BIRD CALLER ID – IDENTIFYING SHORT BIRD CALLS USING MFCC CLASSIFICATION TECHNIQUES

Henry Finston Perry
Department of CSC,
University of Victoria
khan@uvic.ca

Paul Henderson
Department of CSC,
University of Victoria
ptmhende@uvic.ca

Jamila Tomines
Department of CSC,
University of Victoria
jtomines@uvic.ca

ABSTRACT

Music Information Retrieval tools are currently used in the field of environmental conservation, through forest bioacoustics and birdcall recognition. The Bird Caller ID project aims to identify bird species using techniques and tools used by MIR researchers globally, especially the open-source project BirdNET-Lite. Furthermore, Bird Caller ID may be expanded to involve soundscape differentiation, geolocation information, or learning through educational gaming.

1. INTRODUCTION

Environmental conservation efforts that are necessary for ecological sustainability, awareness, and policy creation continue to be threatened by a lack of resources. The local provincial government in British Columbia spends an insignificant amount of funds on preservation efforts [1]. Furthermore, the advent of the COVID-19 pandemic has significantly dampened efforts of citizen-scientist birdwatchers due to restrictions [2]. Additionally, most of the general population lack the specific technical knowledge and education to effectively collect data from a local ecosystem. These complications indicate some areas in which current conservation efforts are lacking.

We propose a solution that monitors birds using digital signal processing, sound information retrieval techniques, and machine learning. Proposed work to identify bird species from audio recordings has been done involving noisy environments, but with little practical applications about larger data sets [3]. Other solutions have included spatio-temporal bird distribution model analyses [4]. Previous work can be combined and improved upon.

Methods to alleviate the financial and technical effort to manually monitor local habitats are important to decrease the shortage of data required for positive environmental change. Policy makers and environmental advocates need convincing data to strengthen their assertions. Indicator species help measure environmental conditions at a given location [5]. Therefore, observing indicator species helps conduct wildlife preservation efforts [6]. Birds are environmental indicators that can be observed by their unique appearances and sounds [4].

A method to process bird sound data can help monitor local habitats. Bird sounds have different meanings and may be categorized into calls and songs. Bird songs are important to portray territorial defense and mate attraction [7]. These songs have been used to model trait elaboration to predict sexual selection and reproductive success [8]. Enabling ease of bird observation through an intuitive and convenient method can encourage more people to bird watch. Birdwatching tourism can improve the environmental and financial aspects of the community through biodiversity education and stimulate incentives for successful natural habitat protection and preservation efforts [9]. Being able to record, filter, recognize key aspects, locate, and characterize bird sounds can help account for and monitor them.

2. TEAM MEMBERS & ROLES

Everyone contributes approximately one third to project documents each/edits other team members work/finds 3-5 relevant sources and cites them properly in the reference doc.

2.1 Henry Finston Perry

2.1.1 Process Analyst

Responsible for keeping the contribution document up to date and organized. Makes sure that team members are doing a fair amount of work throughout the different deliverables.

2.1.2 Devil's Advocate

Responsible for asking harder questions, thinking outside the box, and always considering whether there is another way of doing things. Should not over-purposefully hinder the project, but at least once per meeting should offer if another way of doing things might be explored.

2.2 Paul Henderson

2.2.1 Energizer

Checks in with the team at minimum weekly, describes goals for that week, makes sure everyone gets their opinion in during team meetings and keeps a positive attitude and environment.

2.2.2 Recorder

Takes meeting minutes when the team has project meetings, shares those minutes within 2 days of the meeting to the rest of the team on Notion.

2.3 Jamila Tomines

2.3.1 Timekeeper

Manages deadlines for the project using Notion, announces to the team both 1 week and 2 days before a project task is due, can/should be direct with team members who are falling behind.

2.3.2 Communicator

Organizes emails to George. After a project task is complete, makes sure the team is happy with a final product before being the person to submit the task on Brightspace. Confirms with the team after submitting.

3. TOOLS & LITERATURE

The topic of bird sound identification is well documented, yet we are interested in testing both new and established techniques to identify bird species in isolated recordings as well as longer soundscape recordings. This project will use Python 3.8 because of its readability, familiarity, and use in reviewed literature.

We will be referencing work from the BirdNET[1] project throughout our implementation. Stefan Kahl is a creator and researcher with BirdNET and has provided an open-source version of the app on GitHub called BirdNET-Lite[2], which uses the TensorFlow Lite[3] machine learning platform for fast, on-device inference. The existing research and suite of open-source tools currently available will provide a solid starting point and serve as a reference for the performance of our implementation.

