Herpetological Conservation and Biology 15(1):1–15.
Submitted: 14 August 2019; Accepted: 16 December 2019; Published: 30 April 2020.

Improving Automated Detection of Frog Calls in Noisy Urban Habitats Using Narrow-banded Recognizers

Nolan Bielinski[1,3], Jennifer Pajda-De La O[2], Adrianna Gorniak[1], and David Wise[1]

[1]*Department of Biological Sciences, University of Illinois at Chicago, 845 West Taylor Street, 3272 SES, Chicago, Illinois 60607, USA*
[2]*Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, 851 South Morgan Street, 322 SEO, Chicago, Illinois 60607, USA*
[3]*Corresponding author, e-mail: nbieli2@uic.edu*

Abstract.—**Tracking behavioral and demographic changes of anuran populations in urban landscapes presents difficulties due to the high amount of noise interference from anthropogenic sources. In this study, we used Song Scope software to build narrow-banded recognizers that only cover a limited portion of the full spectral range of a call and tested if these recognizers can improve automated call-detection capabilities at noisy sites. We built recognizers for two species with naturally broad-spectrum calls, the Green Frog (*Lithobates clamitans*) and American Bullfrog (*L. catesbeianus*) and tested them at five noisy ponds in the suburbs of Chicago, Illinois, USA. Narrow-banded recognizers had greater percentages of true positives compared to full-spectrum recognizers. Classification indices used to assess call recognition efficacy showed that narrow-banded recognizers were more effective at all sites for the Green Frog, and at two sites for the American Bullfrog. High-frequency recognizers had 13% fewer errors caused by anthropogenic noise ($P < 0.010$) than other recognizers. Finally, for every recognizer, true positives standardized by the maximum daily value was highly and significantly correlated with the number of calls identified manually, indicating that automated detection data is an accurate proxy for the actual number of calls at noisy sites. For acoustic taxa, we recommend that scientists consider identifying broad-spectrum calls using narrow-banded recognizers to reduce detection problems associated with noise interference between anthropogenic noises and biotic acoustic signals**.

Key Words.—acoustic species; anthropogenic noise; bioacoustics; American Bullfrog; *Lithobates catesbeianus*; Green Frog; *Lithobates clamitans*; soundscape

Introduction

Noise pollution is a significant disturbance to wildlife in urban landscapes. Anthropogenic noise, defined as noise from human technology and activity, represents a major obstacle to anurans because they rely heavily on vocalizations to find breeding grounds, choose mates, and settle territorial disputes (Gerhardt 1994). Anthropogenic noise is a pervasive edge effect characterized by low-frequency intermittent buzzing or humming (Pijanowski et al. 2011a), which may prominently overlap the frequencies of biological sounds including some anuran mating calls (Bee and Swanson 2007). Many natural habitats in urban regions (even interiors of preserves) experience significant noise pollution, with the main sources being traffic from cars, airplanes, and trains. For example, highway noise from 150 m away can reach 70 dB (Warren et al. 2006). Train noise, although more infrequent, can produce noise up to 90 dB within 10 m (Nolan Bielinski, unpubl. data). Areas under airport flyover paths can reach 74 dB (Warren et al. 2006).

Consequently, frogs living in urban landscapes experience the effects of a dramatically altered soundscape. Traffic noise exposure has been shown to increase corticosterone production and decrease antimicrobial peptide production (Tennessen et al. 2018), change the calling attributes of males (Sun and Narins 2005; Parris et al. 2009; Cunnington and Fahrig 2010), limit male participation in chorusing (Kaiser et al. 2011), and alter the response of females to male calls (Bee and Swanson 2007). Looking across all studies investigating the effects of anthropogenic noise on frog calling behavior, no explicit pattern emerges, suggesting that responses may be species- or region-specific. This highlights the need for further monitoring and improved research techniques for studying anurans in noisy habitats. Future studies would benefit from increased sample sizes and improved efficiency of acoustic analysis through the implementation of the latest soundscape software techniques.

Broad-scale, manual (in-person) calling surveys are often used to assess the vulnerability of frog species (Williams et al. 2013) and population trends over time (Gibbs and Breisch 2001) but are inadequate in many ways. For example, the protocol for the North American Amphibian Monitoring Program advises volunteers to collect data for 2 h total, beginning 30 min after sunset, which systematically ignores post-midnight calling activity (Bridges and Dorcas 2000; US Geological Survey. 2019. North American Amphibian Monitoring Program. Available from https://www.usgs.gov/centers/pwrc/science/north-american-amphibian-monitoring-program?qt-science_center_objects=0#qt-science_center_objects [Accessed 2 August 2019]). Moreover, volunteers dedicate only 3 min to each listening point. Considering that volunteers may disturb frogs when moving to each location, more time may be needed for the frog community to revert to natural calling behavior (Crouch and Paton 2002). In general, manual surveys also risk having high instances of misidentification and omission of species, especially when dealing with faint, infrequent, or scattered calls, or during nights when multiple species are calling simultaneously (Genet and Sargent 2003). Consequently, monitoring programs may significantly bias abundance (Lotz and Allen 2007) or presence/absence data (Genet and Sargent 2003).

Most deficiencies associated with manual surveys can be resolved using automated recording devices (ARDs). ARDs can be used to collect data throughout the entire night or even entire 24-h cycles, which increases the chance of recording rare (MacLaren et al. 2017) and cryptic species (Engbrecht 2011) and provides a more accurate picture of true calling phenologies (Bridges and Dorcas 2000). Breeding ponds also experience less disturbance once recording devices are installed (Acevedo and Villanueva-Rivera 2006). After data collection in the field, recorded calls can be reviewed by smaller cohorts and allow for both visual (spectral) and auditory review to identify species in a more comprehensive and consistent manner, and problematic recordings can be replayed for clarification. Furthermore, soundscape software can be used to detect and analyze changes in several call characteristics such as dominant frequency, pulse rate, and sound pressure level, that may be indiscernible without equipment (Cocroft and Ryan 1995). These quantifiable attributes of the soundscape allow researchers to test hypotheses on how noise may alter frog communities by changing favorable calling strategies.

The final benefit of ARDs is the ability to filter through large datasets to automatically identify calls using digital signal processing algorithms called “recognizers.” Recognizers look for patterns within the time-frequency state-space to isolate patterns matching the call structure of a target species (Wildlife Acoustics. 2011. Song Scope bioacoustics software version 4.0 documentation. Available from https://www.wildlifeacoustics.com/ [Accessed 2 August 2019]). Using recognizers dramatically reduces the time required to convert recorded soundscapes into quantifiable data. This pronounced increase in scale of data collection and analysis afforded by recognizers could be particularly beneficial when attempting to document rare species, catching the onset of breeding from explosive breeders, or monitoring changes in range or behavior of species in response to climate change or the spread of disease across a region.

Recognizer performance in noisy soundscapes remains untested in the literature. The purpose of this study was to investigate the capabilities of recognizers with different frequency ranges to identify frog calls in noisy habitats to advance acoustic monitoring efforts for researchers in urban soundscapes. There are several commercially available soundscape software programs that implement recognizer-like functionality, including Song Scope, Kaleidoscope, MonitoR, and RavenPro; a comparison of these programs is offered by Knight et al. (2017). We chose to use Song Scope because it was rated as the best performing program (Knight et al. 2017), has a relatively low learning curve, is used in many soundscape studies, and is available for free.

The efficacy of recognizers has already been investigated with promising results in areas where anthropogenic noise is minimal (Waddle et al. 2009; Eldridge 2011; Brauer et al. 2016; Crump and Houlahan 2017; MacLaren et al. 2017), but it is unclear whether recognizers are useful in noisy environments. For example, the best strategy for implementing automated detection may differ greatly between sites with pristine versus degraded soundscapes. Thus, it would be beneficial for researchers to know if they should alter the parameters of their recognizer depending on the noise profile of their sites, regardless of the soundscape software being used or the taxon being studied.

