Ecological Indicators 119 (2020) 106852

Contents lists available at ScienceDirect

Ecological Indicators

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

Automated species identification of frog choruses in environmental recordings using acoustic indices

Sheryn Brodie[a,*], Slade Allen-Ankins[a], Michael Towsey[b], Paul Roe[b], Lin Schwarzkopf[a]

[a] College of Science and Engineering, James Cook University, Townsville, Qld 4817, Australia

[b] QUT Ecoacoustics Research Group, Science and Engineering Faculty, Queensland University of Technology, Brisbane, Qld 4000, Australia

ARTICLE INFO

Keywords:
Acoustic indices
Acoustic monitoring
Automated species detection
Bioacoustics
Anurans
Frog chorusing

ABSTRACT

Acoustic monitoring provides opportunities for scaling up bioacoustic study of vocal animals to greater temporal and spatial scales. However, the large amounts of audio that can be easily and efficiently collected necessitates automated methods of analysis to extract useful ecological data. Acoustic indices have been used in spectrographic visualisation of long environmental recordings to successfully identify many biological sounds from their acoustic patterns and features. In particular, the choruses of several frog species are conspicuous in these spectrogram images which suggests that acoustic indices may be useful for detecting species in automated sound classification algorithms. The aim of this study was to investigate the use of acoustic indices as predictors in classification models for automated identification of frog species in environmental sound recordings from breeding habitats in north Queensland, Australia. Three types of classification models (random forests, support vector machines and gradient boosting) were trained and validated on a data set of 3274 1-minute audio segments labelled for the presence or absence of calling of 12 target frog species, and a feature set of 11 acoustic indices calculated on frequency bins of bandwidth 43.1 Hz. Classification performance was high for all 12 target species on the validation data set held out from the labelled training data (precision range 0.90–1.00 and recall range 0.83–0.99). However, performance declined for most target species when predicting frog calling on a further test data set taken from unseen recordings from the same sites. Best prediction results on the test data were achieved for species with the most training data, indicating accuracy may be improved by increasing training data, and this method is best suited to predicting chorusing of common species.

1. Introduction

Monitoring of animal populations and communities is fundamental for management and conservation of biodiversity. Monitoring threatened species is necessary to assess if recovery and management efforts are effective, and monitoring ecological communities provides baseline data that enables detection of changes in populations and ecosystem health. However, there is significant cost in collecting data for species monitoring programs, which take considerable time and effort. Surveys must be repeated over sufficient time and locations, so that changes in species abundance, distribution or behaviour patterns can be detected (Field et al., 2007). Automated sensor technologies, such as GPS tracking devices, motion-sensor cameras and sound recorders that can continuously record data on animal presence, movement and behaviour allow researchers and conservation practitioners to increase the temporal and spatial scale of monitoring animal species. However, the increase in data collection brings new challenges in analysing big data sets to answer ecological questions.

Environmental monitoring using sound recorders has become a common method of monitoring vocal species (e.g. Aide et al., 2013; Hagens et al., 2018). Automated sound recording technology allows sampling over increased spatial and temporal scales, which can provide information to ecologists about species distributions, movement and migration patterns, and breeding phenology (Acevedo and Villanueva-Rivera, 2006; Campos-Cerqueira and Aide, 2016; Sanders and Mennill, 2014). As recording technology has improved, the amount of acoustic data that can be collected is less constrained by data storage or power limitations. Portable recorders can record continuously for weeks or even months, and permanent acoustic monitoring networks providing continuous, long term acoustic data are now feasible (Australian Acoustic Observatory, 2019). Long term acoustic monitoring can provide valuable data on common species distributions and behaviour patterns, which is necessary for detecting changes in populations (Frommolt and Tauchert, 2014). Because of the sheer volume of data,

[*] Corresponding author.
E-mail address: sheryn.brodie@my.jcu.edu.au (S. Brodie).

https://doi.org/10.1016/j.ecolind.2020.106852
Received 2 May 2020; Received in revised form 5 August 2020; Accepted 13 August 2020
Available online 28 August 2020

effective analytical tools are required to automate the extraction of useful ecological data from long-duration sound recordings.

Automated animal call recognition and species identification from sound recordings is an active and ongoing field of research. The conceptual approach to automated species call recognition is to develop computer algorithms that scan sound files, accurately detect target calls, and measure some features of the calls that can then be used as criteria to classify the calls to species. Many studies have demonstrated the feasibility of automated call recognisers achieving high classification performance on test data, but these studies are often aimed at testing the discriminative power of different classification methods or call features using libraries of short recording clips (e.g. Bedoya et al., 2014; Ganchev and Potamitis, 2007; Knight et al., 2019; Noda et al., 2016). The limitations of call recognisers have been highlighted in their failure to effectively scale to long-duration recordings that include higher levels of environmental noise (Crump and Houlahan, 2017; Priyadarshani et al., 2018; Waddle et al, 2009). The main technical difficulty limiting automated call recognition is that species diversity, call variability and noise all increase in longer duration recordings, which leads to loss of accuracy of call identification (Gibb et al., 2019; Priyadarshani et al., 2018). Additionally, while accurate call detection and identification is the fundamental goal underpinning automated recognisers, there are few study objectives for which identification of each and every vocalisation is necessary, but the level of correct identification required may be difficult to specify, or may be dependent on the goals of the monitoring.

