Automatic Bird Sound Detection in Audio by Spectral Features

Yi-Chin Tseng

Faculty of Forestry, University of British Columbia (UBC), Canada

Background

Current biological monitoring research highly relies on audio detection of bird sounds since birds are much easier to be heard than seen in the field. Avian bioacoustics research was revolutionized by the introduction of digital audio recorder but the increasingly large datasets make manual inspection infeasible. In this study, binary logistic regression will be applied to predict the presence/absence of bird sound in each sound clip. This simple estimation has practical relevance since the occupancy modelling framework in statistical ecology uses exactly this binary information (i.e., probability that the abundance in a site is zero or not).

Keywords: bird sound, detection, binary, logistic regression, occupancy model

Method and Material

Workflow

The workflow of an automatic bird sound detection system is shown in Fig.1. Recordings will be separated into two groups, namely training recordings and testing recordings. In the model construction stage, binary logistic regression will be applied to reveal the relationship between spectral features (continuous predictor variables) and the given presence/absence of bird sound (binary dependent variable). In the model evaluation stage, testing recordings will be used to evaluate the accuracy of the model.

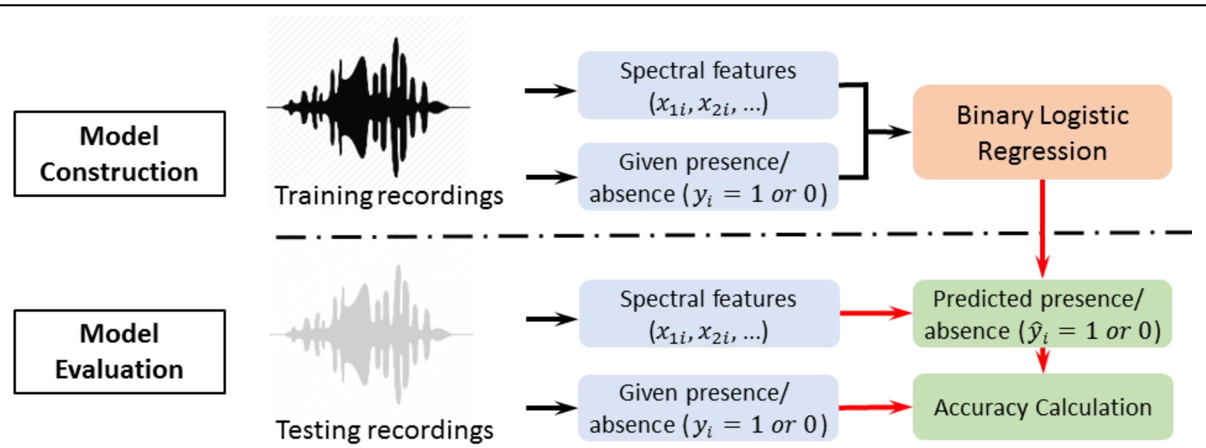

Fig.1. The block diagram of the automatic bird sound detection process. Binary logistic model is built based on training recordings. The accuracy of the regression is made by comparison between the predicted presence/absence and the given presence/absence.

*Bird Audio Detection challenge: http://machine-listening.eecs.qmul.ac.uk/bird-audio-detection-challenge/

Datasets

The datasets from an IEEE Bird Audio Detection challenge* will be used in this study. Around 16,000 ten-second recordings are provided with each of them annotated with the presence/absence of bird sound. The recordings include weather noise, traffic noise, human speech, and even mammal and insect noise. All recordings are formatted into 44100 Hz and mono PCM WAV file type. Recordings will be randomly grouped into two groups for training and testing purpose, respectively.

Statistical Approaches

1. Spectral Features
   Most of the birds have their sound frequency ranging from 1000 to 8000 Hz, which distinguishing them from other daily sounds such as human speech (85-255 Hz) and traffic noise (63-1000 Hz). Based on this significant difference in sound frequency, the 95%, 97%, and 99% quantile of frequency distribution of a recording will be extracted as its sound features, as shown in Fig. 2.
2. Binary Logistic Regression
   Binary logistic regression will be used to examine the relationship between the sound features (continuous predictor variables) and the presence/absence of bird sound (binary dependent variable). The probability of presence event will be explained as a function of predictor variables, as shown in Eq.1. Parameters of the regression is fitted by maximum likelihood decision (MLD) with likelihood function shown in Eq.2. Statistical computation will be performed using R language.

$$\pi_i = \Pr(Y_i = 1 | X_i = x_i) = \frac{\exp(\beta_0+\beta_1 x_1+\cdots+\beta_{p-1}x_{p-1})}{1+exp(\beta_0+\beta_1 x_1+\cdots+\beta_{p-1}x_{p-1})} \tag{1}$$

$$L(\beta_i) = \prod_{i=1}^{n} (\pi_i)^{y_i}(1-\pi_i)^{1-y_i} \tag{2}$$

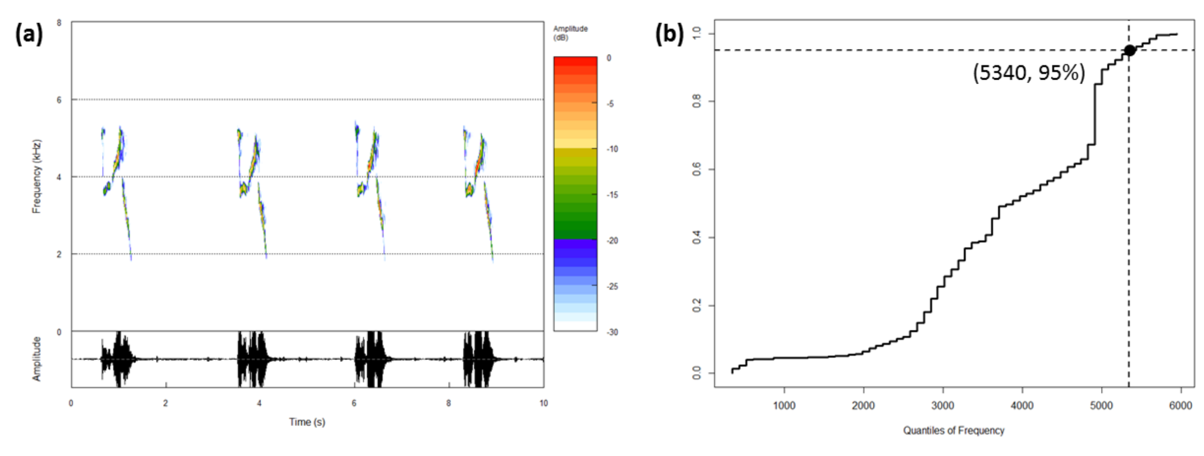

Fig.2. Diagrams demonstrating the extraction of sound spectral feature. (a) The spectrogram of a recording is derived by applying Fast Fourier Transform. (b) Empirical cumulative distribution showing the quantiles of frequency. Here the 95% is marked.

*Bird Audio Detection challenge: http://machine-listening.eecs.qmul.ac.uk/bird-audio-detection-challenge/