

Aquatic Soundscape Recordings Reveal Diverse Vocalizations and Nocturnal Activity of an Endangered Frog

Sam Lapp,^{1,*} Thomas C. Smith,² Roland A. Knapp,² Alexa Lindauer,² and Justin Kitzes¹

1. Department of Biological Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania 15260; 2. Sierra Nevada Aquatic Research Laboratory, University of California, Mammoth Lakes, California 93546

Submitted April 17, 2023; Accepted November 14, 2023; Electronically published March 13, 2024

Online enhancements: supplemental PDF, sound files.

ABSTRACT: Autonomous sensors provide opportunities to observe organisms across spatial and temporal scales that humans cannot directly observe. By processing large data streams from autonomous sensors with deep learning methods, researchers can make novel and important natural history discoveries. In this study, we combine automated acoustic monitoring with deep learning models to observe breeding-associated activity in the endangered Sierra Nevada yellow-legged frog (*Rana sierrae*), a behavior that current surveys do not measure. By deploying inexpensive hydrophones and developing a deep learning model to recognize breeding-associated vocalizations, we discover three undocumented *R. sierrae* vocalization types and find an unexpected temporal pattern of nocturnal breeding-associated vocal activity. This study exemplifies how the combination of autonomous sensor data and deep learning can shed new light on species' natural history, especially during times or in locations where human observation is limited or impossible.

Keywords: acoustic monitoring, *Rana sierrae*, anuran, vocalization, machine learning, automated detection, Sierra Nevada Aquatic Research Lab.

Introduction

Our understanding of a species' natural history is inherently based only on the information that is available to us as human observers. For many species, obtaining new insights into natural history will require that researchers find new methods of observing previously hidden aspects of the species' biology and ecology. In recent years, autonomous sensor technologies such as satellite imagery, autonomous cameras, and autonomous acoustic recorders have demonstrated the potential to measure phenomena that cannot be directly observed by humans. For example,

global-scale satellite imagery, decadal-scale climate data, and hyperspectral imaging can sample processes that are difficult or impossible to observe directly. Terrestrial autonomous sensors, including automated camera traps and acoustic recorders, produce data analogous to visual or aural human observation methods with orders-of-magnitude increases in temporal coverage. In augmenting traditional direct human observations, autonomous sensors hold the potential to shed light onto aspects of species' natural histories that would otherwise remain hidden.

However, the value of autonomous sensors is critically limited by our ability to extract useful insights from the raw data that they produce. Compared with the types of data collected through traditional survey methods (e.g., biometric measurements, species checklists, counts of individuals), the unstructured digital data captured by autonomous sensors are generally not directly interpretable as biological information. Given that the scale of autonomous sensor data is often too large for humans to process through manual inspection, automated data processing techniques are essential for capitalizing on the full potential of autonomous sensor data.

Recently developed machine learning techniques can extract biologically meaningful information from autonomous sensor data, providing a promising avenue to address this bottleneck (Weinstein 2018). Deep learning algorithms such as convolutional neural networks (CNNs) are a subset of machine learning algorithms that are particularly well suited to processing autonomous sensor data. These algorithms learn to recognize patterns and features in unstructured data by iteratively optimizing their performance on annotated training data. Deep learning algorithms have been successfully applied in ecology to detect wildlife in camera trap images (Vélez et al. 2023), to

* Corresponding author; email: sam.lapp@pitt.edu.

map habitat types from satellite imagery (Talukdar et al. 2020), and to recognize species-specific vocalizations of birds, frogs, and bats in audio recordings (Stowell 2022).

In this study, we combine autonomous acoustic sensing and deep learning to gain new insights into the natural history of an endangered amphibian, the Sierra Nevada yellow-legged frog (*Rana sierrae*). Threats from introduced fishes and the lethal amphibian chytrid fungus *Batrachochytrium dendrobatidis* have extirpated this species from more than 90% of its historical habitats in California's Sierra Nevada mountains (Vredenburg et al. 2007). Extensive visual surveys (Bradford 1989, p. 19; Bradford et al. 1994; Drost and Fellers 1996; Knapp and Matthews 2000; Knapp 2005; Knapp et al. 2007, 2011, 2016; Vredenburg et al. 2010) and capture-recapture surveys (Joseph and Knapp 2018) have documented *R. sierrae* population declines and recoveries over four decades and informed conservation efforts.

Understanding of this species' biology and population dynamics could be enhanced by quantifying breeding activity, which provides insights into processes that drive populations to grow or decline (Cruikshank et al. 2021). For example, populations with little to no breeding activity are likely to have lower persistence probabilities than populations with more breeding activity (Cruikshank et al. 2021). However, quantifying *R. sierrae* breeding activity is difficult for at least two reasons. First, *R. sierrae* call and breed underwater while lakes are still partially covered in thick winter ice. Second, although counts of egg masses provide a direct measure of breeding activity, female *R. sierrae* often deposit egg masses in inaccessible habitats, such as under overhanging banks and in rock crevices. These conditions make egg masses virtually undetectable during visual surveys. As a result, current survey methods do not generally quantify breeding activity, leaving an important gap not just in our current knowledge of this species' natural history but in our ability to rapidly assess or predict population trends.