The problem of finding useful and effective analytical methods for long-term recordings of the natural environment has led to recent work investigating the use of acoustic indices for detecting and identifying the vocal activity of species. Acoustic indices are numeric summaries of the energy distribution in a recording based on amplitude and spectral content (Sueur et al., 2014). Indices, such as the acoustic entropy index (Sueur et al., 2008) and the acoustic complexity index (Pieretti et al., 2011) were initially conceived and developed as summary metrics to very broadly characterise biodiversity and assess human disturbance at community and landscape scales. More recent work has used acoustic indices calculated on short time segments to develop a method of visualising the acoustic content of long-duration sound recordings (Towsey et al., 2014b). This visualisation tool combined three acoustic indices to generate a compressed 'false-colour' spectrogram of long-duration (up to 24 h) continuous recordings which highlighted the main acoustic features and events. Some features in the false-colour spectrograms highlighted the vocal activity of some species, in particular birds and frogs (Towsey et al., 2018b). The utility of this visualisation method in highlighting the calls of some species suggests that combinations of acoustic indices may be useful as predictive features for automated detection of those species. Several recent studies have demonstrated that acoustic indices can be used to detect single bird species in continuous recordings with high accuracy (Gan et al., 2018; Dema et al., 2018; Towsey et al., 2018b). Frog choruses are particularly conspicuous in the long-duration spectrograms, since acoustic indices effectively capture the acoustic patterns of the consistent and repetitive calls of chorusing frogs (Fig. 1), but the use of acoustic indices for automated detection of frog species in sound recordings has not been tested. Indraswari et al. (2018) showed values of the acoustic complexity index and the temporal entropy index could distinguish the calls of three frog species in 30-second recordings using ordination. The usefulness of acoustic indices as features in a classification model is therefore worthy of consideration for automation of frog species identification in environmental sound recordings.

The aim of this study was to investigate the potential use of acoustic indices as predictors in classification models for automated identification of frog species in environmental sound recordings. Specifically, we wanted to test if acoustic indices could be used to detect which species of frogs were calling in each minute of audio recorded in a tropical savanna environment where multiple species often call simultaneously in large choruses, generating large amounts of noise and call overlap. Data at this resolution on the calling of patterns of multiple frog species would be useful for studies of the temporal patterns of chorusing of frog populations, but also to study broad-scale acoustic interactions among species (i.e. nightly or seasonally). An accurate automated method of detecting species in environmental recordings would provide invaluable data for long-term monitoring of species, animal communities and biodiversity (Towsey et al., 2014a).

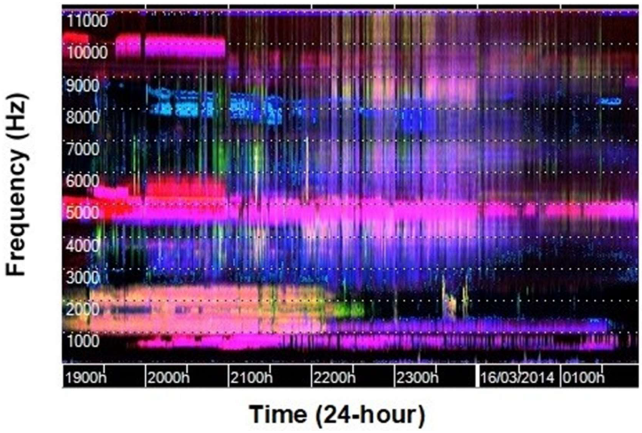

Fig. 1. Example false-colour spectrogram of a 7-hour recording used in this study. Horizontal dotted lines delineate 1000 Hz frequency intervals (0–11025 Hz). 1 pixel represents 1 min of audio and approximately 43 Hz frequency range. Colours derive from 3 acoustic indices mapped to the 3 colour channels (ACI – red; ENT – green; EVN – blue; refer Table 1). Biotic noises featured in this image are: Cane toad (*Rhinella marina*), pink < 1000 Hz; Northern laughing tree frog (*Litoria rothii*), pink/yellow 1000–2500 Hz; Eastern sedge frog (*Litoria fallax*), indigo/green/pink > 2000 Hz; and insects ~5000 Hz, 8000 Hz and 10000 Hz.

2. Materials and methods

2.1. Environmental recordings

Sound recordings were made at eight frog breeding sites in northern Queensland, Australia, from October 2012 to April 2014, as part of a study monitoring frog communities in this region. The study sites were waterbodies (artificial dams or natural creek empondments) at Townsville (19.332° S, 146.761° E) and Hervey Range (19.357° S, 146.454° E). Recordings were made in MP3 file format (128 kbps bit rate; 22.05 kHz sampling rate). Frogs in this region are typically only vocally active at night, and so recorders were set to record continuously from 1800 h each night to 0700 h the following morning. In total, 3965 continuous sound recordings of up to 13 h duration each were made for the monitoring study.

2.2. Audio processing and calculation of acoustic indices

Pre-processing of the audio files and generation of acoustic indices followed that described in Towsey (2017) and were performed using the QUT Ecoacoustics Audio Analysis Software v17.06.000.34 (Towsey and Truskinger, 2017). Recordings were processed into non-overlapping frames of 512 samples per frame (~23.2 ms per frame). Each recording was divided into one-minute segments consisting of 2584 frames. Eleven acoustic indices (Table 1) were calculated for each minute in each of 256 frequency bins from 0 to 11025 Hz (the frequency range of each bin was approximately 43.1 Hz).

