Biodiversity and Conservation
https://doi.org/10.1007/s10531-020-02107-1

ORIGINAL PAPER

Deep learning improves acoustic biodiversity monitoring and new candidate forest frog species identification (genus *Platymantis*) in the Philippines

Ali Khalighifar[1,2,4] · **Rafe M. Brown**[1,2] · **Johana Goyes Vallejos**[1,3] ·
A. Townsend Peterson[1,2]

Received: 7 September 2020 / Revised: 16 December 2020 / Accepted: 29 December 2020

Abstract

One significant challenge to biodiversity assessment and conservation is persistent gaps in species diversity knowledge in Earth's most biodiverse areas. Monitoring devices that utilize species-specific advertisement calls show promise in overcoming challenges associated with lagging frog species discovery rates. However, these devices generate data at paces faster than it can be analyzed. As such, automated platforms capable of efficient data processing and accurate species-level identification are at a premium. In addressing this gap, we used TensorFlow Inception v3 to design a robust, automated species identification system for 41 Philippine frog species (genus *Platymantis*), utilizing single-note audio spectrograms. With this model, we explored two concepts: (1) performance of our deep-learning model in discriminating closely-related frog species based on images representing advertisement call notes, and (2) the potential of this platform to accelerate new species discovery. TensorFlow identified species with a ~ 94% overall correct identification rate. Incorporating distributional data increased the overall identification rate to ~ 99%. In applying TensorFlow to a dataset that included undescribed species in addition to known species, our model was able to differentiate undescribed species through variation in "certainty" rate; the overall certainty rate for undescribed species was 65.5% *versus* 83.6% for described species. This indicates that, in addition to discriminating recognized frog species, our model has the potential to flag possible new species. As such, this work represents a proof-of-concept for automated, accelerated detection of novel species using acoustic mate-recognition signals, that can be applied to other groups characterized by vibrational cues, seismic signals, and vibrational mate-recognition.

Keywords Bioacoustics · Biodiversity inventory · Ceratobatrachidae · Convolutional neural networks · Inception v3 · TensorFlow · Transfer learning

---

Communicated by Dirk Sven Schmeller.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10531-020-02107-1.

✉ Ali Khalighifar
a.khalighifar@colostate.edu

Extended author information available on the last page of the article

Introduction

Species new to science are continuously being described, and therefore many evolutionary, ecological, and behavioral phenomena and processes remain to be discovered (Scheffers et al. 2013; Tonini et al. 2020). However, habitat destruction is triggering the rapid loss of species unknown to science (Bryan et al. 2013; Tapley et al. 2018). Recent efforts to overcome this arms race between species discovery and extinction (González-del-Pliego et al. 2019; Klein et al. 2015) have focused on development of automated monitoring devices, such as passive recorders for monitoring species with acoustic mate-recognition signals (i.e., bats, birds, frogs, crickets; Chen and Wiens 2020). Research using conventional recordings and new automated devices, however, has generated exorbitant quantities of data, at a pace faster than they can be analyzed (Brabant et al. 2018). As such, a need for automated processing tools and acoustic species identification has emerged.

Patterns of seasonal phenology, diel activity, habitat use, and focal species monitoring have been the subject of acoustic signal inventories (Sugai et al. 2019) and biodiversity assessments (Wimmer et al. 2013). The use of advertisement calls for integrative amphibian species delimitation and identification has increased steadily (Brown and Stuart 2012; Philippe et al. 2017; Vieites et al. 2009), but automating species discovery from environmental recordings has not been applied widely.

The advertisement calls of amphibians are primary phenotypes for mate-recognition (Gerhardt 1994; Wells and Schwartz 2007) and analyses of temporal and spectral acoustic data have been used widely to assign populations to species (Feinberg et al. 2014; Gerhardt 1978; Vignal and Kelley 2007). Once species-specific signals have been characterized quantitatively, automated classification methods can be employed for identification and assignment of species from natural soundscapes (Aide et al. 2013; Zhao et al. 2017). However, fundamental challenges arise when species new to science are recorded: classification is prevented by an absence of their temporal/spectral signal properties in training data sets.

The successful application of deep-learning algorithms to automated species identification systems (e.g., images from camera traps (Villa et al. 2017) and dorsal/ventral views of insects (Khalighifar et al. 2019; Rzanny et al. 2017; Villa et al. 2017)), suggests that the unique anuran signals (Gerhardt 1994; Narins and Capranica 1977) characterized in sound spectrograms (sensu Wells 2010) could be exploited for image-based species identification. To explore this possibility, we selected the Philippine frog genus *Platymantis* (family Ceratobatrachidae), for three reasons. First (A) *Platymantis* has been the focus of intensive surveys of advertisement call variation (Brown et al.1997a, b, 1999a, b), and the fast pace of species description (Brown and Stuart 2012; Brown and Gonzalez 2007; Diesmos et al. 2015), and identification of candidate species for future taxonomic studies (Brown et al. 2015a,b) demonstrate the prevalence of considerable underestimated species diversity (Brown et al. 2013, 2015b). Additionally, (B) as the focus of recent molecular phylogenetic analyses, *Platymantis* is a demonstrably monophyletic, Philippine-endemic clade (Brown et al. 2015b), sister to the similarly-diverse Papuan genus *Cornufer*. Thus, the geographic and systematic understanding of *Platymantis* is much improved over earlier work (Inger 1954). Third, (C) calls of nearly all recognized *Platymantis* species (AmphibiaWeb 2020, https://amphibiaweb.org) are available in the public domain. The combination of available call resources (large samples of calls from recognized and undescribed species), and a robust phylogeny as an historical framework render *Platymantis* a promising focal system for research in automating species identification, with intended future expansion of the system to incorporate comparative phylogenetic methods.

Automation offers a solution for tasks requiring repetition when experienced workers are lacking or cost-prohibitive (Gaston and O'Neill 2004). Recent computer science developments in image classification and signal processing provide new tools that may improve biodiversity assessment in taxa representing visual identification challenges (Guirado et al. 2019; MacLeod et al. 2010). Although automated species identification systems using handcrafted feature extraction has shown promising results (Gurgel-Gonçalves et al. 2017; Holmgren et al. 2008; Kumar et al. 2012), it requires advanced expertise and programming to accomplish robust performance.