Acoustic surveys of breeding-associated vocalizations provide an indirect measurement of breeding activity for many anuran species. In most anurans, males produce advertisement calls to convey their species, sex, reproductive state, and position to conspecific females during breeding activity (Wilczynski and Chu 2001; Wells and Schwartz 2006). In natural environments, these acoustic signals are often easier to detect than the organisms themselves. Although such signals can often be detected by traditional human surveys, automated acoustic monitors can allow for the observation of sound-producing organisms across large temporal scales: a single deployment of acoustic recording units can capture audio recordings across weeks or months. Furthermore, because automated recording units can survey terrestrial and aquatic soundscapes across any hour of the day for weeks at a time, acoustic sensors can

record amphibian vocal activity even when an environment is inaccessible to humans. These advantages are particularly important in the case of *R. sierrae*, which vocalizes while underwater in high mountain lakes (Ziesmer 1997) and, as a result, is rarely detected acoustically by human observers.

In this study, we conducted surveys of breeding-related behaviors of *R. sierrae* by recording vocal activity through a novel combination of two emergent techniques for ecological monitoring: low-cost autonomous hydrophones and deep learning classification models. By developing and applying a deep learning method to detect specific breeding-associated *R. sierrae* vocalizations in hydrophone recordings spanning an entire breeding season, we characterize the vocal activity of a wild *R. sierrae* population across daily and seasonal timescales. This combination of autonomous sensing and deep learning enables us to discover previously unknown *R. sierrae* life history traits that may help to inform conservation planning and further research. We see this study as an illustrative example of an emerging pattern in which combinations of autonomous sensing and deep learning will open new windows of observation into the natural world.

Methods

Study Site

We collected acoustic data for this study at an alpine lake in Yosemite National Park (site ID: 70550; Knapp et al. 2020) in the Sierra Nevada of California. Acoustic data collection at this site was conducted under a research permit issued by Yosemite National Park (permit YOSE-2022-SCI-0075). Because of the sensitive nature of this federally protected species' conservation, we are unable to publish the name and coordinates of the lake. Lake 70550 is a typical Sierran alpine lake: elevation of 3,200 m, maximum depth of 11 m, and surface area of 4.7 ha. The lake generally remains ice-covered from November to June. Lake 70550 contains a population of *Rana sierrae* that occupies the site year-round and does not contain introduced fish. Although *R. sierrae* was absent from the lake in 2001 (R. Knapp, unpublished data), it was successfully reintroduced in 2006. In 2017 the population was estimated to include approximately 400 adults (Joseph and Knapp 2018) and thousands of tadpoles and subadult frogs (R. Knapp and T. C. Smith, unpublished data).

Acoustic Monitoring

We used five AudioMoth acoustic recorders (Hill et al. 2019; ver. 1.2.0 with firmware ver. 1.5.0) in waterproof cases

(Lamont et al. 2022) to record underwater soundscapes at lake 70550. The number of devices was constrained by hardware availability, as the availability of the waterproof case was limited at the time of this study. We programmed the devices to record for 1 min starting every 15 min for 24 h per day. We used a 32-kHz sampling rate and medium gain with 16-bit sample depth. We deployed devices underwater by attaching the devices to rocks with cable ties and placing them on the bottom of the lake (fig. 1) within 2 m of the shoreline and at depths of less than 1 m (device locations in table S1, available online). The minimum spacing between any two devices was 110 m. We deployed the devices on June 2, 2022, when most of the lake was covered in ice, and recovered the devices on July 22, 2022.

Investigating the Diversity of R. sierrae Vocalizations

Aquatic soundscapes typically contain biotic sounds from amphibians, invertebrates, and fish (Desjonquères et al. 2020). Besides *R. sierrae*, lake 70550's amphibian commu-

nity includes only *Anaxyrus canorus* (Yosemite toads), which generally call in or near snowmelt pools adjacent to the lake but are rarely found within the lake. Because the vocalizations of *A. canorus* are audible in air, their vocalizations are known and well documented (California Herps 2023a). The lake does not contain fish. The only other taxa in the lake known to produce acoustic signals are members of the family Corixidae, which stridulate underwater. Corixid stridulations are pulsed, wideband sounds typically occupying frequencies from 2 to 12 kHz (Theiss and Prager 1984), making them qualitatively distinct from vocalizations of frogs in the genus *Rana*.

To investigate the diversity of *R. sierrae* vocalizations recorded in our dataset, we reviewed a random subset of the audio and sorted call types into distinctive groups. We first randomly selected 500 recordings from the complete set of 23,931 60-s recordings generated by the audio recorders, then randomly selected 10 s from each 60-s recording. We listened to and viewed a spectrogram of each clip and took descriptive notes on all sounds that



Figure 1: AudioMoth acoustic recorder deployed underwater at the study site.

we suspected were biotic. We then grouped *R. sierrae* vocalizations into qualitatively similar categories and named each category as a call type (A–E; see “Results”).

