Approaches to Multiple Concurrent Species Bird Song Recognition

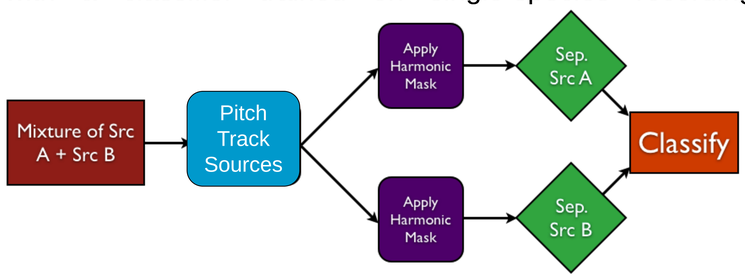
Jonathan Springer, Zhiyao Duan, Bryan Pardo
Northwestern University, Evanston, Illinois 60208, U.S.A.

Problem Statement

Most existing work in automatic bird species identification from audio recordings has assumed a single species is present. We explore two approaches to identifying multiple concurrent bird species from an audio recording .

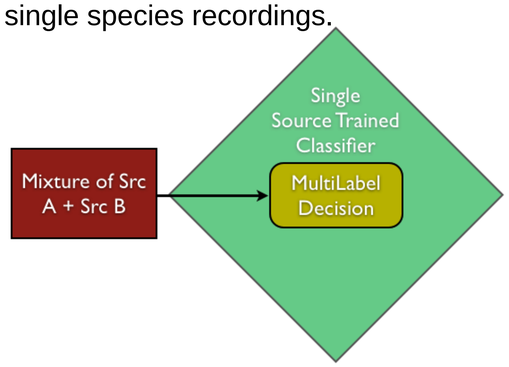
Source-Separation Approach

Idea 1: 1) estimate pitches contours of all birds; 2) separate sources by harmonic masking; 3) classify separated sources with a classifier trained on single-species recordings



Multi-label Classification Approach

Feed a mixture of multiple birds (no source separation) to a multi-label classifier chain [2] built from binary classifiers trained on single species recordings.



Dataset

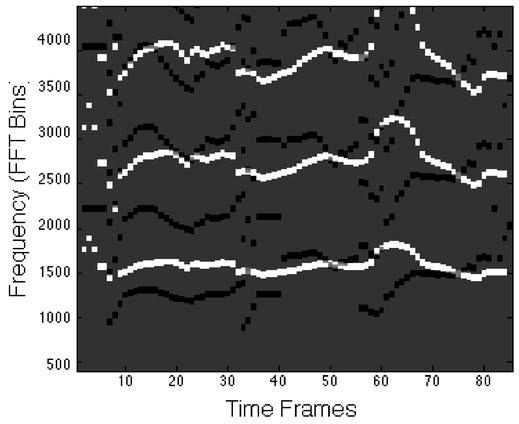
797 mixtures were created from individual bird song recordings of 10 different species [3].

- 111 2-species, 311 3-species, and 375 4-species mixtures;
- contains little noise;
- mixed with equal root-mean-square power.

YIN Pitch-Based Source Separation

For each bird, a time-frequency mask is built by allowing in energy at the harmonics of the fundamental frequency returned by the pitch tracker.

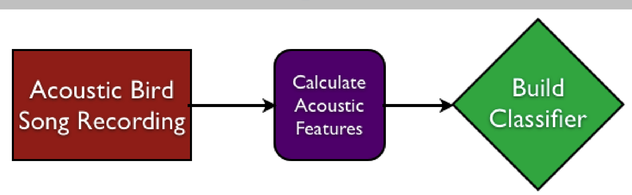
Apply this mask to the mixture spectrogram to separate out an individual bird's signal.



Example time-frequency mask of a recording of two birds. White pixels indicate bird 1. Black pixels are bird 2. Gray are background.

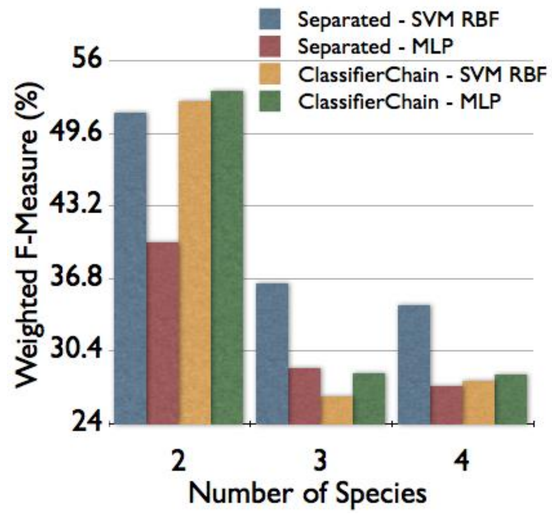
NOTE: In this study, we wanted to examine the best-case scenario for recognition from audio using pitch-track-based separation. Therefore, rather than apply a multi-pitch tracker to the mixture, pitch contours of each bird are tracked from the original single-bird recordings, prior to mixing. This was done with YIN [1].

Bird Song Classifier



Acoustic Features	Machine Learning Algorithm
First 12 MFCCs	Support Vector Machine using Radial Basis Kernel
Zero Crossing Rate	Artificial Neural Network - Multilayer Multilayer Perceptron
Spectral centroid, flux, and rolloff	

Experimental Results

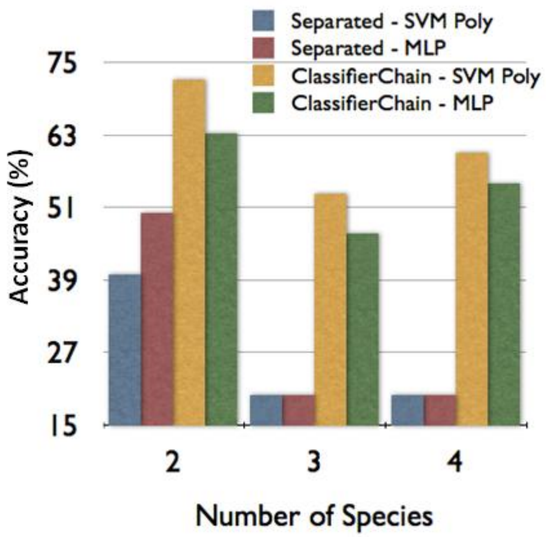


Frame-level

Average weighted F-measures for classification of the testing set of 100 mixtures of different (2,3, or 4) numbers of species in a file.

File-level

Accuracy (i.e. percentage of correctly predicted species over all present species) obtained by taking majority vote of frame-level predictions for each file. Averaged over 100 files for each polyphony.



References

[1] Cheveigné, A.; Kawahara, H. "Yin, a fundamental frequency estimator for speech and music," J. Acoust. Soc. Amer., vol. 111, pp. 1917–1930, 2002.

[2] Read, J.; Pfahringer, B.; Holmes, G.; Frank, E. Classifier Chains for Multi-label Classification. Machine Learning Journal. Springer. Vol. 85(3), pp 333-359. (May 2011).

[3] Cornell Lab of Ornithology – Macaulay Library, http://birds.cornell.edu/MacaulayLibrary/