In recent years, a robust group of classifiers has been introduced (Deep Neural Networks, DNNs; Schmidhuber 2015), which outperform existing methods in various classification tasks (Ramcharan et al. 2017; Smith et al. 2019). One of the state-of-the-art DNN platforms is TensorFlow (Abadi et al. 2016), an open-source software platform designed by the Google Brain Team (https://www.research.google.com/teams/brain). One crucial advantage of applying TensorFlow is the Transfer Learning technique, which is a shortcut for achieving high-performance classification. This approach involves using a large dataset to train a model, and then re-training with a new calibrating dataset both to improve identification rates at lower computational cost.

Here, we applied TensorFlow Inception v3 (Szegedy et al. 2016), implemented in the Linux environment (Ubuntu, version 18.04; https://www.ubuntu.com), to explore the challenge of automating frog species identification. We explored two major challenges in this study: (1) whether TensorFlow is able to discriminate among species of Philippine forest frogs, based on simple visual-image representations of auditory signals, and (2) whether TensorFlow could go beyond simple inventory to accelerate new candidate species discovery by objectively identifying undescribed species. To address these questions, we designed an automated, DNN-based species identification system for 41 described *versus* undescribed *Platymantis* species, which we tested using single-note call spectrograms. This study thereby lays a foundation for automated identification capabilities in biodiversity conservation and assessment via auditory signals.

Methods and materials

Data processing

We obtained and analyzed frog recordings from two sources: (1) a large collection of Philippine frog advertisement calls collected, archived (by RMB and colleagues), and made publicly available via Cornell University's Laboratory of Ornithology and Macaulay Library of Natural Sounds (https://www.macaulaylibrary.org), and (2) recent collections (2005–2019) of numerous undescribed species (review: Brown et al. 2015b; Diesmos et al. 2015), to augment sample sizes of previously described species and add distinctive new candidate species identified with genetic and phenotypic characters (RMB, unpublished data). We consider candidate species as unconfirmed (awaiting validation and formal description) but likely to eventually be recognized as new species, if they are demonstrably diagnosable via at least three out of four independent data streams: discrete phenotypic differences (traditional taxonomic characters), discrete spectral (frequency-related) or structural (presence/absence of syllables or notes) bioacoustic traits, genetically divergent monophyletic haplotype clades (we do not adopt arbitrary cut-offs or genetic distance thresholds), and allopatry on different Philippine

Pleistocene Aggregate Island Complexes (PAICs). The majority of species we characterize as such, actually possess all four sets of attributes. Additional collection-associated natural history information, frog microhabitats, community composition, recording methodology (device information, digitization specifications), and metadata are available via the Macaulay Library portal and the KU Herpetology online Specify database, as well as via GBIF, iDigBio, and other aggregators; behavioral context of calls, and qualitative descriptions of calls are available in original descriptions (e.g., Brown et al. 2015a; Brown and Gonzalez 2007; Siler et al. 2007, 2010).

We surveyed 175 recordings, representing 20 species (Fig. 1) using the cross-platform audio editor Ocenaudio (https://www.ocenaudio.com). This software is based on Ocen Framework, a powerful library to simplify and standardize manipulation and analysis of audio files. We clipped 20 high-quality single notes per each species, and saved each as 32-bit, single-channel WAV files (44.1 kHz sampling rate). To standardize temporal scale across comparisons, we designated a duration of one second; all known *Platymantis* species' single notes fit this range. To do so, we added silence in equal length to the beginning and the end of each clipped single note. Then, we used R packages warbleR (Araya-Salas and Smith-Vidaurre 2017) and Seewave (Sueur et al. 2008) to generate spectrograms across a standardized range of frequencies, 1.0–5.5 kHz; all known *Platymantis* calls fall within this range (Fig. 1). To generate oscillograms, we chose a fast-Fourier transformation (FFT) of 512 points, with 90% overlap between two successive windows. We saved all spectrograms as Portable Network Graphics (PNGs).

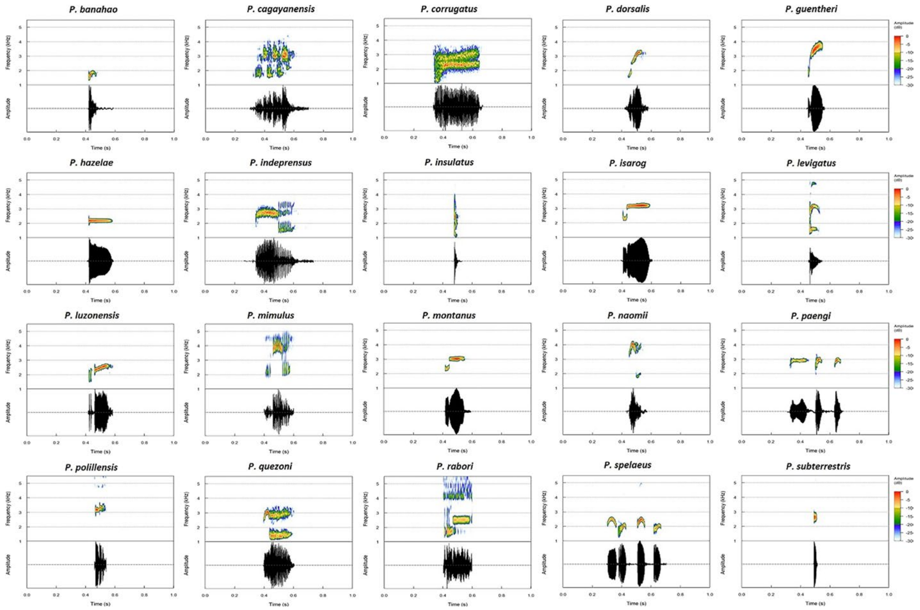


Fig. 1 Examples of images used for identification of 20 currently recognized species of Philippine forest frogs, genus *Platymantis* (Brown et al. 2015b), available from the Cornell Lab of Ornithology Macaulay Library. Each spectrogram has the same time duration (one second), and frequency limits (1–5.5 kHz)

Model architecture

Convolutional neural networks (CNNs) are a subset of DNNs that are specialized for image classification tasks and pattern recognition. One of the main advantages of CNNs is the ability to perform automated feature extraction, eliminating the need for hand-crafted feature extraction. CNN architecture is built on three types of layers: (1) convolutional layers, which are the most important because they apply hierarchical feature extraction and decomposition of input images; (2) pooling layers, which carry out operations to reduce numbers of parameters and necessary computation; and (3) fully connected layers, which perform the actual classification at the end of the pipeline.

