ENGINEER - Vol. LIV, No. 02, pp. [07-13], 2021

DOI: http://doi.org/10.4038/engineer.v54i2.7386

Novel Feature Extraction Algorithm for Classification of Multiple Occurrence of Flight Calls

D.N. Egodage and S.J. Sooriyaarachchi

Abstract: Acoustic monitoring of migratory birds is becoming a demand with respect to public policy related to wind power because wind mills are responsible for the death of a large number of migratory birds. Acoustic monitoring is associated with three main processes, namely pre-processing, feature extraction and classification. An improved algorithm that can extract features has been developed in this research by combining well known MSER technique with traditional techniques. Extracted features from the said algorithm and other algorithms were combined to create three different feature sets. Classification techniques, including kNN, RF, SVM and DNN, were used to evaluate a real-world dataset in terms of the extracted features. The feature extraction technique proposed in this research. namely SMSER, performs better than SATF feature set alone and combination of SATF and SIFS feature sets with the highest performing classifier DNN with an accuracy of 87.67%.

Keywords: Similar acoustic event, Maximally stable extremal region, Feature extraction

1. Introduction

Birds use several vocalisations in various behavioural contexts for different purposes such as to maintain communication with a respective social group, to warn about predators and to request parental care. Birds do long sustained flights with the help of wind during migration. The vocalisations made by migratory birds to keep contact with the flock are known as flight calls. Flight calls have specific characteristics such as frequency modulated, tonal and monosyllabic [1] sounds. Flight calls generally have 50 to 300 ms duration, and their frequencies range from 1 to 11 kHz [1]. Flight calls are different from songs and alarm calls because these are relatively simple vocalisations that present a pattern of rapid frequency sweeps.

Consequently, flight calls have similarities compared to other complex vocalisations that birds generally generate. Better understanding of these flight calls is useful for several applications, in particular, to reduce the impact of migratory birds on wind energy production, understand the movement of species during seasonal migrations and estimate the density of birds in migration from vocalisation counts. Moreover, public policies that govern the establishment and operation of windmills require to ensure minimal damage to fauna. The main reasons to impose such policies are to save the migratory bird collisions against windmills. Turning off turbines and adjusting the rotor blades to minimise their surface relative to the main direction of migration could help reduce collision extent. Automatic classification of bird flight calls has addressed to study migration patterns and monitor areas of human interactions such as wind farms due to the above reasons.

More effective analysis tools will improve the large-scale monitoring of flight calls of migratory bird species. Therefore, the insight of count of flight calls could use as an estimate of the impact. Acoustic classifiers in previous research in the field of flight call classification are both in the manual [8]-10] and automatic [11]-[21] approaches. Automatic acoustic classifiers approach flight call classification as an N-class problem [11]-[15], a binary open-set problem for specific species [16]-[20], and a binary open-set problem for several species [21]. Training a model to classify a given clip to N classes assuming that the dataset contains clips of only N number of species is known as N-class problem. Classification accuracy of 97.6% has been obtained in [12] for N class problem scenario owing to the use of feature learning techniques. The Binary open-set problem for specific species is to classify whether a vocalisation of a particular species is present in a set of sound clips or not. Finally, the binary open-set problem for several species is to classify whether

Eng. D.N. Egodage, *AMIE(SL), B.Sc. Eng. (Moratwa), Post Graduate Student, Department of Computer Science and Engineering, University of Moratuwa.*
Email:dinethegodage.13@cse.mrt.ac.lk
ORCID ID: http://orcid.org/0000_0001_7675_2725
Eng. (Dr.) S.J. Sooriyaarachchi*, AMIE(SL), B.Sc.Eng. (Peradeniya), PhD (Moratuwa), M.Sc. (Moratuwa), Lecturer, Department of Computer Science and Engineering, University of Moratuwa.*
Email:sulochanas@cse.mrt.ac.lk
ORCID ID: http://orcid.org/0000_0003_0905_7196

there exists a flight call or not, in a set of sound clips in the presence of vocalisations of multiple species.

A general classifier that can identify the flight calls irrespective of species needs to be implemented to solve the problem of the binary open set for several species. Therefore, 'binary open-set problem for multiple species' kind of classifier can be used to determine the flight call count of a flock of birds. That will help to estimate a rough number of birds involved in a non-invasive manner.