We conducted this research in the suburbs of Chicago, Illinois, USA. In urban landscapes like the Chicago region, few natural habitats are completely shielded from noise pollution (U.S. Department of Transportation [DOT]. 2019. National Transportation Noise Map. Available from https://maps.bts.dot.gov/arcgis/apps/webappviewer/index.html?id=a303ff5924c9474790464cc0e9d5c9fb [Accessed 2 August 2019]). In Chicago the gridded streets and large highways bisect green spaces, airplanes crowd the skies from the second and twenty-fifth busiest airports in the U.S. (Federal Aviation Administration. 2019. Air traffic by the numbers. https://www.faa.gov/air_traffic/by_the_numbers/ [Accessed 18 February 2019]), and train traffic is high enough to designate Chicago as busiest rail freight gateway in the USA. We specifically chose our study sites based on high noise levels from the above-mentioned sources.

We recorded Green Frog (*Lithobates clamitans*) and American Bullfrog (*L. catesbeianus*; hereafter, bullfrog) calls to test these soundscape methods because both species are widespread in urban areas. Additionally, these species have calls with low dominant frequencies,

Table 1. Distances to noise sources and measured noise levels during equipment setup at each study site. Ambient noise is defined as the noise level without any identifiable or distinguishable sources. Peak noise levels were used for ambient and acute noises.

Site	Nearest road (m)	Nearest train tracks (m)	Ambient noise (dBA)	Acute noises (dBA)
A	110	3,885	50	cars (65); truck (70); motorcycle (75)
B	402	1,832	51	train (64); airplane (59)
C	180	119	50	cars (52); train (81); airplane (60)
D	176	174	46	cars (48); train (72); airplane (61)
E	383	37	44	cars (45); train (83); airplane (69)

meaning that spectral interference with anthropogenic noise should be greater for them than other local species (Phillips et al. 1999; Pijanowski et al. 2011b). For both species, we created and compared multiple recognizers with different spectral parameters. Additionally, we characterized and measured background noise from samples of true positives (TPs), false positives (FPs), and false negatives (FNs) to describe the specific strengths and weaknesses of each recognizer. Finally, we compared manual counts to the respective recognizer output to determine if recognizer data can be used as a reliable proxy for actual call rate (defined here as the number of calls over time) at noisy sites.

Materials and Methods

Audio surveys.—We deployed ARDs (SongMeter model SM4, Wildlife Acoustics, Maynard, Massachusetts, USA) at five breeding sites (A-E) between 25 June 2016 and 15 August 2016 on forest preserves and private land with ponds in the suburbs of Chicago (Supplemental Table S1). We secured the ARDs to trees approximately 1 m away from the edge of each pond. We programmed ARDs to record the first 5 min of every hour from 1800 to 0100 for four nights. Throughout the study, the ARDs did not require any battery or memory card changes, which means we avoided all physical disturbance around the pond edge.

Study sites.—Site A is a road-side permanent pond located in Cherry Hill Woods (Cook County Forest Preserve). The surrounding area has a mixture of forest, open woodlands, and savanna. It is dominated by oak (*Quercus* spp.) and hickory (*Carya* spp.). Site B is a large slough located in the center of Wolf Road Woods (Cook County Forest Preserve). This area has rolling hills and a mixture of forest and open woodlands with interspersed ephemeral pools. It is dominated by oak and hickory. Sites C and D are artificial ponds located on private property in Palos Hills, Illinois. The neighborhood has scattered homes adjacent to patches of oak-dominated forest on rolling hills. Sites C and D are close to each other (118 m) so they share the same anthropogenic noise sources; however, calls from one pond could not be heard or detected at the other, meaning that recorded calls from these sites were from different individuals. Site E is a small semi-permanent pond located in Van Patten Woods (Lake County Forest Preserve). It is bordered on one side by a train track, and on all other sides by oak forest.

Noise profiles at sites.—During installation and removal of the ARDs, we spent an hour collecting data on sound levels using a sound pressure level meter (Model DS-HWCJ04, Koolertron, Shenzhen, China). We recorded the distance to noise sources and the sound pressure level of ambient noise (general ongoing sounds) and acute noise, defined here as distinct instances of punctuated anthropogenic sounds from cars, trucks, motorcycles, trains, and airplanes (Table 1). Every site is close to anthropogenic noise sources, with site B being the farthest away from the nearest road or train track, at 402 m. The measured noise levels from acute instances of anthropogenic noise reflect these proximities. Putting these noise levels in a biological context, at 1 m Green Frogs can call at 84 dB (Bee and Perrill 1996) and bullfrogs call at 80 dB (Simmons 2004). Using the inverse square law for sound intensity, this is an equivalence of 64 dB and 60 dB, respectively, at 10 m, approximately the distance where females can interpret multiple calls from potential mates, based on male territory size (Wells 1977). In our recordings we also encountered acoustic signals from insects including crickets (family Gryllidae), Dog-day Cicadas (*Neotibicen canicularisi*), and katydids (family Tettigoniidae), and vocalizations from birds including Red-winged Blackbirds (*Agelaius phoeniceus*), Great Blue Herons (*Ardea herodias*), and several others. We did not identify insect and bird species for our noise categorization analysis. Instead, we classified them as biological noise.

Supplemental recordings.—During the same field season, we collected additional training data (i.e., data required to build recognizers; see below) from study sites via supplemental recordings of extra nights, and from Hegewisch Marsh (Chicago Park District) and Hickory Creek (Will County Forest Preserve).

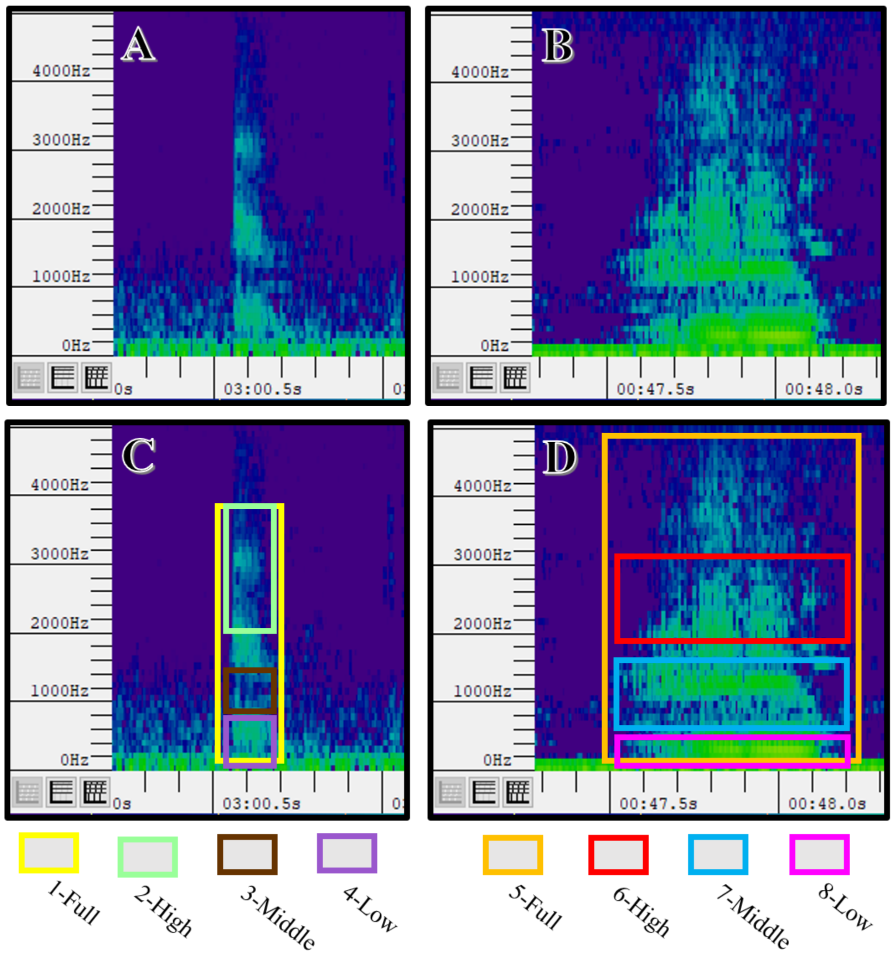


Figure 1. Spectrograms of mating calls for our target species with minimal background noise interference. (A) The Green Frog (*Lithobates clamitans*) has a quick "banjo-strumming" call with several simultaneous frequency harmonics. (B) The American Bullfrog (*Lithobates catesbeianus*) has a longer, sometimes vibrating call, with several simultaneous frequency harmonics. (C) Green Frog recognizers and (D) bullfrog recognizers are visually depicted over their respective calls to show the varying frequency ranges of each recognizer.