CNNs require large training datasets to achieve accurate classification rates. Although training on a large dataset provides a powerful framework, building and training a CNN from scratch is both computationally expensive and time consuming. To overcome these limitations, we used a transfer learning technique. Transfer learning means using experience acquired from classification task A in classification task B. This technique allows the user to retrain the final layer of an existing model on the training set associated with a new classification task. One of the most successfully implemented models of transfer learning is Inception v3 (Szegedy et al. 2016)—a CNN, implemented in TensorFlow (Abadi et al. 2016). This CNN consists of 48 layers, and is trained on > 1 M images from the ImageNet database (http://www.image-net.org). Inception v3 is widely recognized for outperforming other models in challenges involving classifying images into thousands of classes (Russakovsky et al. 2015).

Classification challenges

We designed four classification challenges using single-note spectrograms and we assessed whether TensorFlow (Inception v3) is capable of successfully identifying *Platymantis* species based on frequency distributions of individual call notes of each species. For the training process, we modified two parameters of the model: (1) number of training steps, and (2) validation percentage. We explored different numbers of training steps and compared results to find an optimum balance between computing time and classification efficiency. Given our limited number of images per species (n = 20), we increased the validation parameter to 20%. For the same reason, we used a leave-one-out cross validation technique (Molinaro et al. 2005) to evaluate model performance in Challenges 1, 3, and 4. The four classification challenges we explored are as follows:

1. *Applying TensorFlow to identify species available from the Cornell Library of Natural Sounds* We applied TensorFlow to data from the Macaulay Library, including 20 recognized species of *Platymantis* (from among 33 described forms; Brown et al. 2015b), as classes for input. We generated 400 spectrograms (20 per species) for the identification challenge.
2. *Challenging TensorFlow with species not in the reference library* We trained TensorFlow on all images (i.e., 400 spectrograms) from Challenge 1 as an image reference library. Then, we applied the trained model to a test dataset from 22 robustly-identified species from recent field surveys by RMB. We addressed two questions: (1) could TensorFlow identify species existing in its reference library among the unknown species when they constitute new recordings obtained from different individuals? And, more importantly,

(2) how does TensorFlow perform when it encounters species that do not exist in its reference library? To answer these questions, we generated 20 spectrograms per species for this new dataset, resulting in a total of 440 spectrograms, to be subjected to identification using the CNN developed in Challenge 1.

3. *Challenging TensorFlow to identify all 41 species for which recordings are available* We increased the number of species in the TensorFlow reference library to 41 by adding 21 new, undescribed species (one species in the new dataset was already present in reference library, owing to a recent taxonomic change; Brown et al. 2015b). Then, we followed the same procedure as in Challenge 1, to test performance on a reference library that is twice as large as the original (i.e., 41 classes).
4. *Faunal region-based identification in natural species pools* We used distributional data (Brown et al. 2015b; Diesmos et al. 2015) to create subsets of species, with the goal of generating separate classification tasks with lower numbers of classes (species) per task (local species pools, reflecting documented patterns of co-distributed species from the archipelago's faunal regions). First, we grouped the 41 species (20 described species, plus 21 undescribed candidate species) based on Philippine islands they inhabit, resulting in 15 subsets of co-occurring species in "communities" of 3 to 27 species (Fig. 2). Then, we trained and tested the classifier employing the set of samples from the species

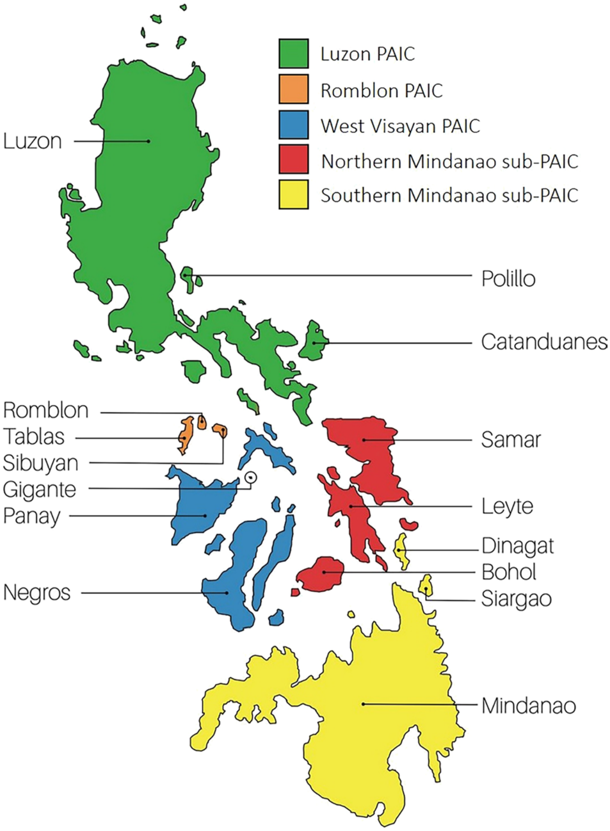


Fig. 2 Map of the Philippines, with Pleistocene Aggregate Island Complex (PAICs) faunal regions (colored shading) used to create realistic species pools (Challenge 4) to enhance identification. Map simplified to include only PAICs relevant to this study; see Diesmos et al. (2015) for full enumeration of species pool composition for each PAIC

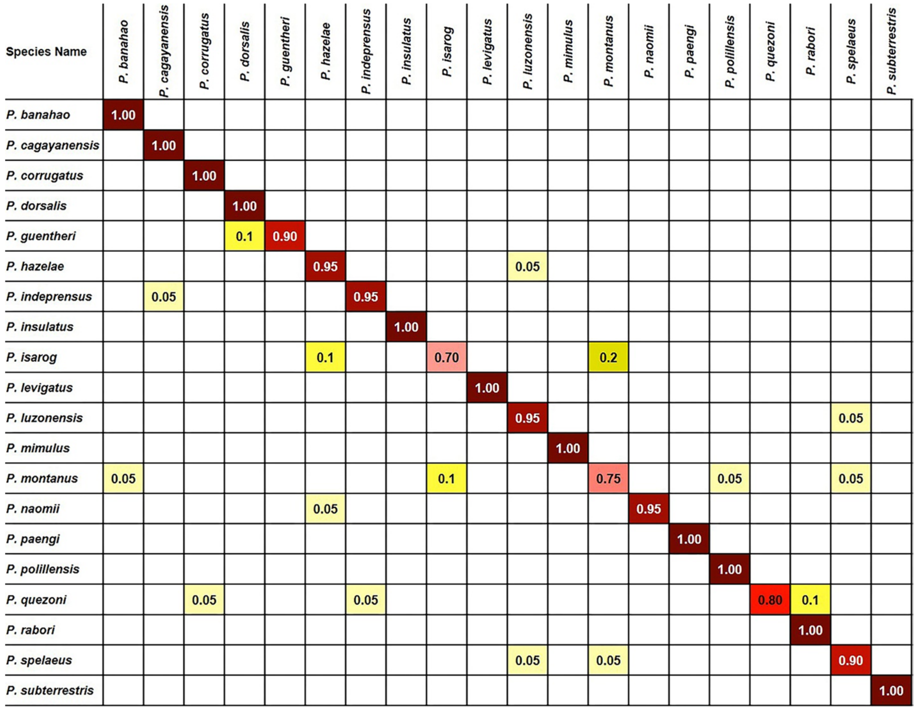
found on those islands. Finally, we calculated the overall correct identification rate across 15 islands to compare with that based on the full reference library.

