

Mel-log energies analysis of authentic audible intrusion activities in a Malaysian forest

Amirul Sadikin Md Afendi¹, Marina Yusoff², Megawati Omar³

¹Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Malaysia

²Advanced Analytic Engineering Center (AAEC), Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Malaysia

³Academy of Language Studies, Universiti Teknologi MARA, Malaysia

Article Info

Article history:

Received Oct 20, 2019

Revised Dec 28, 2019

Accepted Feb 8, 2020

Keywords:

Acoustic features

Anti-poaching

Audio processing

Sound event detection

Wildlife protection

ABSTRACT

Wildlife has been endangered due to illegal activities. This requires more effective surveillance measures. Felling timber and poaching are regular illegal activities but challenging to detect. Hence authorities should resort to modern technologies such as employing autonomous surveillance to stop them. The Malaysian forest audio data were recorded to lay a foundation in initiating a cheaper and practical approach. Hence this paper reports the collection, processing and analysis of audio data in preparation to develop an autonomous sound event detection system. The recording was an emulation of possible illegal activities in a reserved forest. Sounds of chainsaw and hand hatchet cutting tree trunks were taken. It was found that there was a distinct pattern in the Mel-log energies audio feature of the sound, which could be used to identify illegal activities. Thus, it is believed that a detection through audio is a possible approach to be employed as one of the methods to stop illegal activities in the tropical reserve forests like those in Malaysia.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Marina Yusoff,
Advanced Analytic Engineering Center (AAEC),
Faculty of Computer and Mathematical Sciences,
Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia.
Email: marinay@tmsk.uitm.edu.my

1. INTRODUCTION

There has been a surge of threats to wildlife globally that requires more sophisticated efforts to counter. Malaysia's wildlife conservation enforcement faces many challenges in effort to protect flora and fauna [1]. These intrusion activities include illegal tree felling and poaching. Recently, the perpetrators of illegal logging are becoming too comfortable [2]. The poachers are hunting animals to the brink of extinction [3]. It is tough to catch the poachers in the act and prevent them, only after wildlife casualties the poachers were captured [4].

The felling and other sounds in these activities, such as those of chainsaw and gunshots can be heard afar but authorities fail to locate the source in time [3]. Hence the perpetrators can escape easily. The common gunshot noise could travel over a mile in certain circumstances [5]. Malaysian forest environment is unique with its composition of climate, trees and animals. Sound travels much faster in warm air near the ground than in colder air higher up [6]. The research collects audible intrusion activities sound that could be heard in range of 100 m. Thus, the bigger picture of this research was an attempt to record the illegal activity sounds to explore new feasible methods to prevent illegal activities in tropical forest

reserves. The approach of audio detection has not been attempted in Malaysia. However, it is believed that by employing it the authority can save cost in protecting wildlife.

Past studies in speech recognition has successfully been used in our daily lives. Now the research in Sound Event Detection (SED) is showing possibilities for application [7]. The detection of rare sounds such as baby crying, glass breaking, and gunshot were tested by numerous researchers in the DCASE 2017 [8]. The DCASE 2017 challenge results show accuracies upto 95% detection of these rare sounds [9, 10]. This shows the applicability of SED in autonomous surveillance is almost ready for the industry. The present research aims to apply SED as poacher intrusion in wildlife reserves. In the big picture the study aims in setting up the foundation of sound-based wide-area autonomous surveillance in a Malaysian forest environment.

The objective of this research is to record and analyse the audio of the emulated illegal activities. It is to discover the audio feature patterns between different audible illegal activities in reserve forest regions. The patterns will contribute to the possibility of SED to monitor the said illegal activities. The main concern to record the sound of audible illegal activities in a reserved forest environment. To analyse patterns of audio features between illegal activities and the ambience. This paper demonstrates the collection, processing and analysis of audio data. The data analysis should show the applicability to develop an autonomous sound event detection system for surveillance in a Malaysian Forest.