Recognizer development.—We performed all soundscape data analysis on Song Scope bioacoustics software (version 4.1.5; Wildlife Acoustics, Inc., Concord, Massachusetts, USA). To build recognizers, the software must be fed training data, where one manually identifies confirmed signals from a target species within a spectrogram and collects them in a Build Recognizer page, and then recognizer parameters can be adjusted and filtered to properly match the vocalizations in the training data (Wildlife Acoustics. 2011. *op. cit.*). Song Scope then uses hidden Markov models to construct a model call to compare to candidate vocalizations in new recordings. More detail on recognizers is available elsewhere (Agranat, I. 2009. Automatically identifying animal species from their vocalizations. Wildlife Acoustics, Concord, Massachusetts, USA. Available from https://pdfs.semanticscholar.org/7129/78f16ef0d1d4e81fcf3dc6bab77406b54d1e.pdf [Accessed 1 June 2019]).

Normally, frog calls are structurally simple enough to cover using a single recognizer. Most calls lack the spectral and temporal complexities that may warrant splitting up a call into multiple sub-signals, which is sometimes implemented to identify more complex avian songs (Gelling 2010). The Green Frog and the bullfrog have short, simple breeding calls usually consisting of one syllable, with most of the acoustic power at a low frequency; however, both calls have simultaneous medium- and high-frequency harmonics, as seen by the large signal range displayed along the y-axis of a spectrogram (Fig. 1A and B). We took advantage of these naturally large spectral ranges by building four recognizers for each species that focused on a different spectral range of their call (Table 2). We considered frequency range as the most important parameter because anthropogenic noise often overlapped part of the natural frequency range of our target calls (for a full list of every parameter, see Appendix 1).

Considering the high levels of anthropogenic noise at our sites, calling males experienced spectral interference in the low frequency range. Moreover, the soundscapes

Table 2. Recognizer information: Recognizers 1–4 are for Green Frog (*Lithobates clamitans*) mating calls; Recognizers 5–8 are for American Bullfrog (*Lithobates catesbeianus*) mating calls.

Recognizer Name	Target Species	Type	Frequency range (Hz)
1-Full	Green Frog	Conventional	187.5 to 3,875
2-High	Green Frog	Narrow	2,000 to 3,875
3-Middle	Green Frog	Narrow	812.5 to 1,562.5
4-Low	Green Frog	Narrow	125 to 875
5-Full	bullfrog	Conventional	187.5 to 5,250
6-High	bullfrog	Narrow	2,062.5 to 3,000
7-Middle	bullfrog	Narrow	562.5 to 1625
8-Low	bullfrog	Narrow	187.5 to 500

also included avian and insect signals, which overlap with frog calls at higher frequencies. Thus, for the Green Frog and the bullfrog, we built recognizers that encompassed the full frequency range (a conventional recognizer), and three narrow-banded recognizers covering the highest portion, a middle portion, and the lowest portion of the frequency range. We named Green Frog recognizers 1-Full, 2-High, 3-Middle, and 4-Low; we named bullfrog recognizers 5-Full, 6-High, 7-Middle, and 8-Low (Fig. 1C and D).

Recognizer analysis.—We investigated whether the narrow-banded recognizers could outperform the conventional full recognizers in noisy environments. For each 5 min recording (of 200 recordings total), we conducted a manual count of the number of calls from Green Frogs and bullfrogs, providing a reference, or a condition positive with which we could compare recognizer results. Next, for recordings with Green Frog calls (112 total) we ran recognizers one through four, and for recordings with bullfrog calls (98 total) we ran recognizers five through eight. We counted TPs (calls correctly identified), FPs (sounds that were wrongly identified as calls), and FNs (missed calls). We chose to not count true negatives (correctly ignored noises) because the constancy of anthropogenic noises makes distinguishing independent units of noise very difficult. We recorded the amount of time it took make these counts.

To determine if recognizers produced accurate estimations of the actual call rate at a site, we compared relationships between TPs and manual counts over the same time-series. To do this we standardized the TPs and manual counts by their respective maximum values for each day by site. We then ran a correlation analysis between the standardized manual count and TP values per recognizer. Because our data did not meet the normality assumptions for a Pearson correlation (most likely because calling behavior varied greatly with the weather from each experimental night), we used a Kendall rank correlation test to calculate Tau-b, a statistic that tests the strength of association in ranked data while also making adjustments for ties (McLeod 2011). For each correlation, x represents the standardized manual counts from every date by hour, and y represents the standardized TPs from every date by hour.

To investigate how background noise affects recognizer accuracy, we took a random subsample of a maximum of five TPs, five FPs and five FNs from every hour during the first night. Then, for each sample, we categorized the background noise type as either biological, geological, anthropogenic, no noise, or recorder error. To test if the type of background noise affects the ability of recognizers to correctly identify calls, we ran a PERMANOVA of the Bray-Curtis dissimilarities of noise category counts for TPs, FPs, and FNs across recognizers, stratified by site (Oksanen et al. 2019). We also ran a difference in proportions test to determine if certain recognizers were less impeded by anthropogenic noise interference. Next, we measured the power (peak-to-peak voltage) at each of these samples and calculated 95% confidence intervals to compare noise volume levels across recognizers.

Finally, at every site, we assessed the efficacy of each recognizer by calculating the following indices: (1) True positive rate (TPR) = True Positives / Manual Count; (2) Precision (PPV) = True Positives / (Manual Count + False Positives); (3) False negative rate (FNR) = False Negatives / Manual Count; and (4) False discovery rate (FDR) = False Positives / (False Positives + True Positives). These are common indices used in classification scenarios. In this case, they are describing the ability of a recognizer to correctly identify a sound in the recording as a frog call of a target species. We created our graphs in Tableau 2019.1.3 (Tableau Software, Inc., Seattle, Washington, USA). We conducted the statistical analyses in R version 3.5.0. (R Core Team 2019).

Results

For site A, a programing error resulted in the loss of data for the 1800 hour. After this omission, we recorded 156 audio files totaling 780 min from our five sites.

Green Frog recognizer performance.—From all sites we manually identified 11,608 Green Frog calls. Each of the four Green Frog recognizers varied in performance based on the research site (Fig. 2). When comparing the ratio of TPs to the manual counts per recognizer, the conventional recognizer 1-Full was never the top-identifying recognizer.

For each recognizer, we ran separate Kendall's Tau-b correlations on standardized manual counts and TPs. All recognizer TP's were correlated with actual calling taking place. The level of correlation was highest for

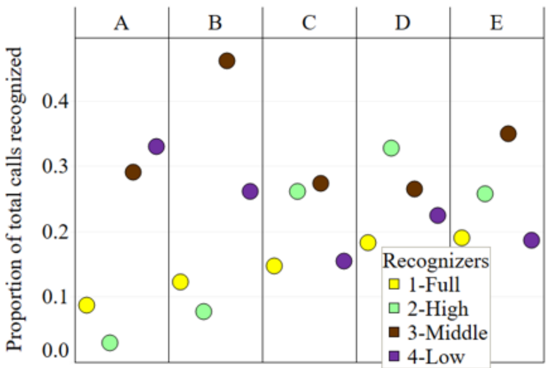


Figure 2. The proportion of correctly identified calls over manual counts for the Green Frog (*Lithobates clamitans*) per recognizer and site (A-E). Recognizer 1-Full (yellow) is the conventional recognizer, and all other recognizer are narrow-banded, meaning that only part of the full frequency range of typical Green Frog calls were scanned in the spectrograms.

recognizer 4-Low (Tau-b = 0.721, $P < 0.001$) followed by 3-Middle (Tau-b = 0.712, $P < 0.001$), 1-Full (Tau-b = 0.607, $P < 0.001$), and 2-High (Tau-b = 0.538, $P < 0.001$).

