Using Convolutional Neural Networks to Identify Sounds Sources in Audio Recordings for Biodiversity Research

Liam Quaas, Kristen M. Bellisario, PhD, David Savage, Diego Gomez Morales, Bryan C. Pijanowski, PhD

Introduction

Strategies for understanding and characterizing biodiversity are challenges faced by ecologists. Soundscape ecology is the study of the relationship between the sounds produced in an environment (including biophony, geophony and anthrophony). These three sound components are described respectively as: the collection of sounds produced by all organisms, sounds originating from the geophysical environment, and sounds produced by stationary and moving human made objects or humans themselves [3]. These intertwine deeply as their interaction builds and creates patterns in the sonic environment [2].

Figure 1: An acoustic sensor deployed at Purdue's Ross Biological Reserve. Similar sensors were used at other sites for this study.

Soundscape ecology recognizes that sounds in the environment could help us understand and track changes in biodiversity. One difficulty in the ecological analysis of soundscapes is the tagging and classification of the sounds to allow content analysis.

The conceptual model of soundscape content analysis includes data ingestion, data processing, and content labeling, time intensive process [1]. Supervised learning, a method in data mining, uses pattern recognition algorithms that train on datasets to predict an outcome [1].

We hypothesize that spectrograms can be used to automatically label broad taxonomic classes of sound sources using a convolutional neural network (CNN). To test this, we conducted a research study using common sounds collected in Indiana.

REFERENCES: [1] Bellisario, K. M., Broadhead, T., Savage, D., Zhao, Z., Omrani, H., Zhang, S., … Pijanowski, B. C. (2019). Contributions of MIR to soundscape ecology. Part 3: Tagging and classifying audio features using a multi-labeling k-nearest neighbor approach. Ecological Informatics, 51, 103–111. doi: 10.1016/j.ecoinf.2019.02.010 - [2] Farina, A. (2014). Soundscape Ecology. doi: 10.1007/978-94-007-7374-5 - [3] Pijanowski, B. C., Farina, A., Gage, S. H., Dumyahn, S. L., & Krause, B. L. (2011). What is soundscape ecology? An introduction and overview of an emerging new science. Landscape Ecology, 26(9), 1213–1232. doi: 10.1007/s10980-011-9600-8 - [4] Servick, K. (2014). Eavesdropping on Ecosystems. American Association for the Advancement of Science.

Methodology

File Name	Date Recorded	Bird	Insect	Amphib	Mammal	Other_bio	Wind	Rain	Thunder	Car/motor	Airplane	Window Duration	Purity	Confidence
PWA_20130507_063000.flac	5/7/2013	1	0	1	0	0	0	0	0	0	0	3		
PWA_20130507_063000.flac	5/7/2013	1	0	1	0	0	0	0	0	0	0	10		
PWA_20130507_063000.flac	5/7/2013	1	0	1	0	0	0	0	0	0	0	60	1	3

Figure 2: Example of presence-absence table for taxonomic classes with 1 being present and 0 being absent. Each recording also received a viewer purity and confidence level score, rated from 1 (least pure or least confident) to 3 (most pure or very confident).

- The full dataset for this study includes 1904 audio samples from Tippecanoe County (2013) and 600 samples from Hamilton County (2015), Indiana at 9 different sites (all Temperate Forest locations).
- A stratified random sample of the dataset was labeled using a manual identification and visual identification method in two separate phases. Two reviewers labeled the audio recordings for content (Phase 1: Hamilton County (n=1305) – KB, LJ, Phase 2: Tippecanoe (n=599) - DG, LQ).
- In Phase 1 and 2, labels were assigned to each recording into taxonomic level by manual inspection (listen and visual inspection of spectrogram) for biological sounds (e.g. bird, mammal, etc.) and broad classes for abiotic sounds (e.g., wind, rain, or car).
- In Phase 2, two reviewers (DG, LQ) labeled qualitative scores for **purity** (how clear was the sound) and **confidence** (how certain they were of the label assignment)