Another useful resource is the bank of research from participants of the annual BirdCLEF[4] conference and challenge, which focuses on developing machine learning algorithms to identify avian vocalizations in continuous soundscape data to aid conservation efforts worldwide. BirdCLEF uses the Xeno-canto[5] sound library as a standard reference point for a near-complete collection of bird sounds, which we also plan to incorporate into our training dataset.

One of the results of the later BirdCLEF conferences was the discovery that integrating metadata, particularly geolocation and date/time of recording, drastically improved classification accuracy [10, 11, 12]. Our original idea also incorporated geographic filtering, though this has been left as a possible advanced topic once the fundamental application is built and tested.

4. ADVANCED TOPICS

Bird Caller ID has the potential to scale into several different variations. The base functionality would involve recognizing bird sounds using MIR techniques stated above in Tools and Literature such as using spectral peak tracks [13] or sinusoidal tonals [14]. Further implementations could be more expansive to discern and determine bird calls existing in a soundscape of varying background sounds and noises. Other iterations could also involve an educational component where users can play a game involving identifying bird sounds within a soundscape themselves [15, 16]. This game could help users appreciate the bioacoustics involved with a healthy, thriving natural environment [17]. Bird Caller ID could also involve GIS (Geographic Information System) functionality, where bird sounds are represented as an overlay on Google Maps or an equivalent interface helping to create an immersive bird sound experience while exploring different areas of the world.

5. PROGRESS REPORT (03/22)

The Bird Caller ID project is under development and aspects have been addressed. A GitHub repository has been created with a Jupyter notebook and a data set. The data set was downloaded from Xeno-Canto's web repository.

5.1 Accomplishments

Initially, the scope of the bird sounds downloaded began with those from British Columbia. Once accumulated into a directory, information was extracted from these MP3 files with the Python package *mutagen.mp3*. Among the information gathered from the MP3 files were audio length (in seconds), channels, bitrate, sample rate, layer, bitrate mode, protected, and sketchy values. Some of the MP3 files had a bitrate mode set to *BitrateMode.UNKNOWN* which meant that the audio bitrate was guessed based on the first frame. The sketchy value, if true, meant the file may not be valid MPEG audio. The MP3 objects could be played with *IPython.display* in Jupyter Hub. However, the MP3 objects were not of the correct format to be processed through the methods learned in class.

Classification labels for bird sounds were decided according to tone. Tones include whistled, hooting, clicking, burry or buzzy, nasal, noisy, and polyphonic. Whistled tones appear over a period and only increase in pitch slightly towards the middle of the duration. Hooting tones are lower-pitched whistles. Clicking is a short period of time of multiple pitches at once, similar to a wood-pecker. Burry or buzzy tones appear sinusoidal in pitch and last for a moderate duration. Nasal tones appear like multiple simultaneous whistles at different pitches. Noisy tones are varying pitches at different times and have no clear pattern. Polyphonic tones are made by multiple separate blending sounds simultaneously, usually one from each lung. These polyphonic sounds are dissimilar in shape, irregularity, and space. They may also be simultaneously rising and falling.

[1] https://birdnet.cornell.edu
[2] https://github.com/kahst/BirdNET-Lite
[3] https://www.tensorflow.org/lite
[4] https://www.kaggle.com/c/birdclef-2021
[5] https://xeno-canto.org

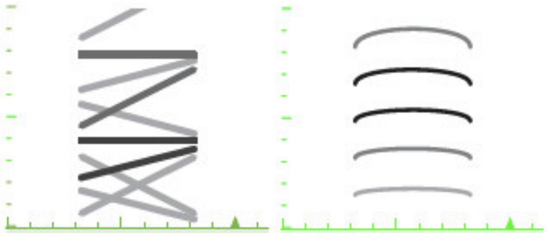

Figure 1. Polyphonic (left) vs. Nasal (right) Tones

Some audio information from the MP3 files was used to create a plot. Plots generated through *numpy* and *matplotlib* were created to visualize audio information. Audio file length and sample rate were used to calculate sample intervals and static frequencies were plotted on various subplots. An additional subplot was made with *numpy.fft.fft*.