Next, we compared the classification indices between recognizers (Table 3). The best-rated value (highest value for TPR and PPV and lowest for FNR and FDR) in each index. Depending on the site, the top performing recognizers were either 2-High or 3-Middle. It is also evident from this table that 4-Low across all sites had too many FPs to be considered effective, considering its remarkably high FDR values. Using the optimum recognizer per site (Table 3), we would achieve a mean TPR of 0.345, compared to a mean TPR of 0.146 from the conventional recognizer 1-Full alone. Furthermore, this optimum set of recognizers would produce a mean FNR of 0.646, compared to a mean FNR of .835 from 1-Full alone.

Noise description during Green Frog recognizer output.**—**Site B had the highest amount of biological background noise recorded, which was from intense insect chorusing. Additionally, sites D and E had higher proportions of geophonies primarily from wind (Appendix 2). Green Frog recognizer output indicated a weak effect of background noise on recognizer identification capability (P[Pseudo-$F_{3,16}$] = 0.025, r^2 = 0.11). There were no general patterns between noise and TP, FP, and FN classifications within recognizers (Supplemental Fig. S1).

All narrow-banded recognizers experienced significantly lower average noise volume levels than the conventional recognizer, with 3-Middle being the lowest, and 2-High being the second lowest (Fig. 3). When considering recognition errors across sites (Fig. 4), the percentage of noise classified as anthropogenic is lowest in recognizer 2-High. A difference in proportions

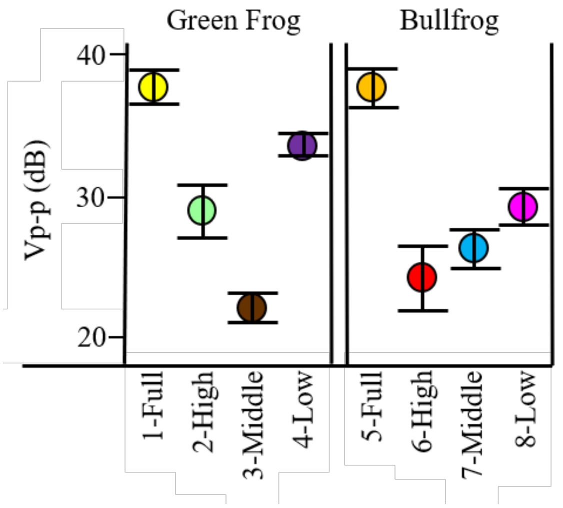


Figure 3. Average relative volume in peak to peak voltage (Vp-p) of background noise samples from our recognizers (solid vertical bars) for the Green Frog (*Lithobates clamitans*) and American Bullfrog (Lithobates catesbeianus). Vertical lines represent 95% confidence intervals. All narrow-banded recognizers were significantly lower than the conventional recognizers (1-Full and 5-Full).

test showed that there were 13% fewer errors attributed to anthropogenic noise for high-frequency recognizers ($P < 0.01$), demonstrating that recognizers that avoid low frequencies can reduce the amount of anthropogenic noise interference that cause identification errors.

Bullfrog recognizer performance.**—**From all sites we manually identified 3,695 bullfrog calls. Site C was not used in the analysis, as no bullfrogs were present. Each of the four bullfrog recognizers varied in performance based on the research site (Fig. 5). When comparing the ratio of TPs to the manual counts per recognizer, narrow-banded recognizers were the top performer in 3 out of 4 sites.

As with the Green Frog, all bullfrog recognizer TPs were correlated with actual calling taking place. The level of correlation was highest for 7-Middle (Tau-b = 0.688, $P < 0.001$) followed by 8-Low (Tau-b = 0.611, $P < 0.001$), 5-Full (Tau-b = 0.531, $P < 0.001$), and 6-High (Tau-b = 0.443, $P < 0.001$). According to the classification indices (Table 4), the conventional bullfrog recognizer (5-Full) had the best scores for two sites, D and E. Recognizer 7-Middle scored best for A, and 8-Low scored best for B. Using the optimum recognizer per site, we would achieve a mean TPR of 0.558, compared to a mean TPR of 0.281 from the conventional recognizer 5-Full alone, and a mean FNR of 0.450 compared to a mean FNR of 0.719 from 5-Full alone (Table 4).

Noise description during bullfrog recognizer output.**—**Site A shows a high level of biological noise

Table 3. Classification indices by site for Green Frog (*Lithobates clamitans*) recognizers. The indices are as follows: true positive rate (TPR); precision (PPV); false negative rate (FNR); and false discovery rate (FDR). The best-rated index value in each category is bolded. The most optimal recognizer per site is bolded and starred.

Site	Recognizer	TPR	PPV	FNR	FDR
A	1-Full	0.09	0.15	0.83	0.85
	2-High	0.03	0.13	0.92	0.87
	3-Middle	0.29	**0.21**	0.71	**0.79**
	4-Low	**0.33**	0.09	**0.67**	0.91
B	1-Full	0.12	0.01	0.88	0.99
	2-High	0.07	**0.04**	0.93	**0.96**
	3-Middle	**0.45**	0.01	**0.55**	0.99
	4-Low	0.25	0.00	0.75	1.00
C	1-Full	0.15	0.23	0.84	0.77
	2-High	0.26	**0.27**	0.70	**0.73**
	3-Middle	**0.27**	0.24	**0.68**	0.76
	4-Low	0.15	0.04	0.80	0.96
D	1-Full	0.18	0.28	0.82	0.72
	2-High	**0.33**	**0.38**	**0.67**	**0.62**
	3-Middle	0.26	0.23	0.74	0.77
	4-Low	0.22	0.07	0.78	0.93
E	1-Full	0.19	**0.78**	0.81	**0.22**
	2-High	0.26	0.70	0.74	0.30
	3-Middle	**0.35**	0.65	**0.65**	0.35
	4-Low	0.19	0.30	0.81	0.70

interference in recognizers 5-Full and 6-High, as their frequency range overlapped regularly with bird calls (Appendix 3). This was not the case for Green Frog recognizers because their call length parameter was shorter than the signals coming from birds, which disqualified them as potential target calls. A large amount of biological noise in site B interfered with the 6-High recognizer, which is from high-frequency insect chorusing. Bullfrog recognizer output indicated a weak effect of background noise on recognizer identification capability (P[Pseudo-$F_{3,16}$] = 0.094, r^2 = 0.22). There were no general patterns between noise and TP, FP, and FN classifications within recognizers (Supplemental Fig. S2).

All narrow-banded recognizers experienced significantly lower noise volume levels than the conventional recognizer, with 6-High being the lowest, and 7-Middle being the second lowest (Fig. 3). When considering recognition errors across all sites (Fig. 4), the percentage of noise classified as anthropogenic is lowest in recognizer 6-High. A difference in proportions test showed that anthropogenic noise produced 13% fewer errors for high-frequency recognizers (P < 0.01), mirroring the results from the Green Frog recognizers.