2.3. Labelled data set

A labelled data set for training classification models was compiled

Table 1
Descriptions of acoustic indices used as predictors in classification models.

Acoustic index	Description	Reference
Acoustic complexity Index (ACI)	the amount of relative change in sound amplitude from one frame to the next	Towsey (2017)
Background Noise (BGN)	the modal decibel value of background noise in each frequency bin	Towsey (2017)
Cover (CVR)	the fraction of spectrogram cells in a given frequency bin where the acoustic energy exceeds 2 dB (dB)	Towsey et al. (2014b)
Entropy (ENT)	a measure of temporal concentration of acoustic energy in a frequency bin.	Towsey (2017)
Event Count (EVN)	the number of acoustic events (exceeding 3 dB) per minute in each frequency bin.	Towsey (2017)
Power minus noise (PMN)	the maximum decibel value in each frequency bin minus the decibel value of the background noise	Towsey (2017)
Ridge Indices - Horizontal (RHZ), Negative (RNG), Positive (RPS), and Vertical (RVT)	average decibel value of ridge cells in one of the four directions identified in the frequency bin, representing presence of harmonics	Towsey (2017)
Spectral peak tracks (SPT)	a measure of spectral peak tracks, or local maxima of amplitude, in a frequency bin	Towsey (2017)

Table 2
Target frog species and the frequency range of typical calls.

Species	Call Frequency Range (Hz)
Litoria rubella	1400 – 3800
Rhinella marina	300 – 1400
Litoria fallax	2300 – 4800
Litoria nasuta	1000 – 4000
Platyplectrum ornatum	400 – 1500
Litoria caerulea	300 – 1500
Litoria rothii	1000 – 3300
Cyclorana novaehollandiae	400 – 1200
Crinia deserticola	3000 – 5000
Cyclorana alboguttata	400 – 2300
Limnodynastes convexiusculus	800 – 2400
Uperoleia mimula	1200–3000

Table 3
Number of minutes in the labelled data (n = 3274) with each level of species richness (0–7 species calling in the same minute).

Species count	0	1	2	3	4	5	6	7
No. of minutes	527	750	464	468	661	271	105	28

by listening to sampled audio segments along with their spectrograms and labelling the presence or absence of 12 target species (Table 2). Playback and spectrogram analysis of the audio segments was done using Audacity sound analysis software (version 2.2.1, http://www.audacityteam.org). Segments of audio were chosen so that the labelled data set would capture variability in acoustic content of the recordings and contained minute segments with and without each species calling, in combination with and without other frog species' calls, other noise, and silence. The number of calling frog species identified in each minute segment ranged from 0 to 7 species (Table 3). The labelled data set comprised a total of 3274 labelled minutes from 125 different nights across all eight sites. The number of labelled minutes varied per night (2–391) to capture enough species presences on active calling nights.

The protocol for labelling a minute with a positive instance of a target frog species was: if there were two or more vocalisations (i.e., single calls can easily be mistaken for other animal sounds, one call was disregarded and labelled as a negative instance). The calls of *Rhinella marina* (cane toad) are continuous trills of up to 12 s in duration, but they also emit short 'release' calls, so *R. marina* was counted as 'present' if the vocalisation was at least one second in duration, and could be confidently identified to this species. Low-quality or distant calls were counted if they could be confidently identified to a species either aurally or visually on the spectrogram, or both.

As multiple frog species often call together within the same minute, a large proportion of presences for one species may coincide with presences for another. To reduce the chance the classifiers would learn features related to non-target species, the frequency range of the indices used for each species' classifier was restricted to the respective frequency range of their calls (Table 2). Therefore, the final labelled data set for each species consisted of 3274 binary labels indicating species presence or absence for each minute of audio analysed, with a corresponding feature set comprised of acoustic index values. The size of the feature sets ranged from 220 to 781 values, depending on the frequency range used (11 acoustic indices over a predetermined range of frequency bins, Table 2).

2.4. Classification models

Three different classification models were trained for each species: random forest, extreme gradient boosting, and a support vector machine. These three methods were chosen because they can handle high-dimensional data sets with many predictor variables. All analyses were performed using RStudio (RStudio Team, 2019) running R version 3.6.0 (R Core Team, 2019) and packages randomForest (Liaw and Wiener, 2018), xgboost (Chen et al., 2019), kernlab (Karatzoglou et al., 2019) and caret (Kuhn et al., 2019). The labelled dataset was randomly split into a training (70%) and validation set (30%), resulting in a training set of 2293 min and a held-out validation set of 981 min which were unseen by the model training process. All species retained the same prevalence in both the training and validation sets.

Tuning of parameters and selection of the best model for each species was performed using 3 × 10-fold cross-validation on the training set (i.e 10-fold cross-validation was repeated 3 times using a different random split of the training data each repeat). Model performance was assessed using the mean Cohen's Kappa statistic of the 30 cross-validation folds. The Kappa statistic, where Kappa = (observed accuracy − expected accuracy)/(1 − expected accuracy), was used rather than accuracy because of the imbalance between presences and absences in the labelled data set (Kuhn and Johnson, 2013). The final model for each species was then trained on the complete training set.