2. MEL-LOG ENERGIES AUDIO FEATURE

Mel-log energies MLE also known as Log Mel-filter energy features (LMFE) or Mel-scaled spectrogram. Mel filtering shows the characteristics of the human auditory system in frequency, the natural logarithm compensates with the non-linear loudness [11]. The feature extraction method employed in this research is based on Mel-log energie. Research has been conducted to find better representations that allow the extraction of useful information from audio [12]. A useful feature has a distinctive consistent pattern and low signal-to-noise ratio. The low signal-to-noise ratio will allow to achieve higher accuracy in detection [13]. The features however have many parameters that could be tweaked and tuned for performance [14]. The extraction of MLE features are based on the following steps and parameters. The following steps are the procedures to extract the said features.

- Take the transformation of (a windowed extract of) a signal from Fourier.
- Using triangular overlapping frames to project the power of the spectrum obtained above to the Mel -scale
- Take the power logs at each frequency of the Mel-scale.

Cepstral features that are computed by taking the FFT of the warped log spectrum, they contain data of the changes in the various spectrum bands [15]. Cepstral features are favourable due to their ability to separate the impact of source and filter in a speech signal [16]. The influence of the source and the filter sound can be differentiated as there are in various regions in the domain of cepstral features [17]. MLE was the chosen feature for this experiment as to popular opinion that it holds more dimensions compared to MFCC.

3. METHODOLOGY

The methodology consists of data collection and analyse of audio features. The details of location, sound collected, equipment used in the present research are explained below.

3.1. Location of authentic audio collection

The location of authentic audio collection was at Endau Rompin National Park, Johor, Malaysia on 4th August 2019. The Endau Rompin National Park region is a tropical rainforest with equatorial weather, hot and humid throughout the year. Its yearly average rainfall is 2000mm to 2500mm [15]. The audio was collected in the environment of a thick forest. A thick forest is a place having high density of trees and vegetation. This location has various obstruction for sound to travel. The obstruction can be trees trunk, thick foliage, and woods growing under the tree canopy. A thick foliage usually dampens the sound waves emitted by the source, which can reduce the audible distance of the said sound [18, 19].

3.2. Sound collected

The collected sounds were suggested by the Wildlife Conservation and Science (WCS) Malaysia. The WCS is an active non-governmental organisation which has been strong effort in wildlife conservation. The sound collected were three types obtained from the following activities:

- Hatchet hitting a tree trunk
- Forest ambience
- Chainsaw revving

The machete and axe were employed because they are often used by the poachers to cut trees to build their camps. A large axe hitting a tree can be heard a mile away in certain cases [18]. Then the chainsaw is a common tool of illegal loggers. The listed sounds are considered rare sounds in forest reserves. The next element considered in the data collection is the ambience of the location. This sound data consists of the forest natural environment. It is important because the sound is considered as the background noise.

3.3. Recording equipment

Recording equipments include, a recorder and microphone. The recorder used is a Zoom H6 Handy Recorder paired with a LR XYH Zoom H6 Capsule Mic. The settings for recording audio are sampling rate of 44.1 kHz with Phantom Power 12V. In digital audio recording it is recommended that the sampling rate must be double the maximum audio frequency generally accepted for humans [19]. Thus, the research employs sampling rate of 44.1 kHz.

3.4. Audio feature extraction

Audio Feature Extraction is a process done on the collected audio data. The features will allow recognition but high dimensional features causes overfitting [20]. The Cepstral features such as MLE mimics the human perception of sound [21]. The feature of MLE has produce well in the past on SED. The features selections plays a crucial role influencing the results [22]. MLE features are extracted from the audio files using python libraies. The research uses python libraries such as matplotlib.pyplot used to plot the feature heatmap representations, speechpy is used to extract features with specific parameters as stated in Table 1. In specific the functions from speechpy applied in research is speechpy.processing.preemphasis and speechpy.feature.lmfe. Feature extraction parameters are the tweakable variables in the extraction of features. The features include frame length, frame stride, number of filters, FFT length, minimum and maximum frequency of Mel-filter banks. The value of variables included in this experiment is stated in Table 1.