Results

We calibrated models for classification challenges using different numbers of training steps, and considered two factors to find optima for training TensorFlow (Khalighifar et al. 2019): correct identification rate and processing time. As a result, for all challenges except Challenge 3, we chose 4000 training steps as an optimum number. For Challenge 3, given the number of species (41 species), 8000 training steps proved to be the optimum number. The details of results associated with each classification task are as follows:

Challenge 1

We created a confusion matrix to depict TensorFlow's initial results with 20 species (Fig. 3). The overall correct identification rate was 94.3%. We achieved 100% correct identification rate for 11, and 90% or above for 17, species. The lowest identification rates were for closely-related species *Platymantis isarog* and *P. montanus*, with 70 and 75% correct



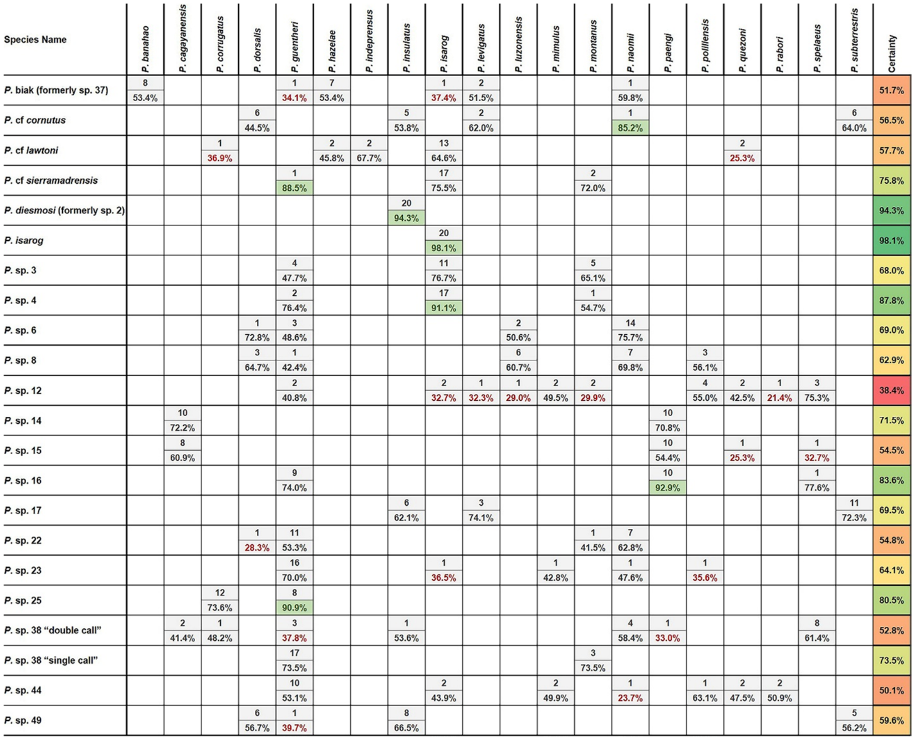
Species Name	P. banahao	P. cagayanensis	P. corrugatus	P. dorsalis	P. guentheri	P. hazelae	P. indeprensus	P. insulatus	P. isarog	P. levigatus	P. luzonensis	P. mimulus	P. montanus	P. naomii	P. paengi	P. polillensis	P. quezoni	P. rabori	P. spelaeus	P. subterrestris
P. banahao	1.00																			
P. cagayanensis		1.00																		
P. corrugatus			1.00																	
P. dorsalis				1.00																
P. guentheri				0.1	0.90															
P. hazelae						0.95					0.05									
P. indeprensus		0.05					0.95													
P. insulatus								1.00												
P. isarog						0.1			0.70				0.2							
P. levigatus										1.00										
P. luzonensis											0.95								0.05	
P. mimulus												1.00								
P. montanus	0.05								0.1				0.75			0.05			0.05	
P. naomii						0.05								0.95						
P. paengi															1.00					
P. polillensis																1.00				
P. quezoni			0.05				0.05										0.80	0.1		
P. rabori																		1.00		
P. spelaeus											0.05		0.05						0.90	
P. subterrestris																				1.00

Fig. 3 Confusion matrix for 20 currently recognized species of Philippine forest frogs, genus *Platymantis* (Brown et al. 2015b), using a leave-one-out cross-validation technique. Red = correct identification; yellow = misidentifications. All values of zero are removed for ease of visualization

classification, respectively. We did not detect any systematic errors in TensorFlow classification, such as repeatedly confusing one species with another.

Challenge 2

TensorFlow provides each image identification task with two elements/features: suggested species names and a certainty rate. Certainty rate can be a factor by which to evaluate classifier performance on test images as well. After applying TensorFlow on a testing dataset consisting of 22 species, we considered two factors to evaluate model performance: (1) number of images per class assigned to a species present in reference library, and (2) average certainty rate associated with those identifications (Fig. 4). Among the 22 species in the test dataset, only one, *P. isarog*, was also present in the reference library; however, the remainder were new to the training set. As a result, it was impossible for TensorFlow to provide a correct answer for the other 21 species. The overall certainty rate for those 21 species was 65.5%. However, in Challenge 1, the overall certainty rate for the 20 species present in reference library was 83.6%. A Mann–Whitney U test (Python 3.8.2) comparing certainty rates associated with species present in the reference library (n = 20) to species