Justin et al. [21] have addressed the same issue with a dataset called Birdvox-full-night, using a Convolutional Neural Network (CNN) with three convolutional layers and two dense layers having 677k parameters and achieved 90.48% accuracy by training the CNN on a Graphics Processing Unit (GPU). Due to the following reasons, this research work claims that a different approach should be found:

- Computational power used by [21] will not be practical to be used in a remote environment
- Developing countries might have constraints to implement such kind of classifiers
- To find a less computationally intensive detection technique

Therefore, the main focus of this paper is to:

- Implement an automatic acoustic-based detection of multiple occurrences of flight calls in real word recordings, using signal processing and classification techniques with less computational power.

The paper has been organised as follows: Section 2 discusses the dataset, Section 3 explains on feature extraction methods, and Section 4 explains methodologies and experiments done. Section 5 shows the results. Section 6 presents the conclusion and related future work.

2. Data Set

2.1 Bird Vox-full-night

This dataset can be used for *Binary open-set scenario for multiple species* classification. It contains six far-field audio clips recorded during a full night, which consists of 35,000 flight calls of 25 species. The dataset has been annotated individually by experts with the time of bird call existence. Every clip in this dataset will either contain a flight call of any species or not.

Since the only dataset that can be used to address the Binary open set problem for multiple species is Bird Vox-full-night, that dataset was selected to implement the general classifier to recognise flight calls in an audio clip and also to compare the effectiveness of the implemented approach with the results of work done by [21].

3. Feature Extraction Methods

This section presents an overview of three different feature extraction methods used in this study. Spectral and Temporal Features are the most commonly used feature sets used in related prior work. Spectrogram-based Image Frequency Statistics have been used in [13] to improve the classification accuracy of the flight calls, claiming that SIFS + MFCC dataset shows a better classification accuracy than using MFCC only. Therefore, these two feature sets are used as state of the art. While using the SIFS algorithm, some drawbacks were noticed. A novel modification to overcame those drawbacks was done in this research. Furthermore, a new feature set called Spectrogram-based Maximally Stable Extremal Regions (SMSER) was implemented. Preprocessing step was conducted to filter out the background noise to increase the quality of features to some extent. To filter out the background noise that occurs from the wind, 1kHz butter-worth high pass filter was used. Signal-to-noise ratio improved using the spectral subtraction method, where an average noise spectrum is subtracted from the signal. The average noise spectrum was estimated from the periods where the signal is not present.

3.1 Spectral and Temporal Features (SATF)

These features include the common features that have been generally used for classification of birds in related work.

- Zero Crossing Rate
- Energy (Root Mean Square)
- Spectral Centroid
- Spectral Bandwidth
- Spectral Contrast
- Spectral Flatness
- Spectral Roll off
- 13 Mel-Frequency Cepstral Coefficients (MFCCs)

Altogether there are 20 features of concern. When extracting these features from a 500 ms sound clip, with 12ms window length, it accounts to a distribution of 41 data points for each feature. Therefore, the distribution of each 20 features was modelled over the segment, using mean and standard deviation. So, the final feature vector used for classification consists of 40 features.

3.2 Spectrogram-based Image Frequency Statistics (SIFS)

Selin et al. [13] have used SIFS features to evaluate the classification of bird flight calls. This feature set has been implemented based on the following assumptions, according to the authors.

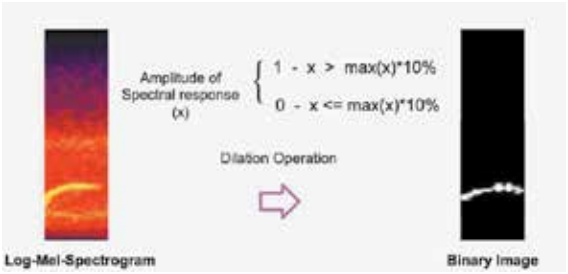

Figure 1 - Conversion of Log-Mel-Spectrogram into a Binary Image

- Flight calls can be characterised by the highest amplitude of the signal.
- In time-frequency domain, it is most obviously distinguishable.
- The highest amplitude noise is present in low frequencies.

