International Journal of Neural Networks and Advanced Applications
DOI: 10.46300/91016.2026.10.1 Volume 10, 2026

Mel-Frequency Cepstral Coefficient-Enabled Analysis of Doppler Effect on Road Traffic Noise from Nairobi City

Ochungo Akech Elisha[1,2,*], Osano Simpson Nyambane[1], Gichaga John Francis[2]
[1] Department of Civil and Construction Engineering,
University of Nairobi, P.O. Box 30197-00100, Nairobi,
Kenya.

[2]Department of Civil, Faculty of Engineering and Technology (FoET),
Multimedia University, P.O. Box 15653- 00503 Nairobi,
Kenya.

**Corresponding Author*

*R*eceived: January 27, 2026. Revised: April 11, 2026. Accepted: May 4, 2026. Published: September 10, 2026.

Abstract:- This study investigated the Doppler effect on road traffic noise levels in an urban setting by deploying Mel-Frequency Cepstral Coefficient (MFCC) to characterize acoustic signatures across vehicle mix classes. Using a dataset of audio recordings from Nairobi city's road network sampled from 42 sites, the study adjusted the equivalent continuous sound levels (Leq) data for the analysis of the Doppler effect based on vehicle pass-by speeds, and extracted MFCC features. This was followed by semantic matching, which aligned the sample site locations with captured audio signal data, hence facilitating accurate speed assignments. Statistical analyses and visualizations revealed variations in noise levels and MFCC features by vehicle class and location. The result of the analysis presents mathematical formulations, comprehensive Doppler effect characterization, and visualizations in the form of graphical diagrams. The study has contributed knowledge on urban noise-scape scenario building using computational methods. This has confirmed that MFCC-enabled analysis of audio signal data is a novel approach for the characterization of road traffic noise hotspots in urban environments. The same can be used to inform active noise control strategies for cities to improve the welfare of urban residents.

Keywords: - Doppler effect, mel-frequency cepstral coefficient, noise, extraction, features, visualization

I. INTRODUCTION

Sound detection has been widely explored in urban settings, for example, to identify emergencies that require attention. Because of increased car usage in urban areas, road traffic noise has become the dominant component of the urban environmental noise-scape, [1]. It has an important research agenda for sounds to be identified, even for an inhomogeneous traffic mix. Typically, an urban area's traffic composition comprises multiple vehicular classes, such as cars, trucks, buses, vans, and motorized 2-3-wheelers, which together form distinct vehicle signatures. The same can be used to detect and classify vehicle types based on their sound frequency types. Vehicle detection while it is in motion is a prerequisite for traffic and speed management, classified vehicle count, traffic signal time optimization, gap/ headway measurement, and military purposes, [2]. Most important is that the road traffic noise (RTN) can be used as audio signal data for the analysis of Doppler effects on urban roads. Therefore, to comprehensively understand the true character of environmental noise, scholars have always deployed traditional sound evaluation methods such as the use of standard noise level measurement instruments, [3]. Newer methods of doing this have, however, emerged.

For example, modelling of sound using an auralisation method is an emerging novel approach. In this, the analysis of the Doppler effect of passing by vehicles on street sections has become a useful tool for urban noisescape quality characterization methods, [4]. This has stemmed from the effort by industrial hygienists, who, for a long time, have been measuring and predicting reverberant noise in buildings. This development builds upon the foundational work in 1895 that corrected the acoustics of Fogg's lecture hall, [5]. Following the cue of noise control in buildings, abatement of noise in the open environment has become a key topic in research to improve the welfare of urban residents. Scholars have mostly used passive noise control techniques like noise barriers and green belts to effectively mitigate noise.

However, these old methods occupy urban space, hence intensifying the scarcity and high cost of already congested city areas. Accordingly, new novel noise reduction strategies known as active noise control (ANC) are emerging to eliminate the need for physical isolation structures and to quieten the noise within specific frequency ranges more effectively, [6]. To this end, we are witnessing a proliferation of audio processing techniques to profile noise data into decision models for use by urban spatial planners, [7]. One such method is

DOI: 10.46300/91016.2026.10.1 Volume 10, 2026

the analysis of Doppler frequency, which the present study is applying to understand the character of passing-by vehicles' sound collected along Nairobi city roads, [8].

The Doppler effect is a method of explaining the relationship between the frequency of the harmonic waves generated by a moving source and the frequency as captured by an observer. In the case of a moving car, its generated sound waves propagate as a group of circles, as shown in Figure 1 (Appendix). The visualization of the Doppler effect can be for situations where (i) both source and observer are stationary, (ii) source moving and observer stationary, (iii) source stationery and observer moving, and (iv) both source and observer are moving. In this research, the classical Doppler effect of mechanical waves, that is, sound waves from road traffic noise data whose speed is relative to an underlying medium (air), has been used to visualize frequency, wavelength, and speed based on case scenario (ii) above.

The Doppler effect is nowadays used to describe the change in wave frequency due to relative motion between a source (a moving car in the present research) and a stationary observer. This is treated as critical since it facilitates the analysis of vehicle noise in urban environments. There are numerous ways of classifying vehicles using signals. Image processing approaches are deployed to classify vehicles in real-time traffic management and for Intelligent Transportation Systems (ITS). Inductive loop-based techniques are commonly used for the determination of vehicle counts. In the present time, acoustic signals based on emitted sounds have emerged as novel methods in classifying vehicles. In this, methods such as autoregressive modeling algorithms, Probabilistic Neural Network (PNN) models, among many other feature extraction approaches, have been discussed, [9]. Researchers have also used Mel-Frequency Cepstral Coefficient (MFCC) to extract features from audio signals, [10]. The method is particularly suited to large dataset environments where machine learning applications are used as part of analytics, [11]. The feature extraction is the process of extracting and tackling hidden information in the raw data signal, [12].

Feature extraction is done to find harmonics and sidebands of the signal in both the time and frequency domains, which are important to any pattern recognition system. Methods such as power spectrum using Fast Fourier Transform (FFT) have been used to capture the harmonics and sidebands of the signal in the time domain. On the other hand, cepstrum, such as Mel Frequency Cepstrum Coefficient (MFCC), Gamma Tone Cepstrum Coefficient (GTCC), is capable of extracting harmonics and sidebands of the spectrum version of the signal. This has seen widespread use of MFCC in many fields, [13], as demonstrated in Figure 2 (Appendix). The present study falls under industrial analysis.

Mel-Frequency Cepstral Coefficients (MFCC) are robust features for capturing spectral characteristics of audio signals, making them ideal for vehicle noise analysis, [14].

This study uses a dataset from Nairobi, Kenya, stored in MFCC stats with doppler.csv, which includes MFCC features, Leq, Doppler-adjusted Leq, and vehicle speeds for various vehicle classes across urban roads. The main objective of the research was to visualize the Doppler effect on road traffic noise. Specifically, the study sought to: (a) quantify the Doppler effect on noise levels, (b) analyze MFCC features across vehicle classes and locations, (c) provide statistical insights and visualizations to support findings, and (d) identify research gaps and limitations. The main contribution of this study is the demonstration of the utility of MFCC in extracting spectral values of audio signals of vehicle noise for creating decision models for quieting urban areas. The rest of this paper is organized as follows: the computation of MFCC is explained in Section 2. Section 3 introduces MFCC applications in vehicle noise feature extraction. Followed by a summarized analysis presentation in section 4, and finally, the conclusion is drawn in section 5.