Species Name	*P. banahao*	*P. cagayanensis*	*P. corrugatus*	*P. dorsalis*	*P. guentheri*	*P. hazelae*	*P. indeprensus*	*P. insulatus*	*P. isarog*	*P. levigatus*	*P. luzonensis*	*P. mimulus*	*P. montanus*	*P. naomi*	*P. paengi*	*P. polillensis*	*P. quezoni*	*P. rabori*	*P. spelaeus*	*P. subterrestris*	Certainty
P. biak (formerly sp. 37)	8 53.4%				1 34.1%	7 53.4%			1 37.4%	2 51.5%				1 59.8%							51.7%
P. cf *cornutus*				6 44.5%				5 53.8%		2 62.0%				1 85.2%						6 64.0%	56.5%
P. cf *lawtoni*			1 36.9%			2 45.8%	2 67.7%		13 64.6%								2 25.3%				57.7%
P. cf *sierramadrensis*					1 88.5%				17 75.5%				2 72.0%								75.8%
P. diesmosi (formerly sp. 2)								20 94.3%													94.3%
P. isarog									20 98.1%												98.1%
P. sp. 3					4 47.7%				11 76.7%				5 65.1%								68.0%
P. sp. 4					2 76.4%				17 91.1%				1 54.7%								87.8%
P. sp. 6				1 72.8%	3 48.6%						2 50.6%			14 75.7%							69.0%
P. sp. 8				3 64.7%	1 42.4%						6 60.7%			7 69.8%		3 56.1%					62.9%
P. sp. 12					2 40.8%				2 32.7%	1 32.3%	1 29.0%	2 49.5%	2 29.9%			4 55.0%	2 42.5%	1 21.4%	3 75.3%		38.4%
P. sp. 14		10 72.2%													10 70.8%						71.5%
P. sp. 15		8 60.9%													10 54.4%		1 25.3%		1 32.7%		54.5%
P. sp. 16					9 74.0%										10 92.9%				1 77.6%		83.6%
P. sp. 17								6 62.1%		3 74.1%										11 72.3%	69.5%
P. sp. 22				1 28.3%	11 53.3%								1 41.5%	7 62.8%							54.8%
P. sp. 23					16 70.0%				1 36.5%			1 42.8%		1 47.6%		1 35.6%					64.1%
P. sp. 25			12 73.6%		8 90.9%																80.5%
P. sp. 38 "double call"		2 41.4%	1 48.2%		3 37.8%			1 53.6%						4 58.4%	1 33.0%				8 61.4%		52.8%
P. sp. 38 "single call"					17 73.5%								3 73.5%								73.5%
P. sp. 44					10 53.1%				2 43.9%			2 49.9%		1 23.7%		1 63.1%	2 47.5%	2 50.9%			50.1%
P. sp. 49				6 56.7%	1 39.7%			8 66.5%												5 56.2%	59.6%

Fig. 4 Confusion matrix resulting from challenging TensorFlow with potential species unknown to the reference library—numeric species identifiers from Brown et al. (2015b). Columns = species in the reference library; rows = potential unknown species, with exception of *P. isarog* (see text). Red numbers = certainty rates below 40%; black = 41–85%; green > 85%. Far right column = average certainty rate for species identifications

new to the library (n = 21) indicated that TensorFlow yielded a significantly lower certainty rate for the latter group (U = 56379.5, *P* = 1.09e–19).

Certainty rates in Challenge 2 ranged from 38.4 to 98.1%. TensorFlow yielded an overall certainty rate ≥ 90% for only 2 of the 22 species in our test dataset, one of which was a correct identification. That is, all images associated with *P. isarog* identified as *P. isarog* with an average certainty rate of 98.1%. In contrast, all *P. diesmosi* images were identified as *P. insulatus* with a 94.3% average certainty rate. Another species with a relatively high average certainty rate was *P.* sp. 4, a taxon originally described as "*P. rivularus*" (Taylor 1923), which is expected to be elevated from synonymy of *P. subterrestris* with ongoing studies (Brown et al. 2015b). Individuals of this population were identified as *P. isarog*, *P. guentheri*, and *P. montanus*, with an average certainty rate of 87.8%. The lowest certainty rates yielded by TensorFlow were for three undescribed species, *P.* sp. 12 (onomatopoeically nicknamed "churink" with a 38.4% average certainty rate), *P.* sp. 44 ("Ee-yow" with a 50.1% average certainty rate), and *P. biak* (with a 51.7% average certainty rate), which were classified as 10, 7, and 6 different species, respectively. Among the species in the reference library, the most frequently suggested species was *P. guentheri*, which was suggested for 15 of 22 species in our test dataset. However, regarding two evaluation factors mentioned above, in none of those cases, could *P. guentheri* be considered as the primary identification for those species (Fig. 4).

Challenge 3

Training TensorFlow on all 41 species, we observed a mere 0.2% decline in overall correct identification rate (i.e., overall correct identification rate for 41 species was 94.1 *versus* 94.3% for 20 species in Challenge 1). Although we added 21 species to the reference library, we observed no negative impact on correct identification rates, even though such impacts were noted in our previous work (Khalighifar et al. 2019). TensorFlow was able to identify 37 of 41 species with ≥ 90% correct identification rate; 17 species were identified with 100% identification rate. The lowest correct identification rates were for *P. isarog* and *P.* sp. 3 (sensu Brown et al. 2015b) with 70% (Supplementary Table S1). Similar to Challenge 1, no species was detected to be mis-identified repeatedly as another species in particular.

Challenge 4

Classifier accuracy generally improves as the number of classes that must be distinguished decreases (Khalighifar et al. 2019). To improve accuracy and reduce numbers of species, we used geographic information (species distributions from Diesmos et al. 2015) to refine identification efforts. Species subsets ranged from 3 on Gigante, Panay, and Romblon to 27 on Luzon (Fig. 5). As expected, TensorFlow was able to increase the overall identification rate from 94.1 to 98.7% by incorporating distributional information.