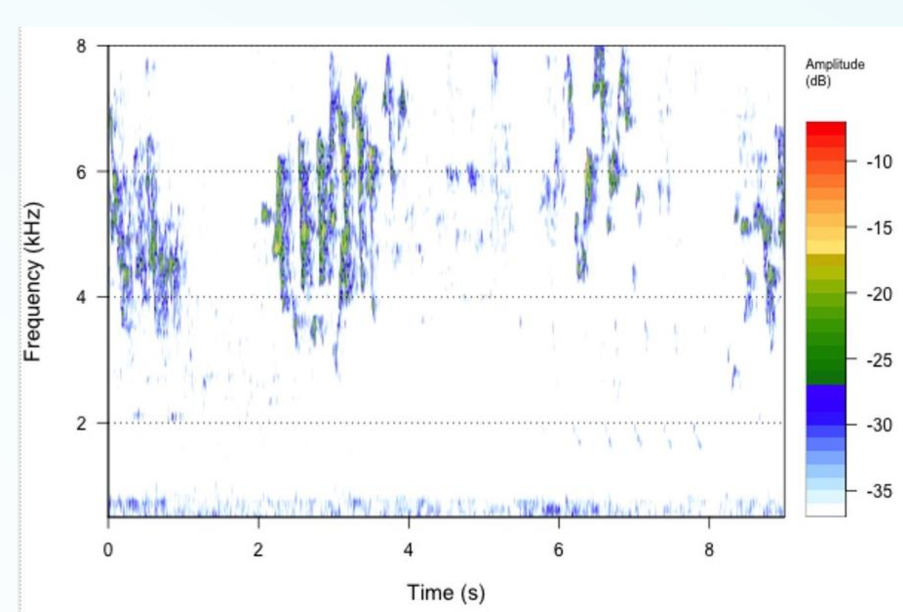

Figure 3a: Spectrogram of Bird for Training Data

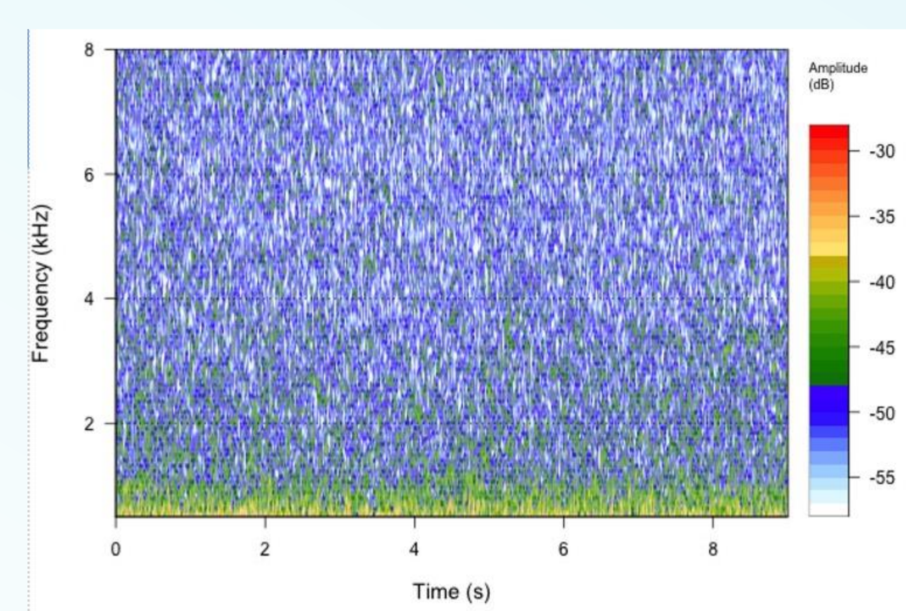

Figure 3b: Spectrogram of Rain for Training Data

- In Phase 3, the pilot phase of CNN, we used a subset (n=126) of validated labels from Phase 1 (bird: n=70, insect: n=25, noise: n=6, rain: n=24, mixed: n=24) for each spectrogram (600 x 400 pixels, png format) for the training dataset in two scales (full and subsections of an image).
- We split the training data into training and testing datasets. Then, we used a convolutional neural network implemented using TensorFlow in R.

