

Articles

Development and Validation of Automated Detection Tools for Vocalizations of Rare and Endangered Anurans

Andrew R. MacLaren,* Shawn F. McCracken, Michael R.J. Forstner

Department of Biology, Texas State University, 601 University Drive, San Marcos, Texas 78666

Abstract

For many rare or endangered anurans, monitoring is achieved via auditory cues alone. Human-performed audio surveys are inherently biased, and may fail to detect animals when they are present. Automated audio recognition tools offer an alternative mode of observer-free monitoring. Few commercially available platforms for developing these tools exist, and little research has investigated whether these tools are effective at detecting rare vocalization events. We generated a recognizer for detecting the vocalization of the endangered Houston toad *Anaxyrus houstonensis* using SongScope® bioacoustics software. We developed our recognizer using a large sample of training data that included only the highest quality of recorded audio (i.e., low noise, no interfering vocalizations) divided into small, manageable batches. To track recognizer performance, we generated an independent set of test data through randomly sampling a large population of audio known to possess Houston toad vocalizations. We analyzed training data and test data recursively, using a criterion of zero tolerance for false-negative detections. For each step, we incorporated a new batch of training data into the recognizer. Once we included all training data, we manually verified recognizer performance against one full month (March 2014) of audio taken from a known breeding locality. The recognizer successfully identified 100% of all training data and 97.2% of all test data. However, there is a trade-off between reducing false-negative and increasing false-positive detections, which limited the usefulness of some features of SongScope. Methods of automated detection represent a means by which we may test the efficacy of the manual monitoring techniques currently in use. The ability to search any collection of audio recordings for Houston toad vocalizations has the potential to challenge the paradigms presently placed on monitoring for this species of conservation concern.

Keywords: Houston toad; *Anaxyrus houstonensis*; audio; bioacoustics; recognizer; SongScope; surveys

Received: May 30, 2017; Accepted: February 19, 2018; Published Online Early: February 2017; Published: June 2018

Citation: MacLaren R, McCracken SF, Forstner MRJ. 2018. Development and validation of automated detection tools for vocalizations of rare and endangered anurans. *Journal of Fish and Wildlife Management* 9(1):144–154; e1944-687X. doi:10.3996/052017-JFWM-047

Copyright: All material appearing in the *Journal of Fish and Wildlife Management* is in the public domain and may be reproduced or copied without permission unless specifically noted with the copyright symbol ©. Citation of the source, as given above, is requested.

The findings and conclusions in this article are those of the author(s) and do not necessarily represent the views of the U.S. Fish and Wildlife Service.

* Corresponding author: arm107@txstate.edu

Introduction

Long-term monitoring of anuran populations is required to understand population dynamics, the causes for growth or decline, and to enable informed conservation assessments and stewardship (Pechmann et al. 1991). Most male anurans vocalize to attract females for breeding, and population monitoring programs often use manual, human-performed auditory call surveys (MCSs) for assessments of breeding site occupancy

(Bridges and Dorcas 2000; Crouch and Paton 2002; Schmidt 2003; Pierce and Gutzweiller 2004; Jackson et al. 2006; USFWS 2007). These surveys can also be used to estimate anuran abundance by indexing the number of individuals being heard (Zimmerman 1994; Weir and Mossman 2005). However, these indices do not show strong correlation with true abundance, and no consensus of approach exists at this time (Corn et al. 2011; Pierce and Hall 2013). MCSs can have a multitude of confounding factors, such as anthropogenic noise



disturbance (Bee and Swanson 2007) or temporal bias (Cook et al. 2011). These challenges are further compounded when monitoring for rare or elusive species (Crouch and Paton 2002; Williams et al. 2013).

Automated methods of audio monitoring offer alternatives to traditional MCS (Digby et al. 2013). Automated recording devices (ARDs) can be less expensive than MCSs, monitor inhospitable and remote sites, avoid bias due to observer disturbance, and avoid temporal bias through rigorous sampling regimes. Recording devices can provide reliable data more rapidly than MCSs (Dorcas et al. 2009). Common applications of ARDs include testing or improving MCS methods for application over a larger spatial scale (Dorcas et al. 2009; Williams et al. 2013), as well as in determining the environmental cues for anuran chorusing (Bridges and Dorcas 2000; Oseen and Wassersug 2002; Acevado and Villanueva-Rivera 2006; Digby et al. 2013; Willacy et al. 2015). Early ARD users were burdened by the necessity to manually listen (O'Neal 2014) or spectrographically review field recordings. Advancements in the emerging science of bioacoustics have provided researchers with many techniques for automated audio detection, alleviating this burden. There are many commercial and open-source pattern recognition platforms available such as RAVEN (Charif et al. 2010), R packages “Seewave” or “monitoR” (Sueur et al. 2008; Katz et al. 2016), the automated detection toolbox in language C# (Towsey et al. 2012; Digby et al. 2013), and SongScope© (Wildlife Acoustics 2011a). See Obrist et al. (2010) for a more detailed account of bioacoustics software. These programs rely on a variety of complex mathematics to achieve recognition of focal audio (i.e., hidden Markov models, mel-frequency cepstral coefficients, neural networks, fuzzy clustering, and decision trees). However, not all studies that describe novel methods for automated detection provide information regarding what data type shall be generated (i.e., abundance, presence/absence) and seldom seek to answer an ecological question regarding focal species.

The efficacy of automated audio detection has been criticized for commonly featuring excessive false-positive detections (Barclay 1999; Swiston and Mennill 2009). However, trade-offs exist between false-positive (type I errors) and false-negative (type II errors) detections (Waddle et al. 2009). Researcher subjectivity, as well as quality and amount of training data, can affect the magnitude of these trade-offs. These limitations are exacerbated among studies focusing on rare or elusive species that might vocalize infrequently (Swiston and Mennill 2009; Goh 2011; Digby et al. 2013). For studies that aim to detect a rare animal's vocalization, false-negative detections are more problematic than false-positive detections. In other words, overlooking the call of a rare animal as a type II error has a greater consequence than simply having to filter through a larger number of type I errors to find what you're looking for. Alternatively, this trade-off may not affect methods for measuring biodiversity, where identifying the maximum number of species vocalizing, but not necessarily

all species that may be present, is prioritized (Hsu et al. 2005; Aide et al. 2013; Bedoya et al. 2014; Noda 2016).

The Houston toad *Anaxyrus houstonensis* (Sanders 1953; Frost et al. 2006) is a rare species of anuran endemic to southeastern central Texas, and is listed as endangered at state, federal, and international levels (Gottschalk 1970; Honegger 1970; U.S. Endangered Species Act [ESA 1973, as amended]; Hammerson and Canseco-Márquez 2004). Ongoing habitat loss and fragmentation throughout its range are major drivers of population declines (Brown 1971; Potter et al. 1984). Houston toads have undergone extirpations in at least three Texas counties, restricting their range to only 10 remaining counties (Potter et al. 1984). Robust populations are documented in only Bastrop and Robertson counties, Texas. However, Houston toad populations in Bastrop County are declining, likely due to an increase in population stressors (e.g., drought, wildfires, development; Gaston et al. 2010, Duarte et al. 2014).

