Deductive Refinement of Species Labelling in Weakly Labelled Birdsong Recordings

Veronica Morfi and Dan Stowell
Machine Listening Lab, Centre for Digital Music (C4DM), Queen Mary University of London, UK

Abstract

Automated species identification is a challenging task due to the complexity of bird song, the noise present in most habitats, and the simultaneous song that occurs in many bird communities. A lot of methods on bird species classification exist, but more work is needed to address the problem of identifying all species and the exact times of their vocalizations.

We propose a method that given a weakly labelled dataset:

- **Detects all bird vocalisations**
- **Classifies them to the proper bird species**

Problem

Wildlife bird sound scenes:

- Multiple birds
- Multiple species
- Overlapping bird vocalisations
- Noisy environments

We will use datasets with weakly labelled recordings (we know the species vocalising but not their exact vocalisations) in order to detect and classify each bird vocalisation.

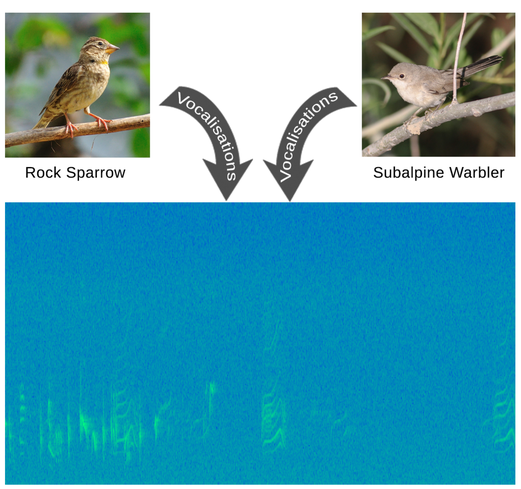

Figure 1: Field recording

Method

Segmentation-Detection

- Unsupervised extraction of vocalisation segments
- Event detection paradigm
- Clears noise
- Reduces noise generated segments
- Detects coordinates of vocalisations

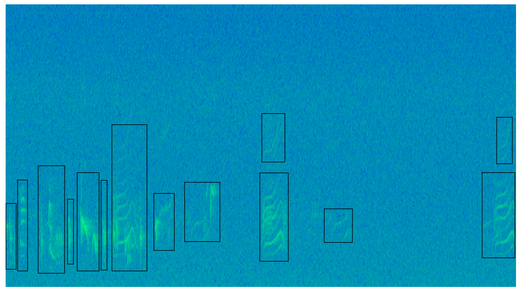

Figure 2: Results of Segmentation-Detection

Classification

- Segments deriving from the Segmentation-Detection step
- No explicit training phase
- Use of weakly labelled recordings
- Deduction of possible label list for each segment
- Performance increases as the number of recordings per each species increases

This process is implemented in three different procedures, that are applied to the recording in order:

1. First-Pass
   - Create groups of recordings containing the weak labels (groups divided by number of weak labels present)
   - Use normalised correlation to find matches between a segment and the recordings
   - Label(s) of best match assigned to segment
   - Segments with no match get label MNF
2. Second-pass
   - Solves unclassified segments problem when there are unallocated labels in a recording by assigning these labels to the MNF segments
3. Third-Pass
   - Solves unallocated labels problem when more than one segments have the same label

Examples

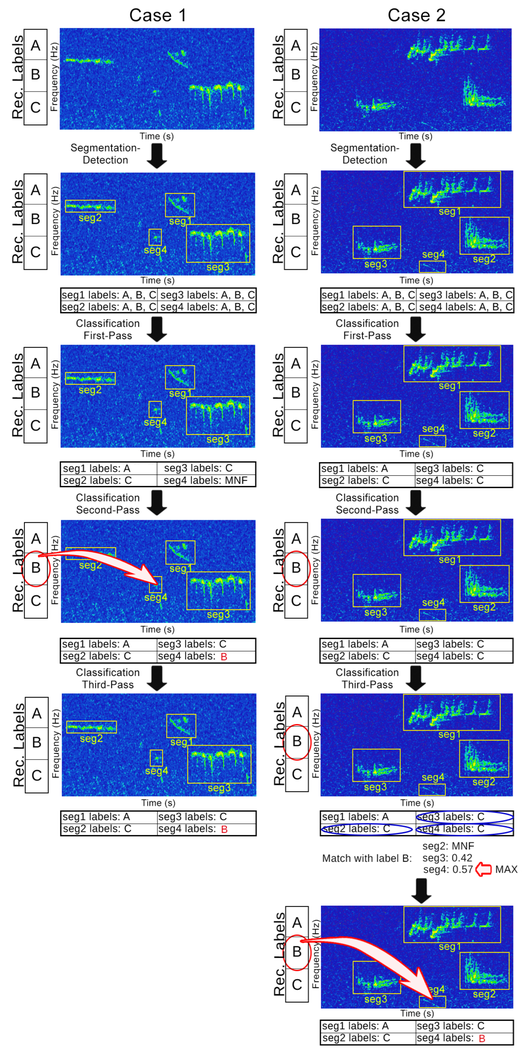

Figure 3: Two different classification scenarios. Case 1: Unassigned label and MNF segment. Case 2: Unassigned label

Contact Information

Email:
g.v.morfi@qmul.ac.uk

Evaluation

Synthetic dataset D:

- single labelled recordings of NIPS4B
- 51 out of 87 labels with single labelled recordings
- 50 synthetic recordings
- duration of 5 seconds
- zero labelled NIPS4B recording as base
- 2-4 randomly picked labels per synthetic recording
- total of 138 segments (mean of 2.76 segments)
- remainder of NIPS4B dataset used for classification

Table 1: Classification Results for D

	Correct	Wrong	Unknown
Chance	36.2%	63.8%	—
First-Pass	68.9%	24.6%	6.5%
Second-Pass	71%	24.6%	4.4%
Third-Pass	75.4%	20.2%	4.4%

Synthetic dataset D_{1000}:

- 50 recordings
- 2-4 labels per recording
- segment size $\geq$ 1000 pixels
- total of 152 segments (mean of 3.04 segments)

Table 2: Classification Results for D_{1000}

	Correct	Wrong	Unknown
Chance	32.89%	67.11%	—
First-Pass	66.5%	21.7%	11.8%
Second-Pass	71%	22.4%	6.6%
Third-Pass	74.3%	19.1%	6.6%

Conclusions

Our method is used to fully annotate weakly labelled recordings at a unit level.
Segmentation-Detection: fully automatic extraction of vocalisations
Classification: deductive refinement of segment labels using normalised correlation and weak labels