Results

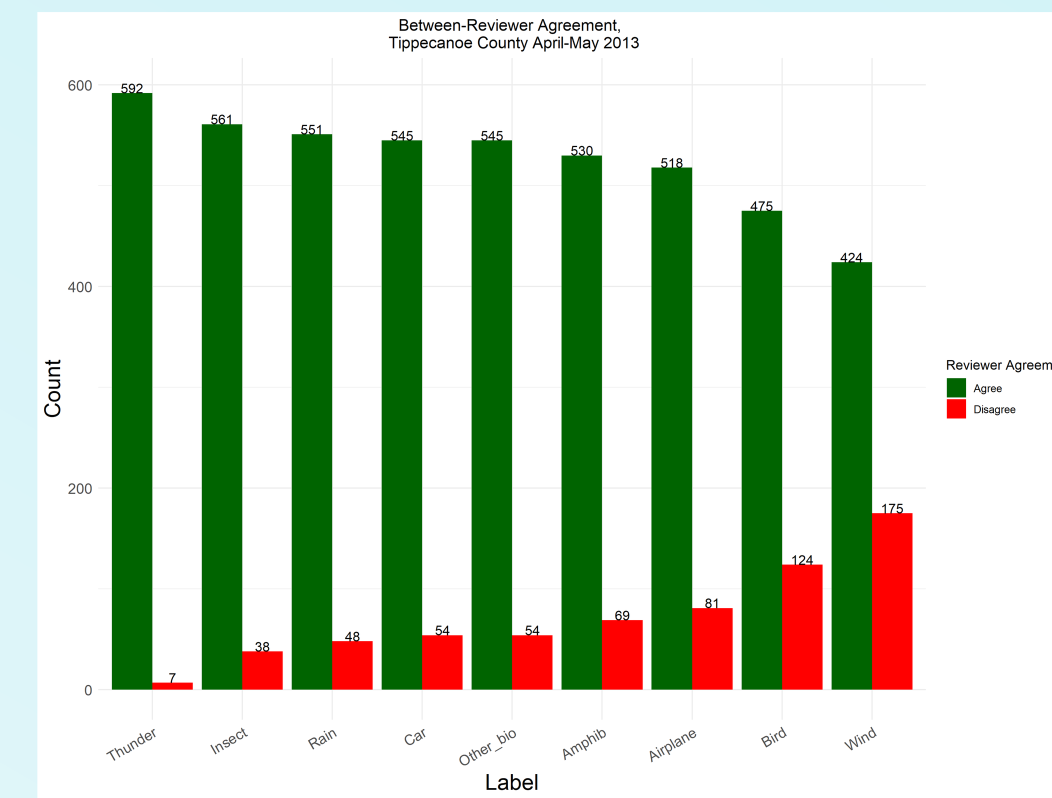

Figure 4: Comparison of Labels Between Reviewers

- A comparison of labels showed that all **class labels** were agreed upon by the two reviewers with 85% agreement or higher.
- The reliability scores for **purity** had a 62% agreement.
- The **confidence** level of the taxonomic class label for both reviewers was 85%.

In the CNN test, we performed 10 epochs with the highest accuracy rating in an epoch reaching 65%, higher than the average chance rate of 50%.

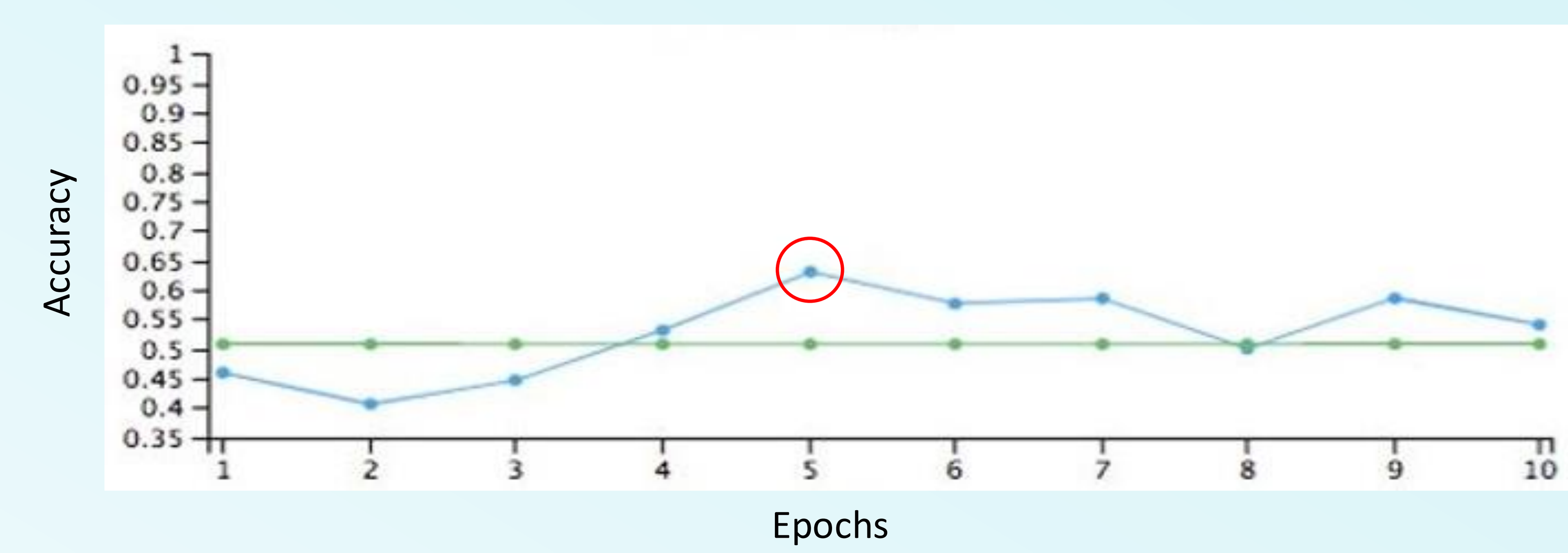

Figure 5: This graph shows accuracy vs epochs (number of passes through the entire training dataset). The green line is the performance threshold and the blue line is tested data. There is a peak at 65% accuracy in epoch 5.

Discussion/Conclusion

Soundscape ecology, recognizes that sounds in the environment could help us understand and track changes in biodiversity trends. Acoustic monitoring results in a massive number of sound files that require manual labeling, which is time intensive to manually tag. With further TensorFlow training, we will be able to tag and sort data in a fraction of the current time it takes. Because we only used a subset of unbalanced data for training, we will improve the model by adding more data in the training dataset. Overall, this study shows that using spectrograms for automatic labeling in a convolutional neural network could benefit ecologists.