Ecological Indicators 82 (2017) 13–22

Contents lists available at ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind

An intelligent system for estimating frog community calling activity and species richness

Jie Xie[a,b,*], Michael Towsey[a], Mingying Zhu[c], Jinglan Zhang[a], Paul Roe[a]

[a] Science and Engineering Faculty, Queensland University of Technology, QLD, Australia
[b] Department of Electrical and Computer Engineering, University of Waterloo, Ontario, Canada
[c] Department of Economics, University of Ottawa, Ontario, Canada

ARTICLE INFO

Keywords:
Frog community calling activity
Frog species richness
Acoustic event detection
Multi-label learning

ABSTRACT

Over the past decade, dramatic declines in frog populations have been noticed worldwide. To examine this decline, monitoring frogs is becoming increasingly important. Compared to traditional field survey methods, recent advances in acoustic sensor technology have greatly extended spatial and temporal scales for monitoring animal populations. In this paper, we examine the problem of monitoring frog populations by analysing acoustic sensor data, where the population is reflected by community calling activity and species richness. Specifically, a novel acoustic event detection (AED) algorithm is first proposed to filter out those recordings without frog calls. Then, multi-label learning is used to classify each individual recording with six acoustic features: linear predictive coding coefficients, Mel-frequency cepstral coefficients, linear-frequency cepstral coefficients, acoustic complexity index, acoustic diversity index, and acoustic evenness index. Next, frog community calling activity and species richness are estimated by accumulating the results of AED and multi-label learning, respectively. Finally, ordinary least squares regression (OLS) is conducted to reveal the relationship between frog populations (frog calling activity and species richness) and weather variables (maximum temperature and rainfall). Experimental results demonstrate that our proposed intelligent system can significantly facilitate the effort to estimate frog community calling activity and species richness with comparable accuracies. The statistical results of OLS indicate that rainfall pattern has a lagged impact on frog community calling activity (significant in the first day after rainy day) and species richness (significant in the fourth day after rainy day). Temperature is shown to affect species richness but is less likely to change calling activity.

1. Introduction

As a widely distributed taxonomic group, frogs are considered excellent indicators of biodiversity. However, frog populations have been experiencing dramatic declines over the past decade due to habitat loss, climate change, and invasive species (Li et al., 2013; Quilodrán et al., 2015; Gillespie et al., 2015). Therefore, long-term monitoring of frog populations is becoming increasingly important to optimise conservation policy. Previous work has demonstrated that vocalising animals can produce species-specific spectral and temporal communication patterns to minimise acoustic interference (Römer et al., 1989; Romer, 1993; Schmidt and Balakrishnan, 2015). Acoustic data, which can reflect spectral and temporal communication patterns of recorded animals, is thus widely explored to monitor animal populations. Compared to traditional field survey methods, recent use of acoustic sensors can monitor frog populations over larger spatial and temporal scales (Haselmayer and Quinn, 2000; Acevedo and Villanueva-Rivera, 2006), but generate large volumes of acoustic data. Therefore, enabling automating species identification in acoustic data has become necessary.

Various methods have been proposed to detect and classify frog calls (Huang et al., 2009b, 2014; Han et al., 2011; Colonna et al., 2015; Xie et al., 2016b; Noda et al., 2016). However, these methods have two limitations in common: (1) each individual recording is assumed to have only one frog species; (2) the time span of analysed recordings is short and often less than one hour. Therefore, more work needs to be conducted on acoustic monitoring of frogs over the long term. Recently, few studies explored acoustic monitoring of frogs over the long term. Canavero et al. (2008) investigated the relationship between calling activity of anuran assemblages and seasonal changes of weather variables, such as temperature and rainfall. However, both frog calling activity and species richness had to be manually calculated, which was time-consuming. Ospina et al. (2013) first detected an individual

* Corresponding author at: Science and Engineering Faculty, Queensland University of Technology, QLD, Australia.
E-mail addresses: xiej8734@gmail.com (J. Xie), m.towsey@qut.edu.au (M. Towsey), mzhu089@uottawa.ca (M. Zhu), jinglan.zhang@qut.edu.au (J. Zhang), p.roe@qut.edu.au (P. Roe).

http://dx.doi.org/10.1016/j.ecolind.2017.06.015
Received 16 November 2016; Received in revised form 3 June 2017; Accepted 10 June 2017
Available online 27 June 2017
1470-160X/

acoustic event (e.g., a frog call) by analysing a time-frequency matrix (i.e., spectrogram). Time and frequency variables based on the acoustic event were then used to build a species identification model with a Hidden Markov Model. However, their method focused only on species richness analysis rather than calling activity. Xie et al. (2015b) used acoustic indices to detect frog calling activity. Four indices were selectively combined to detect frog calling activity with a Gaussian Mixture Model: spectral peak track index, harmonic structure index, oscillation structure index, and Shannon entropy index. However, the explored correlation between frog calling activity and climate information was weak due to the short-term analysis (12 h).