According to the above three assumptions, the algorithm shown in Figure 1 has been implemented. The Log-Mel-Spectrogram of each call is extracted as the first step. Then the bottommost frequency responses have been filtered out. As per Selin et al. [13], because of the position of spectral responses inside the spectrogram was of more interest than the amplitude, all the spectral response amplitudes more significant than 10% of the overall maximum spectral amplitude of the spectrogram were set to a value of one, and all others were set to a value of zero. Further, a dilation operation was performed to enhance the continuity in the call. Figure 1 shows the binary image after conducting the above-explained steps. Then the clip was subjected to feature extraction. The highest, lowest, median and mean frequencies were computed for the first 3/7, second 3/7, third 3/7 and whole of the image respectively as to sixteen features. Sixteen features in total can be extracted from an audio clip using SIFS.

3.3 Spectrogram-based Maximally Stable Extremal Regions (SMSER)

In the computer vision domain, Maximally Stable Extremal Regions (MSER) is frequently used to detect a region in an image. MSER can be used to find objects in an image. Similarly, this method has been used to detect the regions of a flight call in a binary image.

As discussed in the SIFS method in Section 3.2, the Log-Mel-Spectrogram of a sound clip was converted to the binary image after the dilation operation. It was discovered that a 10% static margin which was used in the above step did not apply to all the scenarios in this research. Figure 2 shows that a 10% static margin gives more distraction to detect the flight call in the clip. After further investigation, the flight call was visible and all the distracting noise was removed at a 60% margin for the specific flight call.

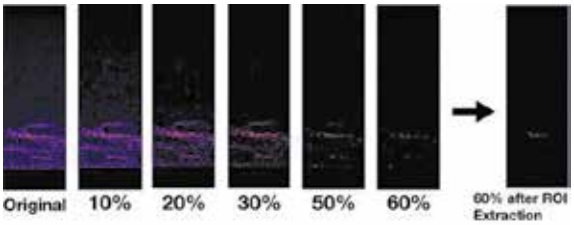

Figure 2 - Static Margin from 10% to 60% and ROI Extracted from 60% Margin

The above mentioned scenario led to the implementation of the SMSER algorithm. The SIFS algorithm works with a fixed percentage threshold where dynamically chosen percentage is used in SMSER algorithm, starting from 100% to 0%, until it find the best percentage considering the contour area which matches a flight call area. Figure 3 shows the simplified version of the algorithm which was used for the feature extraction. The ultimate goal of the algorithm was to determine the best percentage to capture the flight call which has the highest amplitude of the signal. The binary image consisted of 44 pixels along the frequency axis and 257 pixels along the time axis. According to the assumptions stated in [1] regarding flight calls, smallest flight call (area wise) should be varied between a range of 1 kHz and 50 ms long= (44*1/11) * (257*50/500) = 103 square pixel, and largest flight call (area wise) should be varied

between a range of 3 kHz and 300 ms long = (44*3/11) * (257*300/500) = 1848 square pixel. The function 'get Number Of Contours ( )' returns the count of all the contours that has a minimum area of 103 and maximum area of 1848. The algorithm will stop after 100 iterations, if it cannot find a percentage which matches the above condition.

Figure 4 shows a few sample flight calls which have been captured by the algorithm. After this process, the image inside the contour will be isolated by another operation. After that the following features will be calculated:

- Moments (2 features):
  - Center of mass of the object (x, y) coordinates
- Contour area (1 feature)
- Contour perimeter (1 feature)
- Contour bounding rectangle attributes (4 features):
  - Top left coordinates (x, y), height and width
- Aspect ratio of the bounding rectangle (1 feature)
- Extent of the contour area compared to bounding rectangle area (1 feature)
- Leftmost, rightmost, topmost and bottommost (x, y) coordinates of the contour (8 features)
- Radius and center coordinates of the minimum enclosing circle of the contour (3 features)
- Starting coordinate and ending coordinate of the contour fit-line (4 features)
  - Starting coordinates (x, y)
  - Ending coordinates (x, y)

Altogether there are 25 features from SMSER.

4. Methodology and Experiments

This section presents how the three feature sets in Section 2 were evaluated against four classification techniques, namely deep neural network (DNN); Random Forest (RF) algorithm; k-Nearest Neighbours (kNN) algorithm; and Support Vector Machines (SVM) algorithm. These four different classification techniques are the most used techniques of previous research. Therefore, they were used as the classification techniques. To show the effectiveness of the developed feature-sets, the four classification technique was used.

