*Pak. J. Agri. Sci., Vol. 58(4), 1395-1403; 2021*
*ISSN (Print) 0552-9034, ISSN (Online) 2076-0906*
*DOI:10.21162/PAKJAS/21.195*
http://www.pakjas.com.pk

# Automated birdsong clustering and interactive visualization tool

Ayesha Hakim[1,*] and Muhammad Tariq Mahmood[2]

[1]Department of Computer Science, MNS-University of Agriculture, Multan;
[2]Department of Zoology, Cholistan University of Veterinary and Animal Sciences, Bahawalpur
[*]Corresponding author's email: ayesha.hakim@mnsuam.edu.pk

Acoustic recordings of birds have been used by conservationists and ecologists to determine the population density and biodiversity of bird species in a region. However, it is hard to analyze and visualize the presence/absence of a specific bird species by aurally hearing these recordings even by an expert bird song specialist. In this paper, we present a computational tool to cluster and recognize bird species based on their sounds and visualize relationships of within-species and between-species sounds based on their similarity measures. The tool has been evaluated on two datasets of varying complexity containing acoustic recordings of eleven birds' songs and calls using various similarity measures. Principal Component Analysis (PCA) was used for feature selection. Euclidean distance, Mahalanobis distance, and cosine similarity among features was used for pair-wise similarity calculation. The results of similarity measures have been compared using 3-fold cross-validation and validated by spectrograms patterns obtained from frequency representation of acoustic recordings of the selected birds' songs and calls. Cosine similarity performed better to measure underlying patterns of birds' sounds and identify mutual relationship among species. It was concluded that the proposed tool can be used as a novel method for conversationalists, ecologists, ornithologists, and evolutionary scientists as well as tourists and bird watchers to recognize different birds' species, study their mutual relationship, locate the area with highest population density, estimating the predators, and biodiversity in a specific region.
**Keywords:** birdsong, data visualization, population density, biodiversity, IUCN, conservation, d3js.

## INTRODUCTION

Pakistan is enriched with a wide range of ecosystem and habitats hosting a broad diversity of bird species in the central Asia (Baig *et al.*, 2009). Despite the efforts of several international and local bodies to conserve birds of Pakistan, unfortunately they are quickly declining in number due to various threats including residential and commercial development, human intrusion, and pollution (IUCN, 2021). According to IUCN red list of threatened species (iucnredlist, 2021), out of total 659 native bird species of Pakistan, 32 are near-threatened, 26 are vulnerable, 10 are endangered, and 8 are critically endangered. The biggest threats include biological resource use through hunting and trapping, growing non-timber crops in agriculture, and climate change. To conserve threatened birds' species, there is a need to monitor habitats of threatened species and take appropriate actions such as area management, invasive species control, and education.

Birds use their voice as the main method of communication (Arriaga *et al*., 2015) that includes variants of songs and calls. Birds songs are typically loud, regularly repeated, persistent and often complex repertoires that they use to communicate with their mates only in the breeding season (Slater, 2000). On the other hand, calls are the sounds which they learn genetically that are typically less musical and are often just a single note (Boughman and Moss, 2003). The reasons of the bird's calls vary, such as, calling their chicks or mate, claiming territory, staying in touch with flock mates, scolding an intruder of the same species or different species, or announcing the presence of a predator. Bird's songs and calls can further be divided into phrases, syllables, and elements (Catchpole and Slater, 2003).

Birds songs and calls produce slight fluctuations in the air pressure, these vibrations can be turned into voltages that can be digitally recorded by a microphone and analysed by intelligent computer systems (Priyadarshani *et al.* 2018). By 'understanding' and differentiating between birds' sounds and calls, we may be able to detect and recognize their behavior, that may help in their protection and conservation. In this paper, we are presenting an automated system to recognize, and cluster different bird species based on their sounds. For this, we developed an unsupervised machine learning API that takes birdsongs audio files as input, normalize, and transform the data into lower dimension based on their variability and calculates the pair-wise similarity

Hakim, A. and M.T. Mahmood. 2021. Automated birdsong clustering and interactive visualization tool. Pak. J. Agri. Sci. 58:1395-1403.
[Received 7 Feb 2021; Accepted 5 Aug 2021; Published (online) 21 Sep 2021]

matrix depicting the relationship between birdsongs. Interestingly, we observed very distinct within-species and between-species sound variation produced by different bird species. The similarity matrix was used into the web services API to generate the online interactive graphical visualization in order to observe the relationship between different bird sounds that revealed interesting results. The work presented in this paper was performed under a New Zealand based project called Cacophony (cacophony, 2021), that aims to develop technologies to save birds from introduced predators such as, possums, rats, and stoats. However, the techniques presented in this paper might be used in any geographical region in the world.

The visualizations obtained by applying the proposed tool can be mapped on the geographical locations of birds that may be used by the department of conservation and environment to keep track of population density and bio-diversity of birds at a particular location. Our hypothesis is that by observing the change in population density at a specific location, the number of pests or any other factor affecting the bird's growth and well-being, can be tracked. Once the highly vulnerable areas are identified, the department of conservation and environment may take appropriate steps to eradicate pests/invasive species from that location without affecting other locations with a relatively healthier ecosystem. The visual representations may also be used by the department of tourism to construct new walking trails where there are more population of diverse bird species. These walking trails might be used by bird watchers and tourists to enjoy unique bird diversity at that region, and also by ornithologists to study maximum number of different birds sounds at one location.