II.NOTATION AND SYMBOL DEFINITIONS

In the following sections, the mathematical formulation of the Mel-Frequency Cepstral Coefficient (MFCC)-based Doppler analysis for road traffic noise is presented. To ensure clarity, consistency, and reproducibility, the key symbols used throughout this study are defined below. The analysis adopts standard digital signal processing notation, where continuous-time signals are denoted by $x(t)$ and discrete-time signals by $x[n]$. Frequency-domain representations are obtained using Fourier-based transforms.

A. Signal and Transform Notation

$x(t)$represents the continuous-time acoustic road traffic noise signal, while $x[n]$denotes the discrete-time sampled audio signal. $w[n]$is the analysis window function (for example, Hamming or Hanning window), and N represents the number of samples per analysis frame. N_l denotes the length of the discrete sequence or window. The Discrete-Time Fourier Transform (DTFT) of $x[n]$is represented as $X(e^{j\omega})$, where ω is the normalized angular frequency in radians per sample within the range $[-\pi, \pi]$. The Discrete Fourier Transform (DFT) coefficient at frequency bin $k$is denoted by $X[k]$, where $k = 0,1,2,\dots,N-1$. The twiddle factor matrix element is defined as $W_N = e^{-j\frac{2\pi}{N}}$. The Short-Time Fourier Transform (STFT) representation is denoted as $X[n,k]$, while $P[k]$ represents the power spectrum at the frequency bin k. The forward and inverse Fourier transform operators are denoted by $\mathcal{F}$ and $\mathcal{F}^{-1}$, respectively.

B. MFCC and Mel-Filter Bank Parameters

The Mel-filter bank analysis consists of M triangular filters, where $m = 1,2,3,\dots,M$. $H_m[k]$ represents the frequency response of the m^{th} Mel filter at bin k, and

$f(m)$ denotes its center frequency in Hertz. $S[m]$ represents the filter-bank energy output for the m^{th} filter, while $c[n]$ denotes the Mel-Frequency Cepstral Coefficient (MFCC) of order n, where n is the cepstral index.

C. Doppler and Acoustic Parameters

The emitted source acoustic frequency is denoted by f_s, while the observed frequency at the receiver is denoted by f_o. The relative velocity between the source and observer is represented by v_r in meters per second (m/s), while $v_{\text{km/h}}$ represents vehicle speed in kilometers per hour. The speed of sound in air is denoted by c, approximately 343 m/s.

D. Sound Level and Empirical Parameters

The equivalent continuous sound level is denoted by L_{eq} in dBA, while ΔL_{eq} represents the Doppler-induced correction in sound level in decibels (dB). The adjusted equivalent continuous sound level is denoted by $L_{eq,\text{adj}}$. The parameter α represents an empirical coupling coefficient capturing Doppler–noise interaction effects and is dimensionless.

III. CEPSTRAL IMPLEMENTATION VIA MFCC

MFCC is one of the most commonly used extraction methods in cepstral analysis, particularly in signal processing of audio data records, [15]. Cepstrum begins with the computation of the Fourier transform of a signal's time domain to get its frequency spectrum, [16]. Afterwards, the logarithm of the same is taken and subjected to the inverse Fourier Transform of the log-spectrum; see Equation 1.

$$C_\tau = \mathcal{F}^{-1}\left[log|\mathcal{F}\{x(t)\}|\right] \quad (\mathbf{1})$$

Where $C(\tau)$ is the cepstrum of a signal $x(t)$, $\mathcal{F}$ is the Fourier transform, $\mathcal{F}^{-1}$ is the inverse Fourier transform, and τ is the quefrency variable. $|\mathcal{F}\{x(t)\}|$ is the magnitude spectrum of the signal.

Normally, MFCC uses five consecutive processes, which include: (i) signal framing, (ii) computing the power spectrum, (iii) applying a Mel filter bank to the obtained power spectra, (iv) calculating the logarithm values of all filter banks, and (v) finally applying the Discrete Cosine Transform (DCT). Figure 3 (Appendix) presents the processes of the MFCC procedure as elaborately explained by [17].

A. Signal Processing

Signal processing approaches were used in developing computational methods as guided in the work by [18]. Just to explain the basics regarding signals and different types of signals, the study uses the example of a Low-Pass Filter (LPF) and the Fourier transform. To begin, a signal is described as a function (or a clue, a message, an indicator) that expresses information, [19], which can manifest in a wide variety of fields. These may include, but are not limited to, communications, aeronautics and astronautics, circuit design, acoustics, seismology, biomedical engineering, energy generation and distribution systems, chemical process control, and speech processing. When a signal is described as a function, the independent variable(s) can be continuous or discrete. Continuous time signals or analog signals refer to signals defined along continuous independent variable(s). An example of a one-dimensional continuous-time signal is shown in Figure 4a (Appendix).

Discrete-time signals refer to signals defined at discrete independent variables, commonly integer variables. Discrete-time signals are often represented by sequences of numerical values. A discrete-time signal can be generated from a continuous-time signal by the approach of sampling. A one-dimensional discrete-time signal example is given in Figure 4b (Appendix). The amplitude or the dependent variable of a signal can also be continuous or discrete. Digital signals are signals in which both independent variable(s) and dependent variable(s) are discrete. A digital signal can be generated from a discrete-time signal by the approach of quantization, as shown in Figure 4c (Appendix). The term signal processing refers to the representation, transformation, manipulation, analysis, and synthesis of signals, [19]. A transformation that maps an input sequence x[n] into an output sequence y[n] is often referred to as a discrete-time system. In this study, the discrete-time signal processing used is the Fourier transform (FT).

B. Pre-emphasis

Pre-emphasis is a form of pre-processing activity in the signal processing field for compensating for the high frequency of the signal, which is to be suppressed during the signal production. It is the first step in the MFCC feature extraction in the High-Pass Filter with a typical setting, thus (1, -0.97). The filtering flow changes energy distribution across frequencies, alongside the overall energy level, [20].

C. Signal framing and windowing

In this stage, the signals are split into unique 'frames' simply to break down the raw data signal into frames, making it more stationary. For stable acoustic characteristics, audio signals for examination should be over a sufficiently short period of time. Regarding the audio signal, the period of data capture needs to be a Quasi-Stationary Segment (QSS) because signals in the real world are always nonstationary, but locally stationary. They have stable second-order statistical properties within a short period of time and differ from any other frame, [21]. However, the capturing should be sequenced in convenience in readiness for this step. Hence, short-term spectral measurements are typically carried out over short frame windows, and each frame is overlapped with the next one. Such overlaps of frames enable the temporal characteristics of the audio signal to be tracked. With overlapping frames of audio signals, the sound representation would be approximately centered at some frame. For each frame, a window is applied to narrow the signal towards the border of the frame. In general, Hanning and Hamming windows are the most well-known nominees as applied in the research study by [22]. The Hamming window is a discrete-time finite impulse response (FIR) system that is often used as a filter. Such windows can enhance harmonics, smooth edges, and diminish edge effect while taking a Discrete Fourier Transform (DFT) on the signal. Figure 5 (Appendix) demonstrates the rectangular, Hamming, and Hanning windows in both the time and frequency domains. Filters are linear time-invariant systems that select or modify signals according to their frequency components. Specifically, the low-pass filter (LPF) passes the low-frequency component of a signal and attenuates the high-frequency components of the signal. As discrete-time systems can be categorized into infinite impulse response (IIR) and finite impulse response (FIR), discrete-time filters can also be categorized into IIR filters and FIR filters. IIR filters are commonly designed by applying a transformation from continuous-time filters into discrete-time IIR filters, whereas FIR filters are commonly designed by approximating the desired frequency response and impulse response. The simplest FIR filter designing method is based on windowing. The rectangular window, the Bartlett (triangular) window, the Hanning window, the Hamming window, and the Blackman window are windows that are frequently used as filters. The rectangular window, Hanning window, and Hamming window with a length of M_{w+1} can be expressed as Equation 2, Equation 3, and Equation 4, respectively.