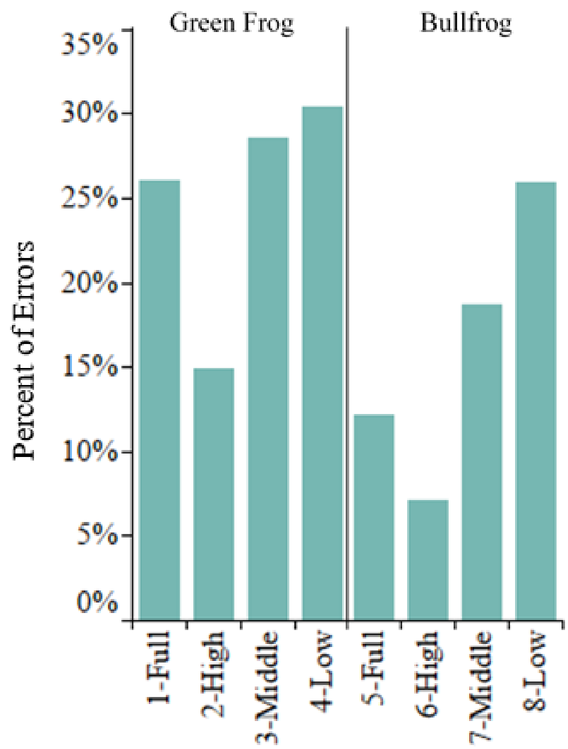


Figure 4. Percentage of errors (FPs + FNs) from our recognizers classified as anthropogenic noise for both Green Frogs (*Lithobates clamitans*) and American Bullfrogs (*Lithobates catesbeianus*).

Discussion

In contrast to previous studies using recognizers (Waddle et al. 2009; Eldridge 2011; MacLaren et al. 2017), this study implemented soundscape techniques at sites that were particularly noisy from multiple anthropogenic sources. Frequency overlap between the frog calls and noise was common. It was a challenge to build recognizers for our target species because their calls are very short relative to other acoustic signals, for

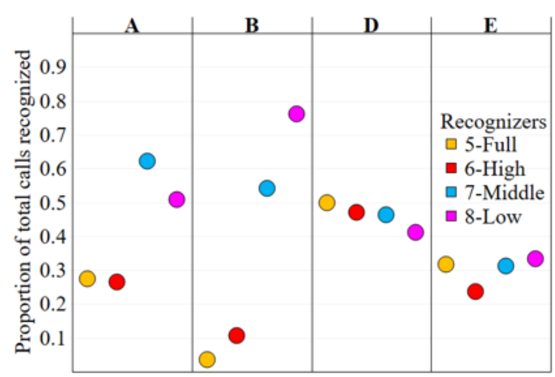


Figure 5. The proportion of correctly identified calls over manual counts for the American Bullfrog (*Lithobates catesbeianus*) per recognizer and site (A, B, D, and E). Recognizer 5-Full (orange) is the conventional recognizer, and all other recognizer are narrow-banded, meaning that only part of the full frequency range of typical bullfrog calls were scanned in the spectrograms.

Table 4. Classification indices by site for American Bullfrog (*Lithobates catesbeianus*) recognizers. The indices are as follows: true positive rate (TPR); precision (PPV); false negative rate (FNR); and false discovery rate (FDR). The best-rated index value in each category is bolded. The most optimal recognizer per site is bolded and starred.

Site	Recognizer	TPR	PPV	FNR	FDR
A	5-Full	0.27	0.19	0.73	0.81
	6-High	0.26	0.07	0.74	0.93
	7-Middle	**0.62**	0.26	**0.38**	0.74
	8-Low	0.51	**0.35**	0.49	**0.65**
B	5-Full	0.03	0.08	0.97	0.92
	6-High	0.11	0.06	0.89	0.94
	7-Middle	0.54	0.10	0.46	0.90
	8-Low	**0.76**	**0.16**	**0.24**	**0.84**
D	***5-Full***	**0.50**	0.57	**0.50**	0.43
	6-High	0.47	0.44	0.53	0.56
	7-Middle	0.46	**0.59**	0.54	**0.41**
	8-Low	0.41	0.46	0.59	0.54
E	***5-Full***	0.32	**0.22**	0.68	**0.78**
	6-High	0.24	0.11	0.76	0.89
	7-Middle	0.31	0.11	0.69	0.89
	8-Low	**0.35**	0.16	**0.65**	0.84

the Green Frog especially. Due to this, the recognizers have limited information to work with along the time axis, making it harder for the recognizers to identify patterns in the state-space of a spectrogram (Brauer et al. 2016). Some noises, like splashes in the water or snaps of branches, mimicked the shape of Green Frog calls on the spectrogram, resulting in some FPs (Appendix 5). Moreover, Green Frog and bullfrog calls have broad frequency ranges and limited pulsations compared to other frogs. Therefore, the recognizers cannot home in on any specific pure tone or pulsation pattern, which are two major parameters incorporated into recognizer builds. This limited the overall accuracy of the recognizers compared to the ideal scenario: a noiseless site with more structurally complex calls. Recognizer 4-Low had so many FPs, that reviewing the recognizer output essentially approached conducting a manual count over the full recording. Thus, any promising 4-Low output in terms of TPs was not considered too beneficial.

Error trade-off and time investment.—The consequence of surveying noisy sites with ARDs is an increase in error rate, either through FPs or FNs depending on how the confidence parameters of the recognizer are adjusted. Thus, the recognizers built for this study were more error-prone compared to previous studies (Waddle et al. 2009; Eldridge 2011; MacLaren et al. 2017). It is important to adjust error trade-offs to match the goal of the study. For our study, we prioritized a reduction in FNs. Using Song Scope software, it is quick and easy to review identified calls and manually remove FPs from the recognizer data output, so a priority in automatically reducing FNs may generally be the best route when building recognizers because FN's can only be checked by listening to full recordings (Eldridge 2011).

Recognizers represent a potentially noteworthy jump in data collection efficiency. For a single recording the average time for the manual collection of data took 6 min 19 s, whereas the automated collection plus manual removal of FPs only took 3 min 28 s. Large jumps in efficiency have also been recorded for previous recognizer studies (Knight et al. 2017; MacLaren et al. 2017). As alluded to by Waddle et al. (2009) and Eldridge (2011), a combined approach where recognizers are used but then closely monitored manually may be best, especially for noisy sites.

Recognizer performance.—According to our classification indices for Green Frog recognizers, by using either the high- or medium-frequency recognizers we were able to improve TPR while simultaneously reducing the error rate. Because 1-Full was never the top-identifying recognizer, the broad frequency range of a conventional recognizer may be less effective in noisy environments. For bullfrog recognizers, the conventional recognizer performed better, which may be attributable to the fact that bullfrog calls are longer than Green Frog calls and therefore less limiting in the time dimension, meaning that the recognizer has more state-space to work with to make proper identifications.

For the Green Frog recognizers, we found that 3-Middle avoided the most noise according to noise level, but 2-High best avoided anthropogenic noise specifically. For the bullfrog recognizers, 6-High avoided the highest noise levels, and best avoided anthropogenic noise. Therefore, instead of relying solely on conventional recognizers at noisy sites, choosing a recognizer with appropriate frequency bands based off noise profile for each site can improve identification performance.

Noise profiles and calling effort estimation.—Our five study sites, although all chosen because of their high levels of anthropogenic noise, have different noise profiles. Interestingly, the results of our PERMANOVAs indicated only small differences in the type of noise interfering with each recognizer. This may be a result of certain noise categories having broader spectral effects on the soundscape. It is only when looking at anthropogenic noise alone that we could see that higher-frequency recognizers avoided interference better than lower-frequency recognizers.

We found that using conventional and narrow-banded recognizers at noisy sites can produce results that correlate to the true calling behavior of frogs, meaning that TP values from recognizers can be used as a reliable proxy for call rate. For both species, the correlation was strongest for a narrow-banded recognizer. Thus, scientists and land managers looking to monitor and study changes of acoustic urban species should consider narrow-banded recognizers to collect data.

Conclusion.—With urbanization posing a major obstacle for amphibian biodiversity, more emphasis will likely be placed on tracking frog population changes in noisy areas. Considering the noise profile of a site by using narrow-banded recognizers could prove useful for this purpose. As seen in this study and previous research, anuran recognizers can be error-prone (Waddle et al. 2009; Engbrecht 2011) and building recognizers has some subjective parameters (e.g., complexity and resolution parameters); however, they offer such a substantial boost in time efficiency that they can still be useful. Additionally, if the goal of the research is to compare calling effort across sites or over time, recognizer output can be used as a proxy because the overall pattern of call rate from automated surveys still resembles the true call pattern.