Average identification rates for more than half of the islands (8 of 15) were 100%. Luzon Island hosts 27 species, and scored the lowest overall identification rate, 94.6%. We noticed that another factor affecting overall identification rate for each island was species composition. That is, for some islands with lower numbers of species, we nonetheless found a lower overall identification rate as well. For example, species from Siargao and Dinagat islands, with the same species composition (*Platymantis corrugatus*, *P. dorsalis*, *P. guentheri*, and

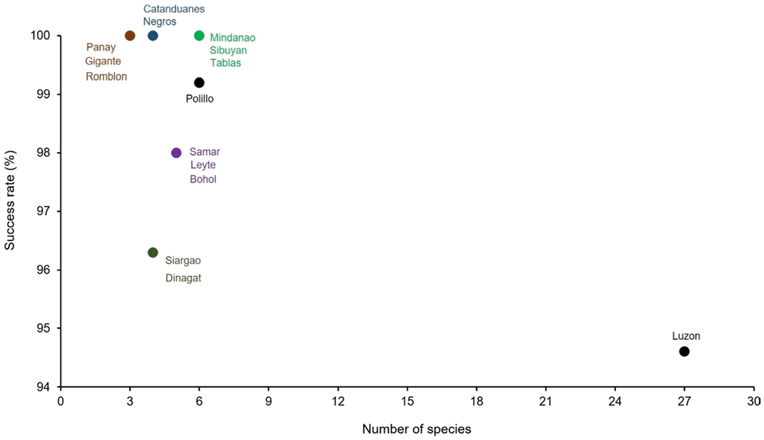


Fig. 5 Island-based identification for 20 currently-recognized species of Philippine forest frogs (Brown et al. 2015b) and 21 new, undescribed *Platymantis* species, incorporating biogeographic information (island bank-based faunal regions; Brown et al. 2013) and species distribution information (Diesmos et al. 2015). Note that although we incorporated distributional information (PAICs) in our analyses, this figure summarizes individual island results. Each point is only associated with two elements; (1) number of species per island (*x*-axis), and (2) correct identification rates (*y*-axis)

P. rabori) had a lower overall correct identification rate than Mindanao Island, a far larger landmass with at least six species (96.3% *versus* 100%).

Discussion

To improve bioacoustic monitoring and to automate rapid candidate species discovery, we explored CNN Inception v3 as a means to automate several key inferential tasks. We successfully identified *Platymantis* species with an impressive overall identification rate (94.3%) based on single call note characteristics, a surprising result for field biologists (RMB and colleagues) who are accustomed to discovering and describing *Platymantis* species based on temporal patterns of note repetition, complex call elements, and rich spectral properties of many calls (Brown and Gonzalez 2007; Brown et al. 2015b; Siler et al. 2007, 2010). Incorporating distributional data (grouping species by Pleistocene island bank-based faunal regions) allowed us to create realistic subsets of species pools. By referencing relevant, naturally-occurring species pools, we further increased overall identification rates to 98.7%. In addition, our model was able to flag species new to the reference library, indicating potential to recognize novel species in the field. This suggests that our model could provide a powerful framework for automated species identification and new species discovery in hard-to-access regions, as well as those with high rates of undiscovered biodiversity.

Deep neural networks

TensorFlow's flexible architecture allows easy adaptation and deployment on different platforms including desktops, clusters of servers, edge devices, and mobile phones (https://

www.tensorflow.org/). Inception v3 is a sophisticated network given the number of layers (48), and is already trained on more than a million images.

However, one major challenge was the number of call notes (input spectrograms) available for each species. To address this, future studies should access more recordings per species, particularly those represented now by recordings of few individuals. Despite this limitation, our study is a novel use of a deep-learning platform to distinguish between closely-related species of frogs using simple, single-note, two-dimensional depictions of primary mate-recognition cues (mating calls). A second challenge, also related to sample size, was model validation. We would have been more satisfied with our evaluation if we had more recordings from numerous individuals to utilize *k*-fold cross-validation (Kohavi 1995) instead of leave-one-out approaches, because the former offers greater independence between calibration and evaluation datasets.

Taxonomic identification

In this study, we demonstrate the efficacy of deep learning technology for reliably identifying—and distinguishing among—closely-related frog species, as exemplified by single-note call segments (Fig. 1). Given known phylogenetic relationships (Brown et al. 2015b), we were surprised that multiple closely-related species pairs were distinguished from one another perfectly (100% success). For example, species pairs *P. indeprensus* and *P. mimulus* (both members of the subgenus *Lupacolus*), *P. hazelae* and *P. montanus* (subgenus *Tirahanulap*), and *P. levigatus* and *P. insulatus* (subgenus *Lahatnanguri*) could each be predicted, by virtue of their close phylogenetic relationships, to have similar spectral (frequency-related) and temporal (time-related) call properties—which they do (Fig. 1). Still, with only a single isolated note per species, TensorFlow is able to distinguish them and correctly classify species' identity, when presented with a large sample of positively identified individual notes (i.e., known populations of confidently-identified species, based on fully documented voucher specimens deposited in biodiversity repositories).

That said, illustrative examples of how the methods failed in our study—cases where identification was problematic, attributed to multiple species, or when a sample of notes were classified to wrong species—are worthy of consideration. In these cases (Supplementary Table S1), two categories of identification errors emerge: Type A stems from closely-related species, with brief, pure-tone, constant-frequency calls, whose calls are exceptionally simple, intra-specifically invariant, and even inter-specifically quite difficult to distinguish. Referred to as "cloud frogs," members of subgenus *Tirahanulap* (former "P. hazelae Group" species; Brown et al. 2015b; Diesmos et al. 2015) are all diminutive (1–3 g body mass), primarily higher-elevation moist, closed-canopy shrub frogs. Their close phylogenetic relationships (Brown et al. 2015b) and remarkably similar microhabitat preferences render it no surprise to us that TensorFlow had difficulty distinguishing *P. isarog, P. montanus, P.* sp. 3, *P. sierramadrensis,* and *P. lawtoni.* Another category of error was exemplified (Type B) by instances of apparent convergence in frequency modulation, exemplified by unrelated species such as *P. dorsalis* (subgenus *Lupacolus*), *P. guentheri* (*Tahananpuno*), and the amplitude-modulated, rapidly-repetitive pulse-train calls of taxa like *P. luzonensis, P.* sp. 6, and *P.* sp. 8. (*Tahananpuno*). In these taxa, it is little surprise that single-note call components are occasionally mis-specified by TensorFlow, given that they are essentially homologous call elements (Brown et al. 2015b), temporally arranged to differ only by numbers of notes per call and calling rate (Brown et al. 2015b; Brown et al. 1997a, b, 1999a, b). However, given that only a few closely-related species pairs exhibit overlapping, sympatric geographic ranges (Diesmos et al. 2015), our confidence is further bolstered with the confirmation that automated discrimination can be enhanced by limitation of species classes to realistic species pools (Brown et al. 2013; see below). In summary, we take these results as encouraging in that the efficacy of automated species identification can be improved with biogeographic information.