Another potential use of the visualization tool is the behavioral analysis of birds. The communication patterns of different bird species are considered such as they belong to discrete categories (Barlow, 1977). However, close examination reveals some degree of variation even among and between the same species (Podos et al., 1992). Same as humans, the male and female birds have some degree of variations in the song types, duration, and tone. The female mate chooses the male based on their singing behavior (Ballentine *et al.*, 2004). The song variation among the same species might also be based on age, environment, health, immune function, parasitism, and secondary sexual characters (Garamszegi *et al.*, 2005). Interestingly, birds tend to sing at a higher pitch in urban regions than in the quite forest (Slabbekoom and den Boer-Visser, 2006). These variations play an important role for both functional and evolutionary analysis of birds' behavior.

We analyzed the song repertoire of two endangered and nine common bird species including North island brown kiwi (both male and female, scientific name: *Apteryx*), kakapo (*Strigops habroptilus*), Morepork (Māori name: ruru, scientific name: *ninox novaeseelandiae*), North island robin (*Petroica longipes*), Tui (*Prosthemadera novaeseelandiae*), North island kaka (*Nestor meridionalis*), Hihi (*Notiomystis cincta*), North island saddleback (*Philesturnus rufusater*), Marsh wren (*Cistothorus palustris*), Western meadowlark (*Sturnella neglecta*), and Horned lark (*Eremophila alpestris*) from a bird population in New Zealand. We extracted their audio features and used these features to cluster birds of the same species based on the similarity in their sounds. An interactive chordial graph has been generated that shows similarities within and between species through interesting patterns.

There has been previous research on automated recognition of birdsong such as, Sprengel *et al.* (2016) used spectrogram representation of 999 bird species as feature vector and used Convolutional Neural Networks as birdsong recognition methods. They tested the system on around 35,000 audio recordings and attained 69% mean average precision (MAP). Potamitis *et al.* (2014) used Hidden Markov Model on a feature set of Linear Prediction Cepstrum Coefficients (LPCC) to recognize birdsong of two bird species Robin and Vouliagmeni. They achieved 77% recall and 85% precision on Robin songs and 85% recall and 85% precision on Vouliagmeni songs on the audio recordings of around 42 hours duration. Murcia and Paniagua (2013) used Neural Network on MFCC features reduced by linear discriminant analysis (LDA) to recognize birdsong of 35 species. Dufour *et al.* (2013), Ganchev *et al.* (2012), and Ventura *et al.* (2015) used SVM, statistical log-likelihood ratio estimator, and HMM on MFCC features to recognise 35, 1, and 40 bird species respectively. Ulloa *et al.* (2016) used Spectrogram cross-correlation to identify birdsong of Screaming Piha (*Lipaugus vociferans*) based on a mean template derived from 10 standardized audio samples. They achieved 35% recall and 100% precision on audio recordings of 5 hours and 36 minutes duration. All these attempts tend to identify bird species based on their sounds, but it is hard to uniformly compare the outputs of these studies as some of them are evaluated on the data based on the number of bird calls/songs, while others are based on the number of hour of recordings. According to Priyadarshani *et al.* (2018), the different ways of data collection affect the outputs based on this data.

In this paper, we present an automated bird species recognition based on their sounds in open field with relatively low background noise. The analysis of result comes up with very interesting patterns that requires further research related to the relationship of birds sounds. In the literature, there has been no such attempt of developing automated tools to study relationship among different bird species based on the similarity of their songs and calls.

## MATERIALS AND METHODS

***Dataset*:** Two different bird song datasets of varying complexity in terms of duration of recordings and environmental noise has been used for experimental evaluation. The first dataset contains songs of two endangered

and one relatively common New Zealand bird species, including North Island brown kiwi (both male and female), kakapo (both booming and changing sounds), and Morepork. Most of the recordings were collected using automated recorders, but a few were recorded manually. This dataset contains 700 syllables of 10 different birds from seven basic call types and is available at (Priyadarshani *et al.* 2018). The average duration of each syllable-level audio recording is around 1 second. This dataset is of relatively good sound quality and is segmented into syllable level components.

The second dataset contains sounds of birds' species that are different from the primary dataset. The eight species (North Island robin, Tui, North Island kaka, Hihi, North Island, Saddleback, Marsh wren, Western Meadowlark, Horned Lark) in this dataset comprise of seven song birds and one parrot, which have complex songs and significant song diversity. This dataset has whole song segments instead of syllable level components. A series of unsegmented recordings comprising of consecutive calls from each species mentioned in the first dataset was also used for analysis. The average duration of each audio recording in this dataset is around 3.6 seconds. The second dataset is relatively harder in terms of noise and complexity of data.