In September 2011, the Bastrop County Complex Fire burned > 13,700 ha of habitat, including 96% of Bastrop State Park, then believed to be the species' last stronghold (Price 2003). Because of the Houston toad's endangered status and local extirpations, numerous groups are interested in conducting consistent, long-term monitoring for this species. Researchers, environmental consultants, and federal agencies rely primarily on MCSs to determine breeding site presence of Houston toads (USFWS 2007), which is common among endangered anurans of North America (USFWS 1999, 2005, 2006). Continuous monitoring across its remaining range is essential for tracking population trends, and determining appropriate local management strategies (e.g., forest restoration, population supplementation). However, because of time, personnel, and financial constraints, thorough long-term, range-wide monitoring has been impossible to achieve.

This study illustrates how to develop a robust and reliable tool for recognizing the vocalizations of rare and endangered anurans within SongScope (version 4.1.3A) bioacoustics software (Wildlife Acoustics, Maynard, Massachusetts), using the Houston toad as a model species. Additionally, we describe how to validate the effectiveness of a recognition tool through detailed manual review. Because of the endangered status of the Houston toad, we prioritized minimization of missed calls (i.e., type II errors, false-negative detections).

Methods

Audio surveys

We deployed ARDs (SongMeter models SM2, SM2+, and SM3) at 35 potential breeding locations from 3 January to 12 July 2014 in two counties of east central Texas ($n = 11$ in Bastrop County and $n = 24$ in Robertson County). We secured ARDs to structure objects < 10 m from pond, drainage, or water-body edge. We programmed each to record the first 10 min of each hour from 1800 hours to 0500 hours the following morning. This resulted in 12 10-min segments (120 min) of audio

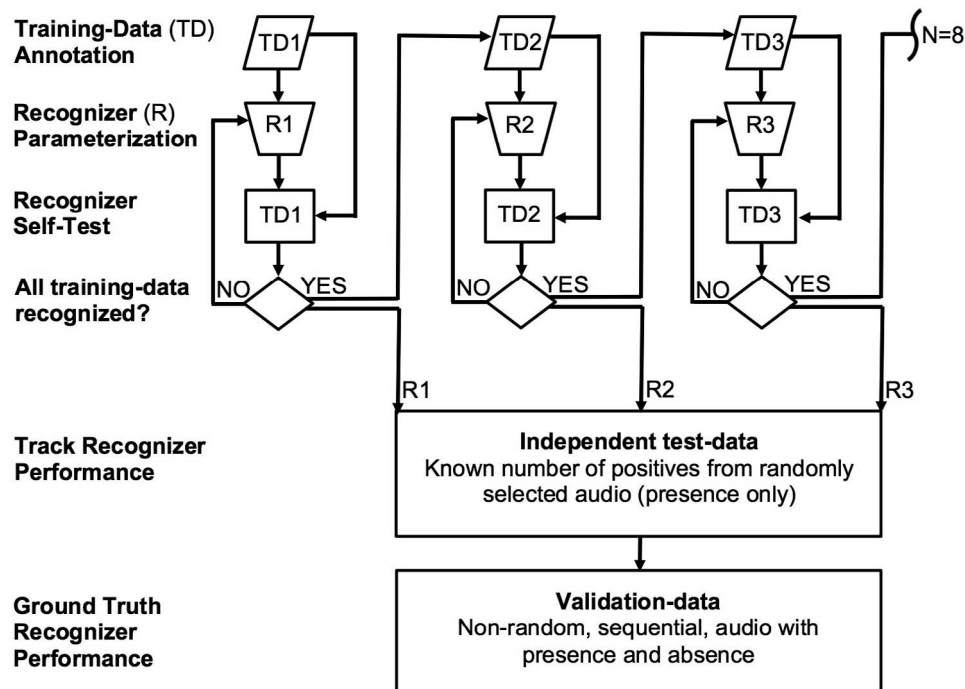


Figure 1. Flowchart detailing steps of development and validation of our recognizer for the call of the Houston toad *Anaxyrus houstonensis* using SongScope® bioacoustics software. We provided three iterative and recursive steps as an example, although we used a total of eight within this study. Each step includes new training data (TD) that become a new recognizer (R). If the recognizer is capable of identifying all the vocalizations used to create it (i.e., recursive self test), then it is used against an independent set of test data, and also passed along to the next iteration (TD2). Once all batches of training data have been included, the final recognizer is ground truthed using a third independent set of validation data that possesses audio examples of both Houston toad presence and absence. We collected all Houston toad vocalizations used as training or test data during spring 2014 from Bastrop and Robertson counties, Texas.

per device per survey night. To reduce file size we selected the proprietary WAC format, and reduced sample rate to 16 kHz. This lowered the maximum frequency recorded to 8 kHz, which is appropriate for detection of most North American anuran vocalizations (Narins et al. 2004). Under these settings ARDs required battery changes approximately every 40 d. During these visits SD cards containing field recordings were swapped for blank replacements. We also carried out MCSs following the guidelines provided by the U.S. Fish and Wildlife Service (USFWS) at a subset of the same sites that were monitored with ARDs (USFWS 2007). Surveys occurred on nights that met or exceeded the environmental conditions prescribed by the USFWS to ensure that no chorusing events would be missed. During MCSs, we monitored each site once per night, for 5 min, without the explicit goal of overlapping with audio recorded by ARDs.

Recognizer development

We used SongScope to spectrographically review audio. We cross-referenced dates and locations of MCS detections to find audio files containing high-quality vocalizations for recognizer training data ($n = 7$ files; Audio S1–S7; Figure 1). Ideal vocalizations are visible within a spectrograph, audible when played back, and

do not overlap with other animal vocalizations. We ensured that these files encompassed multiple sites across both counties surveyed. We annotated between 13 and 61 Houston toad vocalizations from each file. Annotating vocalizations in SongScope is a “click and drag” highlighting process used to define the bounds of a vocalization within the viewable spectrograph, that is, the moments in time when a vocalization starts and ends, and which frequencies it occupies, for the purpose of incorporating the sound into a recognizer (Figure 2).

To track recognizer performance as it was built, we arranged randomly selected data into a simulation of one full survey night of detections (i.e., 12 files; 120 min), referred to hereafter as “test data” (Figure 1). To accomplish this we generated a population of 105 files containing Houston toad calls, independent of the 7 training data files, using the same cross-referencing technique described above. We randomly selected 12 files from this population using Program R (R Core Team 2014). We obtained the number of Houston toad calls within these test data ($n = 186$) by visually inspecting their spectrographs.

