Acoustic Pipeline

Classifier Technical Specifications: Eurasian Curlew

1. Introduction

The BTO Acoustic Pipeline's **Eurasian Curlew** classifier (version 1) detects display and alarm calls of Eurasian Curlew *Numenius arquata*. This document provides technical notes on the derivation of the classifier, its use and how to interpret results. The rest of this document is arranged into five parts:

- Training data sample sizes
- Precision and Recall statistics on withheld data
- False positive rates on independent data
- Known issues

2. Training data sample sizes

This classifier is trained on strongly labelled audio recordings. The Curlew call types data were collected and labelled by David Jarrett. Data for other species and ambient sounds were collected, collated and labelled by BTO, with some additional recordings from xeno-canto. We are grateful to our collaborators and the sound recordists who share recordings via xeno-canto. The following table gives the number of audio samples used to train and evaluate the classifier. The Background class encompasses ambient and anthropogenic sounds plus non-target wildlife (i.e. other bird species, mammals, amphibians and insects). During training the classifier includes calls for four other wader species. These are not currently exported but merely used for training purposes for refining Curlew detections. In time we aim to include call-level reporting for other waders too.

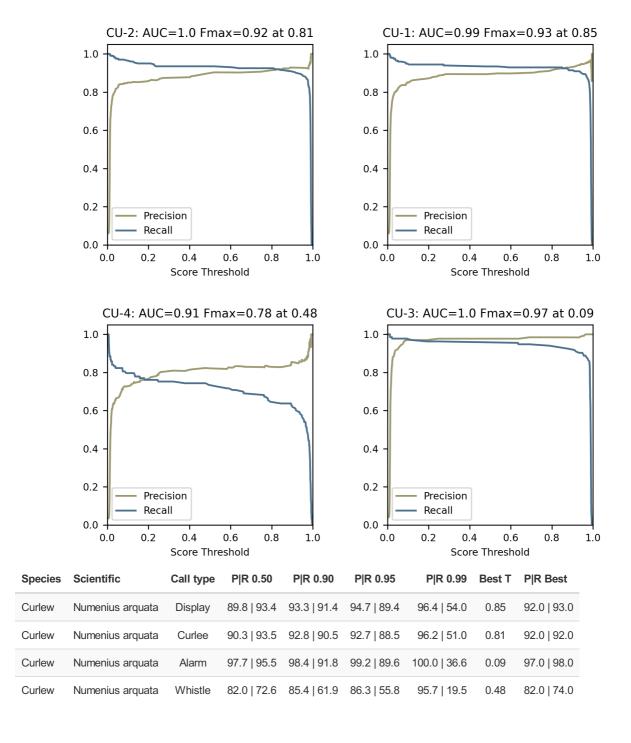| Class	Train sample size	Evaluation sample size
Oystercatcher - All	1500	200
Lapwing - All	1500	151
Curlew - Display	1500	198
Curlew - Curlee	1500	200
Curlew - Alarm	1500	134
Curlew - Whistle	1500	113
Snipe - All	1500	200
Redshank - All	1500	200
Background	15000	2000

3. Precision and Recall

Classifier performance is typically evaluated using the metrics *Precision* and *Recall*:

- **Precision** is the % of detections returned by a classifier that truly are of that species. If the classifier thinks 100

clips contain Curlew alarm calls, Precision is the percentage of these that actually do contain Curlew alarm calls.

- **Recall** is the percentage of true instances that are detected. For example, if 100 clips contain Curlew alarm calls, Recall is the percentage of these that are found by the classifier.

These metrics are *threshold dependent*. If we say that all classifier scores greater than 0.5 constitute a detection we will get different Precision and Recall values than if we use a more stringent score threshold of 0.9. Figure 1 shows how Precision and Recall vary with score threshold for each call type: CU-1 = Display, CU-2 = Curlee, CU-3 = Alarm, CU-4 = Whistle. This figure is based on application of the classifier to withheld training data. The subsequent table provides Precision and Recall values for commonly used score thresholds. Additionally, *Best T* gives the threshold at which the F-score is maximised, which is a way of optimising both Precision and Recall. The Precision and Recall statistics for that threshold (*P|R Best*) are also shown. Detections exported from the Acoustic Pipeline by default use a conservative threshold of 0.9 to increase precision.

Species	Scientific	Call type	P|R 0.50	P|R 0.90	P|R 0.95	P|R 0.99	Best T	P|R Best
Curlew	Numenius arquata	Display	89.8 | 93.4	93.3 | 91.4	94.7 | 89.4	96.4 | 54.0	0.85	92.0 | 93.0
Curlew	Numenius arquata	Curlee	90.3 | 93.5	92.8 | 90.5	92.7 | 88.5	96.2 | 51.0	0.81	92.0 | 92.0
Curlew	Numenius arquata	Alarm	97.7 | 95.5	98.4 | 91.8	99.2 | 89.6	100.0 | 36.6	0.09	97.0 | 98.0
Curlew	Numenius arquata	Whistle	82.0 | 72.6	85.4 | 61.9	86.3 | 55.8	95.7 | 19.5	0.48	82.0 | 74.0

4. False positive rates

The false positive rate indicates how often the classifier suggests a species is present when it is not. For this test we

use a benchmark dataset of 21,000 ambient sound clips that have been manually checked to confirm they contain no bird records. We run the classifier against this dataset and summarise the percentage of clips that are falsely assigned to a species. As for performance metrics, this measure is threshold dependent, with typically fewer errors when a more stringent score threshold is applied. False positive rates for each Curlew call type are shown below. Note that this is a simple test: in reality false positive rates may be higher in natural soundscapes, for example, where a distant call of one species is mistaken for another species.

Species	Scientific	Call type	0.50	0.90	0.95	0.99	Best
Curlew	Numenius arquata	Display	0.00000	0.00000	0.00000	0.00000	0.00000
Curlew	Numenius arquata	Curlee	0.00000	0.00000	0.00000	0.00000	0.00000
Curlew	Numenius arquata	Alarm	0.00000	0.00000	0.00000	0.00000	0.00000
Curlew	Numenius arquata	Whistle	0.00000	0.00000	0.00000	0.00000	0.00000

5. Known issues

This classifier is still in active development and some false positives are to be expected whilst we build the training data to reflect different environments. Nevertheless, performance of this classifier is generally good.

6. Acknowledgements

This classifier relies heavily on training data collected and labelled by David Jarrett.