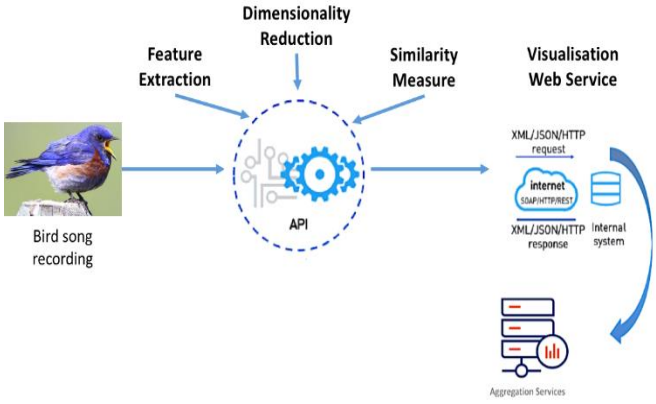



**Figure 1. Methodological framework for birdsong recognition and online interactive visualization.**

The methodological framework of the proposed system is presented in Fig. 1. The whole process is divided into two sub processes. The first subprocess is focused on data cleaning, normalization, feature extraction, and dimensionality reduction. In the second process, the web services used for visualization have been developed. Both processes are combined as a standalone aggregated service that is able to cluster bird species and visualize relationship within and between-species birdsongs.

***Feature Extraction*:** The short-term and mid-term features of each audio recording have been extracted using the approach similarly to (Ginnakopoulos, 2015). As birdsong and calls are temporal, it is important to select a suitable time window. We split each audio signal into time windows (frames). In the literature, the most widely accepted short-term frame (time window) size is 20 to 100msecs. We used a frame size of 50msec and a frame step of 25msec using overlapping framing. For each frame, 34 short-term features were extracted from each frame for robust representation of audio signal in varying environmental noises Alim *et al.*, 2018). As a result, each frame is represented as a feature vector of 34 elements each as shown in Table 1.

**Table 1. Complete list of 34 audio vectors extracted as short-term feature vectors for signal processing**

| Index | Name | Description |
|---|---|---|
| 1. | Zero Crossing Rate | The rate of sign-changes of the signal |
| 2. | Energy | The sum of squares of the signal values, normalized by the respective frames |
| 3. | Entropy of Energy | The entropy of sub-frame's normalized energies, interpreted as a measure of abrupt changes |
| 4. | Spectral Centroid | The center of mass of the spectrum |
| 5. | Spectral Spread | The measure of average speed of the spectrum |
| 6. | Spectral Entropy | The measure of spectral power distribution for a set of sub-frames |
| 7. | Spectral Flux | The squared difference between magnitude of two successive frames to measure how quickly the power spectrum of a signal is varying. |
| 8. | Spectral Rolloff | The fraction of bins at which the frequency lies below 85% of the magnitude distribution in the power spectrum |
| 9-21 | MFCCs | Mel Frequency Cepstral Coefficients forms a cepstral representation where frequency bands are distributed according to the mel-scale. |
| 22-33 | Chroma Vector | A 12-element facture vector indicating energy of each pitch class present in the signal. |
| 34 | Chroma Deviation | The standard deviation of the 12 Chroma coefficients |

Once the short-term features are extracted, the mid-term features are computed by calculating the two statistics of each short-term feature. The following statistics are computed: (a) the average value ($\mu$), (b) the standard deviation ($\sigma^2$). As a result, each frame is represented as a 68-dimensional feature vectors, where the first half of the values (in each frame) corresponds to the average value, while the second half to the standard deviation of the respective short-term feature. A long-term average is calculated with respect to all frames resulting in a one feature vector for each audio recording. Each of these long-term averages of mid-term feature vectors is fed into the dimensionality reduction technique to extract the most varying features out of the whole audio signal.

***Dimensionality Reduction:*** A 68-dimensional feature vector has been computed for each audio recording. The number of

feature vectors depend on the number of audio recordings. Each feature vector is normalized to 0-mean and 1-standard deviation. To improve the efficiency of the algorithms, the first step is to find a simplified representation of high-dimensional data in order to visualize and understand the relationships among multiple variables. Generally, in a multivariate dataset there is more than one variable measuring the same kind of behavior. The problem may be simplified by replacing such redundant groups of variables by a single new variable.

A standard technique to model data variation and analyze sets of datapoints in high dimensional spaces is Principal Component Analysis (PCA) (Jolliffe, 2011). PCA finds a new set of variables, called principal components (PCs), by identifying a linear transformation (translation, rotation, and scaling) of the original variables in the dataset. All principal components are mutually orthogonal, such that ideally there is no redundant information. In this case, no redundancy means that the principal components are uncorrelated with each other. Each component accounts for a maximal amount of variance in the observed variables that was not accounted for by the preceding components and is therefore uncorrelated with all of the preceding components.

The principal components are statistically independent to each other only for normal (Gaussian) random variables (Jolliffe, 2011). As a whole, the set of principal components form an orthogonal basis for the space of the original dataset. The resultant basis has maximum variance of the dataset along the first basis vector, and successively less variance amongst the following basis vectors. A scree graph was generated to select the suitable number of principal components in order to transform the data into low dimensional space while preserving the variation in the data. The first ten PCs were selected to transform the data leading to 10-dimensional feature vectors corresponding to each audio recording.

