AUTOMATED ROBUST ANURAN CLASSIFICATION BY EXTRACTING ELLIPTICAL FEATURE PAIRS FROM AUDIO SPECTROGRAMS

Marcello Tomasini[2], Katrina Smart[2], Ronaldo Menezes[2], Mark Bush[1], and Eraldo Ribeiro[2]

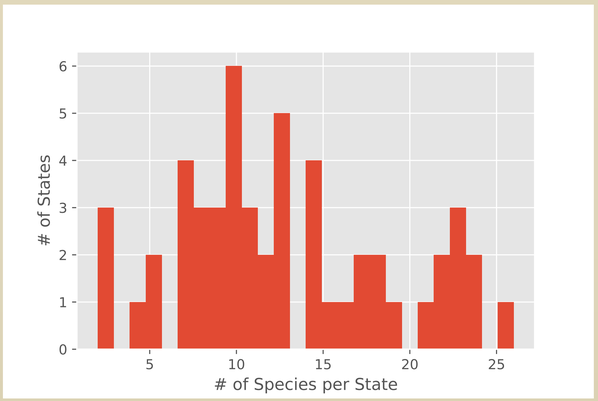
[1]Department of Biological Sciences, [2]School of Computing
Florida Institute of Technology, USA

Abstract

Ecologists can assess the health of wetlands by monitoring populations of animals such as Anurans (i.e., frogs and toads), which are sensitive to habitat changes. But, surveying anurans requires trained experts to identify species from the animals' mating calls. This identification task can be streamlined by automation. To this end, we propose an automatic frog-call classification algorithm and a smartphone application that drastically simplify the monitoring of anuran populations.

The data

- 48 species represented by 736 calls from the 50 U.S. states, the District of Columbia, and the Virgin Islands. Calls last from 1 to 40 seconds.
- Recordings come from a mixture of web-based collections and field recording made in Florida by our team.

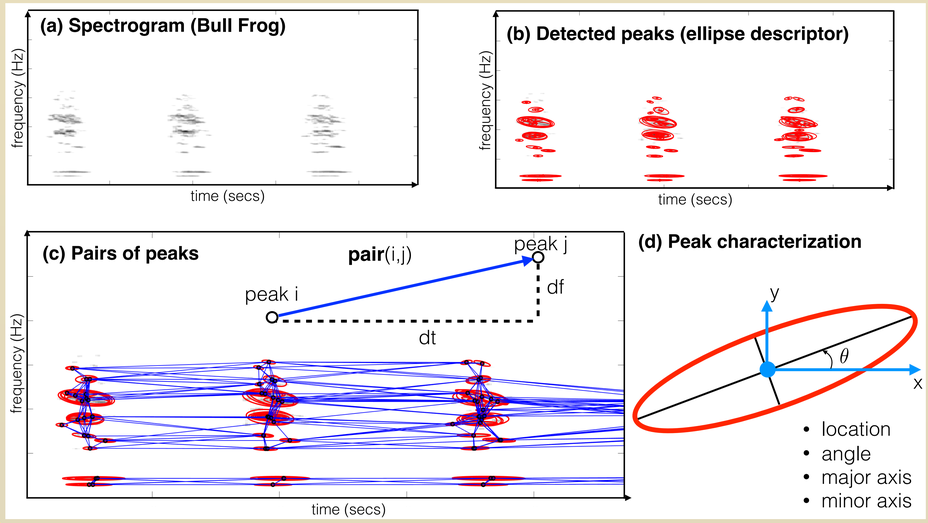


Experiments

- We trained and tested the classifiers using a 10-fold cross-validation. For the GPS-enabled classification, we ran a 10-fold cross validation per state/region.
- The average classification accuracy was computed as the average of the classification accuracy of the states/regions. For the no-GPS classifier, the average accuracy was the average accuracy of the 10-cross-validation runs.

Research supported in part by the National Science Foundation under **Grant No. 1152306**.