Using data sets with unbalanced classes can cause problems when fitting classification models (Chawla et al., 2004). Therefore, three different methods of subsampling were used for the four species with the lowest prevalence in the training data set (prevalence < 0.15; Table 4) and compared with results using the full unbalanced training set. Subsampling was performed using down-sampling of the most prevalent class (here absences), as well as SMOTE and ROSE algorithms, two methods that both down-sample the most prevalent class and synthetically synthesize new data points for the least prevalent class (Chawla et al., 2002; Lunardon et al., 2015; Menardi and Torelli, 2014; Torgo, 2015). Model performance was greatest using the unbalanced training set for all four species and was therefore used for predictions.

The predictive performance of the best performing classification model for each species was assessed on the held-out validation set (n = 981). As the minutes in the validation set came from the same nights and segments of recordings as the training set, judging model performance based on the validation set alone may produce over-

Table 4
Species prevalence (number of minutes with presence/absence) in the training data set (n = 2293) used to tune and select the optimal classification model for each species, and the held-out validation set (n = 981), and performance on the validation set. Species are listed in order of prevalence of minutes with species presence. Best model abbreviations are: svm - support vector machine; xgboost - gradient boosting model.

Species	Training (no. minutes)		Validation (no. minutes)		Prevalence in training set	Best model	Performance on validation set		
	Presence	Absence	Presence	Absence			Kappa	Precision	Recall
L. rubella	1045	1248	442	539	0.46	svm	0.88	0.95	0.92
R. marina	738	1555	304	677	0.32	xgboost	0.85	0.90	0.89
L. fallax	669	1624	288	693	0.29	xgboost	0.85	0.94	0.86
L. nasuta	612	1681	265	716	0.27	svm	0.83	0.91	0.84
P. ornatum	501	1792	214	767	0.22	svm	0.92	0.94	0.94
L. caerulea	451	1842	192	789	0.20	svm	0.87	0.93	0.85
L. rothii	437	1856	186	795	0.19	svm	0.85	0.93	0.83
C. novaehollandiae	358	1935	144	837	0.16	xgboost	0.90	0.93	0.89
C. deserticola	231	2062	116	865	0.10	svm	0.96	0.95	0.98
C. alboguttata	149	2144	66	915	0.06	svm	0.97	0.98	0.95
L. convexiusculus	132	2161	64	917	0.06	svm	0.96	1.00	0.92
U. mimula	75	2218	27	954	0.03	svm	0.90	0.96	0.85

optimistic results. Therefore, model performance was also measured on an additional ‘test’ set of 1173 min randomly selected from nights not used in the initial training and validation stage. This should allow a more realistic estimate of how well each species’ classifier performs when exposed to novel acoustic patterns.

The predictive performance of the classification models was evaluated using three measures:

(i) Kappa statistic – agreement of predicted and observed classes above that expected by chance, as defined above in this section (Kuhn and Johnson, 2013).
(ii) Precision – the proportion of positive classifications made by the model which were correct (True Positives/(True Positives + False positives)).
(iii) Recall – the proportion of minutes containing the species’ calls which were detected by the classifier (True Positives/(True Positives + False Negatives)).

3. Results

3.1. Classifier performance

The support vector machine and extreme gradient boosting classification models outperformed random forest for all species on the cross-validated training data (Table 4; Appendix A Fig. A1). Model performance was very high on the validation data for all species (Kappa 0.83–0.97, Table 4), and varied little across a wide range of probability threshold values suggesting the acoustic indices used allowed a strong separation between presence and absence of calls (Appendix A Fig. A2). Precision was 90% or higher for all species, and recall ranged from 83% to 98% (Table 4). False-positive classifications were low relative to false-negative classifications for all species except *P. ornatum*, for which these were similar, and *C. deserticola*, for which there were more false-negatives than false-positives (Appendix A Table A1).

Model performance varied greatly on the test data set (Kappa 0–0.84) but was highest for species with the highest prevalence in the training set (Table 5). Best performance was achieved for *R. marina* (precision 82%; recall 96%), and for *L. fallax* (precision 85%; recall 76%). The model for *L. rubella* achieved good recall of 82%, and precision of 60%. Classification performance was moderate for *L. nasuta* (precision 62%; recall 63%) and *L. rothii* (precision 50%; recall 53%) despite having sufficient training and test cases. Our method of randomly selecting test case minutes affected our ability to measure model accuracy for all species, as the prevalence of positive instances of some species in the test set was very low (Table 5; Appendix A Table A2).

Table 5
Best model performance measures on the test data set and number of minutes each species is present or absent in the test data (n = 1173).

Species	No. minutes		Model performance on test set		
	Presence	Absence	Kappa	Precision	Recall
L. rubella	121	1052	0.65	0.60	0.82
R. marina	279	894	0.84	0.82	0.96
L. fallax	345	828	0.72	0.85	0.76
L. nasuta	138	1035	0.57	0.62	0.63
P. ornatum	2	1171	0	0	0
L. caerulea	3	1170	0	0	0
L. rothii	119	1054	0.46	0.50	0.53
C. novaehollandiae	19	1154	0.29	0.24	0.42
C. deserticola	29	1144	0.15	0.21	0.14
C. alboguttata	12	1161	0.43	0.45	0.42
L. convexiusculus	76	1097	0.12	0.55	0.08
U. mimula	3	1170	0	0	0

Table 6
Categories of noise features in a sample of 119 instances of misclassifications from the test data. Observations were made of noise features in the frequency band of the target frog call that may have contributed to the misclassification.