Table 1. MLE feature extraction parameters

Parameters	Value
frame length: The length of each frame in seconds.	100ms
frame stride: The step between successive frames in seconds.	50ms
Number of filters: The number of mlf to extract,	40
Fast Fourier Transform length: Number of FFT points	512
Min frequency: Minimum band edge of Mel-filters.	20 Hz
Max frequency: Maximum band edge of Mel-filters.	20kHz

The parameters employed in this research were based on common practices of past studies. The minimum and max band edge for Mel filters are set based on the human hearing capabilities [22]. The frame length of 100ms is believed fine enough for the 10-second audio file analysis. The frame stride of 50ms also means that the features are overlapping 50% from the previous frame. This provide more detail of feature changes in between windows. The features should gather information in the whole spectrum to allow maximum performance [23]

4. RESULTS

A significant pattern was found on the Mel-log energy audio features between audible illegal activities. The patterns are believed to be good fingerprints in training artificial intelligence for intrusion detection. Figure 1 shows the feature extracted as a heat map of forest ambience. The heat map is 10-second cut of forest ambience with 40 MLE a frame. A frame is 40 MLE obtained from 0.1 seconds of audio. Each frame consists of overlapping audio features of 50% using a hop(stride) size of 0.05 seconds on each frame. Whereas each frame is fitted with a hamming window to reduce spectral leakage. The overlapping of 50% was important to detect changes in between frames more effectively compensating for the hamming window [24]. This was to avoid losing meaningful features in between frames as hamming windows are focused on the centre of the audio frame.

Ten seconds of audio produced 198 frames and each frame contained 40 MLE. The total features extracted were 7920 floating numbers. These features are visualized in the heat map representation using python plot library. As a heatmap visualizations is an effective method in pattern analysis [25]. The following Figures 1, 2 and 3 shows ths heatmap of extacted features of their respective audible intrusion activity. The brighter or hot colours, which is yellow, shows the presence of a certain audio event. The darker or cold colours, which is blue, shows the less intense banks. Figure 1 show the MLE heatmaps of the forest ambience the natural sound of the forest.

Figure 2 shows the extracted features of chainsaw activity. It can be seen that the banks of 0-6 and 20-40 are more intense from the rest. Its pattern too is consistent throughout the 10-second audio. As such this can be used as the audio event fingerprint for this particular activity, which is the sound of a chainsaw being used to cut logs. Figure 3 shows the heat map of hatchet activity, a higher intensity (shown as yellow) on the Mel-log filter banks in the range of 29-40. It is consistent throughout the 10-second audio. The pattern consistency of features can be employed on the classification of the audio event.