$$\omega_{rectangular}n = \begin{cases} 1, 0 \leq n \leq M_\omega \\ 0, \ otherwise \end{cases} \quad (2)$$

$$\omega_{Hanning}n = \begin{cases} 0.5 - 0.5\, cos\left(\frac{2\pi n}{M_\omega}\right), \\ \quad 0 \leq n \leq M_\omega \\ 0, \ otherwise \end{cases} \quad (3)$$

$$\omega_{Hanning}n = \begin{cases} 0.5 - 0.46\, cos\left(\frac{2\pi n}{M_\omega}\right), \\ \quad 0 \leq n \leq M_\omega \\ 0, \ otherwise \end{cases} \quad (4)$$

Compared with the rectangular window, the Hamming and Hanning windows have lower side lobes in the frequency domain and a smoother transition in the time domain, [19], [23]. The Hamming window does not touch zero at the endpoint in the time-domain as does the Hanning window. The representations of these windows in the time and frequency domains are illustrated in Figure 5 (Appendix).

D. *Fourier Transform*

The Fourier analysis is typically used to convert a signal or a system from the time domain to the frequency domain. The use of the Fourier analysis in discrete-time systems is affected by the sampling approach (the approach of obtaining a discrete-time signal/system from a continuous-time signal/system). The study explains the fundamentals of the Nyquist frequency, which is the maximum frequency that can be described by a discrete-time signal. It is followed by the fundamentals of several Fourier analyses, including the discrete-time Fourier transform (DTFT), the discrete Fourier transform (DFT), the fast Fourier transform (FFT), as well as the short-time fast Fourier transform (STFFT).

The frequency $2\omega_M$ is often referred to as the Nyquist rate and the frequency ω_M is the Nyquist frequency. The discrete-time Fourier transform (DTFT) of a sequence, $x[n]$, provides information about the composition of various -frequency complex exponentials. This DTFT is often referred to as the spectrum and is described in Equation 5, [18].

$$Xe^{j\omega} = \sum_{n=-\infty}^{+\infty} x[n]e^{-j\omega} \quad (5)$$

The resulting Fourier transform $X(e^{j\omega})$ is a periodic function with a period of 2π. The high-frequency components of x[n] are represented by the values of $X(e^{j\omega})$ with ω near the odd multiples of π. When x [n] is real, the DTFT satisfies conjugate symmetry, meaning the magnitude spectrum is even, such that $\left|X(e^{j\omega})\right| = \left|X(e^{-j\omega})\right|$.

Moreover, the energy-density spectrum of x[n] is defined by$\left|X(e^{j\omega})\right|^2$.

A DTFT version of finite-length signals is represented by the discrete Fourier transform (DFT). The DFT of an N_l -The length sequence is described in Equation 6 as per, [19]. The spacing between DFT frequencies is $\frac{2\pi}{N_l}$ and therefore, the frequencies related to those frequencies are $\omega_k= \frac{2\pi k}{N_l}$. The computation of all the values of X[k] requires N_l^2 complex multiplications and $N_l\,(N_l\ -1)$ complex additions.

$$X[k] = \sum_{n=0}^{N_l-1} x[n]W_N^{kn}, 0 \leq k \leq N_l\ - 1 \quad (6)$$
$$W_N = e^{-j\left(\frac{2\pi}{N_l}\right)}$$

W_N is the complex exponential "twiddle factor" that represents a fixed phase rotation used to combine signal samples into different frequency components in the DFT. The fast Fourier transform FFT is an efficient computational algorithm to compute the DFT. By exploiting its symmetry and periodicity properties, the computation of the FFT requires roughly $N_l \log_2 . N_l$ complex multiplication. One of the FFT algorithms is the Cooley-Turkey algorithm, [24], [25], which has several variations, such as the decimation in time and the decimation in frequency, [19].

The time-varying Fourier transform was introduced by considering that the Fourier transform is not directly applicable for the analysis of a signal that has properties that change as a function of time. For example, in the analysis of speech processing, the short-time analysis principle is seen as a valid approach, [26]. Here, the temporal properties are assumed to be fixed over a short time, and the spectral properties are assumed to change relatively slowly. The time-varying Fourier transform is described as in Equation 7, where $\omega[m]$ is a real window sequence and n is a particular time index. The length of the window $\omega[m]$ is often finite, [26]. If $\omega[m]$ is a rectangular window with a length of $M_{\omega+1}$, the time-varying Fourier transforms by Equation 8

$$X_n(e^{j\omega}) = \sum_{m=-\infty}^{+\infty} x[m]\omega[m-n]e^{-j\omega m} \qquad (\mathbf{7})$$

$$X_{n,r^{j\omega}} = \sum_{m=n}^{n+M_\omega} x[m]e^{-j\omega m} \qquad (\mathbf{8})$$

The efficient computation algorithm to compute a time-varying Fourier transform is called the short-time fast Fourier transform (STFFT)

E. Power spectrum

A power spectrum can be described as the distribution of the power of the frequency components that compose the signal. Traditionally, the Discrete Fourier Transform (DFT) is utilized to compute the power spectrum, [27]. The power spectrum of each of the obtained frames must be determined based on the following Equation 9:

$$P(k) = \left|\sum_{n=0}^{N-1} x(n)e^{-j\frac{2\pi}{N}kn}\right|^2 \qquad (\mathbf{9})$$

Where k = 1, 2, 3,…, N-1. x(n) is a discrete signal, and N is the length of the signal.

F. Mel filter bank

The Mel band-pass filter is a bank of filters that is constructed based on pitch perception. The Mel filter was originally developed for speech analysis, and like the human ear's perception of speech, it targets extracting the non-linear representation of the speech signal. The conventional Mel filter bank is constructed of triangular filters. The transfer function (TF) of each of the m-th filter can be computed via Equation 10.

$$H_m(k) = \begin{cases} 0, k < f(m-1) \\ \frac{k-f(m-1)}{f(m)-f(m-1)}, \\ f(m-1) \le k \le f(m) \\ \frac{f(m+1)-k}{f(m+1)-f(m)}, \\ f(m) \le k \le f(m+1) \\ 0, k > f(m+1) \end{cases} \qquad (\mathbf{10})$$

Where f(m) is the centre frequency of the triangular filter, and the filter bank is normalized such that: $\sum_m^{M-1} H_m(k) = 1$

The Mel scale to the response frequency and vice versa is computed by Equations 11 and 12

$$m = 2595 \log_{10}\left(1 + \frac{f}{700}\right) \qquad (\mathbf{11})$$

$$f = 700\left(10^{\frac{m}{2595}} - 1\right) \qquad (\mathbf{12})$$

where $f$is the frequency in Hertz and $m$is the corresponding value on the Mel scale.