Category	False positives (species incorrectly predicted)	False negatives (species presence missed)
Other frog species	29	32
Birds or insects	6	0
Wind or Rain	4	2
Vehicles or Aircraft	16	0
Audio distortion	13	0
Unidentified noise	3	3
Short calling bout of target species	na	2
Low quality/distant call of target species	na	9
Totals	71	48

3.2. Analysis of misclassifications

Analysis of a sample of misclassified cases in the test data set revealed that the majority of false-positive detections occurred for minutes in which other frogs were calling (Table 6). Passing vehicles or aircraft, noise caused by audio distortion, birds, insects, wind and rain were also factors causing misclassification. Most false-negative detections also included other frog species, wind or rain. However, several false-negative detections were cases observed as having ‘low-quality calls’, i.e., these were minutes in which the target species was calling but distant to the microphone, and there was no other dominant source of noise in that bandwidth. Two of the missed detections contained short calling bouts (i.e. short sequence of only a few individual vocalisations).

3.3. Times to conduct different aspects of the analysis

It took four observers on average 2.2 min to manually analyse and label each minute segment and record species presence in a spreadsheet to compile the training and test data, which comprised data on a total of 4447 one-minute segments of audio. Some recordings were more time consuming than others to label, because more species called simultaneously and had to be replayed more than once. Other recording segments with few or no species calling could be analysed visually on the spectrogram, and in these cases analysis of the minute segment took only a few seconds.

Calculation of the acoustic indices was the most time consuming and computer-intensive component of this method. With a 16-core computer processor, each recording (of up to 13 h) took 13–17 min to process, and with batch parallel processing it took approximately 2 weeks to process the entire set of 3965 recordings. This process included generation of false-colour spectrogram images in addition to calculation of acoustic indices.

The number of tuning variables for each classification model varied, and therefore influenced the time required for model training. Parameter tuning and fitting of the final classification models on the training data took an average of 127 min for each species using random forest (17 tuning values), 55 min using extreme gradient boosting (40 tuning values), and 16 min using support vector machine (12 tuning values) using the caret package on a personal computer with 16 GB RAM and 2.6 GHz processor.

4. Discussion

We found that acoustic indices can be used as predictive features for automated detection of frog chorus activity to species level in environmental recordings. High levels of predictive performance were obtained for all species on the held-out validation data set, demonstrating that acoustic indices were effective at capturing the acoustic characteristics of frog calls in minute segments of audio. Good performance was also achieved on the additional test data set for the most prevalent species in the training data (*R. marina*, *L. fallax* and *L. rubella,* Table 5). Therefore, this method is suited to detection of commonly calling species for which sufficient training examples can be obtained. The method shown here demonstrates reliable automated detection of frog species using acoustic indices can be achieved at a one-minute scale. At this resolution, data on the vocal activity of multiple frog species can be obtained to monitor species presence and absence, and to investigate temporal patterns of within-night chorus behaviour as well as to describe seasonal breeding patterns.

This study has demonstrated an approach to automated species identification in acoustic monitoring studies where individual call recognisers may not be suitable. For the most prevalent species, we achieved precision and recall rates greater than 80%. This is comparable to reported results of tests of individual call recognition on long field recordings, reflecting the real scenarios encountered in acoustic monitoring of wildlife in natural habitats. Typically successful automated recognition of bird and frog calls is greater than 80% precision or recall (e.g. Bardeli et al., 2010; Corrada Bravo et al., 2017; Potamitis et al., 2014). Other studies have demonstrated that, while high detection accuracy can be achieved when the target calls are high in quality, accuracy declines with decreasing signal-to-noise ratio (Bardeli et al.s, 2010; Digby et al.s, 2013). The lack of scalability of individual call recognisers to field recordings with high noise levels and species richness means automated detection methods for species which call in choruses remains a challenge. Our results support the efficacy of acoustic indices as predictive features for identifying chorusing frog species in sound recordings where multiple species and many individuals overlap in their calls.

The majority of classification errors made by our models on the test data were for minutes in which other noise (i.e., other frog species, vehicles or distortion) was present, or the target species calls were of low quality (Table 6). The training data were systematically chosen to challenge the classifiers and provide a range of difficult cases that would occur in most long-duration environmental recordings. Audio segments were selected for labelling that included varying levels of frog chorus activity and call overlap with many individual frogs from multiple species calling. These cases will present a challenge to any detection method - it is difficult even for an experienced human observer to distinguish species in the background in high intensity chorus activity of multiple species. Including additional labelled minutes with non-target noise in the training data could conceivably improve the predictive performance of the models. Likewise, including more examples of low-quality calls (i.e. the calls of target species are low in amplitude because individuals are distant from the microphone) may decrease false-negative detections.