Our refinement of the method, using biogeographically-relevant species sets and limiting species identifications to co-distributed taxa, resulted in a dramatic improvement in method performance (Fig. 4), particularly when considering caveats discussed above. By limiting the possible universe of a species' identification to the biogeographically-relevant species pools, i.e., we both (1) improved performance of identifications of known taxa, and (2) drew attention to (analytically singled out) unknown, new, or undescribed taxa (Figs. 2, 3). These features will be valuable in identifying taxa for subsequent 'validation' of unconfirmed candidate species, using independent data streams (phenotypic data, genetic information, bioacoustics, biogeographical information, ecological characteristics, etc.).

Across broader taxonomic scales and phylogenetic relationships (e.g., Chan and Brown 2017), other means (biogeographic realm, ectomorph type, classification, etc.) of restricting/limiting candidate species pools may prove useful for 'fine-tuning' of TensorFlow's automation of species recognition. Additional caveats for future consideration include (1) single notes per species and (2) sample sizes, which will be limited for rare species, those that occur at naturally low abundances, or taxa characterized by reduced detection probabilities due to cryptic microhabitat preferences, narrow activity patterns, or seasonally-limited reproductive cycles (Wells 2010). Avenues for future development of these methods in our immediate plans include application to additional taxonomic groups (e.g., insects, birds), and automation of call detection from environmental sound samples as a precursor step to automated species identification.

Acknowledgements Philippine *Platymantis* frog calls were collected with support from the U. S. National Science Foundation's former Doctoral Dissertation Improvement Grant (DEB 0073199; 2001–2003) and a Biotic Surveys and Inventories Grant (DEB 0743491; 2008–2012). Collection and co-curated voucher specimens and their associated digital media specimens, archival digitization, data verification, and online serving of digital media specimens was made possible by a NSF Thematic Collections Network (TCN) program Grant (DEB 1304585; 2013–2018). Recent extended specimen collection and curation has been supported by NSF DEB 1654388 and 1557053, with further support from the KU Biodiversity Institute's Rudkin Research Exploration (REX) Fund and KU College of Liberal Arts and Sciences Docking Scholar Fund. We thank A. Diesmos, J. Fernandez, C. Siler, and C. Meneses for help recording frogs.

Funding No funding was used for this research.

Data availability Data are publicly available via Cornell University's Laboratory of Ornithology and Macaulay Library of Natural Sounds (https://www.macaulaylibrary.org).

Compliance with ethical standards

Conflict of interest All authors declare that they have no conflicts of interest.

References

Abadi M et al. (2016) Tensorflow: a system for large-scale machine learning. Paper presented at the 12th USENIX Symposium on Operating Systems Design and Implementation, Savannah, USA, Aide TM, Corrada-Bravo C, Campos-Cerqueira M, Milan C, Vega G, Alvarez R (2013) Real-time bioacoustics monitoring and automated species identification. PeerJ 1:e103

Araya-Salas M, Smith-Vidaurre G (2017) warbleR: an R package to streamline analysis of animal acoustic signals. Methods Ecol Evol 8:184–191

Brabant R, Laurent Y, Dolap U, Degraer S, Poerink BJ (2018) Comparing the results of four widely used automated bat identification software programs to identify nine bat species in coastal Western Europe Belgian. J Zool 148:119–128

Brown RM, Gonzalez JC (2007) A new forest frog of the genus *Platymantis* (Amphibia: Anura: Ranidae) from the Bicol Peninsula of Luzon Island, Philippines. Copeia 2007:251–266

Brown R, Stuart B (2012) Patterns of biodiversity discovery through time: an historical analysis of amphibian species discoveries in the Southeast Asian mainland and island archipelagos. In: Gower D et al (eds) Biotic evolution and environmental change in Southeast Asia. Cambridge University Press, Cambridge, pp 348–389

Brown WC, Alcala A, Diesmos AC (1997a) A new species of the genus *Platymantis* (Amphibia: Ranidae) from Luzon Island, Philippines. Proc Biol Soc Wash 110:18–23

Brown WC, Alcala A, Diesmos AC, Alcala E (1997b) Species of the *guentheri* group of *Platymantis* (Amphibia: Ranidae) from the Philippines, with descriptions of four new species. Proc Calif Acad Sci 50:1–20

Brown W, Alcala A, Ong P, Diesmos A (1999a) A new species of *Platymantis* (Amphibia: Ranidae) from the Sierra Madre Mountains, Luzon Island, Philippines. Proc Biol Soc Wash 112:510–514

Brown WC, Alcala AC, Diesmos AC (1999b) Four new species of the genus *Platymantis* (Amphibia: Ranidae) from Luzon Island, Philippines. Proc Calif Acad Sci 51:449–460

Brown RM et al (2013) Evolutionary processes of diversification in a model island archipelago. Annu Rev Ecol Evol Syst 44:411–435

Brown RM, De Layola LA, Lorenzo A, Diesmos MLL, Diesmos AC (2015a) A new species of limestone karst inhabiting forest frog, genus *Platymantis* (Amphibia: Anura: Ceratobatrachidae: subgenus *Lupacolus*) from southern Luzon Island, Philippines. Zootaxa 4048:191–210

Brown RM, Siler CD, Richards SJ, Diesmos AC, Cannatella DC (2015b) Multilocus phylogeny and a new classification for Southeast Asian and Melanesian forest frogs (family Ceratobatrachidae). Zool J Linnean Soc 174:130–168

Bryan JE, Shearman PL, Asner GP, Knapp DE, Aoro G, Lokes B (2013) Extreme differences in forest degradation in Borneo: coamparing practices in Sarawak, Sabah, and Brunei. PLoS ONE 8:e69679

Chan KO, Brown RM (2017) Did true frogs 'dispersify'? Biol Lett 13:20170299

Chen Z, Wiens JJ (2020) The origins of acoustic communication in vertebrates. Nat Commun 11:1–8

Diesmos AC et al (2015) Amphibians of the Philippines, part I: checklist of the species. Proc Calif Acad Sci 62:457–539

Feinberg JA et al (2014) Cryptic diversity in metropolis: confirmation of a new leopard frog species (Anura: Ranidae) from New York City and surrounding Atlantic coast regions. PLoS ONE 9:e108213

Gaston KJ, O'Neill MA (2004) Automated species identification: why not? Philos Trans R Soc Lond B 359:655–667

Gerhardt HC (1978) Mating call recognition in the green treefrog (*Hyla cinerea*): the significance of some fine-temporal properties. J Exp Biol 74:59–73

Gerhardt HC (1994) The evolution of vocalization in frogs and toads. Annu Rev Ecol Syst 25:293–324