Call characterization and Classification



Characterization

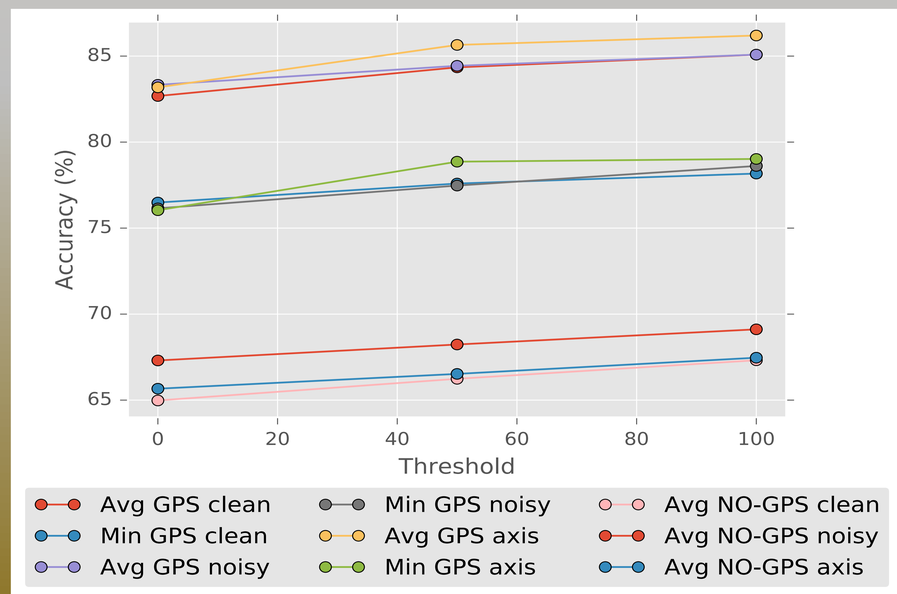
1. Convert call to spectrogram (short-time Fourier transform).
2. Filter spectrogram (keep the 20% highest values).
3. Detect spectral peaks.
4. Fit an ellipse around each peak.
5. Use the ellipse (location, angle, length of axes) to characterize peak.
6. Collect pairs of peaks that fall within a rectangular sliding window.
7. Generate feature vectors
   $f = (f_i, \lambda_i, \theta_i, f_j, \lambda_j, \theta_j, \Delta t, \Delta f)$

Voting-based classifier (k-Nearest Neighbor)

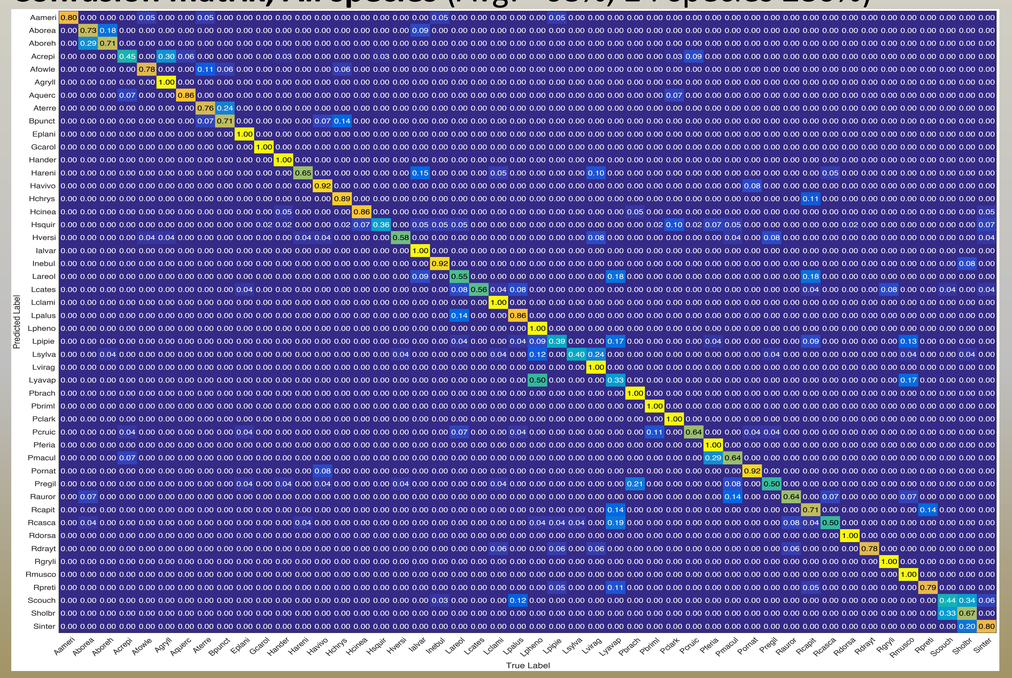
1. Train a state-*k*NN classifier for the species of each state/region (*k*=1).
2. Train an all-species-*k*NN classifier for all the species (*k*=1).
3. **If** GPS location is available **then**
   Use state-*k*NNs to associate feature pairs with species
   else
   Use all-species-*k*NN to associate feature pairs with species
   End
4. Select the species with the most votes.

Results

- The state-specific classification (i.e., GPS) reached a minimum accuracy of 79% and an average accuracy of 85%, greatly outperforming the single-classifier approach (i.e., no GPS) whose maximum accuracy was 70%.
- When multiple calls overlap, the loudest will be identified. Results will be best when one frog is clearly louder than all others.
- Trained citizen scientists correctly identify calling frogs about 80% of the time (Genet & Sargent, 2003).



Confusion matrix, All species (Avg: ~68%, 24 species ≥80%)



Conclusions

- The quality of the recording is critical. Phone microphones are adequate, but distant calls will not be identified.
- Adding more calls to the training library of each species would improve performance.
- Multiple species in a single record make reliable classification hard.
- The algorithm can be extended to support incremental learning.

References

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