G. Discrete Cosine Transforms (DCT)

A Discrete Cosine transform (DCT) expresses a finite sequence of data points as a summation of cosine functions oscillating at different frequencies. The DCT was introduced by [28]. In the MFCC process, the DCT is applied on the Mel filter bank to select the most accelerative coefficients or to separate the relationship in the log spectral magnitudes from the filter bank, [29]. The DCT is computed by Equation 13

$$X(k) = \sum_{n=0}^{N-1} x_n \cos\left[\frac{\pi}{N}\left(n + \frac{1}{2}\right)\mathrm{k}\right] \qquad (\mathbf{13})$$

Where where $\mathrm{k} = 1,2,3,\ldots,\mathrm{N}-1$, x_n is a discrete signal, and N is the length of the signal

IV. MFCC APPLICATIONS

City roads get jammed, [30]. An acoustic approach to monitoring the roads for the arrival of rescue vehicles such as ambulances, fire engines, and police cruisers has become a necessity, hence the deployment of MFCC in this study. Road traffic noise data from Nairobi city were subjected to MFCC for feature extraction from the audio signals as per the earlier work by [31], [32], [33], [34]. Acoustic feature extraction is taken as a challenging endeavor because environmental noises usually cause disturbances in the extracted acoustic features. However, scholars have, over the years, come up with algorithms to extract the acoustic feature for vehicle target, such as eigenfaces, fast Fourier transform (FFT), power spectral density (PSD), and wavelet packet transform (WPT). The Mel-frequency cepstral coefficient (MFCC) feature extraction method has become the best in audio applications. It thrives in a nonlinear scale rather than a linear scale known as the "Mel'' scale, which is similar to the human auditory system.

Acoustic signals of moving vehicles are usually contaminated by unknown interference, such as wind noise, resulting in the pollution of the corresponding MFCC features. However, with the acoustic signatures of moving vehicles, we find that both the spectrums and the MFCC features of the vehicle signals mainly reside in the low-frequency part. In this study, sound recordings were collected from 42 strategically selected urban locations across Nairobi city in Kenya. Each recording location yielded multiple 6-second audio segments captured at a sampling rate of 44,100 Hz using calibrated measurement-grade equipment. The data set comprised 2,185 segments in total, providing robust statistical power for analysis. Recording sessions were conducted during peak traffic periods (6:00 AM – 6:00 PM on weekdays) to capture representative flow conditions and vehicle mix.

A. Data Collected

The dataset was saved as MFCC stats with doppler.csv, and it contained 2,185 audio recordings from urban roads in Nairobi, with columns: relative path, road, class ID, class, Mean MFCC, Std MFCC, Leq dBA, Adjusted Leq dBA, and Speed kmh. Locations were code-named, for example, “Around Arya School” and “Winners Chapel (Likoni Rd)” with coordinates. Vehicle classes included motorcycles, SUVs, heavy trucks, buses, PSVs, and others, as shown in Table 1 (Appendix).

B. Data Processing

In general, the dataset was taken through a computation of MFCC as earlier summarized in Figure 3 (Appendix), beginning with framing the signal into short frames, calculating the power spectrum, finding the mel filter bank for the power spectra, taking the logarithm of all filter banks, taking the IDCT of the log filter banks, and keeping DCT coefficients. The Python script processes the data as follows:

(i) Audio Preprocessing: Audio files were resampled to 22,050 Hz, padded/truncated to 4 seconds, and MFCCs were extracted (40 coefficients, 1024-point FFT, 512-sample hop length).

(ii) Location Matching: Road names were matched semantically using simplified name codes to remove coordinates and prefixes/suffixes, and Sequence Matcher with a 0.55 threshold for fuzzy matching.

(iii) Doppler Effect Adjustment: Leq was adjusted based on vehicle speed using the Doppler formula.

(iv) MFCC Feature Extraction: Mean and standard deviation of MFCCs were computed for each audio clip.

C. Statistical Analysis

The following descriptive attributes were computed from the dataset (i) Descriptive Statistical values Mean, median, standard deviation, and range of Mean MFCC, Std MFCC, Leq dBA, Adjusted Leq dBA, and Speed km/h by vehicle class, (ii) Correlation Analysis values: Pearson correlation between Mean MFCC, Adjusted Leq dBA, and Speed km/h, (iii) ANOVA: One-way ANOVA to test if Mean MFCC and Adjusted Leq dBA differ significantly across vehicle classes, and (iv) Location-Based Analysis: Summary statistics of Adjusted Leq dBA and Speed km/h by location.

D. Visualization

The following visualization diagrams were generated; (i) Box Plot of Mean MFCC by Vehicle Class, showing variability with heavy trucks at lower MFCCs, (ii) Scatter Plot of Adjusted Leq dBA vs. Mean MFCC, illustrating negative correlation, (iii) Bar Plot of Average Adjusted Leq dBA by Location, highlighting noise hotspots, Correlation Heatmap, visualizing relationships between variables, and (iv) Histogram of Speed kmh by Vehicle Class, showing speed distributions.

E. Mathematical Formulations

In outputting the above, the study deployed mathematical formulations, including (i) the Doppler Effect. The Doppler effect adjusted the observed frequency (f_o) based on the source frequency (f_s) and relative velocity ($v_r = \frac{v_{\text{km/h}}}{3.6}$ m/s) is as shown in Equations 14 and 15.

$$f_o = f_s \frac{c}{c - v_r} (approaching) \quad \mathbf{(14)}$$

$$f_o = f_s \frac{c}{c + v_r} (receding) \quad \mathbf{(15)}$$

where c = 343 m/s.

The change in sound level is expressed as Equation 16:

$$\Delta Leq = 10.log_{10}(1 + \alpha v_r) \quad \mathbf{(16)}$$

Where α is an empirically derived coupling coefficient representing the sensitivity of sound level variation to vehicle velocity.

The adjusted Leq is as in Equation 17:

$$Adjusted\ Leq = \max(Leq + \Delta Leq, Leq - 5dB) \quad \mathbf{(17)}$$

L_{eq} is the original equivalent continuous sound level and ΔL_{eq} represents the correction term applied to account for measurement or environmental adjustments. The lower-bound constraint (−5 dB) was introduced to prevent unrealistic attenuation values caused by extreme model corrections.

MFCC Extraction, which involved transformation computations as discussed in section 2 above;

STFT: as shown in Equation 18.

$$Xk = \sum_{k=0}^{N-1} x(n)w(n)e^{-j2\pi kn/N} \quad \mathbf{(18)}$$

Mel-Filter Bank: as shown in Equation 19.

$$S(m) = \sum_{N-1} |X(k)|^2 H_m(k), \quad \mathbf{(19)}$$

where the Mel frequency mapping is given by Equation 20:

$$mel(f) = 2595.log_{10} \quad \mathbf{(20)}$$

Log and DCT: as shown in Equation 21.

$$c_n = \sum_{m=0}^{M-1} \log S(m) \cos\left(\frac{\pi n(m + 0.5)}{M}\right) \quad \mathbf{(21)}$$

Amplitude to dB: as shown in Equation 22.

$$\boldsymbol{MFCC\ dB} = 10Log_{10} \frac{MFCC^2}{\max(MFCC)} + 80 \quad \mathbf{(22)}$$

F. Semantic Matching

Similarity score: as shown in Equation 23.

$$Similarity = \frac{2M}{T} \quad \mathbf{(23)}$$

where *M* is the number of matching characters, and *T* is the total number of characters.