The pre-processing step was conducted to extract audio clips of duration 500 ms from the continuous recordings with annotations. Since flight calls are generally 50 ms to 300 ms long, 500 ms long clips were extracted.

```
#getFeatures function finds the percentage that exactly has one contour and extract
features from that percentage
def getFeatures (File):
   lowerBound = 100
#lowerBound holds last known lower bound percentage where zero contours found
   upperBound = 0
#upperBound holds last known upper bound percentage where >=1 contours found
   difference = 10
#difference holds the value that will increased or reduced in every iteration according to t
direction
   direction  = 1
#there are two directions {100% ===> 0%} = 1 and {100% <=== 0%} = 0
   no_of_contours = 0
#use to store no. of contours returned by MSER which has area of a flight call
   percentage = lowerBound
#percentage will be started with having the value of the lower bound
   while(no_of_contours != 1):
#use to find percentage using no_of_contours
      no_of_contours = getNoOfContours(File, percentage)
#using MSER algorithm to find contours which match the area of flight call
      if (no_of_contours == 1):
         break
      else:
         if (direction):
            if (no_of_contours == 0):
               percentage = percentage - difference
               lowerBound = percentage
#while the lower bound is always with zero contours this will execute
            else:
               upperBound = percentage
               direction  = 0
               difference = difference / 10
               percentage = percentage + difference
#if percentage find more than zero contours the direction will be changed and difference
will be reduced 10 times to find closer point of interest
         else:
            if (no_of_contours != 0):
               percentage = percentage + difference
               upperBound = percentage
#while the upper bound is always with more than zero contours this will execute
            else:
               lowerBound = percentage
               direction  = 1
               difference = difference / 10
               percentage = percentage - difference
#if percentage find zero contours the direction will be changed and difference will be
reduced 10 times to find closer point of interest
   return extractFeatures(File, percentage)
```

Figure 3 - Pseudo Code of SMSER Algorithm

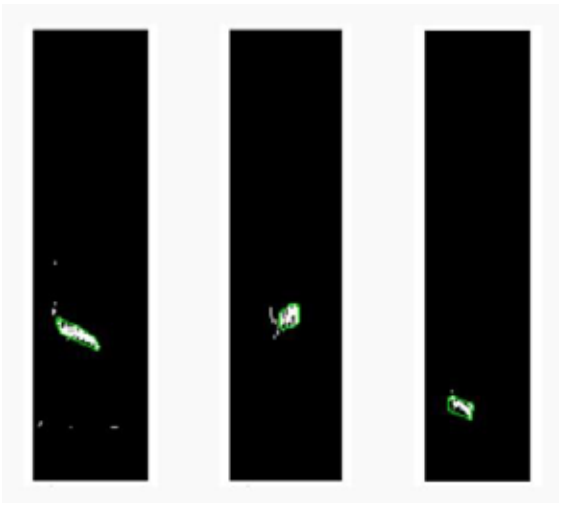

Figure 4 - Samples of Detected Contours using the SMSER Algorithm

It was possible to generate 70,804 clips from all continuous recordings in which 35,402 clips contained flight calls. Dataset was separated as 57,126 clips for the training set and 13,678 clips for the test set. Afterwards, noise reduction was implemented by imposing a 1 kHz Butterworth High Pass filter to remove the noise. Further, spectral subtraction was used to remove the recurrent noise. Log-Mel-Spectral features have been shown to work better than spectral features [22], [23]. Lastly, Log-Mel-Spectrogram was generated for all the clips using a Hann window with a window length of 256 samples and hop length of 32 samples.

4.1 Classifiers

Classification is another crucial part of the system. RF, kNN, SVM, and DNN classifiers were used in the study for the classification process. A k-NN classifier was tested with several nearest neighbour values, while an RF classifier was tested with a different number of trees. An SVM was tested with three different kernels, namely Polynomial, Gaussian, and Gaussian radial basis function (RBF). Lastly, a DNN classifier with Adam optimisation was tested by varying the learning rate, the number of hidden units and the number of layers. The DNN consisted of five layers, and three layers out of them were dense layers. Altogether the network consisted of 10k hidden units in total. The input layer had 81 nodes, and the output layer had two "softmax" activation nodes.

4.2 Feature Sets

Three feature sets were prepared for this research:

- SATF: 40 features
- SATF and SIFS: 56 features (40 + 16)
- SATF, SIFS and SMSER: 81 features (40 + 16 + 25)

It is mandatory to keep Spectral and Temporal features to identify a sound event in an audio clip, and the three feature sets were created by preserving both these feature types. As per Bastas et al. [13], SATF and SIFS together give more accuracy than when they are alone. Therefore, the second feature set was created by including both of them. According to the same assumption, SMSER features were added to the third feature set.