SongScope offers two proprietary filters used for removing unwanted results: quality and score. Quality represents a statistical distribution of parameters within the training data used to build a recognizer, and ranges

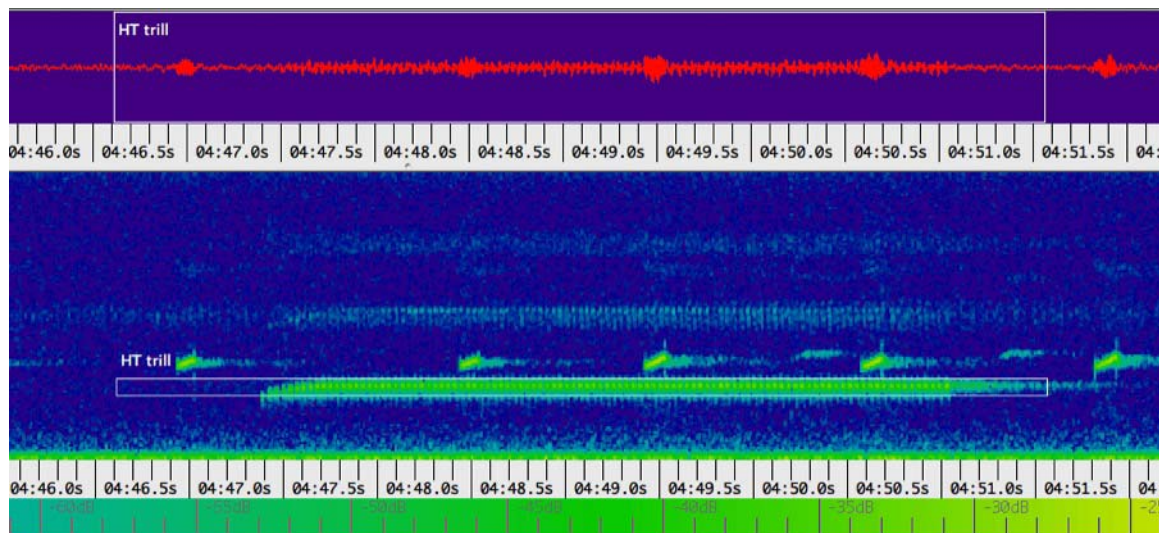


Figure 2. Annotation of a Houston toad *Anaxyrus houstonensis* vocalization in SongScope© bioacoustic software (Wildlife Acoustics). Example of a Houston toad vocalization, collected during spring 2014 from Robertson County, Texas, being annotated (i.e., selected for analysis), indicated by the white box surrounding the vocalization of interest.

from 0.00 to 99.99, higher values indicating greater confidence (Wildlife Acoustics 2011b). Score can range from 0 to 100, and measures the statistical fit of a vocalization to the model estimated by the recognizer (Wildlife Acoustics 2011b). For this experiment, each recognizer scanned the test data with filters disabled to determine the lower threshold for true detections (lowest true positive). This ensures a zero tolerance for false-negative detections.

We incorporated the first training data files' annotations, and adjusted parameters to summarize all annotations entirely. Because batches of training data were divided by file for this study, we scanned the file(s) from which the training data were gathered recursively (Figure 1). This ensured that recognizers could, at minimum, identify the calls from which they were built. We adjusted parameters until the recognizer could accurately identify all training data incorporated. We manually reviewed the results from these self tests to verify detection. For each self test we counted positive detections, the number of true vocalizations each positive detection represents (accounting for overlap), the total number of detections made with filters fixed at zero, and the total number of detections made with filters adjusted to the value of the lowest true-positive detection (eliminating false positives only). Once we constructed, parameterized, and self-tested the recognizer, we scanned the test data with filters fixed at zero (Table 1). Again, we manually reviewed the results of these scans to confirm detections and ensure that no false negatives occurred. We repeated this for each set of annotations incorporated into the recognizer (Figure 1). Once we incorporated all files of training data, we removed unwanted annotations with potentially negative effects (e.g., short bursts, weak signals). Overall, we performed eight iterations of this process.

Recognizer validation

To quantify the failure rate of the final recognizer (step no. 8; Tables 1 and 2), we manually estimated the number of calls in each recording taken in 2014 from a single location ($n = 1,945$). We analyzed these recordings with filters set to lowest true-positive values (Table 1; Figure 1). The location chosen represented the highest probability of Houston toad chorusing on the basis of >10 y of MCS data (M.R.J. Forstner, personal observation). To estimate recognizer failure, we quantified true-positive and false-negative detections. We reanalyzed any files containing false-negative detections with filters disabled to investigate the source of error.

Results

Audio recording

The 35 ARDs recorded between 1,465 and 2,272 audio files each, totaling 657,350 min. Data loss occurred at several ARDs because of inconsistent battery life, which is expected when managing a large collection of rechargeable cells under demanding environmental circumstances. Of the 35 ARDs deployed, 11 yielded detections of the Houston toad.

Recognizer development

During recognizer self tests (Table 2), 68% of the detections were true positives and 32% were false positives, with filters adjusted to the value for lowest true-positive detection. False-positive detections decrease by 27.8% by applying filters in this way. With the incorporation of each new training data batch, the filter thresholds decreased (i.e., excluded fewer false positives), with the exception of the final step in which we removed imperfect annotations.

Table 1. Parameters used in creation of an audio recognizer for the vocalization of the Houston toad *Anaxyrus houstonensis* within SongScope© bioacoustic software. The following table summarizes all parameters estimated by SongScope bioacoustics software, or chosen by the authors, given for each step of audio recognizer development (columns 1–8), when used to scan 120 min (12 files) of audio test data containing a known number of vocalizations ($n = 186$) made by Houston toads, collected during spring 2014 from Bastrop and Robertson Counties, Texas. Refer to the SongScope bioacoustic software user guide for specific definitions of terminology (Wildlife Acoustics 2011b).

Parameter	Step number							
	1	2	3	4	5	6	7	8
Cross-training (%)	0 ± 0	75.93 ± 18.88	76.77 ± 28.76	73.72 ± 12.02	73.81 ± 12.23	72.95 ± 10.92	74.2 ± 10.62	73.05 ± 11.62
Total training (%)	83.3 ± 5.38	76.39 ± 13.57	77.88 ± 31.37	75.67 ± 11.53	75.08 ± 14.63	70.97 ± 12.13	73.91 ± 12.24	73.91 ± 13.72
Model states	38	30	40	45	44	40	40	46
State usage	33 ± 7	12 ± 8	16 ± 10	21 ± 7	21 ± 9	20 ± 8	19 ± 7	24 ± 9
Feature vector	8	8	13	13	13	13	13	13
Mean symbols	675 ± 271	432 ± 377	471 ± 457	594 ± 411	514 ± 413	525 ± 392	514 ± 371	578 ± 399
Syllable types	8	12	7	10	11	7	12	6
Mean duration (s)	8.75 ± 3.51	5.92 ± 4.9	6.06 ± 5.33	8.99 ± 4.66	9.42 ± 5.1	9.51 ± 4.82	9.48 ± 4.62	10.45 ± 4.47
Max. resolution	8	8	13	13	13	13	13	13
Freq. min. (Hz)	30	30	26	26	26	26	26	26
Freq. range (Hz)	5	6	11	11	11	11	11	11
Max. syllable (ms)	200	176	328	1,504	1,504	1,504	1,504	1,504
Max. syl. gap (ms)	96	64	256	1,048	1,048	1,048	1,048	1,048
Max. song (ms)	16,960	16,960	16,544	16,544	16,544	16,544	16,544	16,544
Annotations	25	58	77	117	130	191	219	170
Playback speed	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal
Max. sample delay	64	64	64	64	64	64	64	64
FFT ^a size	256	256	256	256	256	256	256	256
FFT overlap	1/2	1/2	1/2	1/2	1/2	1/2	1/2	1/2
Sample rate (Hz)	16,000	16,000	16,000	16,000	16,000	16,000	16,000	16,000
Amplitude gain (Db)	0	0	0	0	0	0	0	0
Dynamic range (Db)	20	20	20	20	20	20	20	20
Background filter	1 s	1 s	1 s	1 s	1 s	1 s	1 s	1 s
Algorithm	2	2	2	2	2	2	2	2
Max. complexity	48	48	48	48	48	48	48	48
Quality & score = 0								
Total detections	6,829	9,990	4,400	797	797	797	797	797
Detections > 1 s	234	262	692	642	642	642	642	642
Quality & score = lowest true positive								
Total detections	6,829	8,377	360	206	210	204	208	219
Detections > 1 s	234	244	206	187	189	188	189	191
Quality	0	1.3	1.3	1.3	1.3	1.9	1.3	2.7
Score	0	19	46	57	57	5	56	54