V. RESULTS

A. Descriptive Statistics

Table 2 (Appendix) summarizes statistics by vehicle class.

The distribution of Mean MFCC values across vehicle classes is shown in Figure 6 (Appendix), revealing significant variability.

B. Correlation Analysis

Pearson correlation coefficients:

Mean MFCC vs. Adjusted Leq dBA: -0.45 (moderate negative correlation, louder sounds have lower MFCCs).

Mean MFCC vs. Speed kmh: -0.12 (weak correlation).

Adjusted Leq dBA vs. Speed kmh: 0.38 (moderate positive correlation).

The correlation matrix is visualized in Figure 7 (Appendix), which confirms the relationships between variables.

The relationship between adjusted Leq and Mean MFCC is further illustrated in Figure 8 (Appendix).

C. ANOVA

One-way ANOVA results:

Mean MFCC across vehicle classes:

$F = 15.32$, $p < 0.001$, indicating significant differences.

Adjusted Leq dBA across vehicle classes:

$F = 8.45$, $p < 0.01$, showing significant variation.

D. Location-Based Analysis

High-noise locations (mean Adjusted Leq dBA > 67 dBA):

Total Energies Outering: 67.49 dBA, Speed = 55.75 km/h.

Around Baba Dogo Rd: 68.13 dBA, Speed = 51.35 km/h.

Figure 9 (Appendix) highlights the top 10 locations with the highest average adjusted Leq values.

E. Speed Distribution

The speed distribution across vehicle classes is shown in Figure 10 (Appendix).

VI. DISCUSSION

The Doppler effect increases Leq by 0.1–0.2 dBA, with greater impacts at higher speeds (e.g., 76.33 km/h at Nyayo LA Rd). MFCC features effectively distinguish vehicle classes, with heavy trucks showing distinct low-frequency signatures. Visualizations highlight noise hotspots and feature variations, supported by statistical significance from ANOVA. Future work should incorporate dynamic speed data and advanced audio features.

VII. CONCLUSION

This study demonstrates the integration of Doppler effect adjustments with MFCC features for vehicle noise analysis in Nairobi. Visualizations and statistics reveal significant patterns, enhancing urban noise understanding. MFCC methodology is robust in extracting Doppler patterns of road traffic noise.

Acknowledgement

The authors are grateful to the University of Nairobi management for extending good support to the research team administratively.

Declaration of conflict of interest

The research did not receive any funding support from any organization, and none of the authors has a personal interest

AI -assisted content creation

Authors declare that no content was created through AI generation.

REFERENCES

[1] Damiano, Stefano; Dietzen, Thomas; and van Waterschoot, Toon (2024). Frequency Tracking Features for Data-Efficient Deep Siren Identification. arXiv preprint arXiv:2409.08587. https://doi.org/10.48550/arXiv.2409.08587.

[2] George, Jobin; Cyril, Anila; Koshy, Bino I.; and Mary, Leena (2013). Exploring Sound Signature for Vehicle Detection and Classification Using Artificial Neural Networks. International Journal on Soft Computing, 4(2), pp. 29–36. https://doi.org/10.5121/ijsc.2013.4203.

[3] Beranek, Leo L. (1988). Noise and Vibration Control (Revised Edition). Institute of Noise Control Engineering of the USA, Washington, DC. Available at: https://pdhonline.com/courses/m104/m104.pdf. Accessed June 16, 2026

[4] Acosta, Oscar Esneider; Montenegro Marín, Carlos Enrique; and González Crespo, Rubén (2021). Sound Measurement and Automatic Vehicle Classification and Counting Applied to Road Traffic Noise Characterization. Soft Computing, 25, pp. 12075–12087. https://doi.org/10.1007/s00500-021-05766-6.

[5] Quinn-Vawter, Christopher (2016). Evaluation of Sabine's Formula on the Prediction and Control of Reverberant Noise in a Modern LEED Platinum Certified Research Building. Master of Science Thesis, Colorado State University, Fort Collins, Colorado, USA. Available at: https://api.mountainscholar.org/server/api/core/bitstreams/8a3db343-6131-40a7-b13b-e2ef6493333d/content. Accessed June 16, 2026.

[6] Yang, B., Yin, J., Ye, Z., Yang, S., & Wang, L. (2024). Development and Testing of an Active Noise Control System for Urban Road Traffic Noise. Applied Sciences, 14(1), 175. https://doi.org/10.3390/app14010175.

[7] Karim Mohammed Rezaul, Md. Jewel, Md Shabiul Islam, Kazy Noor e Alam Siddiquee, Nick Barua, Muhammad Azizur Rahman, Mohammad Shan-A-Khuda, Rejwan Bin Sulaiman, Md Sadeque Imam Shaikh, Md Abrar Hamim, F.M Tanmoy, Afraz Ul Haque, Musarrat Saberin Nipun, Navid Dorudian, Amer Kareem, Ahmmed Khondokar Farid, Asma Mubarak, Tajnuva Jannat and Umme Fatema Tuj Asha. "Enhancing Audio Classification Through MFCC Feature Extraction and Data Augmentation with CNN and RNN Models". International Journal of Advanced Computer Science and Applications (IJACSA) 15.7 (2024). http://dx.doi.org/10.14569/IJACSA.2024.0150704.

[8] Katzberg, Frank; Maass, Michael; and Mertins, Alfred (2024). Doppler Frequency Analysis for Sound-Field Sampling With Moving Microphones. Frontiers in Signal Processing, 4, Article 1304069. https://doi.org/10.3389/frsip.2024.1304069.

[9] Mayvan, A. D.; Beheshti, S. A.; and Masoom, M. H. (2015). Classification of Vehicles Based on Audio Signals Using Quadratic Discriminant Analysis and High Energy Feature Vectors. International Journal on Soft Computing, 6(1), pp. 53–64. https://doi.org/10.5121/ijsc.2015.6105.

[10] *(text in Chinese)* Li Chengjuan; Yi Qiang; Li Baoqing; and Wang Guohui (2021). Application of Mel-Frequency Cepstral Coefficients Based on Fundamental Frequency in Vehicle Recognition. Journal of Chongqing University, 44(11), pp. 17–23. https://doi.org/10.11835/j.issn.1000-582X.2021.11.003.

[11] Ma, Q., Ma, L., Liu, F., & Sun, D. (2022). A Novel Driving Noise Analysis Method for On-Road Traffic Detection. Sensors, 22(11), 4230. https://doi.org/10.3390/s22114230.

[12] Krishna, Gautam; Tran, Co; Carnahan, Mason; Han, Yan; and Tewfik, Ahmed H. (2020). Generating EEG Features from Acoustic Features. arXiv preprint arXiv:2003.00007. https://doi.org/10.48550/arXiv.2003.00007.

[13] Z. K. Abdul and A. K. Al-Talabani (2022). "Mel Frequency Cepstral Coefficient and its Applications: A Review." IEEE Access, vol. 10, pp. 122136–122158. https://doi.org/10.1109/ACCESS.2022.3223444.

[14] Wieczorkowska, Alicja; Kubera, Elżbieta; Słowik, Tomasz; and Skrzypiec, Krzysztof (2018). Spectral Features for Audio-Based Vehicle and Engine Classification. Journal of Intelligent Information Systems, 50, pp. 265–290. https://doi.org/10.1007/s10844-017-0459-2.