Developing and Applying an Automated Recognizer

Because autonomous sensors often generate datasets too large to process through direct inspection, automated data processing approaches are critical to relieving the bottleneck between raw autonomous sensor data and biological inference (Stowell 2022). In bioacoustics, the rising popularity of acoustic monitoring (made possible by affordable sensors) has been accompanied by the development of automated algorithms to detect sounds of interest in acoustic data (Knight et al. 2017; Gibb et al. 2019). In particular, deep supervised learning algorithms such as CNNs have been applied with great success to bioacoustic recognition tasks, especially for birds (Kahl et al. 2017; Kahl 2019). Supervised deep learning approaches, including CNNs, are well suited to recognition problems where the target sound is highly variable, provided that hundreds of labeled samples can be obtained. By automating the detection of *R. sierrae* vocalizations in soundscape recordings with a CNN, we were able to measure temporal patterns of vocal activity across soundscape recordings from an entire season of *R. sierrae* breeding activity.

To automate the detection of *R. sierrae* vocalizations in the soundscape recordings, we created an annotated set of audio clips and then trained a CNN (for detailed methodology, see the supplemental PDF). We first annotated a subset of the field data from one recording device during one week of the deployment. Using Raven Pro software (Bioacoustics Research Program 2019), we placed a frequency-time box around each *R. sierrae* vocalization. We randomly split the 672 10-s annotated clips produced by this effort into a training set (90%) and a validation set (10%). Next, we trained a CNN to recognize two *R. sierrae* call types, A and E (for call type descriptions, see “Results”). Call type B was excluded because it was rare in the annotated data, while call types C and D were excluded because they did not closely resemble vocalizations described as advertisement calls (Vredenburg et al. 2007). We used the open-source Python package OpenSoundscape (ver. 0.7.1; Lapp et al. 2023) to train a CNN with a ResNet18 architecture on spectrograms of 2-s audio samples. We used standard augmentation routines for audio preprocessing and standard CNN training hyperparameters (details in the supplemental PDF). For each input sample representing 2 s of audio, the CNN generates a score between 0 and 1 representing the algorithm’s confidence that *R. sierrae* vocalizations (types A and E only) are present in the sample.

We evaluated the performance of the CNN on the validation set following the recommendations of Knight

et al. (2017). We calculated the area under the precision-recall curve (average precision score) and receiver operating characteristic (ROC) curves at the 2-s sample level. We selected a threshold score by choosing the lowest score that achieved a precision exceeding 99% on the validation set, such that 99% of all audio clips scoring above this threshold contained the species of interest while 1% represented false positives. We used this threshold to evaluate precision, recall, and F1 score on 2-s samples. For the remainder of the analyses, 2-s audio samples are considered detections if the CNN scored them above the threshold score. To test the generalizability of the CNN’s precision from the validation set to the entire dataset, we manually reviewed a random subset of 100 detections (or all detections if there were fewer than 100) from each device.

We measured temporal patterns of *R. sierrae* vocal activity in the aquatic soundscape by calculating the detected vocal activity rate (Pérez-Granados and Traba 2021), the proportion of samples over a time interval in which the CNN detected *R. sierrae* vocalizations. To describe daily patterns, we calculated the detected vocal activity rate per 1-min time period and plotted the activity averaged across all days of recording. To describe seasonal patterns, we calculated the detected vocal activity rate per 24-h time period.

Results

Diversity of Call Types

By annotating a randomly selected set of 500 10-s audio clips, we found five distinct *Rana sierrae* vocalization types (fig. 2A–2E; for descriptions of each call type, see the supplemental PDF; audio examples are available as supplemental material online). We refer to the five call types here as A through E, corresponding to their labels in figure 2. Call types A and B likely correspond to the unpulsed and pulsed calls, respectively, from Vredenburg et al. (2007) and Ziesmer (1997). To our knowledge, call types C, D, and E have not been previously documented in *R. sierrae* (but may have analogs in *Rana boylei*; see “Discussion”). Intergrading between the previously observed call types (A and B) and the newly discovered call types (C, D, and E) provides evidence that these call types are produced by *R. sierrae*. The type A vocalization is pitched and extremely plastic in its pitch motion, which can be upward, downward, overslurred, or take more complex trajectories. The type B vocalization is unpitched and has rapid amplitude modulation (pulsing). The type C vocalization is a short, unpitched note that is sometimes doubled or tripled (see further examples in the supplemental PDF). The type D vocalization is a short, downward semipitched note that sometimes grades into the A type. The type E vocalization is a pitched and frequency-modulated call. The