After the feature extraction process, the normalisation process for all the data was conducted.

5. Results

The dataset was divided into two parts as Train set and Test set. Test segments were classified using the created classifiers, which are optimised to train segments. Three models were created for the three feature sets. Classification results of the three feature sets for all the four different classifiers are shown in Figure5. DNN has performed best with all three Feature sets. The best classification accuracy has been reached from Feature set 3. The work presented in this paper aimed to find out a less computationally intensive detection technique, which can classify the flight calls. A comparison between the work done in [21] and current work, is listed on Table 1. According to the results, it is clear that the approach taken in work done in [21] is far better with the existence of high computational power. Even though the current work is behind classification accuracy, that result is reached by a smaller network that requires less computational power. In more specific terms, the current work was executed by an Intel Core i5 3.0GHz central processing unit with 16GB DDR3 Random Access Memory.

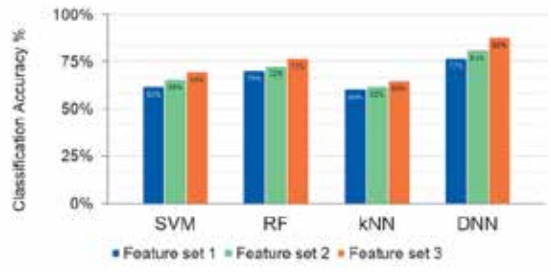

Figure 5 – Accuracy Vs Feature sets along with the Classifier Used

6. Conclusion

Acoustic monitoring is promising in monitoring bird migration, particularly at night time. Flight call identification is crucial, for example to estimate flock size and species, thus guessing flying heights to reduce the number of collisions with windmills. A novel Spectrogram-based Maximally Stable Extremal Regions (SMSER) feature extraction technique was developed in this work. The SMSER feature extraction technique was compared with two different Feature extraction techniques, namely

Table 1 – Comparison of Related and Current Work of Flight Call Classification

	Related Work	Current Work
Dataset	Bird-vox-fullnight	Bird-vox-fullnight
Computational power	GPU	CPU
Classification Technique	CNN (667k hidden units)	DNN (10k hidden units)
Test-set classification accuracy	90.48%	87.67%

SATF alone and a combination of SATF and SIFS. The DNN Classification result in this new feature extraction technique is 87.67%, and it is better than results from other Feature extraction methods. The work proposed in this research will help biologists to implement new methods to mitigate the collisions with wind turbines and get a sense of flight call counts of migratory birds in different continents. The novel feature extraction technique can be modified to account birds found locally according to their flight call duration and the frequency span. Usage of this feature extraction technique might require basic knowledge of how the MSER algorithm works to understand the concept behind it.

Acknowledgement

The authors wish to thank the IntelliSense Lab for supporting the project which was funded by the Senate Research Committee under the grant number SRC/LT/2018/01 of the University of Moratuwa.

Additionally, authors would like to mention the support of Dr. Sampath S. Seneviratne for sharing the domain knowledge regarding bird species and ecology.

References

1. Keen, S., Ross, J. C., Griffiths, E. T., Lanzone, M., & Farnsworth, A. (2014). "A Comparison of Similarity-based Approaches in the Classification of Flight Calls of Four Species of North American Wood-warblers (Parulidae)". *Ecological Informatics*, *21*, 25-33.

2. Farnsworth, A., Sheldon, D., Geevarghese, J., Irvine, J., Van Doren, B., Webb, K., .&Kelling, S. (2014). "Reconstructing Velocities of Migrating Birds from Weather Radar–A Case Study in Computational Sustainability". *AI Magazine*, *35*(2), 31-48.

3. Fink, D., Damoulas, T., Bruns, N. E., La Sorte, F. A., Hochachka, W. M., Gomes, C. P., & Kelling, S. (2014). "Crowdsourcing Meets Ecology: Hemisphere-wide Spatiotemporal Species Distribution Models". *AI magazine*, *35*(2), 19-30.

4. Salamon, J., Bello, J. P., Farnsworth, A., Robbins, M., Keen, S., Klinck, H., &Kelling, S. (2016). "Towards the Automatic Classification of Avian Flight Calls for Boacoustic Monitoring". *PloS one*, *11*(11), e0166866.

