Chirp Checker

A classification model for differentiating between groups of singing insects.

Robert Cass, Caleb Fong, Butovens Mede, Andrew Merwin, Calvin Yost-Wolff, Yang Yang

Overview

The nocturnal soundscapes of late summer and autumn are replete with the familiar chirps, trills, and buzzes of singing insects. But these cryptic performers often remain anonymous and underappreciated. In light of recently documented declines in insect abundance and raising concern for the sustainability of the ecosystem services they provide, spreading knowledge and appreciation for singing insects may be a powerful tool towards broader political support for efforts that protect insects and their habitats.

The goal of this project is to build a model to coarsely categorize sound files as crickets, katydids, or cicadas, as well as 15 families/subfamilies. In the future, a more sophisticated version of this model could be applied to filter large volumes of passively recorded audio from ecological studies of insects and to power phone apps that identify insect songs to the species level. Further, the same modeling approach could be used for stakeholders interested in classifying other types of audio recordings.

Insect Song Dataset and Feature Extraction

We requested access to 13,462 labeled insect audio files through the Macaulay Library of Natural Sounds at Cornell University–approximately 160 GB worth of files. We first cut spoken catalog numbers from the files. Then, to facilitate local processing, audio files were all trimmed to a maximum of 20 sec in length. We further filtered files using voice activity detection (package Silero) to those files that did not contain human speech and which were longer than half a second in length. We were left with a total of 5,899 files.

We extracted distinct features for models using 1D and 2D features. For models using 1D features, K-Nearest Neighbors and Support Vector Machine models, we took the mean and variance of the first 40 Mel Frequency Cepstral Coefficients (MFCCs) for both the entire file (up to 20 seconds in length) and the half-second of audio around the loudest noise within the file. Further, we extracted the highest magnitude frequency (peak frequency), the mean frequency, the frequency with the highest variance, and predominant frequency (see READMe for details). We therefore had a total of 164 features.

For our convolutional neural network model we used a few different feature sets: the first 40 MFCCs for every time stamp for the first 5 seconds(40 x 216 matrix), the first 40 MFCCs for the loudest half second in the first 5 seconds (40 x 22 matrix), average and variance of the MFCC matrix from the half second samples (80 x 1 vector), and spectrogram of the first 5 seconds (257 x 862 matrix).

For all models, MFCCs and frequencies below the range of singing insects (800 hz and below) were not considered when extracting MFCCs and frequency parameters to reduce the influence of background noise.

Modeling Approach and Performance

We used four general modeling frameworks to classify insect singers, including partial least squares discriminant analysis, k-nearest neighbors, support vector classifiers, and convolutional neural networks. To assess the performance of each model, we used the accuracy of predictions for our three groups of interest (cricket, katydid, and cicadas) as well as lower organismal levels of classification (family/subfamily).

We found that a support vector classifier performed best on our 1D dataset, with an **accuracy of 91%** (vs. 67% baseline obtained from the most common sample) on test data when predicting broadly categorized insects as crickets, katydids, and cicadas. It performed less well (67% accuracy vs. 24% baseline) when predicting families/subfamilies. An accuracy of 91% was also achieved when training on family / subfamily labels but predicting only the broader three categories. The convolutional neural network reached a max of 90% when using the 5-second sample MFCCs when predicting crickets, katydids, and cicadas and a 57% when predicting families/subfamilies. The 57% accuracy corresponds to an 86% accuracy after converting the family/subfamily labels to the broader three categories.

Future Work

Access to more audio files of singing insects would likely improve the training and accuracy of our models, especially for the classification of cicadas. Eventually, enough audio files for each species may allow us to differentiate to the species level. Further, files with less background noise would likely improve model training. Although, as an alternative, we may be able to employ noise reduction techniques to increase the signal to noise ratio.

Currently our model only classifies audio known to be from insects. A critical next step would be to develop an algorithm to detect insect sounds from among other background noises.

Eventually, more sophisticated versions of our models could be integrated with a user interface for app development or to automate data processing of passive acoustic monitoring efforts.