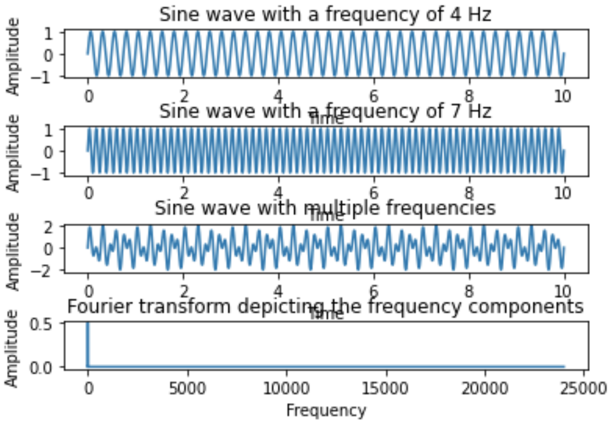

Figure 2. Subplots using sample rate and length

The MP3 files were converted to WAV files through various methods. Since the MP3 objects were found not to be directly usable for plotting and audio processing, the initial solution was to convert them to WAV files. This process began with using *pydub* and *ffprobe*. The *pydub* package has *AudioSegment* which takes MP3 files and enables their export to WAV file formats. The other method of file conversion from MP3 to WAV was through various internet sites and Audacity.

5.2 Challenges

The first main issue faced in the development was handling the massive amount of audio samples available on Xeno-Canto. The scope of the project has thus been scaled down to analyze samples from lower Vancouver Island, rather than the entire province of British Columbia as initially decided. This is because the gigabytes of MP3 files simply could not be dynamically processed on a remote server in a reasonable amount of time. This also led to the decision to batch-process (using FFT) all the audio samples into text files for more efficient queries. The drawback with this technique is the sample library will no longer be dynamic, though the proof-of-concept still holds in this static, scaled down version of the project.

The second challenge faced was how to go about running an FFT process on over 1000 MP3 files in a reasonable amount of time. Initial attempts to use Audacity to batch-export the text files were unsuccessful, so the next step is to write a Python script to process the files. Using Python necessitates converting all the MP3 tiles to WAV format first, but this is a much easier task in Audacity.

5.3 Upcoming Goals

The initial idea of Bird Caller ID to dynamically analyse the database of sounds has proven difficult to accomplish. There are so many files to process, the team has decided the next course of action is to pre-process the audio files and convert that information into text format. This will immensely improve the efficiency of both the computation time and storage of Bird Caller ID. Once this data is pre-processed it will be used to accomplish the primary objective of the project, predicting the species of bird based on a given audio sample input. This will be done using techniques learned in class such as doing FFTs and pitch shifting to determine tonal qualities of the input as well as tempo estimation to determine patterns observed in the samples. This quantitative pitch and tempo data will help determine and associate with more qualitative such as differing tones such as a whistle or hoot and patterns corresponding to whether the sample is a bird call or song. The processing done to the inputted sample will be matched against the pre-processed text information, and a closest match will be determined based on techniques such as RMS, MAP, and the smallest edit distance.

If needed, a more basic level project would be to distinguish between a handful of distinct bird species such as owl vs. robin, still using techniques learned from class at a much smaller scope. Finally, some advanced ideas that the team hopes to accomplish given available resources will be to keep a single recording of each bird species that can be played (this is about 125 out of the 1000 total recordings used for pre-processing), and to generate an image of the bird that is predicted by Bird Caller ID.

6. FINAL REPORT (04/22)

6.1 Work Completed and Results

Several main objectives were completed during this project to process bird sounds and classify them. These were data pre-processing, data aggregation, and bird type classification.

Data pre-processing was done through Audacity. Through Audacity, we converted 1000 MP3 files from Xeno-Canto to WAV files. These audio files contained sounds of birds that originated from Victoria and the surrounding area.

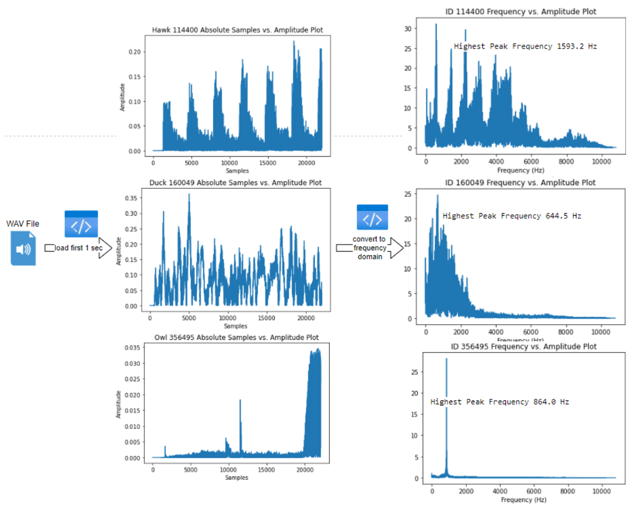

Figure 3. Audio Pre-processing of WAV Files: 114400.wav (Hawk), 160049.wav (Duck), and 356495.wav (Owl) to the Time Domain and Frequency Domain