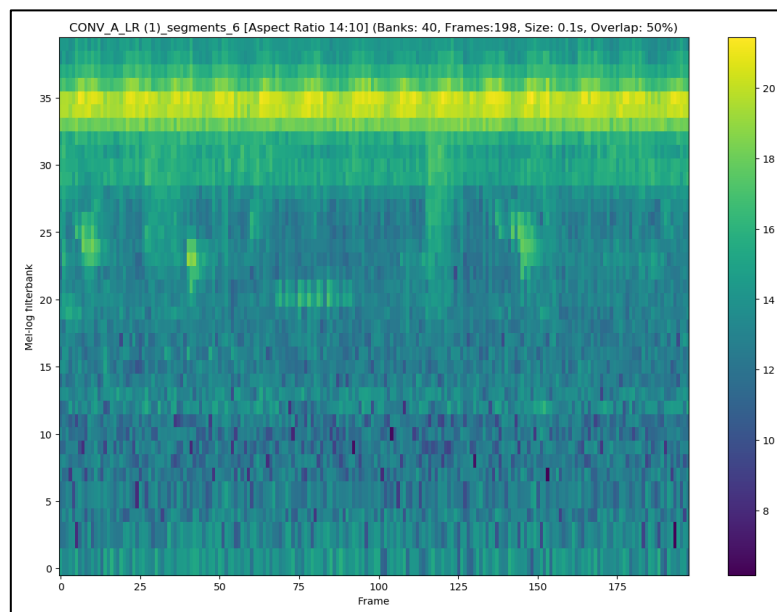


Figure 1. Extracted features heat map of 10-second forest ambience

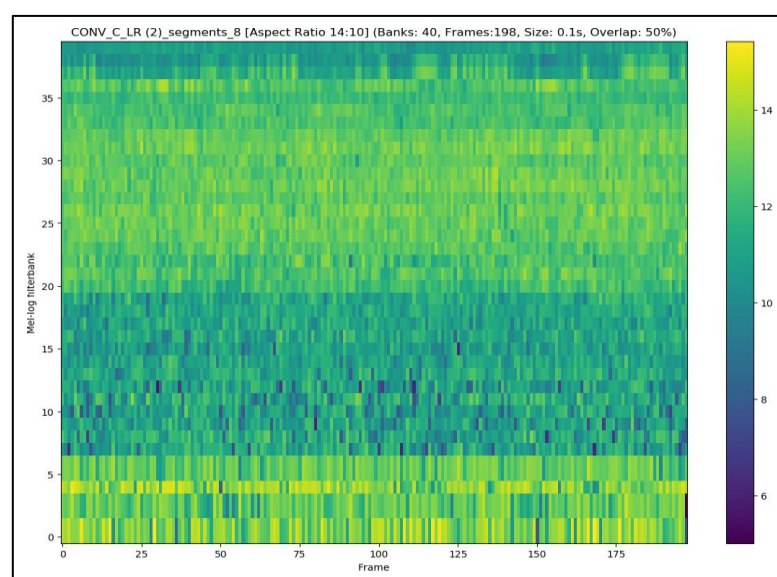


Figure 2. Extracted features heat map of 10-second chainsaw activity recorded

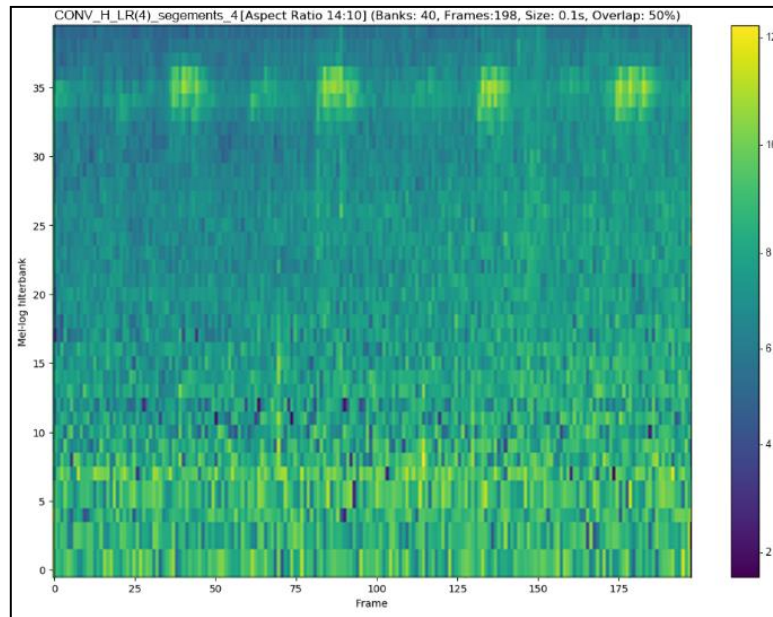


Figure 3. Extracted features heat map of 10-second hatchet activity recorded

The representations in heatmaps eases the pattern analysis task. A collection of heatmaps can be compared side by side to be observed. In search of patterns in the features, it is found that some significant patterns can be easily be found. Table 2 shows an overview of the analysis: First, the intense regions compared to the less intense areas show distinctive pattern that can be observed in specific ranges between MLE. Next, the distribution range of MLE values lower and upper limit. The values vary in different activities.