[15] Liang, Bin; Iwnicki, Simon D.; and Zhao, Yunshi (2013). Application of Power Spectrum, Cepstrum, Higher Order Spectrum and Neural Network Analyses for Induction Motor Fault Diagnosis. Mechanical Systems and Signal Processing, 39(1–2), pp. 342–360. https://doi.org/10.1016/j.ymssp.2013.02.016.

[16] Ferguson, Brian G.; and Ferguson, Eric L. (2025). Understanding the Cepstrum: An Acoustic Signal Processing Method with Quirky Terminology. Acoustics Today (The Acoustical Society of America), 21(2), pp. 20–27. Available at: https://acousticstoday.org/wp-content/uploads/2025/11/FA-Understanding.pdf. Accessed June 16, 2026.

[17] Abdul, Zrar Kh. (2019). Kurdish Speaker Identification Based on One-Dimensional Convolutional Neural Network. Computational Methods for Differential Equations, 7(4), pp. 566–572. Available at: https://cmde.tabrizu.ac.ir/article_9205.html. Accessed: July 3, 2026

[18] Oppenheim, Alan V.; Willsky, Alan S.; and Nawab, S. Hamid (1996). Signals and Systems, 2nd Edition. Prentice-Hall, Englewood Cliffs, New Jersey, USA. Available at: https://www.cedric-richard.fr/assets/files/Signals_and_Systems_2nd_Edition_by_Oppen.pdf. Accessed June 16, 2026.

[19] Oppenheim, Alan V.; Schafer, Ronald W. (2014). Discrete-Time Signal Processing, 3rd Edition. Pearson New International Edition. Available at: https://www-elec.inaoep.mx/~rogerio/Tres/DSPoppenheimetal.pdf. Accessed June 16, 2026.

[20] Zhao, Xiaojia; and Wang, DeLiang (2013). Analyzing Noise Robustness of MFCC and GFCC Features in Speaker Identification. In: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 7204–7208. https://doi.org/10.1109/ICASSP.2013.6639061.

[21] Ma, Wing-Kin; Hsieh, Tsung-Han; and Chi, Chong-Yung (2010). DOA Estimation of Quasi-Stationary Signals With Less Sensors Than Sources and Unknown Spatial Noise Covariance: A Khatri–Rao Subspace Approach. IEEE Transactions on Signal Processing, 58(4), pp. 2168–2180. https://doi.org/10.1109/TSP.2009.2034935.

[22] Timotius, Ivanna Kristianti (2020). Computational Methods for Gait Analysis in Rodents. FAU Studies in Informatics, FAU University Press, Erlangen, Germany. ISBN: 978-3-96147-321-2. https://doi.org/10.25593/978-3-96147-321-2.

[23] Blackman, R. B.; and Tukey, J. W. (1958). The Measurement of Power Spectra from the Point of View of Communications Engineering — Part I. Bell System Technical Journal, 37(1), pp. 185–282. https://doi.org/10.1002/j.1538-7305.1958.tb03874.x.

[24] Cooley, James W.; and Tukey, John W. (1965). An Algorithm for the Machine Calculation of Complex Fourier Series. Mathematics of Computation, 19(90), pp. 297–301. https://doi.org/10.1090/S0025-5718-1965-0178586-1.

[25] Duhamel, Pierre; and Vetterli, Martin (1990). Fast Fourier Transforms: A Tutorial Review and a State of the Art. Signal Processing, 19(4), pp. 259–299. https://doi.org/10.1016/0165-1684(90)90158-U.

[26] Rabiner, Lawrence R.; and Schafer, Ronald W. (1978). Digital Processing of Speech Signals. Prentice-Hall, Englewood Cliffs, New Jersey, USA. Available at: https://ie.u-ryukyu.ac.jp/~asharif/pukiwiki/attach/Acoustic%20Speech%20Signal%20Processing_Prentice%20Hall%20-%20Digital%20Processing%20of%20Speech%20

Signals.pdf. Accessed June 16, 2026.

[27] Feldman, Hume A.; Kaiser, Nick; and Peacock, John A. (1994). Power-Spectrum Analysis of Three-Dimensional Redshift Surveys. The Astrophysical Journal, 426, p. 23. https://doi.org/10.1086/174036.

[28] N. Ahmed; T. Natarajan; and K. R. Rao (1974). "Discrete Cosine Transform." IEEE Transactions on Computers, vol. C-23, no. 1, pp. 90–93. https://doi.org/10.1109/T-C.1974.223784.

[29] Strang, Gilbert (1999). The Discrete Cosine Transform. SIAM Review, 41(1), pp. 135–147. https://doi.org/10.1137/S0036144598336745.

[30] Nithya, T. M.; Dhivya, P.; Sangeethaa, S. N.; and Rajesh Kanna, P. (2024). TB-MFCC multifuse feature for emergency vehicle sound classification using multistacked CNN – Attention BiLSTM. Biomedical Signal Processing and Control, 88, 105688. https://doi.org/10.1016/j.bspc.2023.105688.

[31] Yu, Dong; Deng, Li; Droppo, Jasha; Wu, Jian; Gong, Yifan; and Acero, Alex (2008). A Minimum-Mean-Square-Error Noise Reduction Algorithm on Mel-Frequency Cepstra for Robust Speech Recognition. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 4041–4044. https://doi.org/10.1109/ICASSP.2008.4518541.

[32] Nakagawa, Seiji; Wang, Longbiao; and Ohtsuka, Shinji (2012). Speaker Identification and Verification by Combining MFCC and Phase Information. IEEE Transactions on Audio, Speech and Language Processing, 20(4), pp. 1085–1095. https://doi.org/10.1109/TASL.2011.2172422.

[33] Al-Kaltakchi, Musab T. S.; Woo, Wai Lok; Dlay, Satnam S.; and Chambers, Jonathon A. (2017). Evaluation of a Speaker Identification System With and Without Fusion Using Three Databases in the Presence of Noise and Handset Effects. EURASIP Journal on Advances in Signal Processing, 2017(80), Article 51. https://doi.org/10.1186/s13634-017-0515-7.

[34] Pires, Inês M.; Santos, Ricardo; Pombo, Nuno; Garcia, Nelson M.; Flórez-Revuelta, Francisco; Spinsante, Susanna; Goleva, Ralitsa; and Zdravevski, Emil (2018). Recognition of Activities of Daily Living Based on Environmental Analyses Using Audio Fingerprinting Techniques: A Systematic Review. Sensors, 18(1), 160. https://doi.org/10.3390/s18010160.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

The authors equally contributed in the present research, at all stages from the formulation of the problem to the final findings and solution.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

No funding was received for conducting this study.

Conflict of Interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.