Once converted, WAV file plots in Python using *matplotlib* showed patterns characteristic of the literature tone patterns mentioned previously. For example, there was a repeating pattern in some of the audio signals. In contrast, some of the audio samples had no clear repetition, which was characteristic of other tones. Furthermore, the Fourier Transform implemented by *scipy.fft* was used to convert the first one second of our time domain signals to the frequency domain. From this frequency domain plot, we determined the highest peak frequency by creating a filter for peaks above 80% of maximum peak amplitude and then averaged to get the central peak.

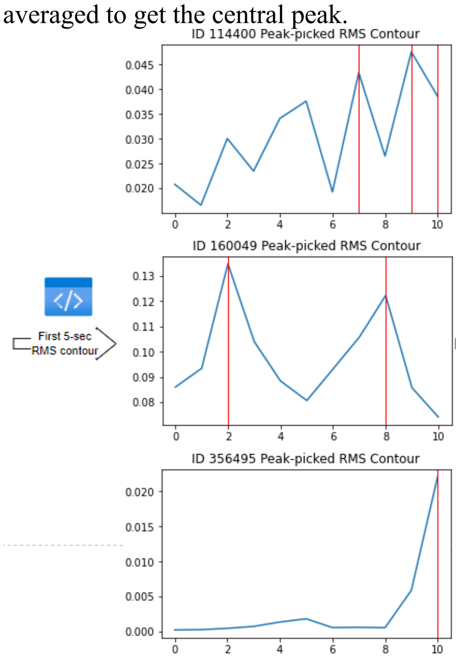

Figure 4. RMS Contour of First 5 seconds of Three out of 1000 WAV Files: 114400.wav (Hawk), 160049.wav (Duck), and 356495.wav (Owl)

Next, we took the first five seconds of the original WAV files and processed it with an RMS contour. The number of peaks were enumerated and averages between the peaks were calculated, with the previously defined threshold. The other approach for data pre-processing involved taking the *Linear SVC* MFCC of the first five seconds of the WAV files.

Data was aggregated into audio feature matrices and class vectors. For the first approach, the audio feature matrix was composed of maximum, number, and average inter-distance of peaks. The second approach had individually averaged MFCC data for proper formatting in the audio feature matrix. The class target vector was used with both approaches individually.

From the data aggregated, models were created to predict birds from 71 distinct types. These types comprised of 'Goose', 'Swan', 'Duck', 'Shoveler', 'Wigeon', 'Mallard', 'Pintail', 'Teal', 'Merganser', 'Quail', 'Grebe', 'Bittern', 'Heron', 'Cormorant', 'Hawk', 'Eagle', 'Rail', 'Sora', 'Coot', 'Oystercatcher', 'Plover', 'Killdeer', 'Surfbird', 'Dunlin', 'Sandpiper', 'Gull', 'Owl', 'Hummingbird', 'Kingfisher', 'Sapsucker', 'Woodpecker', 'Flicker', 'Falcon', 'Phoebe', 'Flycatcher', 'Pewee', 'Vireo', 'Jay', 'Crow', 'Raven', 'Waxwing', 'Chickadee', 'Swallow', 'Martin', 'Bushtit', 'Warbler', 'Kinglet', 'Wren', 'Nuthatch', 'Creeper', 'Catbird', 'Starling', 'Thrush', 'Redwing', 'Robin', 'Sparrow', 'Pipit', 'Finch', 'Crossbill', 'Goldfinch', 'Siskin', 'Longspur', 'Junco', 'Towhee', 'Chat', 'Meadowlark', 'Oriole', 'Blackbird', 'Cowbird', 'Waterthrush', and 'Yellowthroat'.

Figure 5. Prediction from the MFCC SVC approach with a resulting 22.3% probability that input was “Hawk”

The first and second classification approaches had accuracies of approximately 11% and 47.5% respectively. The number of neighbours used for the KNN approach was 22 as this was roughly the square root of the number of samples. The final resulting output of Bird Caller ID’s classification predictor can be seen below with the corresponding confidence probability of the prediction.