Table 2. Patterns of mel-log feature heatmaps

Activity	Intense areas of MLE	Distribution Range (lower-upper)
Chainsaw Revving	0-6 and 20-36	5-15
Forest Ambience	33-36	6-21
Hatchet Cutting	0-10 and 33-36	2-12

5. CONCLUSION

This paper presents Mel-log energies analysis of authentic audible intrusion activities in a Malaysian forest. The objective of this research is to obtain authentic Malaysian forest environment audio and analyse audible intrusion activities MLE audio features. The study has achieved to obtain and analyse MLE features of the audio. Heatmap representations were used to find patterns of illegal intrusion activities in MLE features extracted. It is found that the MLE audio features extracted have distinctive patterns to their respective audio sources. Hence autonomous surveillance is believed to be possible by using MLE audio features and machine learning in a Malaysian forest environment. It is important to mention here the obstacles faced during the data collection. The first obstacle was the equatorial weather, being hot and humid throughout the year with an average rainfall is 2500mm a year. Recording equipment should be waterproof or rainproof to avoid damage. Next is the tropical bugs might be infesting in the equipment in long term recording sessions. An issue that affected the data collection was that the present researcher was not able to employ gunshots as another audible intrusion activity due to Malaysian national park policies and limited provisions. The patterns discovered are deemed to allow artificial intelligence in monitoring forest intrusion activities. It also maybe a feasible approach for wildlife protection in forests environment like Malaysian rainforest. This may improve the security of wildlife in the world. These could be the initial step in employing intelligent security systems in wildlife protection. As has been discussed, this attempt may have practical importance in initiating a new and cheaper solution.

ACKNOWLEDGEMENTS

Institute of Quality and Knowledge Advancement (InQKA) and Universiti Teknologi MARA, Shah Alam, Malaysia for providing essential support and knowledge for the work.