For these techniques to be most reliable, it is vital to choose a recognizer based on the noise profile specific to the site. This could mean building high-frequency recognizers when sites have low frequency interference, such as train noise, or building low-frequency recognizers for sites that have high frequency interference, such as insect chorusing. Our method of splitting up call frequency ranges using multiple recognizers may also be useful with avian species in noisy habitats, but a thorough investigation into this should be conducted. Only bird species with broad-frequency calls (e.g., Red-breasted Nuthatch, *Sitta canadensis*) are suitable for this technique.

Acknowledgments.—We thank members of the Wise Lab at the University of Illinois at Chicago (UIC) for their support and advice. We also thank the undergraduate volunteers who helped with the field work and lab work: Regina Witherell Poe, Victoria Surma, Said Jusic, Arielle Veverka, and Wisam Ali. We are grateful of the Forest Preserve Districts of Lake and Cook County and the Colby family who gave us access to the research sites. Funds for this research were provided by a Provost's Award (UIC), Elmer Hadley Research Grant (UIC), and grants from the Chicago Herpetological Society and the Society for the Study of Amphibians and Reptiles. All data collection was in accordance with the Animal Care Policies of the University of Illinois at Chicago (ACC# 15-117).

Literature Cited

Acevedo, M.A., and L.J. Villanueva-Rivera. 2006. Using automated digital recording systems as effective tools for the monitoring of birds and amphibians. Wildlife Society Bulletin 34:211–214.

Bee, M.A., and S.A. Perrill. 1996. Responses to conspecific advertisement calls in the Green Frog (*Rana clamitans*) and their role in male-male communication. Behaviour 133:283–301.

Bee, M.A., and E.M. Swanson. 2007. Auditory masking of anuran advertisement calls by road traffic noise. Animal Behaviour 74:1765–1776.

Brauer, C.L., T.M. Donovan, R.M. Mickey, J. Katz, and B.R. Mitchell. 2016. A comparison of acoustic monitoring methods for common anurans of the northeastern United States. Wildlife Society Bulletin 40:140–149.

Bridges, A.S., and M.E. Dorcas. 2000. Temporal variation in anuran calling behavior: implications for surveys and monitoring programs. Copeia 2000:587–592.

Cocroft, R.B., and M.J. Ryan. 1995. Patterns of advertisement call evolution in toads and chorus frogs. Animal Behaviour 49:283–303.

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

Crump, P.S., and J. Houlahan. 2017. Designing better frog call recognition models. Ecology and Evolution 7:3087–3099.

Cunnington, G.M., and L. Fahrig. 2010. Plasticity in the vocalizations of anurans in response to traffic noise. Acta Oecologica-International Journal of Ecology 36:463–470.

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

Engbrecht, N.J. 2011. The status of Crawfish Frogs (*Lithobates areolatus*) in Indiana, and a tool to assess populations. M.Sc. Thesis, Indiana State Unviersity, Terre Haure, Indiana. 78 p.

Gelling, D. 2010. Bird song recognition using gmms and hmms. M.Sc. Thesis, University of Sheffield, Sheffield, England. 39 p.

Genet, K.S., and L.G. Sargent. 2003. Evaluation of methods and data quality from a volunteer-based amphibian call survey. Wildlife Society Bulletin 31:703–714.

Gerhardt, H.C. 1994. The evolution of vocalization in frogs and toads. Annual Review of Ecology and Systematics 25:293–324.

Gibbs, J.P., and A.R. Breisch. 2001. Climate warming and calling phenology of frogs near Ithaca, New York, 1900–1999. Conservation Biology 15:1175–1178.

Kaiser, K., D.G. Scofield, M. Alloush, R.M. Jones,

S. Marczak, K. Martineau, M.A. Oliva, and P.M. Narins. 2011. When sounds collide: the effect of anthropogenic noise on a breeding assemblage of frogs in Belize, Central America. Behaviour 148:215–232.

Knight, E., K. Hannah, G. Foley, C. Scott, R. Brigham, and E. Bayne. 2017. Recommendations for acoustic recognizer performance assessment with application to five common automated signal recognition programs. Avian Conservation and Ecology 12:14. https://doi.org/10.5751/ACE-01114-120214

Lotz, A., and C.R. Allen. 2007. Observer bias in anuran call surveys. Journal of Wildlife Management 71:675–679.

MacLaren, A.R., S.F. McCracken, and M.R. Forstner. 2017. Development and validation of automated detection tools for vocalizations of rare and endangered anurans. Journal of Fish and Wildlife Management 9:144–154.

McLeod, A.I. 2011. Kendall: Kendall rank correlation and Mann-Kendall trend test. Version 2.2. https://CRAN.R-project.org/package=Kendall.

Oksanen, J., F.G. Blanchet, M. Friendly, R. Kindt, P. Legendre, D. McGlinn, P.R. Minchin, R.B. O'Hara, G.L. Simpson, P. Solymos, et al. 2019. Vegan: community ecology package. Version 2.5-4. https://CRAN.R-project.org/package=vegan.

Parris, K.M., M. Velik-Lord, and J.M.A. North. 2009. Frogs call at a higher pitch in traffic noise. Ecology and Society 14:1. http://www.ecologyandsociety.org/vol14/iss1/art25/.

Phillips, C.A., R.A. Brandon, and E.O. Moll. 1999. Field Guide to Amphibians and Reptiles of Illinois. Illinois Natural History Survey, Champaign, Illinois, USA.

Pijanowski, B.C., A. Farina, S.H. Gage, S.L. Dumyahn, and B.L. Krause. 2011a. What is soundscape ecology? An introduction and overview of an emerging new science. Landscape Ecology 26:1213–1232.

Pijanowski, B.C., L.J. Villanueva-Rivera, S.L. Dumyahn, A. Farina, B.L. Krause, B.M. Napoletano, S.H. Gage, and N. Pieretti. 2011b. Soundscape ecology: the science of sound in the landscape. Bioscience 61:203–216.

Simmons, A.M. 2004. Call recognition in the Bullfrog, *Rana catesbeiana*: generalization along the duration continuum. Journal of the Acoustical Society of America 115:1345–1355.

Sun, J.W.C., and P.A. Narins. 2005. Anthropogenic sounds differentially affect amphibian call rate. Biological Conservation 121:419–427.

Tennessen, J.B., S.E. Parks, L. Swierk, L.K. Reinert, W.M. Holden, L.A. Rollins-Smith, K.A. Walsh, and T. Langkilde. 2018. Frogs adapt to physiologically costly anthropogenic noise. Proceedings of the Royal Society B-Biological Sciences 285:20182194. http://dx.doi.org/10.1098/rspb.2018.2194.

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

Warren, P.S., M. Katti, M. Ermann, and A. Brazel. 2006. Urban bioacoustics: It's not just noise. Animal Behaviour 71:491–502.

Wells, K.D. 1977. Territoriality and male mating success in the Green Frog (*Rana clamitans*). Ecology 58:750–762.

Williams, P.J., N.J. Engbrecht, J.R. Robb, V.C.K. Terrell, and M.J. Lannoo. 2013. Surveying a threatened amphibian species through a narrow detection window. Copeia 2013:552–561.

Supplemental Information: http://www.herpconbio.org/Volume_14/Issue_3/Bielinski_etal_2020_Suppl.pdf

Nolan Bielinski is a Ph.D. candidate at the University of Illinois at Chicago (UIC), USA, where he is studying how urban soundscapes affect frog calling behavior. He received his Bachelor of Science at Rhodes College in Memphis, Tennessee, USA. He worked for the Forest Preserve District of Dupage County, Illinois, USA, for their deer management team and as a natural recourse technician. He is a member of the Chicago Herpetological Society. (Photographed by Brian Kay).