^a FFT = fast Fourier transform.

Table 2. Performance of audio recognizer built within SongScope© bioacoustic software for the vocalization of the Houston toad *Anaxyrus houstonensis* during self-identification tests. Summary of the performance of our recognizer within SongScope bioacoustic software for each step of development (columns 1–8), that is, the recognizer’s ability to identify the calls from which it is built. All vocalization made by Houston toads were collected during spring 2014 from Bastrop and Robertson counties, Texas. Parameters include the minimum “quality” and “score” assigned to a true-positive detection (det.), the number of results produced when those filters are adjusted to represent the numbers presented in the first two rows (total det.), the number of results that represent positive detections (positive det.), and the number of results produced when filters quality and score are fixed at zero (i.e., not used). Quality and score are filters proprietary to SongScope used to eliminate unwanted detections. Quality represents a statistical distribution of parameters within the training data used to build a recognizer, and ranges from 0.00 to 99.99, higher values indicating greater confidence. Score can range from 0 to 100, and measures the statistical fit of a vocalization to the model estimated by the recognizer.

Parameter	Step number							
	1	2	3	4	5	6	7	8
Min. quality	11.8	3.5	2.3	6.1	8.2	6.6	1.3	3.5
Min. score	78.28	67.63	43.5	47.93	47.18	46.69	40.43	44.89
Positive det.	26	78	90	111	134	163	185	166
Total det.	26	80	122	126	177	216	333	244
Quality & score = 0								
Total det.	492	892	603	221	294	338	337	338

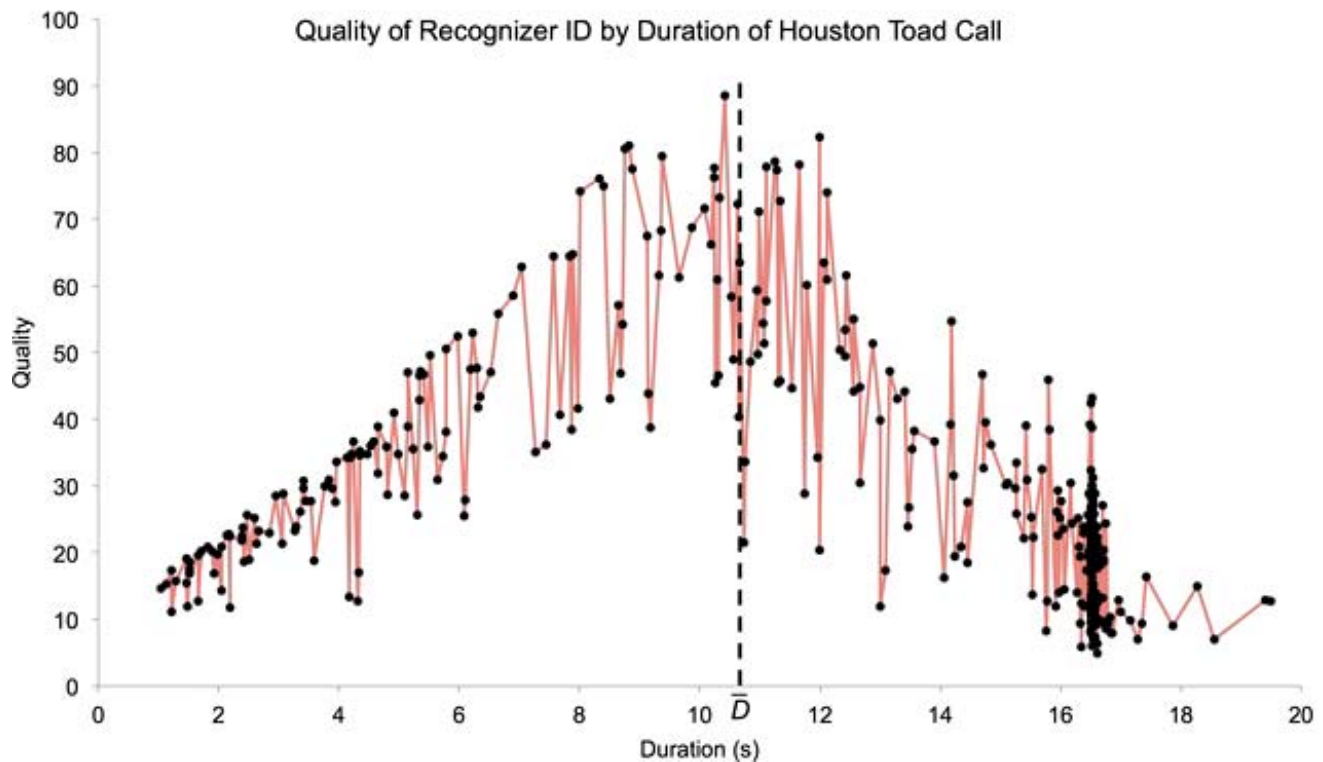


Figure 3. Graph of the quality by duration for detections of Houston toad *Anaxyrus houstonensis* vocalizations. This graph illustrates that Houston toad vocalizations detected by our recognizer that are similar in length to the mean length of all vocalizations used as training data (i.e., duration, \bar{D}) are assigned greater quality in SongScope© bioacoustics software. Quality represents a statistical distribution of parameters within the training data used to build a recognizer, and ranges from 0.00 to 99.99, higher values indicating greater confidence. We collected Houston toad vocalizations used in this analysis during spring 2014 from Bastrop and Robertson counties, Texas.