REFERENCES

- [1] Ariffin, M., "Enforcement against wildlife crimes in west Malaysia: The challenges," *Journal of Sustainability Science and Management*, vol. 10, no. 1, pp. 19-26, January 2015.
- [2] Povera, A., "MACC cracks down on illegal logging and mining," 2019. [online] NST Online. Available: <https://www.nst.com.my/news/nation/2019/09/525563/macc-cracks-down-illegal-logging-and-mining>.
- [3] Lee, S., "Post-mortem finds pygmy elephant was shot multiple times at close range, tusks removed," (2019, September 30). [online]. Available: <https://www.thestar.com.my/news/nation/2019/09/29/post-mortem-finds-pygmy-elephant-was-shot-70-times-at-close-range-tusks-removed>.
- [4] Miwil, O., "Sabah Wildlife Department looking for at least 3 suspects in turtle poaching case," November 2017. [online]. Available: <https://www.nst.com.my/news/nation/2017/10/287055/sabah-wildlife-department-looking-least-3-suspects-turtle-poaching-case>.
- [5] Tsiatis, N. E., "Understanding distance shooting and the type of firearm from the analysis of gunshot sounds," *European Police Science and Research Bulletin*, no. 15, pp. 93-107, 2016.
- [6] Tsiatis, N. E., "Physics and Mathematical Analysis in Gunshot Sounds — Parameters and their Characteristics," *AFFE 2013, 44th Annual Training Seminar*, Albuquerque, New Mexico, pp. 23-28 June, 2013.
- [7] Afendi, A. S., & Yusoff, M., "Review of anomalous sound event detection approaches," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 8, no. 3, pp. 264-269, September 2019.
- [8] Heittola, T., & Mesaros, A., "DCASE 2017 Challenge Setup: Tasks, Datasets and Baseline System," *Detection and Classification of Acoustic Scenes and Events*, 2017.
- [9] Cakir, E., & Virtanen, T., "Convolutional Recurrent Neural Networks for Rare Sound Event Detection," *Detection and Classification of Acoustic Scenes and Events*, November 2017.
- [10] Lim, H., Park, J., & Han, Y., "Rare Sound Event Detection Using {1D} Convolutional Recurrent Neural Networks," *Detection and Classification of Acoustic Scenes and Events*, November 2017.
- [11] Benetos, E., Lagrange, M., Plumbley, M.D., et al., "Detection of overlapping acoustic events using a temporally-constrained probabilistic model," *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, pp. 6450-6454, 2016.
- [12] Adavanne, S., Parascandolo, G., Pertila, P., Heittola, T., Virtanen, T., "Sound event detection in multichannel audio using spatial and harmonic features," *Detection and Classification of Acoustic Scenes and Events 2016*, September 2016.
- [13] Dilber, D., "Feature Selection and Extraction of Audio," *International Journal of Innovative Research in Science, Engineering and Technology*, vol. 5, no. 3, pp. 3148-3155, March 2016.
- [14] Mahkonen, K., "Mel-frequency cepstral coefficients (MFCCs) and gammatone filterbanks," 2015. [Online]. Available: <http://www.cs.tut.fi/~sgn14006/PDF2015/S04-MFCC.pdf>
- [15] Lokman, T., "October a month of flash floods," 2019. [online] NST Online. Available at: <https://www.nst.com.my/news/nation/2019/09/525741/october-month-flash-floods>.
- [16] Wu, Y.-P., Mao, J.-M., & Li, W.-F., "Robust speech recognition by selecting mel-filter banks," *2nd Annual International Conference on Electronics, Electrical Engineering and Information Science (EEEIS 2016)*, vol. 117, pp. 407-416, 2016.
- [17] M. A. Yusnita, M. P. Paulraj, Szali Yaacob, R. Y. and M. N. Fadzilah, "Statistical Descriptors of Mel-Bands Spectral Energy Features with Feature Reduction for Robust Accent Recognition in Malaysian English," *Indian Journal of Science and Technology*, vol. 8, no. 20, pp. 1-10, August 2015.
- [18] Ahmad, S. F., & Singh, D. K., "Automatic detection of tree cutting in forests using acoustic properties," *Journal of King Saud University - Computer and Information Sciences*, pp. 1-7, February 2019.
- [19] Smith S. W., "The scientist and engineer's guide to digital signal processing," San Diego (CA): California Technical Publishing www.dspguide.com/CH28.PDF, pp 351-372, 1997.
- [20] Kamala, R., & Thangaiah, R. J., "An Improved Hybrid Feature Selection Method for Huge Dimensional Datasets," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 8, no. 1, pp. 77-86, March 2019.
- [21] Bezoui, M., "Speech Recognition of Moroccan Dialect Using Hidden Markov Models," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 8, no. 1, pp. 7-13, March 2019.
- [22] Reynolds, R. P., Kinard, W. L., Degraff, J. J., Leverage, N., & Norton, J. N., "Noise in a laboratory animal facility from the human and mouse perspectives," *Journal of the American Association for Laboratory Animal Science (JAALAS)*, vol. 49, no. 5, pp. 592-597, September 2010.
- [23] Noda, J. J., Travieso-González, C. M., Sánchez-Rodríguez, D., & Alonso-Hernández, J. B., "Acoustic Classification of Singing Insects Based on MFCC/LFCC Fusion," *Applied Sciences*, vol. 9, no. 19, pp. 4097, October 2019.
- [24] Podder, P., Zaman Khan, T., Haque Khan, M., & MuktaDir Rahman, M., "Comparative Performance Analysis of Hamming, Hanning and Blackman Window," *International Journal of Computer Applications*, vol. 96, no. 18, pp. 1-7, June 2014.
- [25] Khoo, M., Rozaklis, L., Hall, C., Kusunoki, D., & Rehrig, M., "Heat Map Visualizations of Seating Patterns in an Academic Library," *iSchools*, March 2014.