***Calculating Similarity Matrix:*** Cosine similarity has been used as a measure of similarity between two non-zero vectors of an inner product space normalized by the product of their magnitudes that measures the cosine of the angle between them (Dangeti, 2017). For any pair of real-valued vectors *x* and *y*, *t* is calculated as,

$$\mathrm{SM}(x, y) = 1 - \frac{x.y}{\| x \| \| y \|}$$

In the past, cosine similarity has been used successfully for speaker clustering and verification (Senoussaoui et al., 2014; Dehak et al., 2011). Unlike Euclidean distance, cosine distance regards only to the *shape* of the pattern but not to its magnitude and gives a fair measure to the frames with relatively low power (Vaizman et al., 2014; and Dongen and Enright, 2012). We computed the pairwise cosine similarity between the transformed feature vectors to get a square-form similarity matrix for each audio recording. This similarity matrix was used in web services for online visualization.

***Online Interactive Visualization:*** Extracting meaningful visualization based on the relationships between data variables is useful, especially in large datasets. After transforming the audio signals to a lower-dimensional space, a similarity matrix is calculated based on the pairwise cosine distances of feature vectors in the training set. This similarity matrix was converted into JavaScript Object Notation (json) format to be used in web services for online visualization (Crockford, 2006). Json is a light-weight, text-based data interchange format that makes the similarity matrix language-independent. Based on the similarity matrix, an interactive chord diagram is generated using the powerful data-driven-document (D3) (http://github.com/d3/d3-chord/) approach to visualize similarity among birdsongs in the browsers. D3 provides efficient scene transformation thus providing flexible animation, interaction, complex, and dynamic visualizations for the web. An initial version of the generated visualizations of the birdsongs are available online (https://cacophony.org.nz/birdsong-analysis-and-visualisation).

A chord diagram is a graphical method of displaying the inter-relationships between entities based on a matrix $m[i][j]$ of size $n * n$, representing a directed flow among a network of $n$ nodes. Each element of the matrix, $m[i][j]$, represents the flow of the $i^{th}$ node to the $j^{th}$ node. $m[i][j]$ must be nonnegative, though it can be zero if there is no flow from node $i$ to node $j$. In our case, $m[i][j]$represents similarity of birdsong content in the audio recording i to the audio recording $j$. The matrix is passed to *d3.chord*, that returns an array of chords. Each element of chord array represents bidirectional flow between two nodes $i$ and $j$, and returns zero if there is no flow. Each cord is an object with two sub objects: source and the target. The source and target have the following properties:

- startAngle - the start angle in radians
- endAngle - the end angle in radians
- value - the flow value $m[i][j]$
- index - the node index $i$, and subindex - the node index $j$

The chords are then passed to *d3.ribbon* to display the network relationships using colored ribbons. The returned array includes only the unique chords and the chord objects for which the value $m[i][j]$ is non-zero.

## RESULTS

We implemented our methods in Python with PyCharm IDE using the *pyAudioAnalysis* toolbox (Giannakopoulos, 2015). For visualization of bird species recognition based on their sound's similarities, d3js chord diagram representation has been used. Each node in the graph represents a single audio recording. Fig. 2 presents an interactive chordial graph, drawn on the basis of cosine similarity measures of birdsongs (hihi,