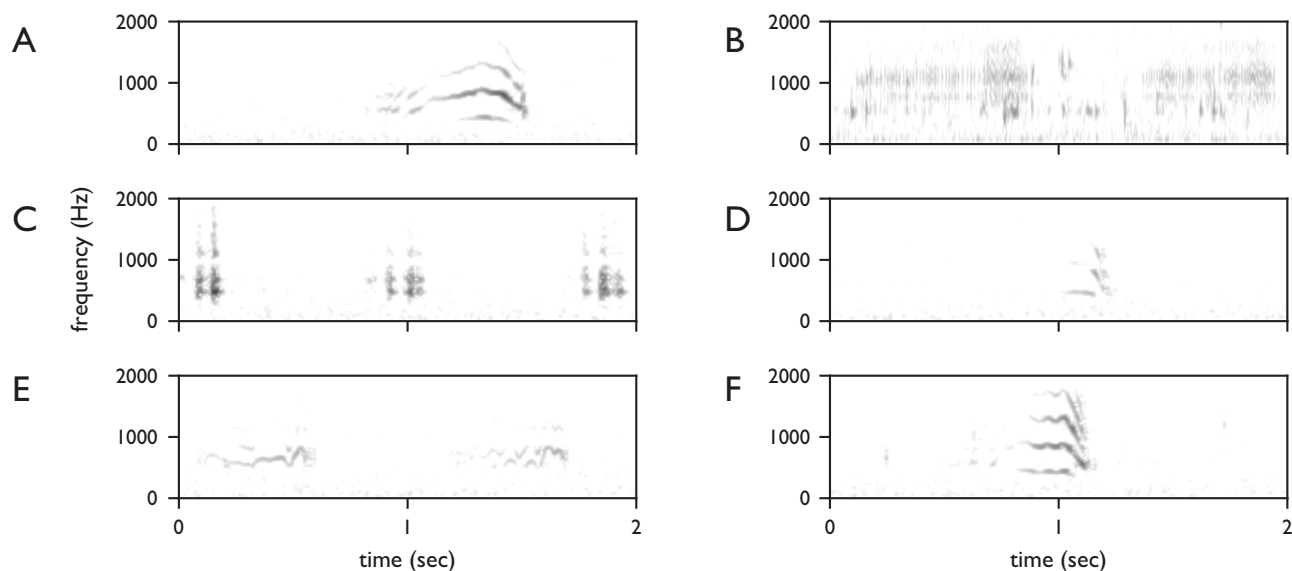


Figure 2: Spectrograms illustrating five distinct vocalization types of *Rana sierrae* (A–E; calls named respectively) and an intermediate call blending the type A and type E calls (F). The convolutional neural network was trained to detect type A and type E calls only. Log spectrograms were generated from 2-s 32-kHz audio files using a window size of 1,024 samples (except B; 512 samples to show temporal detail) and 75% window overlap and were vertically cropped to 0–2 kHz.

type E call can grade into the type A call, as shown in figure 2F. Ziesmer (1997) additionally described a “semi-pulsed” call type of *R. sierrae* that may correspond to the type E vocalization, but we do not have recordings or spectrograms to compare Ziesmer’s semipulsed call type to the vocalizations in our data. Overall, considering that most amphibians produce stereotyped calls (Gerhardt and Bee 2006), the *R. sierrae* call repertoire we documented through autonomous acoustic recorders exhibits a remarkable degree of plasticity.

We did not detect vocalizations of the other amphibian in the community, *Anaxyrus canorus*, in any of the aquatic soundscape recordings. However, the recorders captured high-frequency, wideband amplitude-modulated sounds (fig. S1; figs. S1–S4 are available online), which we suspect are stridulations produced by the aquatic insect family Corixidae (Theiss and Prager 1984).

CNN Performance on the Validation Set

The CNN recognizer effectively and accurately detected *R. sierrae* vocalizations in the validation set. The area under the precision-recall curve (average precision score) on the validation set was 0.92, while the area under the ROC curve was 0.95. The minimum score threshold that achieved 99% precision was 0.99934; at this threshold, precision was 99% and recall was 30% (precision-recall curve in fig. S2), indicating that only 1% of the audio clips above this threshold were false positives and that 30% of all files containing the

target vocalizations were above the threshold. This threshold was used for the analyses of seasonal (fig. 3) and diel (fig. 4) activity patterns. Our model’s high precision generalized across the entire dataset. A review of 100 randomly selected detections from each device (or all detections if fewer than 100) resulted in zero false-positive detections at four out of five devices (devices 1–3: 100 detections each; device 4: 38 detections). Device 5 generated just two detections from the entire season, and these were false positives containing muffled human voices.

The observed seasonal and daily patterns were robust across a range of threshold scores (figs. S3, S4), indicating that a different threshold choice would not have affected the observed patterns in vocal activity. For instance, a low threshold with a precision of 88% and recall of 78% on the validation set produced temporal patterns equivalent to a high threshold with a precision of 100% and recall of 14%. The robustness of temporal patterns to threshold score demonstrates that false positives at lower thresholds (or false negatives at higher thresholds) did not bias the measurement of temporal patterns.

Seasonal and Diel Patterns of Vocal Activity

Seasonal patterns of vocal activity were constrained to 3 weeks of June and differed by recorder (fig. 3). Device 1 showed a short period of intermediate activity levels in the second week of June. Devices 2 and 3 showed high activity levels in the third and fourth weeks of June.

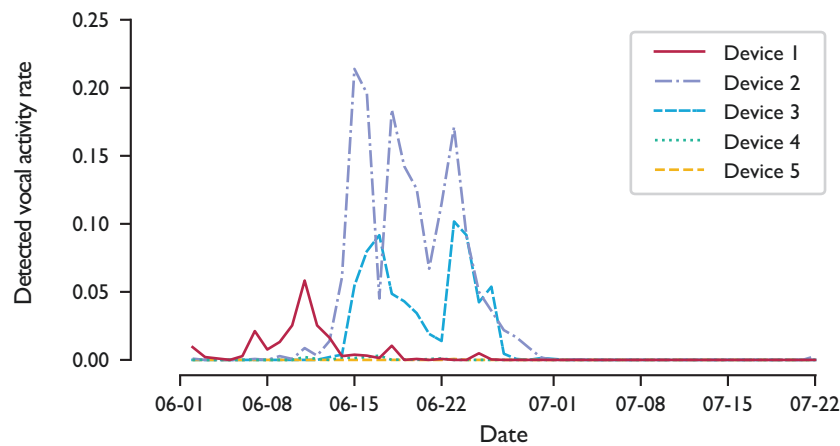


Figure 3: The seasonal pattern of *Rana sierrae* vocal activity varied across the recording devices. Detected vocal activity rate was measured as the proportion of detections from all samples during a 24-h period. Vocalizations were detected by the automated recognizer using a threshold score that obtained 99% precision on a validation set.