When applied to a larger set of recordings, higher variability in the acoustic content is likely to be encountered than is present in the training data, and lower classification performance could be expected. This is supported by our results where classifier performance was higher on the validation set than on the test set. This suggests the labelled training data must capture as much of the variability in the acoustic environment as possible. This requires some familiarity with the study habitats and knowledge of the range of vocal species and environmental noises which occur. The classification model could be iteratively improved in this way, by sampling results to inspect misclassified cases, and including additional training data to better capture the variety of sounds present in the recordings.

Like all machine learning applications, the question arises as to what extent our learned models will be accurate on acoustic recordings obtained from other environments. The acoustic composition of the environment can vary greatly across time and space. Different habitats with different sources of biotic sounds (i.e., species), and abiotic sounds (e.g., vegetation structure, land features, and urbanisation) will vary in their acoustic composition. These factors may affect the value of acoustic indices and therefore training data from recordings in particular environments may not be suitable for use in detecting the same species at other locations. Other factors that could possibly affect the calculation of acoustic indices are the type of recording unit, recording settings and methods of audio pre-processing. For example, recording quality and the method of noise removal employed will affect the absolute signal values used in the calculation of many acoustic indices. Furthermore, Towsey et al. (2015) noted that sound files in compressed format, such as MP3, produced artefacts in spectrograms which affected the values of some acoustic indices. However, the degree to which these factors influence the values of acoustic indices has not been tested. To accurately detect target species in environmental recordings using acoustic indices, training data should be sampled from recordings in the same or similar habitats and using standardised methods of recording and pre-processing.

The few studies that have investigated the use of acoustic indices as predictive features for automated detection of animal calls reported success in detecting the calls of target species, but these have so far been limited to rare or threatened bird species. Towsey et al. (2018b) achieved performance of 90% precision and 67% recall in identifying the calls of Lewin's Rail (*Lewinia pectoralis brachipus*) using a set of 5 acoustic indices. Gan et al. (2018) reported more modest success in detecting calls of the little spotted kiwi (*Apteryx owenii*) using a suite of acoustic indices (precision 85% and recall 53.3%), but their experimental study may have lacked sufficient training data. Dema et al. (2018) achieved classification precision of 97.6% and recall 96.1% on training data in detecting the calls of the endangered white-bellied heron (*Ardea insignis*) in Bhutan, and also tested predictive performance on a larger data set, where precision and recall were reduced to 65.5% and 80.7% respectively. While these detection methods were all tested on recordings from the natural environments of the target bird species, these species have relatively low calling rates and few competing noise sources in the bandwidth of their calls. The results presented here are the first demonstrating the use of acoustic indices to accurately detect frog calls in environmental recordings, and to target the calls of multiple, commonly calling species with calls overlapping in time and frequency. We achieved best prediction results on the target frog species which were vocally dominant in our study system, that is those species calling in loud choruses of many individual persisting over many hours. Researchers interested in applying this method to other target species would need to determine the merits of such an approach over other call recognition methods. For some study objectives it may be more suitable to develop call recognisers, for example for detecting rare species or species that call only occasionally.

The main challenges in developing automated animal call recognition software are noise reduction, call detection, call segmentation, and feature extraction. Using acoustic indices as features sidesteps some of the technical challenges of building call recognisers and treats all segments of the recording as regions of interest. In our approach, the acoustic indices were calculated on the audio segments (i.e., minutes and frequency bins) rather than relying on accurate detection and segmentation of individual animal calls. As we have shown, this approach provides a possible solution to detecting species calling in choruses in which many similar calls are overlapping and difficult to distinguish. As this method detects the presence of vocalisations rather than the vocalisations themselves, there is no requirement other than the ability to identify species' calls in a recording in order to create a labelled data set. This contrasts with traditional call recognisers which require detailed analysis of species-specific call features, as well as expertise in sound analysis.

The software used to calculate the acoustic indices for this study is open-source (Towsey et al., 2018a), performs audio pre-processing and calculates a suite of acoustic indices using the methods of Towsey (2017). Other open-source software such as R packages 'seewave' (Sueur et al., 2008) and 'soundecology' (Villanueva-Rivera and Pijanowski, 2018) are available which take raw sound files as input and compute various acoustic indices. Although processing and analysing very large sets of sound recordings is computationally intensive, the calculation of acoustic indices using available sound analysis software is straightforward in comparison to developing call recognition algorithms. In addition, once calculated, the acoustic indices can be retained as permanent-feature data sets and used for multiple analyses.

The method demonstrated here is a straightforward implementation of a frog call species classifier using open-source software to calculate acoustic indices on minute segments environmental sound recordings, and fit a classification model. The set of predictor acoustic indices was not optimised, that is, 11 indices output by the Ecosounds Audio Analysis program across a wide frequency range were used as classification features. In this sense the classifiers were treated as black-box models to find the optimum criteria from a large set of candidate variables on which to classify recordings. Further work to examine the particular influence that animal call characteristics, and environmental sounds in general, have on the values of acoustic indices is required to find indices most useful for distinguishing target species.