Unlike prior work in the automated acoustic monitoring of frog populations, both frog community calling activity and species richness are investigated in this study. The relationship between frog populations and weather variables is also examined over three months (breeding season for frogs in north-east Australia). Specifically, short-time Fourier transform (STFT) is first used to generate a spectrogram for each 10-s recording, which is sampled from every 10-min recording. A novel acoustic event detection (AED) algorithm is then applied to the spectrogram image to detect individual acoustic events. *Area*, *power*, and *averaged dominant frequency*, which are calculated based on each acoustic event, are employed to filter out those recordings without frog calls using a random forest classifier. Next, frog community calling activity can be estimated based on the accumulated *power* of acoustic events. For those recordings with frog calls, multi-label learning is used to classify them with six acoustic features: linear predictive coding coefficients (LPCs), Mel-frequency cepstral coefficients (MFCCs), Linear-frequency cepstral coefficients (LFCCs), acoustic complexity index (ACI), acoustic diversity index (ADI), and acoustic evenness index (AEI). Based on the multi-label classification results, frog species richness can be estimated by averaging the number of species over all recordings per day. Finally, ordinary least squares (OLS) is used to reflect the relationship between community calling activity or species richness[1] and weather variables (temperature and rainfall).

Experimental results show that our proposed system can automatically estimate both frog community calling activity and species richness over three months with high performance. Also, OLS demonstrates that frog community calling activity is highly related to rainfall of one day lagged, but not associated with temperature. Species richness is highly related to rainfall of four days lagged and temperature.

2. Materials and methods

Our frog call classification system consists of two steps: binary classification and multi-label classification. After binary classification, frog community calling activity and species richness can be estimated using those recordings with frogs (Fig. 1).

2.1. Study site

Audio recordings were collected from three sites in Queensland, Australia using a battery-powered acoustic sensor (stored in a weatherproof metal box) with an external microphone: *Kiyomi dam* (19°23′S, 146°28′E), *Stony Creek dam* (19°23′S, 146°24′E), and *BG Creek dam* (19°27′S, 146°23′E) (Fig. 3), where frogs were likely to be heard. The longest distance between two sites is around 11.4 km, while the shortest distance is around 6.9 km. Recordings collected from February 2014 to April 2014 were selected because it was the breeding season for frogs in Queensland when male frogs made calls to attract females. Nine frog species from five genera, which were widely distributed in Queensland, were selected as the research targets in this study (Fig. 3).

[1] Community calling activity is reflected by the accumulated power of all individual frog calls in a 10-s recording; species richness denotes the number of different frog species in a 10-s recording.

2.2. Data description

All collected recordings were monaural, sampled at 16 kHz and saved in MP3 format. Each individual recording started around 7 pm and finished around 7 am and had a duration of about 12 h since frogs tended to make calls at night. For each 12-h recording, first 10 s from every 10-min recording were sampled; however, due to loss of original data, the number of recordings were 4170, 4980, and 1544 10-s recordings for *Kiyomi dam*, *Stony Creek dam* and *BG Creek dam*, respectively. A representative sample of 342 10-s recordings was constructed for this study. The ground truth of those 342 10-s recordings was generated by a frog expert,[2] who manually tagged each recording with frog species. Fig. 2 indicates the distribution of frog species of sampled recordings, and each recording has an average of 2.57 frog species. This representative sample is used to train and test the multi-label classification model. Once the model is built, it will be applied to those recordings over three months for monitoring frog species richness. Besides, a constructed dataset of 30 recordings with frogs and 30 recordings without frogs is used to evaluate our proposed single-label classification model. Since this study explored the relation between frog community calling activity/species richness and weather variables, weather variables are obtained from one public website (http://www.bom.gov.au/?ref=hdr). Using this website, both temperature and rainfall data are collected through the closest weather station of each data collecting site.

2.3. Acoustic event detection

The aims of acoustic event detection (AED) are to (1) filter out those recordings without frog calls; (2) estimate frog community calling activity based on the accumulated power of detected acoustic events. For each sampled 10-s recording, the regions of interest were first identified using spectrogram image, which was generated by applying STFT to acoustic data. Here the window size and overlap were set at 512 samples and 50%, respectively, which were based on the consideration of computational efficiency and the resolution of spectrogram image. The AED process consists of four main steps:

(1) Gaussian filtering

To reduce the graininess and remove small gaps within each individual acoustic event, a spectrogram image is convolved with a Gaussian kernel. The size of the Gaussian kernel was empirically set at 15 after consideration of the trade-off between removing the background graininess and blurring different acoustic events (Towsey et al., 2012).

(2) Spectral subtraction

Gaussian filtering can successfully reduce the graininess, but some noises, such as wind, insect, motor engine that cover the whole recording cannot be addressed using Gaussian filtering. A modified spectral subtraction is then used to de-noise the spectrogram. A detailed description of this algorithm can be found in our previous paper (Xie et al., 2015c).

(3) Otsu thresholding

Transforming the noise reduced spectrogram into its binary representation is the next step for detecting acoustic events. Here an adaptive thresholding method named *Otsu thresholding* (Otsu, 1975) is employed to find an optimal threshold, which is then used to conduct the transformation.

[2] http://www.reptileecologylab.com/kiyomi-yasumiba.html.