Appendix

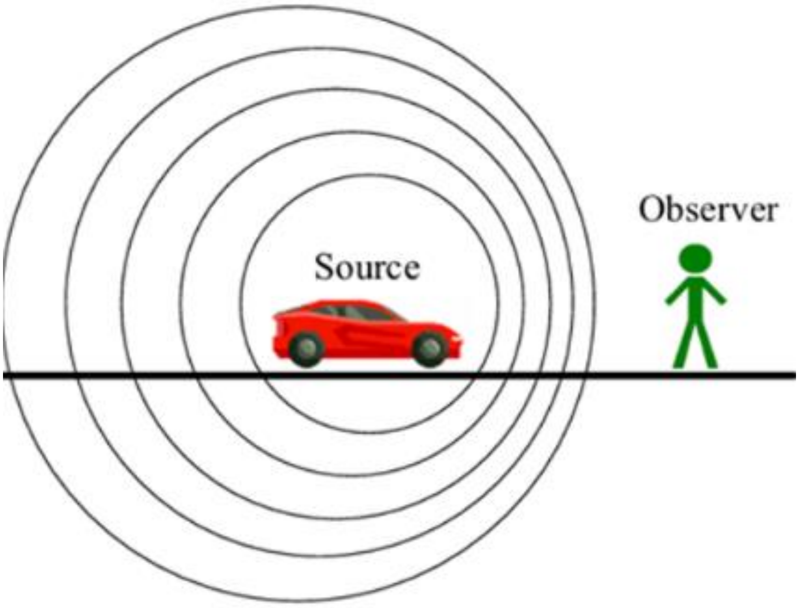

Figure 1: Representation of waves as a group of circles. (Source: Created by the authors.)

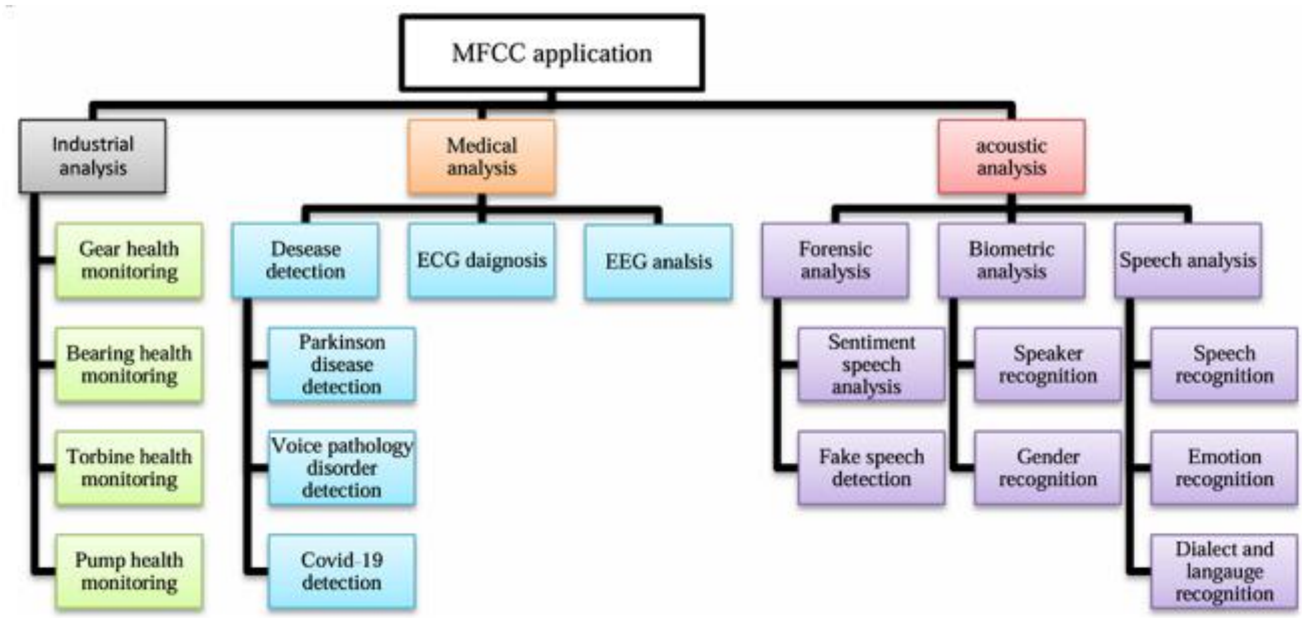

Figure 2: Common application of MFCC (Source: [13])

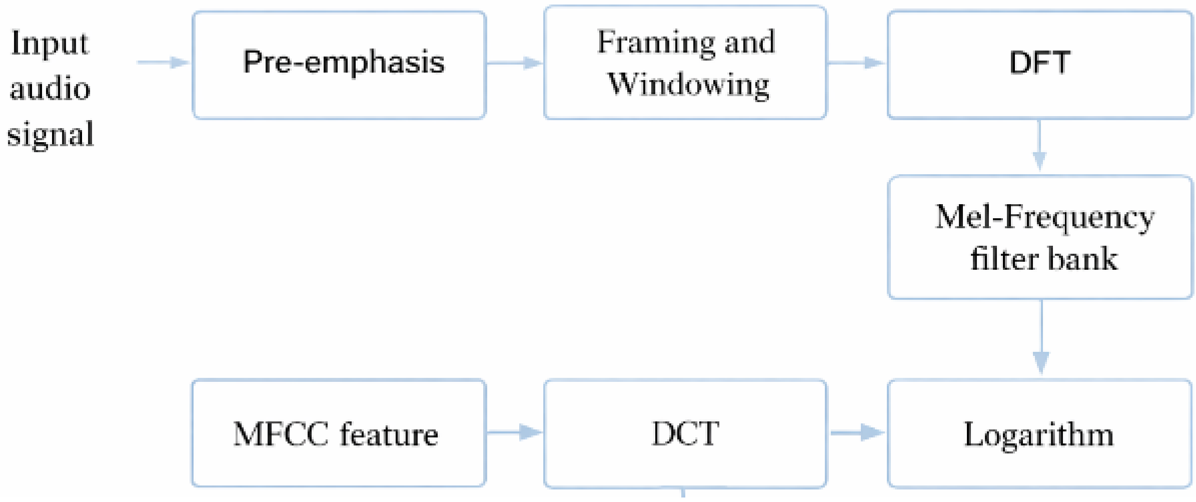

Figure 3: MFCC procedural framework. (Source: [13])