5. Larkin, R. P., Evans, W. R., & Diehl, R. H. (2002). "Nocturnal Flight Calls of Dickcissels and Doppler Radar Echoes Over South Texas in Spring". *Journal of Field Ornithology*, *73*(1), 2-8.

6. Wimmer, J., Towsey, M., Roe, P., & Williamson, I. (2013). "Sampling Environmental Acoustic Recordings to Determine Bird Species Richness". *Ecological Applications*, *23*(6), 1419-1428.

7. Farnsworth, A., & Lovette, I. J. (2005). "Evolution of Nocturnal Flight Calls in Migrating Wood-warblers: Apparent Lack of Morphological Constraints". *Journal of Avian Biology*, *36*(4), 337-347.

8. Emlen, J. T., & DeJong, M. J. (1992). "Countingbirds: The Problem of Variable Hearing Abilities (contando aves: El problema de la variabilidad en la capacidad auditiva)". *Journal of Field Ornithology*, 26-31.

9. Bas, Y., Devictor, V., Moussus, J. P., & Jiguet, F. (2008). "Accounting for Weather and Time-of-Day Parameters When Analysing Count Data from Monitoring Programs". *Biodiversity and Conservation*, *17*(14), 3403-3416.

10. Hochachka, W. M., Fink, D., Hutchinson, R. A., Sheldon, D., Wong, W. K., & Kelling, S. (2012). "Data-intensive Science Applied to Broad-scale

Citizen Science". *Trends in ecology & evolution*, *27*(2), 130-137.

11. Acevedo, M. A., Corrada-Bravo, C. J., Corrada-Bravo, H., Villanueva-Rivera, L. J., & Aide, T. M. (2009). "Automated Classification of Bird and Amphibian Calls Using Machine Learning: A Comparison of Methods". *Ecological Informatics*, *4*(4), 206-214.

12. Damoulas, T., Henry, S., Farnsworth, A., Lanzone, M., & Gomes, C. (2010, December). "Bayesian Classification of Flight Calls with a Novel Dynamic Time Warping Kernel". In *2010 Ninth International Conference on Machine Learning and Applications* (pp. 424-429). IEEE.

13. Bastas, S., Majid, M. W., Mirzaei, G., Ross, J., Jamali, M. M., Gorsevski, P. V., ... & Bingman, V. P. (2012, May). "A Novel Feature Extraction Algorithm for Classification of Bird Flight Calls". In *2012 IEEE International Symposium on Circuits and Systems (ISCAS)* (pp. 1676-1679). IEEE.

14. Stowell, D., & Plumbley, M. D. (2014). "Automatic Large-scale Classification of Bird Sounds is Strongly Improved by Unsupervised Feature Learning". *PeerJ*, *2*, e488.

15. Tanttu, J. T., Turunen, J., Selin, A., & OJANEN, M. (2006). "Automatic Feature Extraction and Classification of Crossbill (Loxia spp.) Flight Calls". *Bioacoustics*, *15*(3), 251-269.

16. 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.

17. Dufour, O., Artieres, T., Glotin, H., &Giraudet, P. (2013). "Clusterizedmel Filter Cepstral Coefficients and Support Vector Machines for Bird Song Identification". In *Soundscape Semiotics – Localization and Categorization* (No. 2013, pp. 89-93). InTech.

18. 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 Recognition Letters*, *31*(12), 1524-1534.

19. Ganchev, T. D., Jahn, O., Marques, M. I., de Figueiredo, J. M., & Schuchmann, K. L. (2015). "Automated Acoustic Detection of Vanellus chilensis lampronotus". *Expert systems with applications*, *42*(15-16), 6098-6111.

20. Digby, A., Towsey, M., Bell, B. D., & Teal, P. D. (2013). "A Practical Comparison of Manual and Autonomous Methods for Acoustic Monitoring". *Methods in Ecology and Evolution*, *4*(7), 675-683.

21. Lostanlen, V., Salamon, J., Farnsworth, A., Kelling, S., & Bello, J. P. (2018, April). "Birdvox-full-night: A Dataset and Benchmark for Avian Flight Call Detection". In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 266-270). IEEE.

22. Mohamed, A. R., Dahl, G. E., & Hinton, G. (2011). "Acoustic Modeling Using Deep Belief Networks". *IEEE transactions on audio, speech, and language processing*, *20*(1), 14-22.

23. Narayanan, A., & Wang, D. (2014). "Investigation of Speech Separation as a Front-end for Noise Robust Speech Recognition". *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *22*(4), 826-835.