González-del-Pliego P, Freckleton RP, Edwards DP, Koo MS, Scheffers BR, Pyron RA, Jetz W (2019) Phylogenetic and trait-based prediction of extinction risk for data-deficient amphibians. Curr Biol 29(1557–1563):e1553

Guirado E, Tabik S, Rivas ML, Alcaraz-Segura D, Herrera F (2019) Whale counting in satellite and aerial images with deep learning. Sci Rep 9:14259

Gurgel-Gonçalves R et al (2017) Automated identification of insect vectors of chagas disease in Brazil and Mexico: the virtual vector. Lab PeerJ 5:e3040

Holmgren J, Persson Å, Söderman U (2008) Species identification of individual trees by combining high resolution LiDAR data with multi-spectral images. Int J Remote Sens 29:1537–1552

Inger RF (1954) Systematics and zoogeography of Philippine Amphibia. Fieldiana Zool 33:181–531

Khalighifar A, Komp E, Ramsey JM, Gurgel-Gonçalves R, Peterson AT (2019) Deep learning algorithms improve automated identification of chagas disease vectors. J Med Entomol 56:1404–1410

Klein DJ, McKown MW, Tershy BR (2015) Deep learning for large scale biodiversity monitoring. Paper presented at the Bloomberg Data for Good Exchange Conference, New York City, USA

Kohavi R (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection. In: International Joint Conferences on Artificial Intelligence, Montreal, Canada, pp. 1137–1145

Kumar N, Belhumeur PN, Biswas A, Jacobs DW, Kress WJ, Lopez IC, Soares JV (2012) Leafsnap: A computer vision system for automatic plant species identification. In: European Conference on Computer Vision, Florence, Italy, pp. 502–516

MacLeod N, Benfield M, Culverhouse P (2010) Time to automate identification. Nature 467:154–155

Molinaro AM, Simon R, Pfeiffer RM (2005) Prediction error estimation: a comparison of resampling methods. Bioinformatics 21:3301–3307

Narins PM, Capranica RR (1977) An automated technique for analysis of temporal features in animal vocalizations. Anim Behav 25:615–621

Philippe J, Felipe L, Celio F (2017) The use of bioacoustics in anuran taxonomy: theory, terminology, methods and recommendations for best practice. Zootaxa 4251:1–124

Ramcharan A, Baranowski K, McCloskey P, Ahmed B, Legg J, Hughes DP (2017) Deep learning for image-based cassava disease detection. Front Plant Sci 8:1852–1859

Russakovsky O et al (2015) ImageNet large scale visual recognition challenge. Int J Comput Vision 115:211–252

Rzanny M, Seeland M, Wäldchen J, Mäder P (2017) Acquiring and preprocessing leaf images for automated plant identification: understanding the tradeoff between effort and information gain. Plant Methods 13:1–11

Scheffers BR, Brunner RM, Ramirez SD, Shoo LP, Diesmos A, Williams SE (2013) Thermal buffering of microhabitats is a critical factor mediating warming vulnerability of frogs in the Philippine biodiversity hotspot. Biotropica 45:628–635

Schmidhuber J (2015) Deep learning in neural networks: an overview. Neural Netw 61:85–117

Siler CD, Linkem CW, Diesmos AC, Alcala AC (2007) A new species of *Platymantis* (Amphibia: Anura: Ranidae) from Panay Island, Philippines. Herpetologica 63:351–364

Siler CD, Diesmos AC, Linkem CW, Diesmos ML, Brown RM (2010) A new species of limestone-forest frog, genus *Platymantis* (Amphibia: Anura: Ceratobatrachidae) from central Luzon Island, Philippines. Zootaxa 2482:49–63

Smith SW et al (2019) A deep neural network learning algorithm outperforms a conventional algorithm for emergency department electrocardiogram interpretation. J Electrocardiol 52:88–95

Sueur J, Aubin T, Simonis C (2008) Seewave, a free modular tool for sound analysis and synthesis. Bioacoustics 18:213–226

Sugai LSM, Silva TSF, Ribeiro JW Jr, Llusia D (2019) Terrestrial passive acoustic monitoring: review and perspectives. Bioscience 69:15–25

Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z (2016) Rethinking the inception architecture for computer vision. Paper presented at the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA,

Tapley B, Michaels CJ, Gumbs R, Böhm M, Luedtke J, Pearce-Kelly P, Rowley JJ (2018) The disparity between species description and conservation assessment: a case study in taxa with high rates of species discovery. Biol Conserv 220:209–214

Taylor EH (1923) Addition to the herpetological fauna of the Philippine Islands III Philippine. J Sci 22:515–557

Tonini JFR, Provete DB, Maciel NM, Morais AR, Goutte S, Toledo LF, Pyron RA (2020) Allometric escape from acoustic constraints is rare for frog calls. Ecol Evol 10:1–10

Vieites DR, Wollenberg KC, Andreone F, Köhler J, Glaw F, Vences M (2009) Vast underestimation of Madagascar's biodiversity evidenced by an integrative amphibian inventory. Proc Natl Acad Sci USA 106:8267–8272

Vignal C, Kelley D (2007) Significance of temporal and spectral acoustic cues for sexual recognition in *Xenopus laevis*. Proc R Soc B 274:479–488

Villa AG, Salazar A, Vargas F (2017) Towards automatic wild animal monitoring: identification of animal species in camera-trap images using very deep convolutional neural networks. Ecol Inform 41:24–32

Wells KD (2010) The ecology and behavior of amphibians. University of Chicago Press, Chicago

Wells KD, Schwartz JJ (2007) The behavioral ecology of anuran communication. Hearing and sound communication in amphibians, vol 28. Springer, New York, pp 44–86

Wimmer J, Towsey M, Roe P, Williamson I (2013) Sampling environmental acoustic recordings to determine bird species richness. Ecol Appl 23:1419–1428

Zhao Z, Zhang S, Xu Z, Bellisario K, Dai N, Omrani H, Pijanowski BC (2017) Automated bird acoustic event detection and robust species classification. Ecol Inform 39:99–108

Authors and Affiliations

Ali Khalighifar[1,2,4] · **Rafe M. Brown**[1,2] · **Johana Goyes Vallejos**[1,3] · **A. Townsend Peterson**[1,2]

[1] Biodiversity Institute, University of Kansas, Lawrence, KS, USA

[2] Department of Ecology and Evolutionary Biology, University of Kansas, Lawrence, KS, USA

[3] Division of Biological Sciences, University of Missouri, Columbia, MO, USA

[4] Present Address: Warner College of Natural Resources, Colorado State University, Fort Collins, CO, USA