 8. Map of East Central Texas, USA, with counties outlined in black. Counties with bold red outline appear among primary literature as those historically inhabited by the Houston Toad (*Bufo houstonensis*). Counties reported to be the site of collection for Houston Toads within natural history collections are highlighted in green. Counties reported to be the site of collection for Dwarf American Toads (*Bufo americanus charlesmithi*) within natural history collections are highlighted in red. Black stars or filled circles reflect reports of Houston Toad and Dwarf American Toad vouchers within Dixon 2013, respectively.
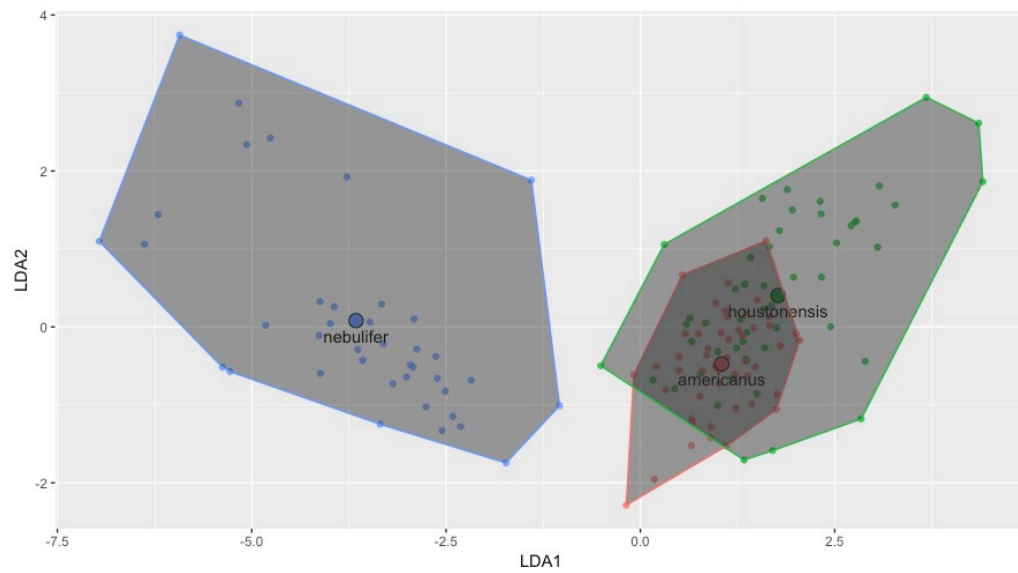
Figure 9. Plot of first two axes (LDA1 and LDA2) of a linear discriminant function analysis using measurements of eight morphological characters among Coastal Plains Toads (*Bufo* [*Incilius*] *nebulifer*; n=39), Dwarf American Toads (*Bufo americanus charlesmithi*; n=50), and Houston Toads (*Bufo houstonensis*; n=51).
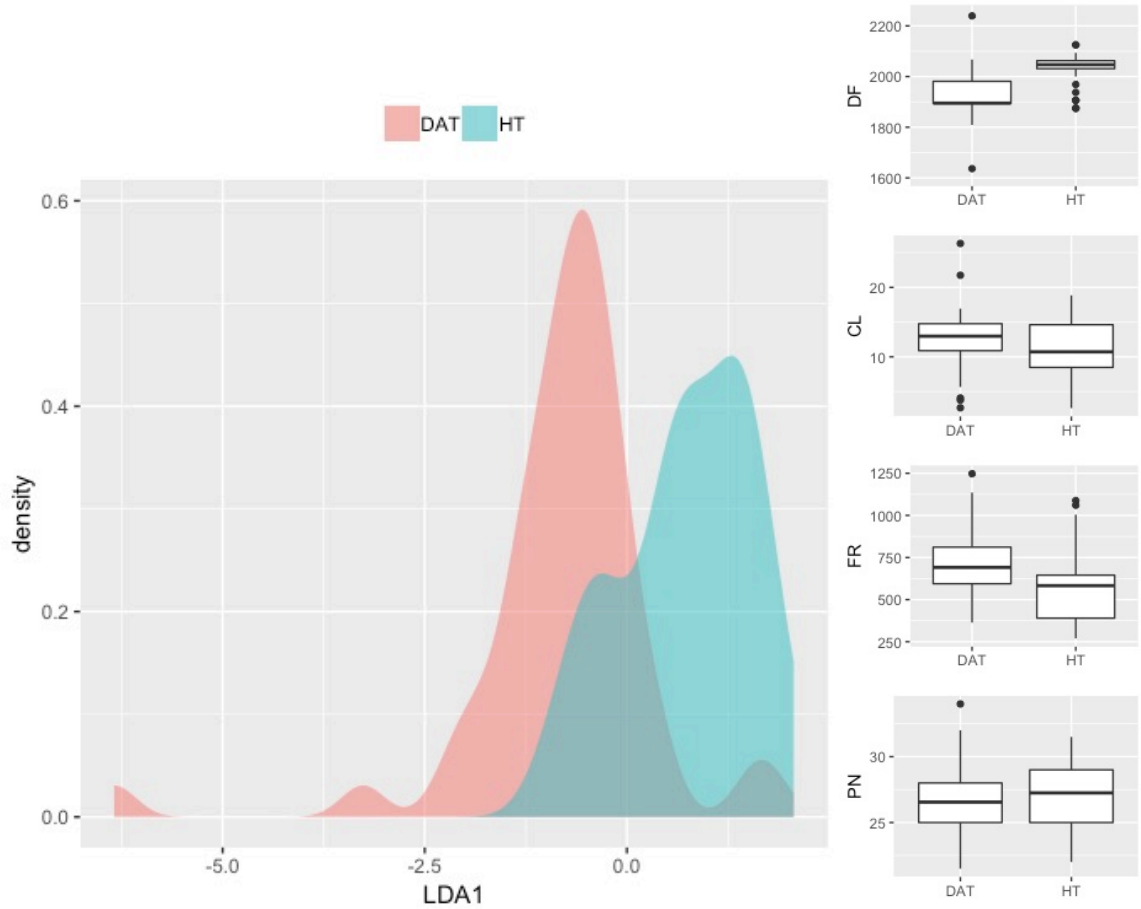
Figure 10. Boxplot of raw call character data and density plot of the first linear discriminant axis summarizing four call characteristics of the Houston Toad (HT; *Bufo houstonensis*) and Dwarf American Toad (DAT; *Bufo americanus charlesmithi*). Overlapping areas indicate a large degree of similarity in call characteristics between species along axis LDA1. CL = Call length, FR = Frequency Range, PN = Pulses per second, DF = Dominant Frequency.
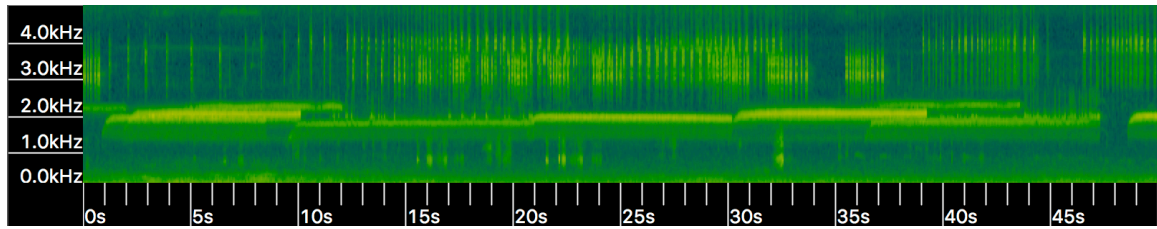
106