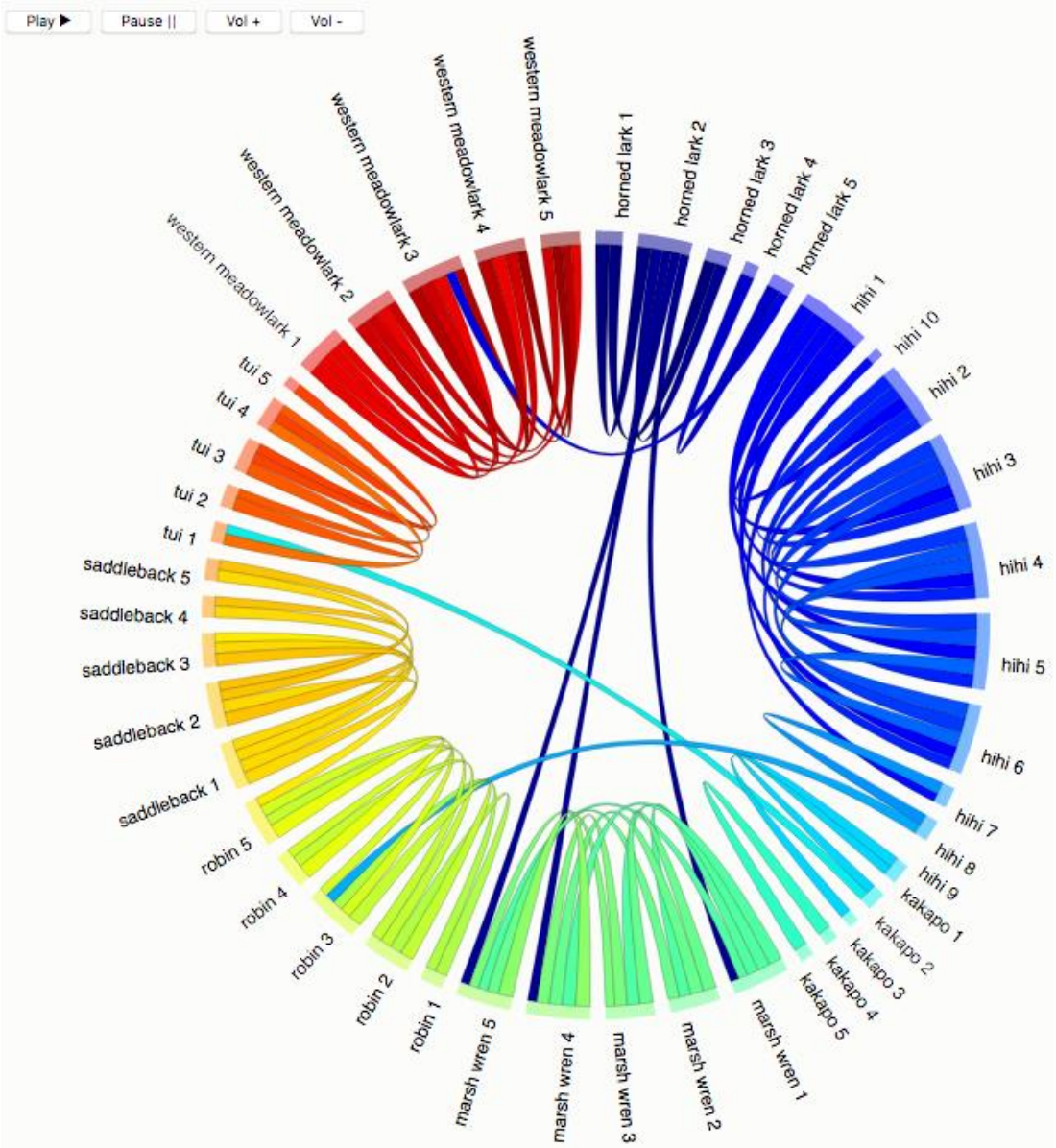



**Figure 2. An interactive chordial graph, drawn on the basis of similarity measures of birdsongs of selected species (hihi, kakapo, marsh wren, robin, saddle back, tui, western meadowlark, and horned lark), presenting clustering mutual relationship based on training data. For details, see the text.**

kakapo, marsh wren, robin, saddle back, tui, western meadowlark, and horned lark). On clicking any name, the image of that bird appears, the sound player plays the sound of the bird from the dataset, and a sound frequency spectrum is generated. The sound visualization uses the open-source API (AudioContext, 2021) to play an audio file, and AnalyzerNode to retrieve the frequency values.

The colors of the ribbons are based on the file names in the dataset such that similar file names generate similar colored ribbons. In the dataset, the name of file is based on the contents of the audio recording, for instance, *kiwi_female.wav* contains the audio recording of a female kiwi. In Fig. 2, the birds of the same species are clustered together, that gives the similar colored ribbons to the whole cluster. Each ribbon represents a relationship between the audio content at both ends. By aurally examining the overlapping ribbons between two different bird species (for instance, hihi and robin; tui and kakapo; and horned lark and marsh wren as shown in Fig. 2), we found the reasons of similarity include some matching notes in the syllable, silence, or some kind of noise in both audio recordings This study may further be extended to study the similarity between two different species based on