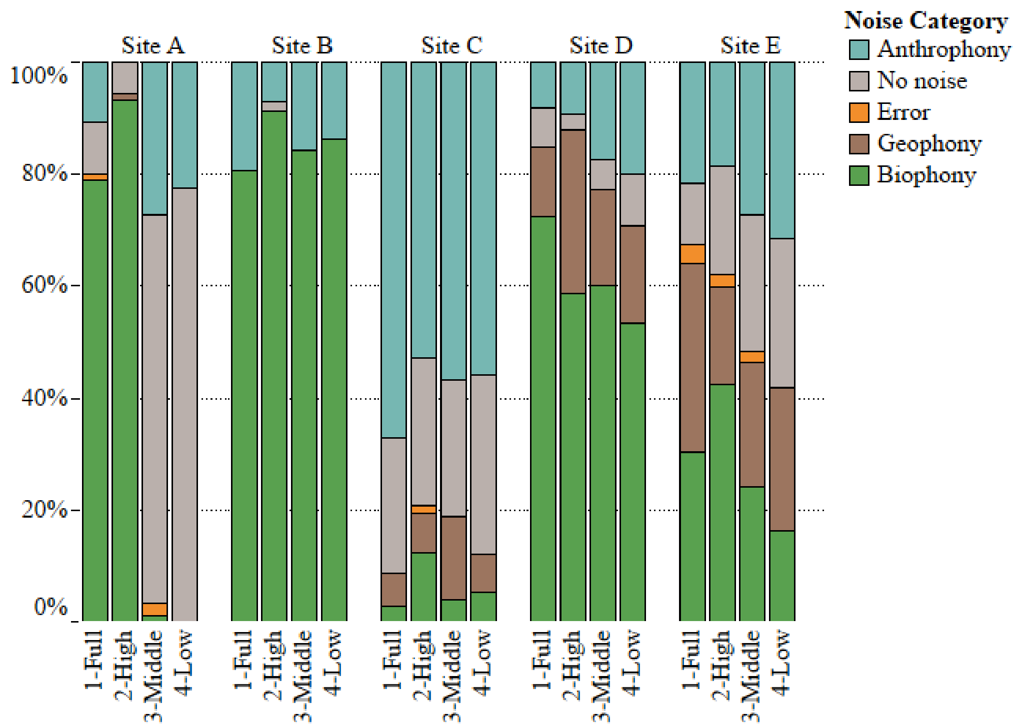
Jennifer Pajda-De La O is a Lecturer at University of Illinois at Chicago (UIC), USA, in the department of Mathematics, Statistics, and Computer Science. She received her Ph.D. in Mathematics with a concentration in Probability and Statistics from UIC. Her areas of research interest include sampling, probability, and ecological applications. (Photographed by Shuai Hao).

Adrianna Gorniak is currently a student at the University of Illinois College of Medicine, Rockford, USA. She was an undergraduate student at University of Illinois at Chicago (UIC), USA, where she received a Bachelor of Science in Biological Sciences. Her undergraduate studies focused on ecology and animal behavior. She was a Research Assistant in the Department of Ecology at UIC. (Photographed by Clarissa Genovese).

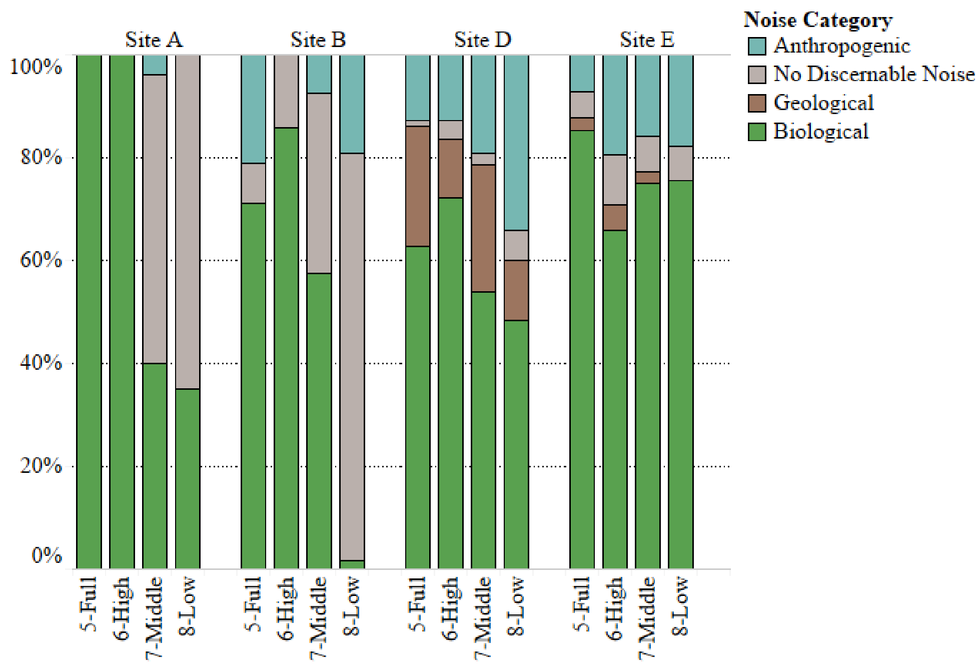
David Wise is a Professor of Ecology and Evolution in the Department of Biological Sciences at the University of Illinois at Chicago (UIC), USA. Previously he held faculty positions at the University of New Mexico, Albuquerque, USA, the University of Maryland Baltimore County, Baltimore, USA, and the University of Kentucky, Lexington, USA. Professor Wise has been an Alexander von Humboldt Research Fellow in Germany and has taught at the University of Buenos Aires, Argentina, as a Fulbright Scholar. He received his M.S. and Ph.D. in Zoology from the University of Michigan, Ann Arbor, USA. (Photographed by Brian Kay).

Appendix 1. The full set of parameters used to create our recognizers in Song Scope. Recognizers 1–4 are for Green Frog (*Lithobates clamitans*) mating calls. Recognizers 5–8 are for American Bullfrog (*Lithobates catesbeianus*) mating calls. For information on the parameters see the Song Scope manual (Wildlife Acoustics. 2011. *op. cit.*).