6.2 Discussion of Bird Classification

In general, the team is highly satisfied with the classifier and the overall learning experience. While the classifier is reasonably accurate at identifying short, repetitive bid calls, the available processing power and time limited the scope of the project quite severely.

In retrospect, our attempt to train the classifier with a population-proportional dataset may have hindered the classifier; instead, a more effective approach would have been to train the classifier with a more limited category space as well as a large, equal number of samples per category with the posterior probability considered after training. That way, the classifier would have been more adept at identifying all species, not just the most populous species in the Greater Victoria area. Based on existing literature, the team perused both MFCC and KNN techniques simultaneously [18].

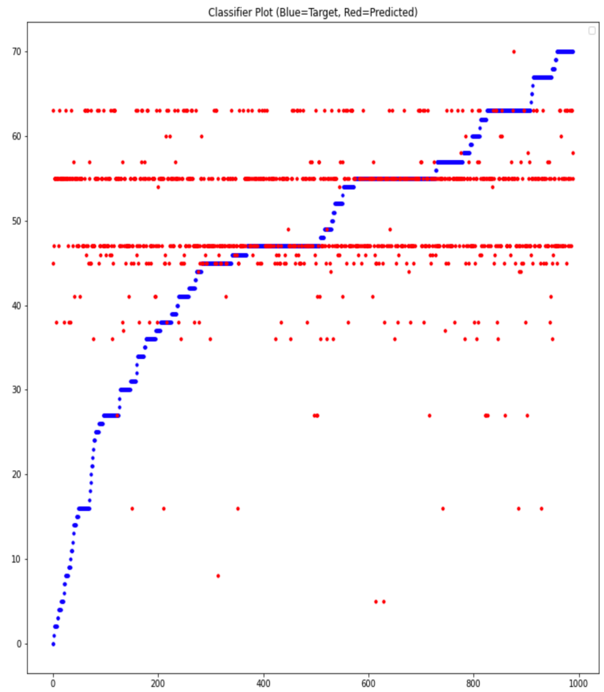

Figure 6. Scatter plot of Sample Number vs. Bird Type of Predicted (red) and Target (blue) KNN Bird Classifications

However, the KNN classifier tended to cluster around targets with the greatest number of samples (Figure 6) and was therefore limited at 11% classification accuracy. The MFCC technique showed better results given the nature of the dataset.

Finally, while any number of classification techniques could have been used (such as Naïve Bayes or Decision Trees), the latest version of the classifier uses Linear SVM. This technique was fast and produced the most accurate results during tests but is by no means the definitive way of making predictions from our dataset; which classification technique is the absolute most effective in this context is yet to be determined.

6.3 Conclusions and Future Work

Bird Caller ID was an interesting exploration of MIR procedures, machine learning techniques, and data visualisation methods. The team was able to learn a lot about the world of MIR, such as identifying useful features to observe in audio data to the importance of using large enough data sets for meaningful data analysis.

There are several avenues Bird Caller ID can explore for future work. Continuing in a similar direction, the project could be expanded by working with larger data sets. This would improve the machine learning techniques used to create classifiers and analyse the accuracy of their predictions. A main limitation of this project is that there were too many categories with limited samples for each, instead it would be more beneficial to decrease the number of categories to around three and make sure that there are at least 100 samples per category. After improving the prediction classifiers' accuracy, a further step would be to create a more enjoyable user interface for the application. This enhanced UI would help with Bird Caller ID's goal for being an educational tool to raise awareness for ecological preservation. Creating an intuitive and easy to use interface would also allow the team to test Bird Caller ID with participants such as amateur and professional bird connoisseurs.

To create a minimal viable product, the team limited the scope of Bird Caller ID to only species of birds found around Victoria, BC, Canada. This project could be further expanded to explore bird species from all over Canada or even the world. This increase in scope would require greater computational power and time, however it would help create a more useful bird classification tool. Currently Bird Caller ID does best at identifying short, rhythmically repetitive bird calls as it takes short snippets of audio and runs MFCC on them. Thus, the project could be expanded to classify bird songs by performing more melodic analysis such as chroma analysis on longer audio segments.

Bird Caller ID provided opportunities for our team to grow as MIR researchers, AI developers, and computer science professionals. With the skills and knowledge gained from this course and project, the sky is the limit.