Parameters estimated by SongScope that underwent the greatest change throughout this study included cross- and total training, model states, state usage, and mean duration (Table 1). Training percentages dropped as variation in annotations increased; however, these percentages ranged from 70.97 to 83.3. Mean duration (range 5.92–10.45 s) seemingly limits the lower threshold of the quality filter considerably. Thus, the eighth iteration of recognizer development was aimed toward increasing the mean duration by removing short vocalizations (Table 1).

Figure 3 shows that quality and duration of detection are approximately normally distributed. This means the closer detections are to the mean duration estimated by the recognizer, the higher the quality. We do not believe this phenomenon to be limited to only effects based on duration; rather, duration is the only parameter having a detectable effect on the efficacy of the recognizer built to identify the call of the Houston toad. Given that the Houston toad's call has a constant frequency that is rather narrow banded (Tipton et al. 2012), length of call varies more than other features. Because quality assesses all the parameters within a recognizer, it is fair to consider the outliers that violate the normal distribution presented in Figure 3 as vocalizations that vary in a metric other than duration.

We noted that false-negative detections occur in the event that a vocalization takes place as a recording begins. In this instance that portion of sound that has been recorded will go unidentified by the software (Figure 4). We believe this to be the only occasion, when using a zero-tolerance approach, in which a false-negative detection may occur. There were false-negative detections in two of the seven training files and in one of the files used to create the test data. We found only one missed call per file, and each occurred at the origin of the file (Figure 4).

Recognizer validation

From the subset of data taken from a single location we processed a total of 1,945 files. Manual review estimated 393 vocalizations in 53 files, ranging from 1 to 30 calls per file, averaging 7.42. The recognizer made 437 true-positive detections in 51 audio files. Of those 51 recordings the number of vocalizations per file ranged from 1 to 26, averaging 8.57 detections per file. There were 11 incidents of false negatives. Six of these incidents were a consequence of vocalizations taking place at the origin of each recording (Figure 4). The five other incidents of false negatives included faint or weak calls. The score assigned for these uncharacteristically faint vocalizations was below the threshold determined

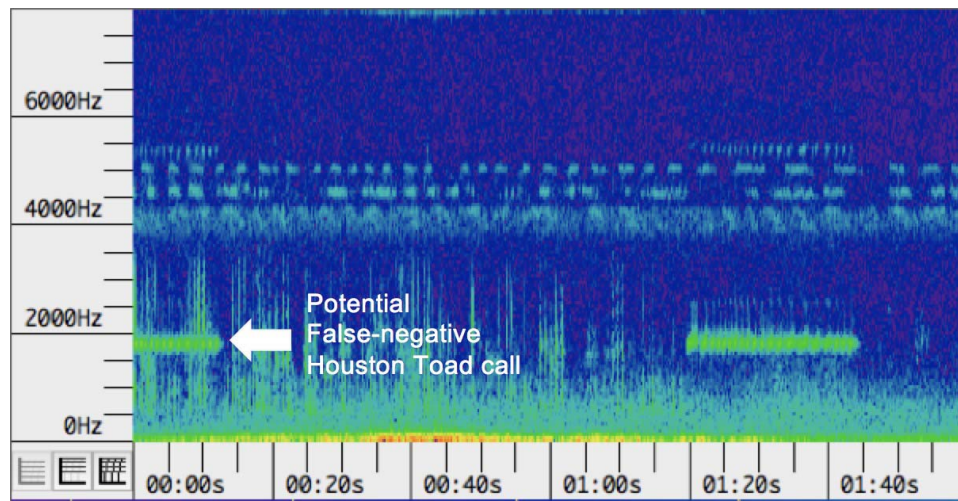


Figure 4. Spectrograph of potential false-negative Houston toad *Anaxyrus houstonensis* vocalization in SongScope© bioacoustics software. Example in which a Houston toad vocalization, collected during spring 2014 in Bastrop County, Texas, is taking place at the beginning of the recording, leading to a false-negative detection, a result of an inherent error within the software.

during recognizer development. Lowering this filter such that it includes these five vocalizations increased the number of false-positive detections by 224% ($n_1 = 1,399$ to $n_2 = 3,133$). Taking all vocalizations into account the recognizer correctly identified 97.2% of the true vocalizations present within the validation audio. False-positive detections consisted primarily of detections shorter than 1 s. These false positives were triggered by the sound of wind, rain, automobile traffic, birds, and other anurans, namely *Hyla versicolor* or *Pseudacris crucifer*.

The number of vocalizations present in test data was underestimated in steps 4, 6, 7, and 8 (Tables 1 and 2). This is due to a single detection accounting for more than one vocalization (Figure 5). The number of vocalizations was overestimated in steps 1, 2, 3, and 5 such that more than one detection was made per

vocalization (Figure 5). Upon the final step of development, self tests show that the finalized recognizer overestimates the number of vocalizations present by 2.7% ($n = 191$), a low deviation from the actual number present ($n = 186$).

Discussion

Error trade-off

Objectivity and attention to detail are important in determining the efficacy of a recognizer. Comparable studies report failure among automated detection procedures that could potentially be eliminated by increasing the rigor of training and validation, resulting in more carefully assembled tools (Eldridge 2011, Waddle et. al. 2009). Our approach for the development and optimization of this recognizer followed a strict criterion

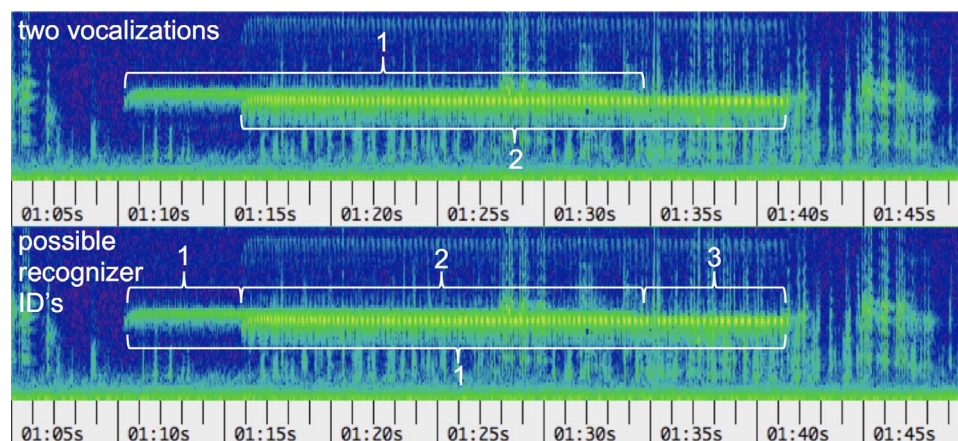


Figure 5. Spectrograph of two overlapping Houston toad *Anaxyrus houstonensis* vocalizations in SongScope© bioacoustics software. Top panel: Spectrographic view of two overlapped Houston toad vocalizations, collected during spring 2014 from Bastrop County, Texas. Bottom panel: Brackets indicate how SongScope bioacoustics software may misrepresent the number of vocalizations by inaccurately detecting overlapping vocalizations.

of zero tolerance for false-negative detections. Although this may not be necessary for recognizers for all species, for the Houston toad this method was effective. Through improving our recognizer iteratively, we reduced false-positive detections without concurrent increase in false negatives, outside of those described in Figure 4 (Table 1). Training-data self-test results were 32% false-positive detections, whereas results from recognizer validation indicate a dramatic increase in false-positive detections. This increase is likely caused by audio containing greater amounts of noise and fewer ideal Houston toad vocalizations.