Devices 4 and 5 showed little to no vocal activity across the season. Across all recorders, vocal activity was minimal before June 5 and after June 30. Contrary to the existing literature and current understanding of *R. sierrae* biology, the diel vocal activity pattern showed increased vocal activity at night and relatively low levels of vocal activity during the day (fig. 4).

Discussion

By combining automated acoustic recording of aquatic soundscapes with deep learning, we provided the first broad temporal-scale monitoring of *Rana sierrae* vocal

activity. *Rana sierrae* has been intensively studied for two decades, primarily using visual encounter and capture-mark-recapture surveys to measure abundance, survival, and recruitment. These survey techniques are critical for monitoring populations and informing species recovery actions. However, these survey methods do not typically quantify *R. sierrae* breeding activity, data that could provide valuable insights into species' biology and mechanisms of population change.

Contrary to current understanding, we found the breeding-associated vocal activity of *R. sierrae* to be primarily nocturnal rather than diurnal. This pattern of nocturnal male advertisement underscores the difficulty of

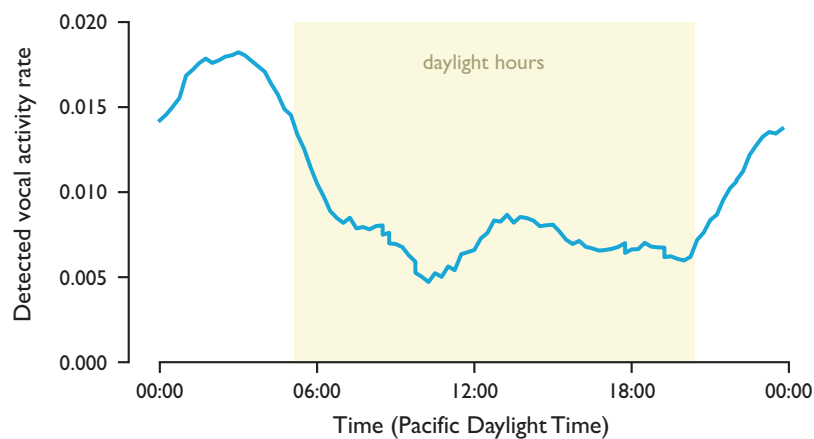


Figure 4: Daily pattern of *Rana sierrae* vocal activity. *Rana sierrae* were most vocally active at night, with an approximately three times higher detected vocal activity rate (the fraction of detections per 1-min time period) than during daylight hours. Vocalizations were detected by the automated recognizer using a threshold score that obtained 99% precision on a validation set. The blue line represents the average activity across all recording dates. The shaded yellow region denotes the approximate daylight hours at the study site.

directly observing breeding activity in this species and suggests that human observation of breeding activity might require nocturnal surveys, which are currently not included in survey protocols. Because our study was limited to one breeding population, further investigation is needed to determine whether each pattern we observed is generalizable across populations. Nonetheless, the results of this study suggest that expanded acoustic monitoring efforts could provide valuable insights into the understanding and management of *R. sierrae*.

Our results demonstrate that aquatic monitoring with AudioMoth recorders and deep learning captured *R. sierrae* vocalizations and provided an effective method of remotely monitoring early-season underwater *R. sierrae* activity that is typically not observed. We found that *R. sierrae* vocal activity is primarily an early-summer phenomenon, which concurs with the predominant view that the species breeds around the time of lake ice-off. Two out of five devices contained no vocal activity, even though visual encounter surveys regularly observe frogs at all five sampling locations used in this study. This unexpected spatial heterogeneity in vocal activity could point to habitat preferences specific to breeding activity, a hypothesis that could be investigated in future work.

Amphibian vocalizations are widely used as a proxy for breeding activity, but this association relies on links between specific vocalization types and biological functions. Thus, understanding the biological functions of *R. sierrae* call types is an important next step toward expanding the use of acoustic recorders to monitor breeding activity. Because our study did not provide a means of observing behaviors associated with vocalizations, our categorization of call types reflects only a qualitative interpretation of the call diversity we observed. Based on direct observations of vocal activity, Ziesmer (1997) posited that the unpulsed (A type) *R. sierrae* call is used for advertisement. Additionally, for each novel *R. sierrae* call type we found (C, D, and E), a similar call has been described for *Rana boylei* (MacTague and Northen 1993; Silver 2017), a closely related species that hybridizes with *R. sierrae* in some regions (Peek et al. 2019). Specifically, the type C call resembles the *R. boylei* “chuckle” (Silver 2017), the type D call resembles the *R. boylei* “unpulsed call” (MacTague and Northen 1993), and the type E call resembles the *R. boylei* “warble” (Silver 2017). These analogous call types may provide clues to the functions of *R. sierrae* call types. For instance, Silver (2017) reported that the warble call type of *R. boylei* was produced only when male frogs were in amplexus with females. Therefore, the similar-sounding call type E of *R. sierrae* represents a promising starting point for uncovering behavioral connections between *R. sierrae* reproduction and their vocalizations.