CRediT authorship contribution statement

Sheryn Brodie: Methodology, Formal analysis, Investigation, Validation, Data curation, Writing - original draft, Project administration. **Slade Allen-Ankins:** Methodology, Formal analysis, Investigation, Writing - review & editing. **Michael Towsey:** Conceptualization, Software, Writing - review & editing. **Paul Roe:** Software, Writing - review & editing, Funding acquisition. **Lin Schwarzkopf:** Writing - review & editing, Supervision, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank Kiyomi Yasumiba, Ross Alford and Richard Duffy for obtaining the original set of field recordings used in this study, and the Fryer and Freestun families for access to the study dams at Hervey Range. We are also grateful to Kayla Mehes, Rishab Pillai and Ryan Searle for their assistance in manual analysis and labelling of audio. Funding for this work was provided by the Australian Research Council (Linkage grant number LP150100675 'Call Out and Listen In: A New Way to Detect and Control Invasive Species') and an Australian Government Research Training Program Scholarship. The Anindilyakwa Land Council were partners in the grant funding this study. This study was conducted under permission of Animal Ethics at James Cook University (permit no. A1838).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecolind.2020.106852.

References

Acevedo, M., Villanueva-Rivera, L., 2006. Using automated digital recording systems as effective tools for the monitoring of birds and amphibians. Wildlife Soc. Bull. 34, 211–214.

Aide, T.M., Corrada-Bravo, C., Campos-Cerqueira, M., Milan, C., Vega, G., Alvarez, R., 2013. Real-time bioacoustics monitoring and automated species identification. PeerJ 1, e103.

Australian Acoustic Observatory. 2019. A20 Australian Acoustic Observatory, Queensland University of Technology (QUT) Ecoacoustics Research Group, Brisbane, Australia, viewed 17 August 2019. https://acousticobservatory.org/.

Bardeli, R., Wolff, D., Kurth, F., Koch, M., Tauchert, K.H., Frommolt, K.H., 2010. Detecting bird sounds in a complex acoustic environment and application to bioacoustic monitoring. Pattern Recogn. Lett. 31, 1524–1534. https://doi.org/10.1016/j.patrec.2009.09.014.

Bedoya, C., Isaza, C., Daza, J.M., López, J.D., 2014. Automatic recognition of anuran species based on syllable identification. Ecol. Inf. 24, 200–209. https://doi.org/10.1016/j.ecoinf.2014.08.009.

Campos-Cerqueira, M., Aide, T.M., 2016. Improving distribution data of threatened species by combining acoustic monitoring and occupancy modelling. Methods Ecol. Evol. 7 (11), 1340–1348.

Chawla, N., Bowyer, K., Hall, L., Kegelmeyer, W., 2002. SMOTE: synthetic minority oversampling technique. J. Artif. Intelligence Res. 16, 321–357.

Chawla, N., Japkowicz, N., Kotcz, A., 2004. Special issue on learning from imbalanced data sets. ACM SIGKDD Explor. Newsletter 6 (1), 1–6.

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho H., Chen, K., Mitchell, R., Cano, I., Zhou, T., Li, M., Xie, J., Lin, M., Geng, Y. and Li, Y., 2019. xgboost: Extreme gradient boosting. R package version 0.82.1. https://CRAN.R-project.org/package=xgboost.

Corrada Bravo, C.J., Álvarez Berríos, R., Aide, T.M., 2017. Species-specific audio detection: a comparison of three template-based detection algorithms using random forests. PeerJ Comput. Sci. 3. https://doi.org/10.7717/peerj-cs.113.

Crump, P.S., Houlahan, J., 2017. Designing better frog call recognition models. Ecol. Evol. 7, 3087–3099.

Dema, T., Towsey, M., Sherub, S., Sonam, J., Kinley, K., Truskinger, A., Brereton, M., Roe, P., 2018. Acoustic detection and acoustic habitat characterisation of the critically endangered white-bellied heron (*Ardea insignis*) in Bhutan. Freshw. Biol.

Digby, A., Towsey, M., Bell, B.D., Teal, P.D., Giuggioli, L., 2013. A practical comparison of manual and autonomous methods for acoustic monitoring. Methods Ecol. Evol. 4, 675–683. https://doi.org/10.1111/2041-210x.12060.

Field, S., O'Connor, P., Tyre, A., Possingham, H., 2007. Making monitoring meaningful. Austral Ecol. 32, 485–491.

Frommolt, K.-H., Tauchert, K.-H., 2014. Applying bioacoustic methods for long-term monitoring of a nocturnal wetland bird. Ecol. Inf. 21, 4–12. https://doi.org/10.1016/j.ecoinf.2013.12.009.

Gan, H., Towsey, M., Li, Y., Zhang, Y., Roe, P., 2018. Animal call recognition with acoustic indices: Little Spotted Kiwi as a case study. IEEE 2018 Digital Image Computing: Techniques and Applications (DICTA). Canberra, Australia.

Ganchev, T., Potamitis, I., 2007. Automatic acoustic identification of singing insects. Bioacoustics 16 (3), 281–328.

Gibb, R., Browning, E., Glover-Kapfer, P., Jones, K.E., Börger, L., 2019. Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. Methods Ecol. Evol. 10, 169–185.

Hagens, S.V., Rendall, A.R., Whisson, D.A., 2018. Passive acoustic surveys for predicting species' distributions: optimising detection probability. PLoS One 13 (7), e0199396. https://doi.org/10.1371/journal.pone.0199396.

Indraswari, K., Bower, D., Tucker, D., Schwarzkopf, L., Towsey, M., Roe, P., 2018. Assessing the value of acoustic indices to distinguish species and quantify activity: a case study using frogs. Freshw. Biol.