Figure 11. Spectrogram (i.e. frequency kHz over time in seconds) of Houston Toad (*Bufo houstonensis*) vocalizations collected within Bastrop County, Texas, USA, illustrating that calls vary around a mean frequency (kHz; Y-axis) when multiple individuals are present. Spectrogram generated in Kaleidoscope 4.3.1 (Wildlife Acoustics).

References

Ballinger, R.E., 1966. Natural hybridization of the toads *Bufo woodhousei* and *Bufo speciosus*. *Copeia*, *1966*(2), pp.366-368.

BenDor, T.K. and Riggsbee, J.A., 2017. Pre-listing conservation of candidate species under the endangered species act: An evaluation of prevalence, accessibility, and market-based conservation efforts. *Environmental Science & Policy*, *74*, pp.68-74.

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

Bishop, P.J., Angulo, A., Lewis, J.P., Moore, R.D., Rabb, G.B. and Moreno, J.G., 2012. The amphibian extinction crisis-what will it take to put the action into the amphibian conservation action plan? *SAPI EN. S. Surveys and Perspectives Integrating Environment and Society*, (5.2).

Blair, A.P., 1957. Amphibians, pp. 211-271 In: W. F. Blair, et al., Vertebrates of the United States. McGraw-Hill Cook Co., New York, pp.ix-819.

Blair, W.F., 1963. Intragroup genetic compatibility in the *Bufo americanus* species group of toads. *Texas Journal of Science*, *13*, pp.163-175.

Bragg, A.N., and Sanders, O., 1951. A new subspecies of the *Bufo woodhousii* group of toad (Salientia: Bufonidae). Wasmann J. Biol. 9:363-378.

Bragg, A.N., 1954. *Bufo terrestris charlesmithi*, a new subspecies from Oklahoma. *Wasmann Journal of Biology*, *12*(2), p.245.

Brown, L.E., 1973. *Bufo houstonensis*. The Catalogue of American Amphibians & Reptiles, p. 133.1

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

Brown, L.E., 1985. The status of the near-extinct Houston Toad (*Bufo houstonensis*) with recommendations for its conservation. Herpetological Review 6(2), pp 37-40.

Bureau of Sports Fisheries and Wildlife. 1970. United States list of endangered native fish and wildlife. Federal Register 35:16047–16048.

Ceballos, G., Ehrlich, P.R. and Dirzo, R., 2017. Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. *Proceedings of the National Academy of Sciences*, *114*(30), pp.E6089-E6096.

Chivers, J.M., 2016. Combining morphological, acoustic, and genetic techniques to better understand hybridization of the most abundant toad in Alabama: *Anaxyrus fowleri*. Thesis, Auburn University.