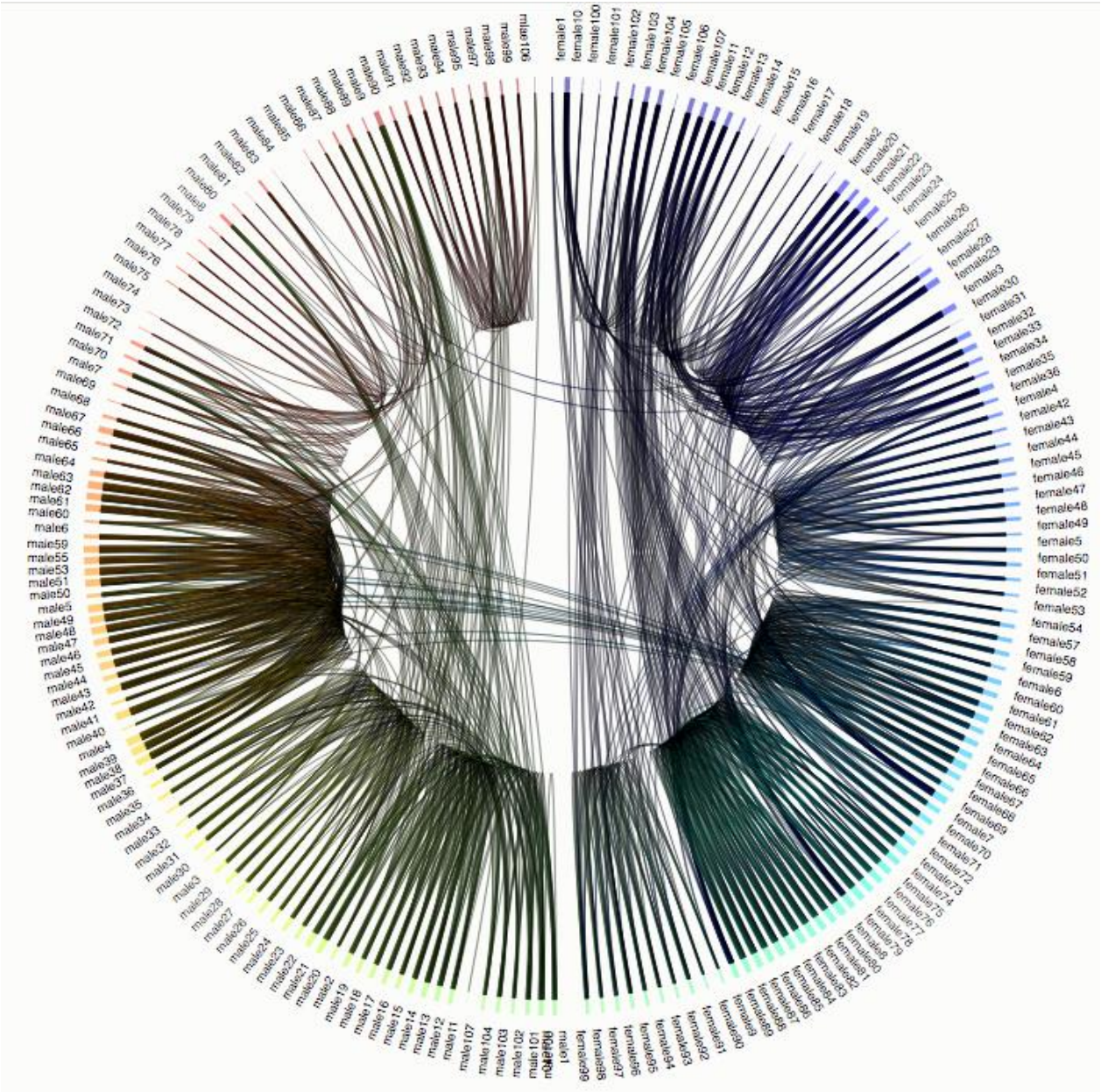


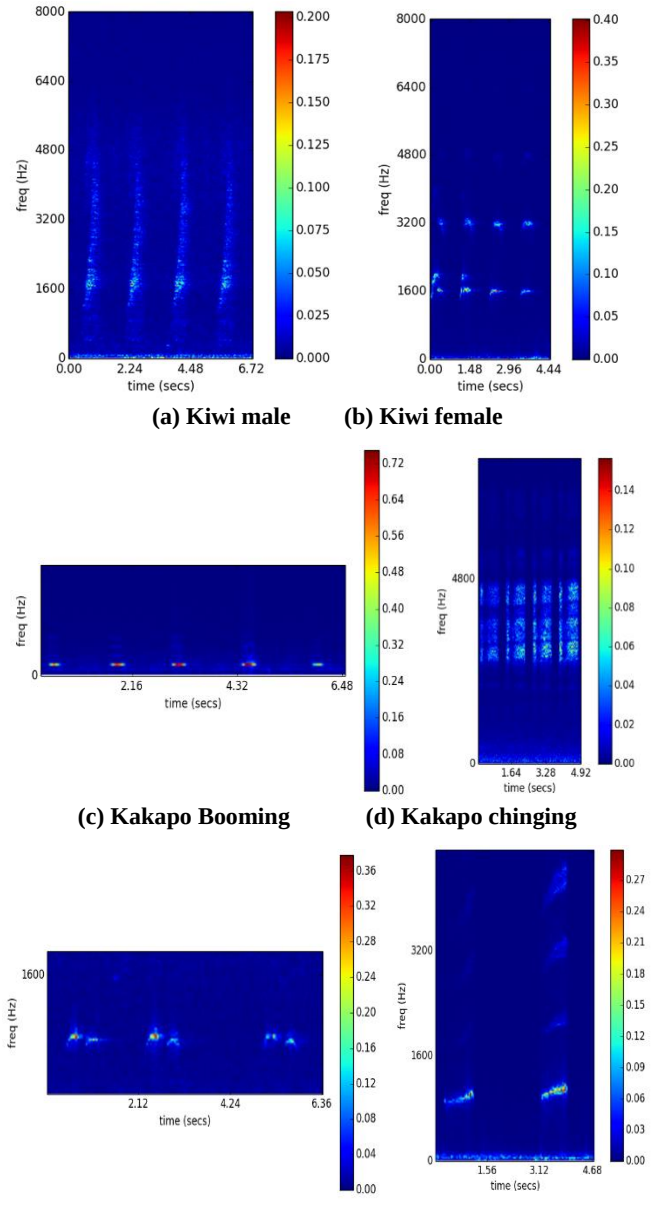
**Figure 3. Within-species song variants presented by distinction between male and female kiwi songs. Naming Convention: kiwi female: *female*; kiwi male: *male.* For details, see the text.**

evolution and differences between the birds of same species based on the difference in sex, age, habitat, climate conditions or other external factors. Fig. 3 presents a closer look at the 106 males and 107 females kiwi syllable-level songs. Interestingly, the songs of male and female kiwis are clustered into separate groups showing within-species song variation.