Estimating the abundance or density of individual organisms remains a largely unsolved challenge in bioacoustic

monitoring (Pérez-Granados and Traba 2021) and is an active area of research (Yip et al. 2017, 2020; Strebel et al. 2021). In this study, estimating the abundance or density of individual frogs vocalizing in a surveyed habitat would require additional knowledge, such as the vocalization rate of individual frogs, the propagation distance of the vocalizations in the aquatic environment, or the ability to identify individuals by the unique characteristics of their vocalizations. Furthermore, knowledge of the calls’ propagation distance would be necessary to rule out the possibility that the same individual’s calls were recorded on multiple devices in our study. Although DeGregorio et al. (2021) tested the detection distance of underwater playbacks of two frog species’ vocalizations from an underwater speaker and detected the calls on hydrophones up to 65 m away, this detection distance is dependent on the playback volume and may not reflect realistic detection distances of vocalizations produced by the organisms themselves. To our knowledge, there is no further literature measuring the maximum detection distance of underwater amphibian vocalizations. In the absence of information about the maximum detection distance for *R. sierrae* vocalizations in our system, we interpret the detected vocal activity rate as an indicator only of overall activity at the lake rather than a measurement of abundance or a count of individuals. The ability of bioacoustic data to provide quantitative population abundance estimates represents a fruitful area for future research, as accurate abundance measurements could facilitate remotely surveying populations with decreased human effort, detecting population decline or growth, and evaluating the need for or success of recovery actions.

We publicly share the results of this project to facilitate further study of *R. sierrae* vocalizations and other aspects of aquatic soundscapes. A public repository of 672 annotated soundscape recordings containing 1,284 annotations of *R. sierrae* vocalizations (Lapp and Kitzes 2023) dramatically expands on the nine previously available online recordings (California Herps 2023b) and provides extensive examples of all call types, vocalization plasticity, and gradation between call types. A GitHub repository (Lapp 2023) provides Python scripts for all steps of model development and analysis. The trained deep learning classifier is also publicly available in the GitHub repository. Through these resources, other scientists may reuse or adapt these data, deep learning models, and bioacoustics methods in applications to the study of *R. sierrae* and to new systems.

Deep learning methods are powerful tools for extracting biological information from autonomous sensor data, but they also carry unique challenges. First, deep learning methods are data hungry in that their accuracy is highest when they are trained with large quantities of annotated data, and models often perform poorly when little training data are available. In our study, generating hundreds

of annotated samples of *R. sierrae* vocalizations was tractable because we focused annotation efforts on a location and time period with high vocal activity. Second, the generalizability or transferability of deep learning models across study systems or domains can be a significant challenge, and these techniques perform best when the data they are trained with closely matches their real-world application (Schneider et al. 2020; Stowell 2022). In this work, our model did not need to generalize to a new domain because our training data was a subset of our field data. Third, no deep learning models applied to wildlife vocalizations to date can perfectly separate sounds of interest from other sounds. In this study, because our classifier had high precision and we did not require high recall, minimal human review was required. In other applications, expert review is often necessary to verify deep learning model outputs.

In summary, remote sensing tools like acoustic recorders offer a cost-effective complement to visual-based survey methods for biological observation, but extracting useful information from vast quantities of remote sensor data remains an outstanding challenge. Fortunately, the qualities of remote sensor data are well suited for emerging deep learning techniques, which have been specifically designed to analyze large, complex, and unstructured data. In this study, we demonstrated that combining acoustic monitoring with deep learning enabled the observation of *R. sierrae* breeding-associated vocal activity. Looking forward, we believe that further intersections of autonomous sensors and deep learning techniques will continue to provide novel insights into the natural world.

Acknowledgments

This work was financially supported by the Department of Biological Sciences at the University of Pittsburgh and the Gordon and Betty Moore Foundation. This material is based on work supported by the National Science Foundation under grants 2120084 and 1935507. This research was authorized by Yosemite National Park (YOSE-2022-SCI-0075). This work was performed in part at the University of California Natural Reserve System Valentine Eastern Sierra Reserve–Sierra Nevada Aquatic Research Laboratory Reserve (<https://doi.org/10.21973/N3966F>) and Valentine Eastern Sierra Reserve–Valentine Camp Reserve (<https://doi.org/10.21973/N3JQ0H>). We thank two reviewers, the editors, Chapin Czarnecki, and Hannah Nossan for their comments on the manuscript.

Statement of Authorship

S.L., T.C.S., R.K., A.L., and J.K. conceptualized the study and participated in experimental design, execution of

the study, and writing, reviewing, and editing the manuscript. S.L., T.C.S., R.K., and A.L. participated in data collection and curation. S.L. developed the machine learning model, analyzed the data, and wrote the original draft. J.K. acquired funding and provided supervision.

Data and Code Availability

The annotated dataset of *Rana sierrae* vocalizations created in this work is publicly available online under an open-source license (<https://doi.org/10.5061/dryad.9s4mw6mn3>; Lapp and Kitze 2023). The full dataset of acoustic recordings collected during this study is available from the corresponding author on reasonable request. Python notebooks with code to reproduce the training of the automated detector, evaluation of model performance, analysis of field data with the automated detector, and exploration of results are available in a public online GitHub repository (<https://github.com/kitzeslab/rana-sierrae-cnn>; Lapp 2023) as well as on Zenodo (<https://doi.org/10.5281/zenodo.10150106>).