Cocroft, R.B. and Ryan, M.J., 1995. Patterns of advertisement call evolution in toads and chorus frogs. *Animal Behaviour*, *49*(2), pp.283-303.

Dixon, J.R., 2013. Amphibians and reptiles of Texas, 3rd edition. Texas A & M University Press, College Station, Texas, USA.

[ESA] U.S. Endangered Species Act of 1973, as amended, Pub. L. No. 93-205, 87 Stat. 884 (Dec. 28, 1973).

Fontenot, B.E., Makowsky, R. and Chippindale, P.T., 2011. Nuclear–mitochondrial discordance and gene flow in a recent radiation of toads. *Molecular Phylogenetics and Evolution*, *59*(1), pp.66-80.

Gottschalk, J.S., 1970. United States list of endangered native fish and wildlife. Federal Register 35: pp.16047-16048.

MacLaren, A.R., and Forstner, M.R.J., 2017. Geographic Distribution. *Bufo* [*=Anaxyrus*] *houstonensis*. Brazos Co., TX. Herpetological Review 48(1):118.

MacLaren, A.R., McCracken, S.F. and Forstner, M.R., 2018. Development and Validation of Automated Detection Tools for Vocalizations of Rare and Endangered Anurans. *Journal of Fish and Wildlife Management*.

Masta, S.E., Sullivan, B.K., Lamb, T. and Routman, E.J., 2002. Molecular systematics, hybridization, and phylogeography of the Bufo americanus complex in Eastern North America. *Molecular Phylogenetics and Evolution*, *24*(2), pp.302-314.

McCoy, M.W., Bolker, B.M., Osenberg, C.W., Miner, B.G. and Vonesh, J.R., 2006. Size correction: comparing morphological traits among populations and environments. *Oecologia*, *148*(4), pp.547-554.

Moriarty, E.C. and Cannatella, D.C., 2004. Phylogenetic relationships of the North American chorus frogs (Pseudacris: Hylidae). *Molecular phylogenetics and evolution*, *30*(2), pp.409-420.

Pauly, G.B., Hillis, D.M. and Cannatella, D.C., 2009. Taxonomic freedom and the role of official lists of species names. *Herpetologica*, *65*(2), pp.115-128.

Potter FEJ, Brown LE, McClure WL, Scott NJ, Thomas RA. 1984. Recovery plan for the Houston toad (*Bufo houstonensis*). U.S. Fish & Wildlife Service.

Powell, R., Collins, J.T., Hooper, E.D. and Powell, R., 2012. *Key to the herpetofauna of the continental United States and Canada*. University Press of Kansas.

R Development Core Team, 2018. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Venables, W. N. & Ripley, B. D., 2002. Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0

Sanders, O., 1953. A new species of toad, with a discussion of morphology of the bufonid skull. Herpetologica, 9:25–47.

Seal US, editor. 1994. Houston toad population & habitat viability assessment. Apple Valley, Minnesota: IUCN/ SSC Conservation Breeding Specialist Group.

Shaffer, H.B., Fisher, R.N. and Davidson, C., 1998. The role of natural history collections in documenting species declines. *Trends in ecology & evolution*, *13*(1), pp.27-30.

Simmons, J.E., 1995. Storage of fluid preserved collections. Pp 161-186 *in* Rose, C.L., C.A. Hawks and H.H. Genoways (eds.). Storage of Natural History Collections: A Preventive Conservation Approach. Society for the Preservation of Natural History Collections, x + 448pp.

Skelly, D.K., Yurewicz, K.L., Werner, E.E. and Relyea, R.A., 2003. Estimating decline and distributional change in amphibians. *Conservation Biology*, *17*(3), pp.744-751.

Strecker, J.K., 1915. Reptiles and amphibians of Texas. Baylor Bulletin, 18(4): 82pp.

Thomas, R.A. and H.C. Dessauer. 1982. Preliminary report. Taxonomy of the Houston Toad, *Bufo houstonensis* Sanders. 11pp.

Tipton, B.L., Hibbitts, T.L. , Hibbitts, T.D., Hibbitts, T.J., and LaDuc, T.J., 2012. Texas
Amphibians: A Field Guide. University of Texas Press. 330pp.

Wake, D.B. and Vredenburg, V.T., 2008. Are we in the midst of the sixth mass
extinction? A view from the world of amphibians. *Proceedings of the National
Academy of Sciences*.

Winston, J.E., 1999. *Describing species: practical taxonomic procedure for biologists*.
Columbia University Press.

Yantis, J.H., and Price, A.H., 1993. Final report. As required by the Endangered Species
Act, Section 6. Texas, Project No: E-1-4. Endangered and Threatened Species
Conservation, Job No. 8, Houston Toad (*Bufo houstonensis*) status survey. Texas
Parks and Wildlife Department, Austin TX.

Zweifel, R.G., 1968. Effects of temperature, body size, and hybridization on mating calls
of toads, *Bufo a. americanus* and *Bufo woodhousii fowleri*. *Copeia*, pp.269-285