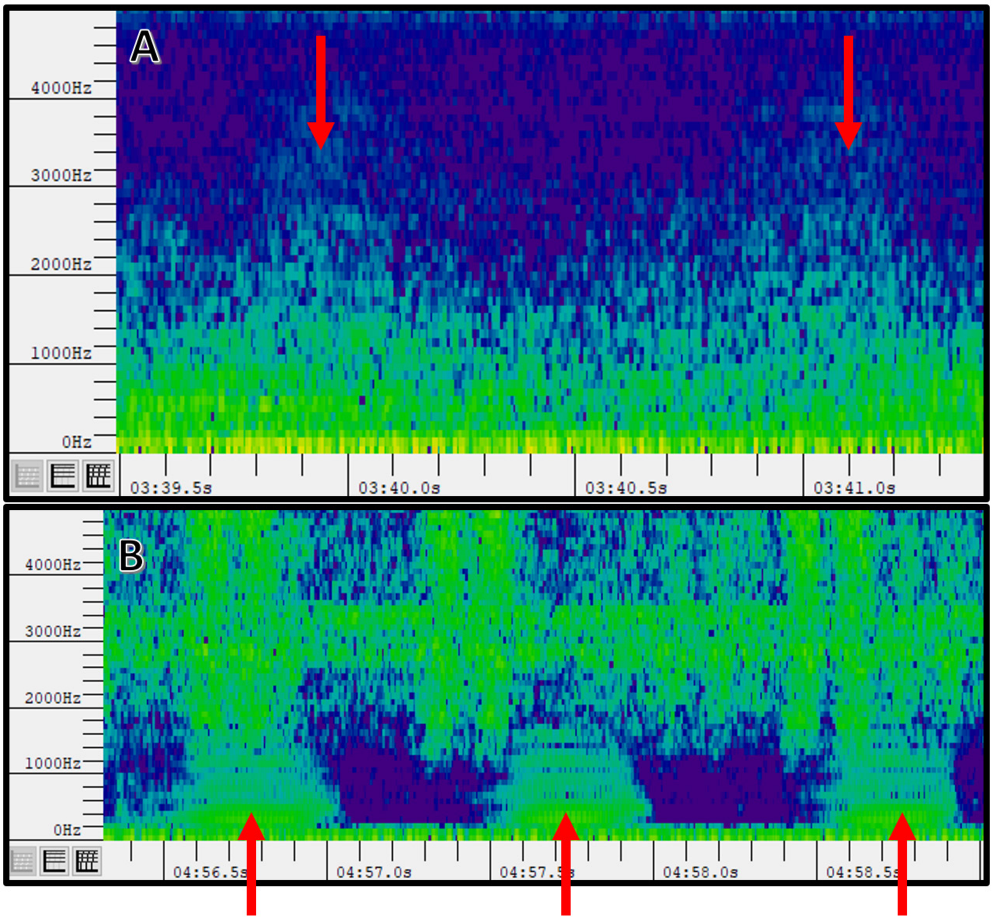
Recognizer Name	1-Full	2-High	3-Middle	4-Low	5-Full	6-High	7-Middle	8-Low
Species	Green Frog	Green Frog	Green Frog	Green Frog	Bullfrog	Bullfrog	Bullfrog	Bullfrog
Type	Conventional	Narrow	Narrow	Narrow	Conventional	Narrow	Narrow	Narrow
Max. Complexity	20	20	20	20	25	25	25	25
Max. Resolution	10	10	10	10	12	12	12	12
Freq. min (Hz)	187.5	2000	812.5	125	187.5	2,062.5	562.5	187.5
Freq. max (Hz)	3,875	3875	1562.5	875	5,250	3,000	1,625	500
Max. syllable (ms)	1,128	1,128	1,128	1,128	504	472	504	496
Max. syl. Gap (ms)	0	0	0	0	232	216	232	256
Max. Song	1,400	1,400	1,400	1,400	2,576	2,064	2,456	2,808
Fast Fourier Transformation Size	512	512	512	512	512	512	512	512
Fast Fourier Transformation Overlap	1/2	1/2	½	1/2	1/2	1/2	1/2	1/2
Sample Rate (Hz)	4,800	4,800	4,800	4,800	4,800	4,800	4,800	4,800
Background filter	1s	1s	1s	1s	1s	1s	1s	1s
Algorithm	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
Quality	40	20	30	30	20	25	20	25
Score	40	65	40	60	50	50	50	65



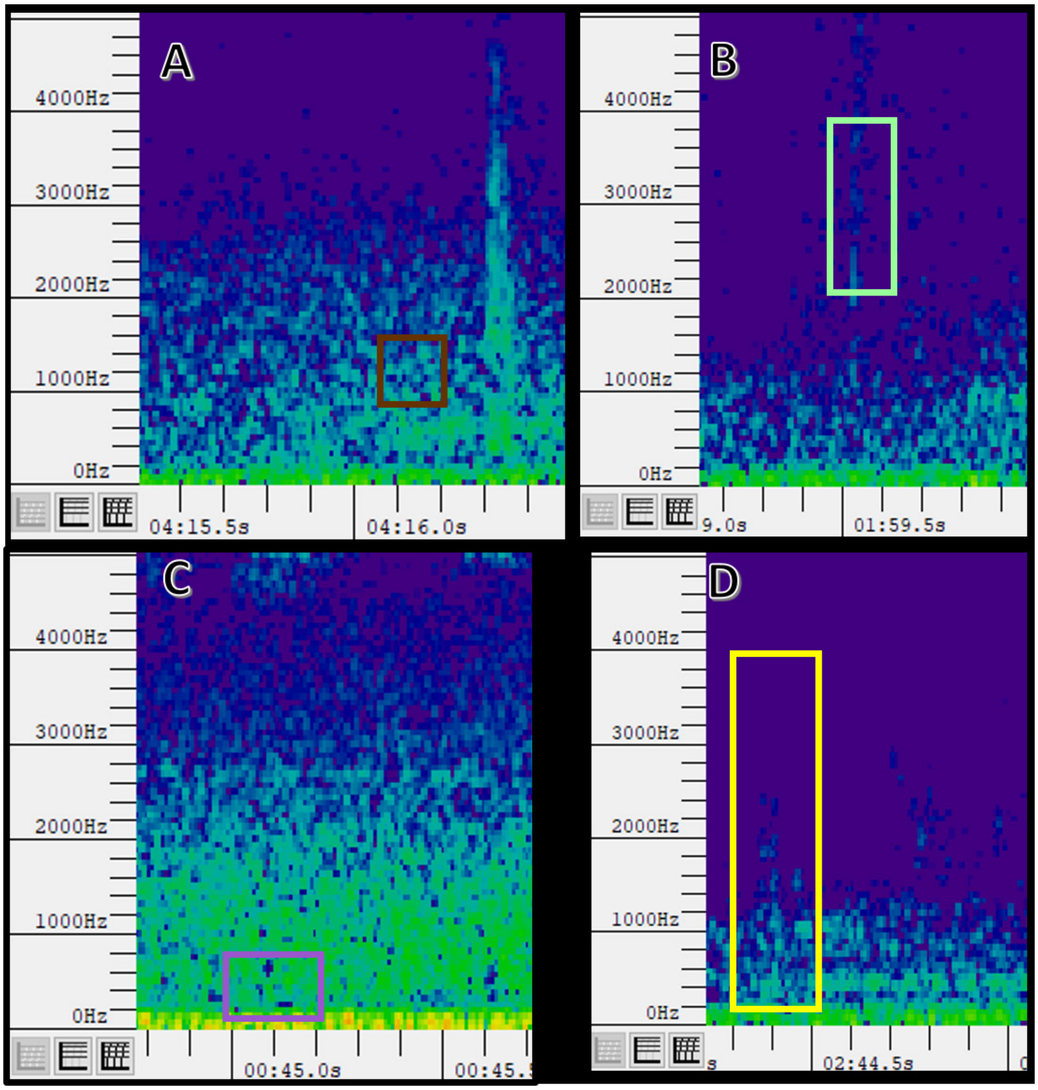
Appendix 2. Noise category percentages from Green Frog (*Lithobates clamitans*) recognizer samples by site. The difference in the noise profiles between sites stresses the importance of considering the noise profile of a research location *a priori* when using automated detection.



Appendix 3. Noise category percentages from American Bullfrog (*Lithobates catesbeianus*) recognizer samples by site. The differences in the noise profiles highlights the value in our approach to independently analyze the performance of recognizers at each site and stresses the importance of considering the noise profile of a research location *a priori* when using automated detection. Site C was removed as there were no calls detected.



Appendix 4. Spectrograms of American Bullfrog (*Lithobates catesbeianus*) calls in noisy environments. Red arrows indicate locations of un-interfered call energy. The benefit of using narrow-banded recognizers is clear in this figure, as only portions of calls are available to be identified. (A) Two calls almost entirely overlapped by airplane noise at site D. A full-frequency recognizer would experience a lot of noise interference, whereas a high-frequency narrow-banded recognizer may be able to identify remnants of the signal. (B) Three calls with higher-frequency overlap from intense insect chorusing at site A. Here, a lower or full-frequency recognizer would be more effective than a high-frequency recognizer to recognize calls.



Appendix 5. Spectrograms of false positives from Green Frog (*Lithobates clamitans*) recognizers. Colored boxes represent the state-space identified incorrectly as frog calls. (A) A false positive from recognizer 3-Middle at site C caused by car traffic. (B) A false positive from recognizer 2-High at site C caused by a snapping branch. (C) A false positive from recognizer 4-Low at site C caused by train traffic. (D) A false positive from recognizer 1-Full at site C caused by a truck.