Literature Cited

- Bioacoustics Research Program. 2019. Raven Pro: interactive sound analysis software. Center for Conservation Bioacoustics, Cornell Lab of Ornithology, Ithaca, NY.
- Bradford, D. F. 1989. Allotopic distribution of native frogs and introduced fishes in high Sierra Nevada lakes of California: implication of the negative effect of fish introductions. *Copeia* 1989: 775–778.
- Bradford, D. F., D. M. Graber, and F. Tabatabai. 1994. Population declines of the native frog, *Rana muscosa*, in Sequoia and Kings Canyon National Parks, California. *Southwestern Naturalist* 39:323–327.
- California Herps. 2023a. Sounds of Yosemite toad—*Anaxyrus canorus*. <https://californiaherps.com/frogs/pages/a.canorus.sounds.html>.
- . 2023b. Sounds of Sierra Nevada yellow-legged frog—*Rana sierrae*. <https://californiaherps.com/frogs/pages/r.sierrae.sounds.html>.
- Cruikshank, S. S., A. Bergamini, and B. R. Schmidt. 2021. Estimation of breeding probability can make monitoring data more revealing: a case study of amphibians. *Ecological Applications* 31: e02357.
- DeGregorio, B. A., P. J. Wolff, and A. N. Rice. 2021. Evaluating hydrophones for detecting underwater-calling frogs. *Herpetological Conservation and Biology* 16:513–524.
- Desjonquères, C., T. Gifford, and S. Linke. 2020. Passive acoustic monitoring as a potential tool to survey animal and ecosystem processes in freshwater environments. *Freshwater Biology* 65:7–19.
- Drost, C. A., and G. M. Fellers. 1996. Collapse of a regional frog fauna in the Yosemite area of the California Sierra Nevada, USA. *Conservation Biology* 10:414–425.
- Gerhardt, H. C., and M. A. Bee. 2006. Recognition and localization of acoustic signals. Pages 113–146 in P. M. Narins, A. S. Feng,