Time investment

In total, the process of preparing and optimizing this recognizer required a comprehensive time investment of approximately 24 h. This is a shorter build time, start to finish, than comparable studies (Eldridge 2011, Waddle et al. 2009). Assembling training data by cross-referencing MCSs decreased the overall time investment required for our study. One other advantage to our approach is the small, yet effective, amount of test data. Time required to process 120 min of audio was < 2 min. Processing times are dependent on a multitude of factors (e.g., computer processing power), and may vary between researchers. We are enabled to utilize fewer files because each file often contains multiple vocalizations because of the Houston toad's explosive breeding strategy (Price 2003).

Manual methods of audio review required approximately 32 h to complete, or approximately 1 min per audio file. For files that contained no Houston toad vocalizations, this was fast and simple. However, for those files that possessed vocalizations, quantification and interpretation required greater effort. Automated methods of detection required < 6 h to complete. It required an additional hour to quantify and interpret results. In other words, 7 h of analytical effort (1 h of active effort) are required to interpret approximately 80 h of audio from a single location. For comparison, the expected MCS effort is between 30 and 60 min per site per season (USFWS 2007). These estimates do not include time or cost invested in the logistics of using ARDs, which are highly variable and difficult to estimate (i.e., deployment, periodic battery and memory card changes, transfer of data from removable memory, and file organization).

Management implications

Errors inherent to traditional MCSs include observer bias, temporal variation, ease of access, right of entry, hazardous roadways, and presence of observer effects (Bee et al. 2007; Crouch and Paton 2002; Cook et al. 2011; Corn et al. 2011; Pierce and Hall 2013). Many, but not all, of these errors can be corrected via the implementation of ARDs. Although ARDs have their own suite of errors (i.e., data loss, battery life, theft, physical right of entry), the advantages they offer may outweigh these short-

comings. Although ARDs are becoming more commonly implemented, we are unaware of any ongoing long-term monitoring program for anurans that uses automated detection in practice. This represents a growing body of anuran chorusing data with limited implementation of available tools for analyzing said data. Given recent advancements, software now offers simple user-friendly foundations for complete development of robust and reliable automated audio pattern recognition tools. Improved methods of creating these tools may enable researchers to better interpret and apply these detection tools, as indicated by this research involving the endangered Houston toad. Furthermore, as the use of ARDs and methods of automated detection become more commonplace, more strict management actions are likely to follow, which inevitably leads to increased interaction, and ideally cooperation, with landowners, minimizing the occurrence of the errors associated with ARDs as outlined above.

More data can be provided by ARDs than by MCSs, and when coupled with automated detection tools, data can be processed quickly and consistently. However, at this time, ARDs and methods of automated detection are not a panacea. During our study ARDs were unable to detect vocalizations emanating from adjacent ponds; thus they are less sensitive to chorusing at great distances than human surveyors. These drawbacks make ARDs less suitable than MCSs for informing regulatory agencies given the current survey guidelines (USFWS 2007). To meet the requirements, as they exist presently, an ARD would be placed at each and every pond within and adjacent to a proposed project area, which is in most cases not feasible. Thus, combining both ARDs and MCSs may be the best solution. Ultimately, our recognizer performed such that no single survey night possessed even a single uncharacteristic call that went overlooked. Thus, any errors resulting from insufficient signal-to-noise ratio had no underlying implications on the recognizer's ability to provide presence/absence data at the site level that are critical to informing state and federal agencies of the current and potentially underestimated occurrence of the rare and endangered Houston toad.

Supplemental Material

Please note: The *Journal of Fish and Wildlife Management* is not responsible for the content or functionality of any supplemental material. Queries should be directed to the corresponding author for the article.

Reference S1. Potter FEJ, Brown LE, McClure WL, Scott NJ, Thomas RA. 1984. Recovery plan for the Houston toad (*Bufo houstonensis*). U.S. Fish & Wildlife Service.

Found at DOI: <http://dx.doi.org/10.3996/052017-JFWM-047.S1>; also available at <http://www.amphibians.org/wp-content/uploads/2013/07/Huston-Toad-Recovery-Plan.pdf> (12,103 KB PDF).

Reference S2. Price AH. 2003. The Houston toad in Bastrop State Park 1990–2002: a narrative. Open- file report 03-0401, Texas Parks & Wildlife Department.

Found at DOI: <http://dx.doi.org/10.3996/052017-JFWM-047.S2> (933 KB PDF).

Reference S3. [USFWS] U.S. Fish and Wildlife Service. 1999. Survey Protocol for the Arroyo toad. Carlsbad & Ventura, California: U.S. Fish and Wildlife Service.

Found at DOI: <http://dx.doi.org/10.3996/052017-JFWM-047.S3>; also available at <https://www.fws.gov/pacific/ecoservices/endangered/recovery/documents/AroyoToad.1999.protocol.pdf> (25 KB PDF).

Reference S4. [USFWS] U.S. Fish and Wildlife Service. 2005. Revised guidance on site assessments and field surveys for the California red-legged frog. Sacramento, California: U.S. Fish and Wildlife Service.

Found at DOI: <http://dx.doi.org/10.3996/052017-JFWM-047.S4>; also available at https://www.fws.gov/arcata/es/amphibians/crlf/documents/20050801_CRLF_survey-guidelines.pdf (144 KB PDF).

Reference S5. [USFWS] U.S. Fish and Wildlife Service. 2006. Chiricahua leopard frog (*Rana chiricahua*) draft recovery plan with appendices. Appendix E:E-1–E-15.

Found at DOI: <http://dx.doi.org/10.3996/052017-JFWM-047.S5>; also available at https://www.fws.gov/southwest/es/Documents/R2ES/DRAFT_Recovery_Plan_for_the_Chiricahua_Leopard_Frog_with_Appendices.pdf (6,427 KB PDF).

Reference S6. [USFWS] U.S. Fish and Wildlife Service. 2007. Section 10(a)(1)(A). Scientific permit requirements for conducting Houston toad presence/absence surveys. Austin, Texas: U.S. Fish and Wildlife Service.

Found at DOI: <http://dx.doi.org/10.3996/052017-JFWM-047.S6>; also available at https://www.fws.gov/southwest/es/Documents/R2ES/Houston_toad_survey_requirements.pdf (29 KB PDF).

Audio S1. Training data used within SongScope© bioacoustics software to build a recognizer for the call of the endangered Houston toad *Anaxyrus houstonensis*.

Found at DOI: <http://dx.doi.org/10.3996/052017-JFWM-047.S7> (21,646 KB PDF)

Audio S2. Training data used within SongScope© bioacoustics software to build a recognizer for the call of the endangered Houston toad *Anaxyrus houstonensis*.