This is also obvious from the spectrogram's patterns obtained from the frequency representation of a discrete recording in a continuous series of male kiwi song (Fig. 4(a)) and female kiwi song (Fig. 4(b)). Fig. 4(c) and 4(d) represents kakapo deep booming call (*booming*) and loud wheezing call (*chinging*) sounds to attract mates, and 4(e) and 4(f) represents morepork song and call (*trill*) respectively. These results are promising that leads towards behavioral studies of within species and between species relationships and differences.

We compared the results of birdsong recognition and visualization with the ground truth data labelled by expert ornithologists for each audio recording. In most occurrences, the similarity in the songs of birds from two different species

is caused by the environmental noise. Since, the cosine similarity is computed on the mean-centered feature vectors, it is reduced to the measure of Pearson correlation coefficient $\rho_{AB}$ (Dangeti, 2017).




**(a) Kiwi male** **(b) Kiwi female**

**(c) Kakapo Booming** **(d) Kakapo chinging**

**(e) Morepork/ruru song** **(f) Morepork/ruru call (Trill)**

**Figure 4. Spectrogram representation of (a) kiwi male (b) kiwi female (c) kakapo booming (d) kakapo changing (e) Morepork song (f) Morepork call (trill) based on the frequency representation of a discrete recording of the continuous bird song.**

Table 2 presents mean accuracy ± standard deviation of bird species classification using 3-fold cross validation based on different similarity measures. Cosine similarly performed similarly to Pearson correlation and Mahalanobis distance measured better than sqEuclidean distance for birdsong clustering. The proposed method got the highest precision accuracy in the shortest processing time on the dataset of relatively low background noise.

**Table 2. Accuracy of bird species classifiers based on similarity measures**

| Sr. | Similarity measures | Mean accuracy (%) |
|---|---|---|
| 1. | Cosine Similarity | 98.5±2.55 |
| 2. | Pearson Correlation | 98.5±2.55 |
| 3. | Mahalanobis Distance | 80.0±4.44 |
| 4. | sqEculidean Distance | 75.0±1.52 |

***Conclusion:*** This paper presented a technique for clustering and recognizing bird species, and generated an interactive online visualization based on their acoustic signals. This tool was produced by aggregating the results obtained by an unsupervised machine learning technique with the web services APIs that took birdsong audio recordings as input and clustered them on the basis of the similarity among the audio features. Since birds' songs and calls are audio signals with some underlying patterns, we extracted useful bird sound features and used these features to detect specific patterns in bird's songs. Interestingly, sound of each species of bird was different; that enabled machine learning techniques to automatically distinguish between different bird species. By computational analysis of these patterns, bird sounds were clustered and identified based on their species and gender. The methods were evaluated on both series and segments of bird sounds that were of good quality with minimal environmental noise. The standard machine learning techniques were combined with data-driven-document visualization technique based on JavaScript to visualize similarity between different sound recordings and their mutual relationship with each other.

The variation in song type between and among species is common, but the visualization graphs revealed interesting results showing some variations *within-species* sounds. By 'understanding' and differentiating between within-species and between-species song and call variation, we might be able to detect and recognize their specific behavior. An online tool detecting within-species song type variation opened up a novel area of investigation for evolutionary and behavioral analysis. This paper described an initial experimental result of an ongoing project aiming at developing smart machines that are able to detect birdsong in the wild in the presence of noises such as wind, leaves, rain, thunderstorm, and other animals; classify an unknown bird species based on its sound similarity with other species; differentiate between birdsong and call; identify the reason of the call; and take appropriate actions if the call is for help.

The biggest challenge in this area was the lack of annotated local birdsongs/calls data, especially the data of rare and

endangered birdsongs. We are currently working on developing the first annotated dataset of Pakistan's native bird species. To facilitate data collection, we are also developing an easy-to-use gaming mobile application that can be used by the community to record images and sound of birds in their regions. The data obtained from this application would be stored on an online central repository that would be used as an input to the tool presented in this article.

***Acknowledgment:*** This project won the 'Most Creative Solution' award at the DataLand NZ, 2018. We acknowledge Prashant Khanna and Reza Rafeh for providing useful insights and feedback on this research. We acknowledge Priyadarshani *et al*. for providing their bird sounds datasets available for research.

## REFERENCES

Alim, S.A., Rashid, N.K.A. 2018. Some commonly used speech feature extraction algorithms. From Natural to Artificial Intelligence: Algorithms and Applications. IntechOpen. pp: 2-19.

Arriaga, J. G., Cody, M. L., Vallejo, E. E., & Taylor, C. E. 2015. Bird-DB: A database for annotated bird song sequences. ECOL INFORM. 27:21-25.

AudioContext, 2021. https://developer.mozilla.org/en-US/docs/Web/API/AudioContext. MDN Webdocs, (last modified on 11 June 2021, accessed on 21 June 2021)

Baig, M. B., Al-Subaiee, F. S. 2009. Biodiversity in Pakistan: key issues. J. Biodivers. 10:20-29.