- R. R. Fay, and A. N. Popper, eds. Hearing and sound communication in amphibians. Springer Handbook of Auditory Research. Springer, New York.
- Gibb, R., E. Browning, P. Glover-Kapfer, and K. E. Jones. 2019. Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. *Methods in Ecology and Evolution* 10:169–185.
- Hill, A. P., P. Prince, J. L. Snaddon, C. P. Doncaster, and A. Rogers. 2019. AudioMoth: a low-cost acoustic device for monitoring biodiversity and the environment. *HardwareX* 6:e00073.
- Joseph, M. B., and R. A. Knapp. 2018. Disease and climate effects on individuals drive post-reintroduction population dynamics of an endangered amphibian. *Ecosphere* 9:e02499.
- Kahl, S. 2019. Identifying birds by sound: large-scale acoustic event recognition for avian activity monitoring. PhD diss. Chemnitz University of Technology.
- Kahl, S., T. Wilhelm-Stein, H. Hussein, and H. Klinck. 2017. Large-scale bird sound classification using convolutional neural networks. CLEF (working notes) 14.
- Knapp, R. A. 2005. Effects of nonnative fish and habitat characteristics on lentic herpetofauna in Yosemite National Park, USA. *Biological Conservation* 121:265–279.
- Knapp, R. A., D. M. Boiano, and V. T. Vredenburg. 2007. Removal of nonnative fish results in population expansion of a declining amphibian (mountain yellow-legged frog, *Rana muscosa*). *Biological Conservation* 135:11–20.
- Knapp, R. A., C. J. Briggs, T. C. Smith, and J. R. Maurer. 2011. No-where to hide: impact of a temperature-sensitive amphibian pathogen along an elevation gradient in the temperate zone. *Ecosphere* 2:art93.
- Knapp, R. A., G. M. Fellers, P. M. Kleeman, D. A. W. Miller, V. T. Vredenburg, E. B. Rosenblum, and C. J. Briggs. 2016. Large-scale recovery of an endangered amphibian despite ongoing exposure to multiple stressors. *Proceedings of the National Academy of Sciences of the USA* 113:11889–11894.
- Knapp, R. A., and K. R. Matthews. 2000. Non-native fish introductions and the decline of the mountain yellow-legged frog from within protected areas. *Conservation Biology* 14:428–438.
- Knapp, R. A., S. Pavaka, E. E. Hegeman, and T. C. Smith. 2020. The Sierra Lakes Inventory Project: non-native fish and community composition of lakes and ponds in the Sierra Nevada, California. Version 2. Environmental Data Initiative.
- Knight, E. C., K. C. Hannah, G. J. Foley, C. D. Scott, R. M. Brigham, and E. Bayne. 2017. Recommendations for acoustic recognizer performance assessment with application to five common automated signal recognition programs. *Avian Conservation and Ecology* 12:14.
- Lamont, T. A. C., L. Chapuis, B. Williams, S. Dines, T. Gridley, G. Frainer, J. Fearey, et al. 2022. HydroMoth: testing a prototype low-cost acoustic recorder for aquatic environments. *Remote Sensing in Ecology and Conservation* 8:362–378.
- Lapp, S. 2023. Automated detection of *Rana sierrae* vocalizations. *American Naturalist*, GitHub, <https://github.com/kitzeslab/rana-sierrae-cnn>.
- Lapp, S., and J. Kitzes. 2023. *Rana sierrae* annotated aquatic soundscapes. *American Naturalist*, Dryad Digital Repository, <https://doi.org/10.5061/dryad.9s4mw6mn3>.
- Lapp, S., T. Rhinehart, L. Freeland-Haynes, J. Khilnani, A. Syunkova, and J. Kitzes. 2023. OpenSoundscape: an open-source bioacoustics analysis package for Python. *Methods in Ecology and Evolution* 14:2321–2328.
- MacTague, L., and P. Northen. 1993. Underwater vocalization by the foothill yellow-legged frog (*Rana boylei*). *Transactions of the Western Section of the Wildlife Society* 29:1–7.
- Peek, R. A., M. Bedwell, S. M. O'Rourke, C. Goldberg, G. M. Wengert, and M. R. Miller. 2019. Hybridization between two parapatric ranid frog species in the northern Sierra Nevada, California, USA. *Molecular Ecology* 28:4636–4647.
- Pérez-Granados, C., and J. Traba. 2021. Estimating bird density using passive acoustic monitoring: a review of methods and suggestions for further research. *Ibis* 163:765–783.
- Schneider, S., S. Greenberg, G. W. Taylor, and S. C. Kremer. 2020. Three critical factors affecting automated image species recognition performance for camera traps. *Ecology and Evolution* 10:3503–3517.
- Silver, C. S. 2017. Population-level variation in vocalizations of *Rana boylei*, the foothill yellow-legged frog. Master's thesis. California State University, Chico.
- Stowell, D. 2022. Computational bioacoustics with deep learning: a review and roadmap. *PeerJ* 10:e13152.
- Strebel, N., C. J. Fiss, K. F. Kellner, J. L. Larkin, M. Kéry, and J. Cohen. 2021. Estimating abundance based on time-to-detection data. *Methods in Ecology and Evolution* 12:909–920.
- Talukdar, S., P. Singha, S. Mahato, Shahfahad, S. Pal, Y.-A. Liou, and A. Rahman. 2020. Land-use land-cover classification by machine learning classifiers for satellite observations—a review. *Remote Sensing* 12:1135.
- Theiss, J., and J. Prager. 1984. Range of corixid sound signals in the biotope. *Physiological Entomology* 9:107–114.
- Vélez, J., W. McShea, H. Shamon, P. J. Castiblanco-Camacho, M. A. Tabak, C. Chalmers, P. Fergus, et al. 2023. An evaluation of platforms for processing camera-trap data using artificial intelligence. *Methods in Ecology and Evolution* 14:459–477.
- Vredenburg, V. T., R. Bingham, R. Knapp, J. A. T. Morgan, C. Moritz, and D. Wake. 2007. Concordant molecular and phenotypic data delineate new taxonomy and conservation priorities for the endangered mountain yellow-legged frog. *Journal of Zoology* 271:361–374.
- Vredenburg, V. T., R. A. Knapp, T. S. Tunstall, and C. J. Briggs. 2010. Dynamics of an emerging disease drive large-scale amphibian population extinctions. *Proceedings of the National Academy of Sciences of the USA* 107:9689–9694.
- Weinstein, B. G. 2018. A computer vision for animal ecology. *Journal of Animal Ecology* 87:533–545.
- Wells, K. D., and J. J. Schwartz. 2006. The behavioral ecology of anuran communication. Pages 44–86 in P. M. Narins, A. S. Feng, R. R. Fay, and A. N. Popper, eds. *Hearing and sound communication in amphibians*. Springer Handbook of Auditory Research. Springer, New York.
- Wilczynski, W., and J. Chu. 2001. Acoustic communication, endocrine control, and the neurochemical systems of the brain. In M. J. Ryan, ed. *Anuran communication*. Chap. 2. Smithsonian Institution Press, Washington, DC.
- Yip, D. A., E. C. Knight, E. Haave-Audet, S. J. Wilson, C. Charchuk, C. D. Scott, P. Sólomos, and E. M. Bayne. 2020. Sound level measurements from audio recordings provide objective distance estimates for distance sampling wildlife populations. *Remote Sensing in Ecology and Conservation* 6:301–315.
- Yip, D. A., L. Leston, E. M. Bayne, P. Sólomos, and A. Grover. 2017. Experimentally derived detection distances from audio recordings and human observers enable integrated analysis of point count data. *Avian Conservation and Ecology* 12:11.

Ziesmer, T. 1997. Vocal behavior of foothill and mountain yellow-legged frogs. Master's thesis. Sonoma State University.

Virtanen, P., R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, et al. 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods* 17:261–272.

References Cited Only in the Online Enhancements

He, K., X. Zhang, S. Ren, and J. Sun. 2015. Deep residual learning for image recognition. arXiv, <https://doi.org/10.48550/arXiv.1512.03385>.

Natural History Editor: Leticia Avilés



“Tufa Domes—Shore of Pyramid Lake.” From the review of the Third Annual Report of the U. S. Geological Survey (*The American Naturalist*, 1885, 19:151–153).