Found at DOI: <http://dx.doi.org/10.3996/052017-JFWM-047.S8> (24,077 KB PDF)

Audio S3. Training data used within SongScope© bioacoustics software to build a recognizer for the call of the endangered Houston toad *Anaxyrus houstonensis*.

Found at DOI: <http://dx.doi.org/10.3996/052017-JFWM-047.S9> (19,956 KB PDF)

Audio S4. Training data used within SongScope©

bioacoustics software to build a recognizer for the call of the endangered Houston toad *Anaxyrus houstonensis*.

Found at DOI: <http://dx.doi.org/10.3996/052017-JFWM-047.S10> (19,407 KB PDF)

Audio S5. Training data used within SongScope© bioacoustics software to build a recognizer for the call of the endangered Houston toad *Anaxyrus houstonensis*.

Found at DOI: <http://dx.doi.org/10.3996/052017-JFWM-047.S11> (23,897 KB PDF)

Audio S6. Training data used within SongScope© bioacoustics software to build a recognizer for the call of the endangered Houston toad *Anaxyrus houstonensis*.

Found at DOI: <http://dx.doi.org/10.3996/052017-JFWM-047.S12> (20,892 KB PDF)

Audio S7. Training data used within SongScope© bioacoustics software to build a recognizer for the call of the endangered Houston toad *Anaxyrus houstonensis*.

Found at DOI: <http://dx.doi.org/10.3996/052017-JFWM-047.S13> (23,369 KB PDF)

Acknowledgments

We are grateful for the support and assistance of Magellan Midstream Partners L.P. and the staff at Zephyr Environmental Corporation for their assistance with sites in Robertson County. We thank Jeff Farrar for his daily assistance in coordinating all of the respective teams, as well as the Boy Scouts of America, Blue Bonnet Energy, the Musgrave Family, and the Texas Department of Transportation, each for right of entry to additional sites. We had exceptional field personnel in the efforts of all those who surveyed for Houston toads throughout the 2014 breeding season, especially Jay Dixon, D.J. Stout, Jim Bell, Tim Clarke, Mike Horvath, and Jennifer Knowles. Finally, we thank Paul Crump, three anonymous reviewers, and the Associate Editor, who each provided comments that improved an earlier version of this manuscript. All work that was conducted to complete this study was performed under scientific permit TE-039544-1 issued to M.R.J.F. by the USFWS.

Any use of trade, product, website, or firm names in this publication is for descriptive purposes only and does not imply endorsement by the U.S. Government.

References

- Acevedo MA, Villanueva-Rivera LJ. 2006. Using automated digital recording systems as effective tools for the monitoring of birds and amphibians. *Wildlife Society Bulletin* 34:211–214.
- Aide TM, Corrada-Bravo C, Campos-Cerqueira M, Milan C, Vega G, Alvarez R. 2013. Real-time bioacoustics monitoring and automated species identification. *PeerJ* 1:e103.



- Barclay RM. 1999. Bats are not birds—a cautionary note on using echolocation calls to identify bats: a comment. *Journal of Mammalogy* 80:290–296.
- Bedoya C, Isaza C, Daza JM, López JD. 2014. Automatic recognition of anuran species based on syllable identification. *Ecological Informatics* 24:200–209.
- Bee MA, Swanson EM. 2007. Auditory masking of anuran advertisement calls by road traffic noise. *Animal Behaviour* 74:1765–1776.
- Bridges AS, Dorcas ME. 2000. Temporal variation in anuran calling behavior: implications for surveys and monitoring programs. *Copeia* 2000:587–592.
- Brown LE. 1971. Natural hybridization and trend toward extinction in some relict Texas toad populations. *Southwestern Naturalist* 16:185–199.
- Charif RA, Waack AM, Strickman LM. 2010. Raven Pro 1.4. Ithaca, New York: Cornell Lab of Ornithology.
- Cook RP, Tupper TA, Paton PW, Timm BC. 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* 6:25–39.
- Corn PS, Muths E, Kissel AM, Scherer RD. 2011. Breeding chorus indices are weakly related to estimated abundance of boreal chorus frogs. *Copeia* 2011:365–371.
- Crouch WB III, Paton PW. 2002. Assessing the use of call surveys to monitor breeding anurans in Rhode Island. *Journal of Herpetology* 36:185–192.
- Digby A, Towsey M, Bell BD, Teal PD. 2013. A practical comparison of manual and autonomous methods for acoustic monitoring. *Methods in Ecology and Evolution* 4:675–683.
- Dorcas ME, Price SJ, Walls SC, Barichivich WJ. 2009. Auditory monitoring of anuran populations. *Amphibian ecology and conservation: a hand book of techniques*. Oxford, UK: Oxford University Press.
- Duarte A, Brown DJ, Forstner MRJ. 2014. Documenting extinction in real time: decline of the Houston toad on a primary recovery site. *Journal of Fish and Wildlife Management* 5:363–371.
- Eldridge JD. 2011. A comparison of current anuran monitoring methods with emphasis on the accuracy of automatic vocalization detection software. Master's thesis. Bowling Green: Western Kentucky University. Available: <http://digitalcommons.wku.edu/theses/1122/> (January 2018).
- Frost DR, Grant T, Faivovich J, Bain RH, Haas A, Haddad CFB, de Sa RO, Channing A, Wilkinson M, Donnellan SC, Raxworthy CJ, Campbell JA, Blotto BL, Moler P, Drewes RC, Nussbaum RA, Lynch JD, Green DM, Wheeler WC. 2006. The amphibian tree of life. *Bulletin of the American Museum of Natural History* 297:1–370.
- Gaston MA, Fuji A, Weckerly FW, Forstner MRJ. 2010. Potential component allee effects and their impact on wetland management in the conservation of endangered anurans. *PLoS One* 5:e10102.
- Goh M. 2011. Developing an automated acoustic monitoring system to estimate abundance of Cory's Shearwaters in the Azores. Masters thesis. London: Imperial College London. Available: <https://www.iccs.org.uk/wp-content/thesis/consci/2011/Goh.pdf> (January 2018).
- Gottschalk JS. 1970. United States list of endangered native fish and wildlife. *Federal Register* 35:16047–16048.
- Hammerson G, Canseco-Márquez L. 2004. *Incilius nebulifer*. In: IUCN red list of threatened species. Version 2009.2. Available: www.iucnredlist.org. (January 2018).
- Honegger RE. 1970. Red data book. Volume 3. Amphibia and Reptilia. Gland, Switzerland: Union for Conservation of Nature and Natural Resources, Survival Service Commission.
- Hsu MY, Kam YC, Fellers GM. 2005. Effectiveness of amphibian monitoring techniques in a Taiwanese subtropical forest. *Herpetological Journal* 15:73–79.
- Jackson JT, Weckerly FW, Swannack TM, Forstner MRJ. 2006. Imperfect detection and number of auditory surveys for Houston Toads. *Journal of Wildlife Management* 70:1461–1463.
- Katz J, Hafner SD, Donovan T. 2016. Tools for automated acoustic monitoring within the R package *monitoR*. *Bioacoustics* 25:197–210.
- Narins PM, Feng AS, Lin W, Schnitzler HU, Denzinger A, Suthers RA, Xu C. 2004. Old World frog and bird vocalizations contain prominent ultrasonic harmonics. *Journal of the Acoustical Society of America* 115:910–913.
- Noda JJ, Travieso CM, Sánchez-Rodríguez D. 2016. Methodology for automatic bioacoustic classification of anurans based on feature fusion. *Expert Systems with Applications* 50:100–106.
- Obriest MK, Pavan G, Sueur J, Riede K, Llusia D, Márquez R. 2010. Bioacoustics approaches in biodiversity inventories. *Abc Taxa* 8:68–99.
- O'Neal B. 2014. Testing the feasibility of bioacoustic localization in urban environments. Masters thesis. Tampa: University of South Florida. Available: <http://scholarcommons.usf.edu/etd/5088/> (January 2018).
- Oseen KL, Wassersug RJ. 2002. Environmental factors influencing calling in sympatric anurans. *Oecologia* 133:616–625.
- Pechmann JHK, Scott DE, Semlitsch RD, Caldwell JP, Vitt LJ, Gibbons JW. 1991. Declining amphibian populations: the problem of separating human impacts from natural fluctuations. *Science* 253:892–895.
- Pierce B, Gutzweiller K. 2004. Auditory sampling of frogs: detection efficiency in relation to survey duration. *Journal of Herpetology* 38:495–500.
- Pierce BA, Hall AS. 2013. Call latency as a measure of calling intensity in anuran auditory surveys. *Herpetological Conservation and Biology* 8:199–206.
- Potter FEJ, Brown LE, McClure WL, Scott NJ, Thomas RA. 1984. Recovery plan for the Houston toad (*Bufo houstonensis*). U.S. Fish & Wildlife Service (see *Supple-*