7. REFERENCES

[1] "Ministry of Environment and Climate Change Strategy," April 2021. [Online]. Available: https://www.bcbudget.gov.bc.ca/2021/sp/pdf/ministry/env.pdf. [Accessed 6 February 2022].

[2] M. Basile, L. F. Russo, V. G. Russo, A. Senese and N. Bernardo, "Birds seen and not seen during the COVID-19 pandemic: The impact of lockdown measures on citizen science bird observations," *National Library of Medicine,* vol. 256, pp. 109079-109079, 2021.

[3] L. Neal, F. Briggs, R. Raich and X. Z. Fern, "Time-frequency segmentation of bird song in noisy acoustic environments," *IEEE International Conference on Acoustics, Speech and Signal Processing,* pp. 2012-2015, 2011.

[4] N. Ferreira, L. Lins, D. Fink, S. Kelling, C. Wood, J. Freire and C. Silva, "BirdVis: Visualizing and Understanding Bird Populations," *IEEE Transactions on Visualization and Computer Graphics,* vol. 17, no. 12, pp. 2374-2383, December 2011.

[5] "Indicator Species," Britannica, [Online]. Available: https://www.britannica.com/science/indicator-species. [Accessed 6 February 2022].

[6] M. Sankupellay and D. Konovalov, "Bird Call Recognition using Deep Convolutional Neural Network, ResNet-50," *ResNet-50,* 2018.
[7] B. Elodie, T. S. Osiejuk, F. Rybak and T. Aubin, "Are bird song complexity and song sharing shaped by habitat structure? An information theory and statistical approach," *Journal of Theoretical Biology,* vol. 262, no. 1, 2010.
[8] M. Soma and L. Z. Garamszegi, "Rethinking birdsong evolution: Meta-analysis of the relationship between song complexity and reproductive success," *Behavioral Ecology,* vol. 22, no. 2, pp. 363-37, April 2011.
[9] C. H. Sekercioglu, "Impacts of birdwatching on human and avian communities," *Environmental Conservation,* vol. 29, no. 3, pp. 282-289, 2002.
[10] S. Kahl, T. Denton, H. Klinck, H. Glotin, H. Goëau, W.P. Vellinga, R. Planqué and A. Joly, "Overview of BirdCLEF 2021: Bird call identifica tion in soundscape recordings." *CLEF 2021 – Con ference and Labs of the Evaluation Forum*, Bucharest, Romania. 2021.
[11] S. Kahl, T. Wilhelm-Stein, H. Hussein, H. Klinck, D. Kowerko, M. Ritter and M. Eibl, "Large-Scale Bird Sound Classification using Convolutional Neural Networks," *CLEF 2017 Conference and Labs of the Evaluation Forum*, Dublin, Ireland. 2017.
[12] S. Kahl, T. Wilhelm-Stein, H. Klinck, D. Kowerko and M. Eibl, "A Baseline for Large-Scale Bird Species Identification in Field Recordings," *CLEF 2018 – Conference and Labs of the Evaluation Forum*, Avignon, France. 2018.
[13] Z. Chen and R.C. Maher, "Semi-Automatic Classification of Bird Vocalizations Using Spectral Peak Tracks," The Journal of the Acoustical Society of America, vol. 120, (5), pp. 2974-2984, 2006.
[14] P. Jančovič and M. Köküer, "Automatic Detection and Recognition of Tonal Bird Sounds in Noisy Environments," EURASIP Journal on Advances in Signal Processing, vol. 2011, (1), pp. 1-10, 2011.
[15] D. Pires, B. Furtado, T. Carregã, L. Reis, L.L. Pereira, R. Craveirinha, and L. Roque, "The blindfold soundscape game: A case for participation-centered gameplay experience design and evaluation," in 2013, DOI: 10.1145/2544114.2544122.
[16] "Soundscape Design in an AR/VR Adventure Game," ProQuest. https://www.proquest.com/openview/144468db02543ff4b791a52102359fcb [Accessed 7 February 2022].
[17] Z. Burivalova, E.T. Game and R.A. Butler, "The sound of a tropical forest: Recording of forest soundscapes can help monitor animal biodiversity for conservation," Science (American Association for the Advancement of Science), vol. 363, (6422), pp. 28, 2019.
[18] Thiruvengatanadhan, R. (2017). Speech/Music Classification using MFCC and KNN. *International Journal of Computational Intelligence Research*, *13*(10), 2449–2452.