Ballentine, B., Hyman, J., Nowicki, S. 2004. Vocal performance influences female response to male bird song: an experimental test. J. Behav. Ecol. 15:163-168.

Barlow, G. W. 1977. Modal action patterns. How animals communicate. *In* SEBEOK, T. A., ed., Indiana University Press. Bloomington. pp. 98-134.

Bostock, M., Ogievetsky, V., Heer, J. 2011. $D^3$ data-driven documents. IEEE Trans Vis Comput Graph.12:2301-2309.

Boughman, J. W., Moss, C. F. 2003. Social sounds: vocal learning and development of mammal and bird calls. *In* Acoustic communication. Springer, New York, NY. pp. 138-224.

Cacophony. 2021. https://cacophony.org.nz (accessed on 21.06.2021)

Catchpole, C. K., Slater, P. J. 2003. Bird song: biological themes and variations. 2nd ed. Cambridge university press.

Crockford, D. 2006. The application/json media type for javascript object notation (json). no. RFC 4627.

Dangeti, P. 2017. Statistics for Machine Learning: Techniques for exploring supervised, unsupervised, and reinforcement learning models with Python and R. Packt Publishing ltd.

Dufour, O., Artieres, T., Glotin, H., Giraudet, P. 2013. Clusterized mel filter cepstral coefficients and support vector machines for bird song identification. *In* Proc. 1st workshop on Machine Learning for Bioacoustics. 951:89-93.

Ganchev, T., Mporas, I., Jahn, O., Riede, K., Schuchmann, K.-L., Fakotakis, N. 2012. Acoustic bird activity detection on real-field data. *In* Maglogiannis, I., Plagianakos, V. and Vlahavas, I. (eds), Artificial intelligence: theories and applications. Springer. pp:190-197.

Garamszegi, L. Z., Heylen, D., Møller, A. P., Eens, M., De Lope, F. 2005. Age-dependent health status and song characteristics in the barn swallow. Behav. Ecol. 16:580-591.

Graciarena, M., Delplanche, M., Shriberg, E., Stolcke, A., Ferrer, L. 2010. Acoustic front-end optimization for bird species recognition. *In* Proc. IEEE Int. Conf. Acoust. Speech Signal Process. IEEE. pp.293-296.

Iucnredlist. International Union for Conservation of Nature and Natural Resources. 2021. ISSN 2307-8235. https://www.iucnredlist.org/ (accessed on 21.06.2021)

Jolliffe, I. 2011. Principal component analysis. *In* Intl Ency Stat Sci. (ISO4). Springer, Berlin, Heidelberg. pp:1094-1096

Murcia, R. H., Paniagua, V. S. 2013. Bird identification from continuous audio recordings. *In* Proc. 1st workshop on Machine Learning for Bioacoustics joint to the 30th ICML. pp: 96-97.

N. Dehak, P. Kenny, R. Dehak, P. Dumouchel, P. Ouellet. 2011. Front-end factor analysis for speaker verification. *In* IEEE-ACM T AUDIO SPE. 19:788-798.

Podos, J., Peters, S., Rudnicky, T., Marler, P., Nowicki, S. 1992. The organization of song repertoires in song sparrows: themes and variations. J. Ethol. 90:89-106.

Potamitis, I., Ntalampiras, S., Jahn, O., Riede, K. 2014. Automatic bird sound detection in long real-field recordings: applications and tools. J. Appl. Acoust. 80:1-9

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

Senoussaoui M, Kenny P, Stafylakis T, Dumouchel P. 2014 A Study of the cosine distance-based mean shift for telephone speech diarization, audio, speech, and language processing. IEEE/ACM Trans. 22:217–227.

Slabbekoorn, H., Den Boer-Visser, A. 2006. Cities change the songs of birds. Curr. Biol. 16:2326-2331.

Slater, P.2012. Bird song and language. In *The Oxford handbook of language evolution*. Oxford. DOI: 10.1093/oxfordhb/9780199541119.013.0008

Sprengel, E., Jaggi, M., Kilcher, Y. and Hofmann, T. 2016. *Audio based bird species identification using deep learning techniques* (No. CONF, pp. 547-559).

Ulloa, J. S., Gasc, A., Gaucher, P., Aubin, T., Réjou-Méchain, M., Sueur, J. 2016. Screening large audio datasets to determine the time and space distribution of screaming piha birds in a tropical forest. Ecol. Inform. 31:91-99.

Vaizman, Y., McFee, B., Lanckriet, G. 2014. Codebook-based audio feature representation for music information retrieval. IEEE-ACM T AUDIO SPE. 22:1483-1493.

Van Dongen, S. and Enright, A.J., 2012. Metric distances derived from cosine similarity and Pearson and Spearman correlations. *arXiv preprint arXiv:1208:3145*.

Ventura, T. M., de Oliveira, A. G., Ganchev, T. D., de Figueiredo, J. M., Jahn, O., Marques, M. I., Schuchmann, K.-L. 2015. Audio parameterization with robust frame selection for improved bird identification. Expert Syst. Appl. 42:8463-8471.