- mental Material*, Reference S1, <http://dx.doi.org/10.3996/052017-JFWM-047.S1>); also available: <http://www.amphibians.org/wp-content/uploads/2013/07/Houston-Toad-Recovery-Plan.pdf> (January 2018).
- Price AH. 2003. The Houston Toad in Bastrop State Park 1990–2002: a narrative. Open-file report 03-0401, Texas Parks & Wildlife Department (see *Supplemental Material*, Reference S2, <http://dx.doi.org/10.3996/052017-JFWM-047.S2>).
- R Core Team 2014. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing,. Available: <http://www.R-project.org/>.
- Sanders O. 1953. A new species of toad, with a discussion of morphology of the bufonid skull. *Herpetologica* 9:25–47.
- Schmidt BR. 2003. Count data, detection probabilities, and the demography, dynamics, distribution, and decline of amphibians. *Comptes Rendus Biologies* 326:119–124.
- Sueur J, Aubin T, Simonis C. 2008. Equipment review: seewave, a free modular tool for sound analysis and synthesis. *Bioacoustics* 18:213–226.
- Swiston KA, Mennill DJ. 2009. Comparison of manual and automated methods for identifying target sounds in audio recordings of Pileated, Pale-billed, and putative Ivory-billed woodpeckers. *Journal of Field Ornithology* 80:42–50.
- Tipton BL, Hibbitts TL, Hibbitts TD, Hibbitts TJ, LaDuc TJ. 2012. Texas amphibians: a field guide. Austin: University of Texas Press.
- Towsey M, Planitz B, Nantes A, Wimmer J, Roe P. 2012. A toolbox for animal call recognition. *Bioacoustics* 21:107–125.
- [ESA] U.S. Endangered Species Act of 1973, as amended, Pub. L. No. 93-205, 87 Stat. 884 (Dec. 28, 1973). Available: <https://www.fws.gov/endangered/esa-library/pdf/ESAall.pdf> (January 2018).
- [USFWS] U.S. Fish and Wildlife Service. 1999. Survey protocol for the Arroyo toad. Carlsbad & Ventura, California: U.S. Fish and Wildlife Service (see *Supplemental Material*, Reference S3, <http://dx.doi.org/10.3996/052017-JFWM-047.S3>); also available: <https://www.fws.gov/pacific/ecoservices/endangered/recovery/documents/ArroyoToad.1999.protocol.pdf> (January 2018).
- [USFWS] U.S. Fish and Wildlife Service. 2005. Revised guidance on site assessments and field surveys for the California red-legged frog. U.S. Sacramento, California: Fish and Wildlife Service (see *Supplemental Material*, Reference S4, <http://dx.doi.org/10.3996/052017-JFWM-047.S4>); also available: https://www.fws.gov/arcata/es/amphibians/crlf/documents/20050801_CRLF_survey-guidelines.pdf (January 2018).
- [USFWS] U.S. Fish and Wildlife Service. 2006. Chiricahua Leopard Frog (*Rana chiricahua*) draft recovery plan with appendices. Appendix E:E-1 – E-15 (see *Supplemental Material*, Reference S5, <http://dx.doi.org/10.3996/052017-JFWM-047.S5>); also available: https://www.fws.gov/southwest/es/Documents/R2ES/DRAFT_Recovery_Plan_for_the_Chiricahua_Leopard_Frog_with_Appendices.pdf (January 2018).
- [USFWS] U.S. Fish and Wildlife Service. 2007. Section 10(a)(1)(A). Scientific permit requirements for conducting Houston Toad presence/absence surveys. Austin, Texas: U.S. Fish and Wildlife Service (see *Supplemental Material*, Reference S6, <http://dx.doi.org/10.3996/052017-JFWM-047.S6>); also available: https://www.fws.gov/southwest/es/Documents/R2ES/Houston_toad_survey_requirements.pdf (January 2018).
- Waddle JH, Thigpen TF, Glorioso BM. 2009. Efficacy of automatic vocalization recognition software for anuran monitoring. *Herpetological Conservation and Biology* 4:384–388.
- Weir LA, Mossman MJ. 2005. North American Amphibian Monitoring Program (NAAMP). Pages 307–313 in Lanoo M, editor. Amphibian declines: conservation status of United States amphibians. Berkeley: University of California Press.
- Wildlife Acoustics. 2011a. SongScope: bioacoustics software, version 4.1.1. Maynard, Massachusetts: Wildlife Acoustics.
- Wildlife Acoustics. 2011b. SongScope user manual: bioacoustics software (version 4.0) documentation. Maynard, Massachusetts: Wildlife Acoustics.
- Willacy RJ, Mahony M, Newell DA. 2015. If a frog calls in the forest: bioacoustic monitoring reveals the breeding phenology of the endangered Richmond Range mountain frog (*Phyllorhina richmondensis*). *Austral Ecology* 40:625–633.
- Williams PJ, Engbrecht NJ, Robb JR, Terrell VC, Lannoo MJ. 2013. Surveying a threatened amphibian species through a narrow detection window. *Copeia* 2013:552–561.
- Zimmerman BL. 1994. Audio strip transects. In measuring and monitoring biological diversity: standard methods for amphibians. Washington, D.C.: Smithsonian Institution Press.