Karatzoglou, A., Smola, A., Hornik, K., National ICT Australia, Maniscalco, M., and Teo, C., 2019. kernlab: Kernel-based machine learning lab. R package version 0.9-29. https://CRAN.R-project.org/package=kernlab.

Knight, E.C., Poo Hernandez, S., Bayne, E.M., Bulitko, V., Tucker, B.V., 2019. Pre-processing spectrogram parameters improve the accuracy of bioacoustic classification using convolutional neural networks. Bioacoustics 1–19.

Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, R Core Team, Benesty, M., Lescarbeau, R., Ziem, A., Scrucca, L., Tang, Y., Candan, C. and Hunt, T., 2019. caret: Classification and regression training. R package version 6.0-84. https://CRAN.R-project.org/package=caret.

Kuhn, M., Johnson, K., 2013. Applied Predictive Modeling. Springer, New York.

Liaw, A. and Wiener, M., 2018. randomForest: Breiman and Cutler's Random Forests for Classification and Regression. R package version 4.6-14. https://CRAN.R-project.org/package=randomForest.

Lunardon, N., Menardi, G. and Torelli, N., 2015. ROSE: Random over-sampling examples. R package version 0.0-3. https://CRAN.R-project.org/package=ROSE.

Menardi, G., Torelli, N., 2014. Training and assessing classification rules with imbalanced data. Data Min. Knowl. Disc. 28 (1), 92–122.

Noda, J.J., Travieso, C.M., Sánchez-Rodríguez, D., 2016. Methodology for automatic bioacoustic classification of anurans based on feature fusion. Expert Syst. Appl. 50, 100–106.

Pieretti, N., Farina, A., Morri, D., 2011. A new methodology to infer the singing activity of an avian community: the Acoustic Complexity Index (ACI). Ecol. Ind. 11, 868–873.

Potamitis, I., Ntalampiras, S., Jahn, O., Riede, K., 2014. Automatic bird sound detection in long real-field recordings: applications and tools. Appl. Acoust. 80, 1–9. https://doi.org/10.1016/j.apacoust.2014.01.001.

Priyadarshani, N., Marsland, S., Castro, I., 2018. Automated birdsong recognition in complex acoustic environments: a review. J. Avian Biol. 49.

R Core Team, 2019. R: A Language and Environment for Statistical Computing. Version 3.4.3. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

RStudio Team, 2019. RStudio: Integrated Development for R. Version 1.2.1335. RStudio, Inc., Boston, MA. http://www.rstudio.com/.

Sanders, C.E., Mennill, D.J., 2014. Acoustic monitoring of nocturnally migrating birds accurately assesses the timing and magnitude of migration through the Great Lakes. Condor: Ornithol. Appl. 116 (3), 371–383.

Sueur, J., Aubin, T., Simonis, C., 2008. Seewave: a free modular tool for sound analysis and synthesis. Bioacoustics 18, 213–226.

Sueur, J., Farina, A., Gasc, A., Pieretti, N., Pavoine, S., 2014. Acoustic indices for biodiversity assessment and landscape investigation. Acta Acustica United Acustica 100, 772–781. https://doi.org/10.3813/aaa.918757.

Torgo, L., 2015. DMwR: Functions and data for 'Data Mining with R'. R package version 0.4.1. https://CRAN.R-project.org/package=DMwr.

Towsey, M, Truskinger, A, Roe, P., 2017. QUT Ecoacoustics Audio Analysis Software (Version 17.06.000.34) [Computer software]. Brisbane: QUT Ecoacoustics Research Group.

Towsey, M., Truskinger, A., Cottman-Fields, M., and Roe, P., 2018. Ecoacoustics Audio Analysis Software v18.03.0.41 (Version v18.03.0.41). Zenodo. DOI:10.5281/zenodo.1188744.

Towsey, M., Parsons, S., Sueur, J., 2014a. Ecology and acoustics at a large scale. Ecol. Inf. 21, 1–3.

Towsey, M.W., Truskinger, A.M., Roe, P., 2015. The navigation and visualisation of environmental audio using zooming spectrograms. ICDM 2015: International Conference on Data Mining. IEEE, Atlantic City, NJ.

Towsey, M., Zhang, L., Cottman-Fields, M., Wimmer, J., Zhang, J., Roe, P., 2014b. Visualization of long-duration acoustic recordings of the environment. Procedia Comput. Sci. 29, 703–712. https://doi.org/10.1016/j.procs.2014.05.063.

Towsey, M., Znidersic, E., Broken-Brow, J., Indraswari, K., Watson, D.M., Phillips, Y., Truskinger, A., Roe, P., 2018b. Long-duration, false-colour spectrograms for detecting species in large audio data-sets. J. Ecoacoustics 2.

Towsey, M., 2017. The calculation of acoustic indices derived from long-duration recordings of the natural environment. Available: https://eprints.qut.edu.au/110634. Accessed 20 Jul 2019.

Villanueva-Rivera, L. and Pijanowski, B., 2018. Soundecology: Soundscape Ecology. R package version 1.3.3. https://CRAN.R-project.org/package=soundecology.

Waddle, J.H., Thigpen, T.F., Glorioso, B.M., 2009. Efficacy of automatic vocalization recognition software for anuran monitoring. Herpetol. Conserv. Biol. 4 (3), 384–388.