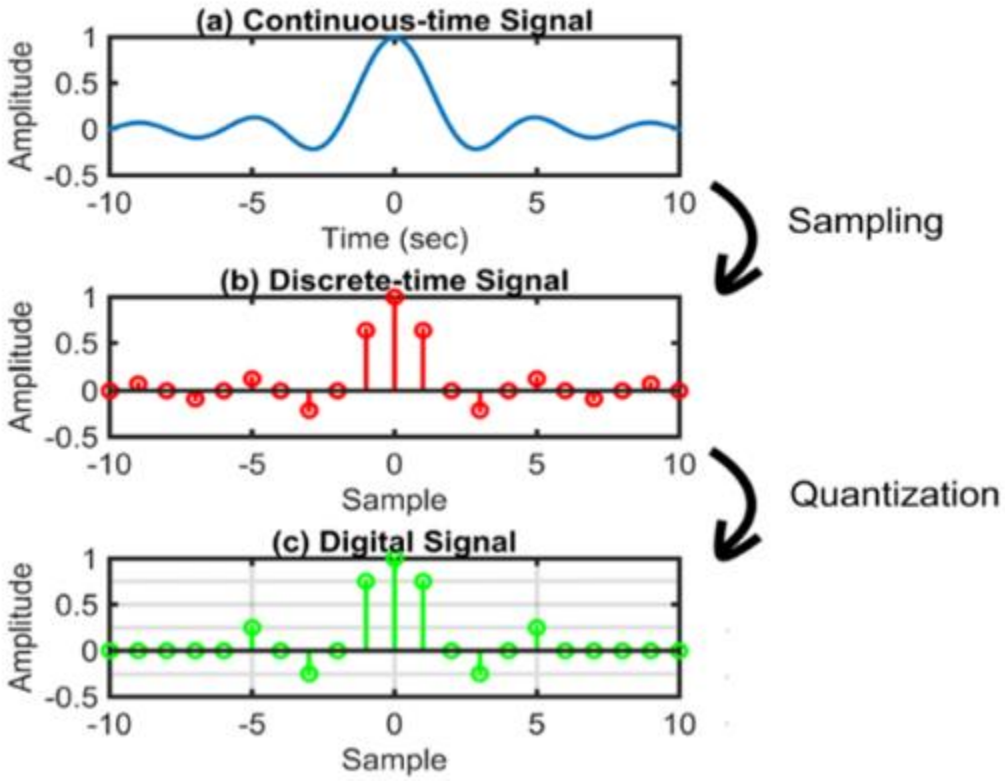

Figure 4: Examples of signals (a) Continuous-time signal (b) Discrete-time signal, which can also be represented as a sequence of numerical values, [0, 0.07, 0, -0.09, 0, 0.13, ...] (c) Digital signal (Source: Created by the authors.)

DOI: 10.46300/91016.2026.10.1 Volume 10, 2026

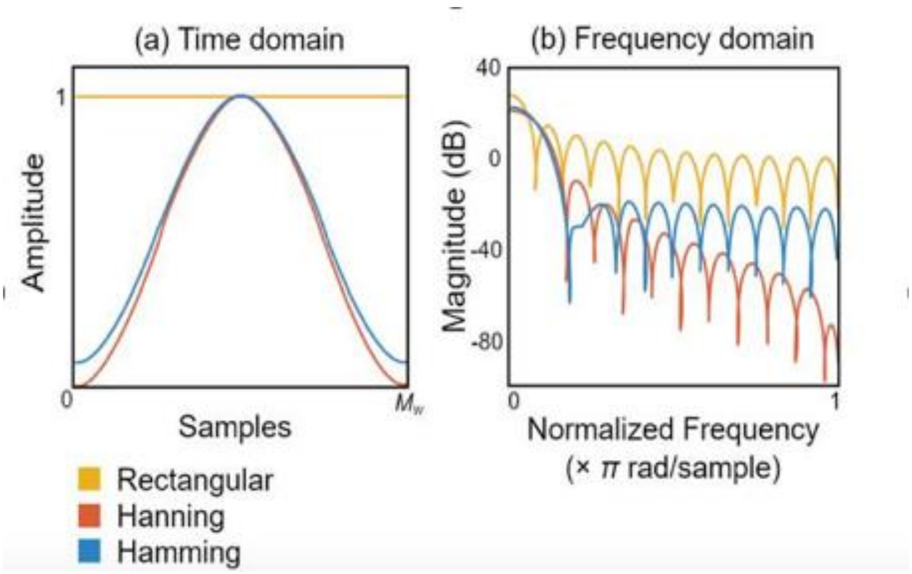

Figure 5: The Rectangular, Hanning, and Hamming windows in the time and frequency domains (Source: [11])

Table 1: Sample Data (Source: Created by the authors.)

Road	**Class**	**Mean MFCC**	**Std MFCC**	**Leq dBA**	
	Adjusted Leq Speed kmh				
Around Arya School 37.71	Heavy truck	-34.77	19.53	64.54	64.68
Winners Chapel (Likoni Rd) 48.23	PSV	-35.31	18.23	66.43	66.60

Table 2: Descriptive Statistics by Vehicle Class (Source: Created by the authors.)

Class	**Mean MFCC (dB)**	**Std MFCC (dB)**	**Leq dBA**	**Adjusted Leq dBA**	**Speed kmh**
Motorcycle	-35.62 ± 2.31	19.32 ± 1.87	66.12	66.29 ± 0.17	47.85
SUV	-34.89 ± 2.45	18.76 ± 1.54	66.54	66.71 ± 0.18	48.32
Heavy Truck	-36.01 ± 2.67	19.85 ± 1.23	65.98	66.15 ± 0.16	47.61
PSV	-35.45 ± 3.12	19.01 ± 1.98	66.72	66.89 ± 0.19	49.03
Bus	-36.78 ± 2.89	19.56 ± 1.45	66.45	66.62 ± 0.17	48.15

DOI: 10.46300/91016.2026.10.1

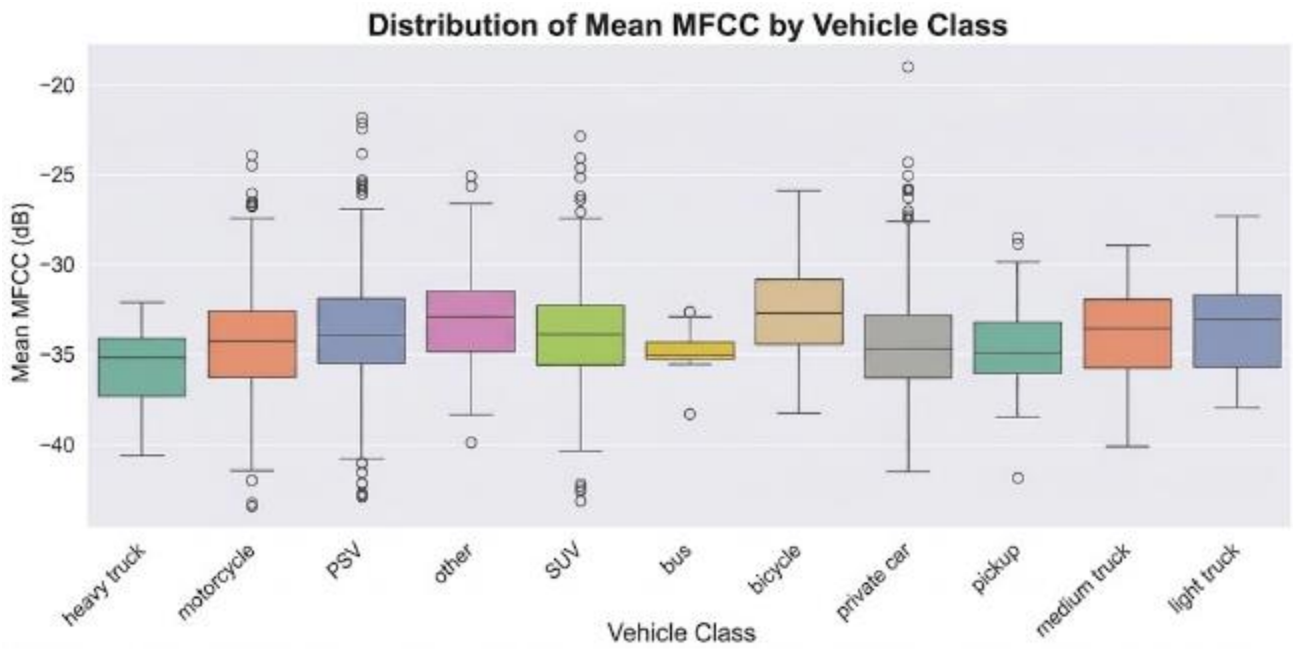

Figure 6: Distribution of Mean MFCC by Vehicle Class. Shows variability in MFCC features, with heavy trucks exhibiting lower mean MFCCs (−36 to −40 dB) due to low- frequency noise. (Source: Created by the authors.)

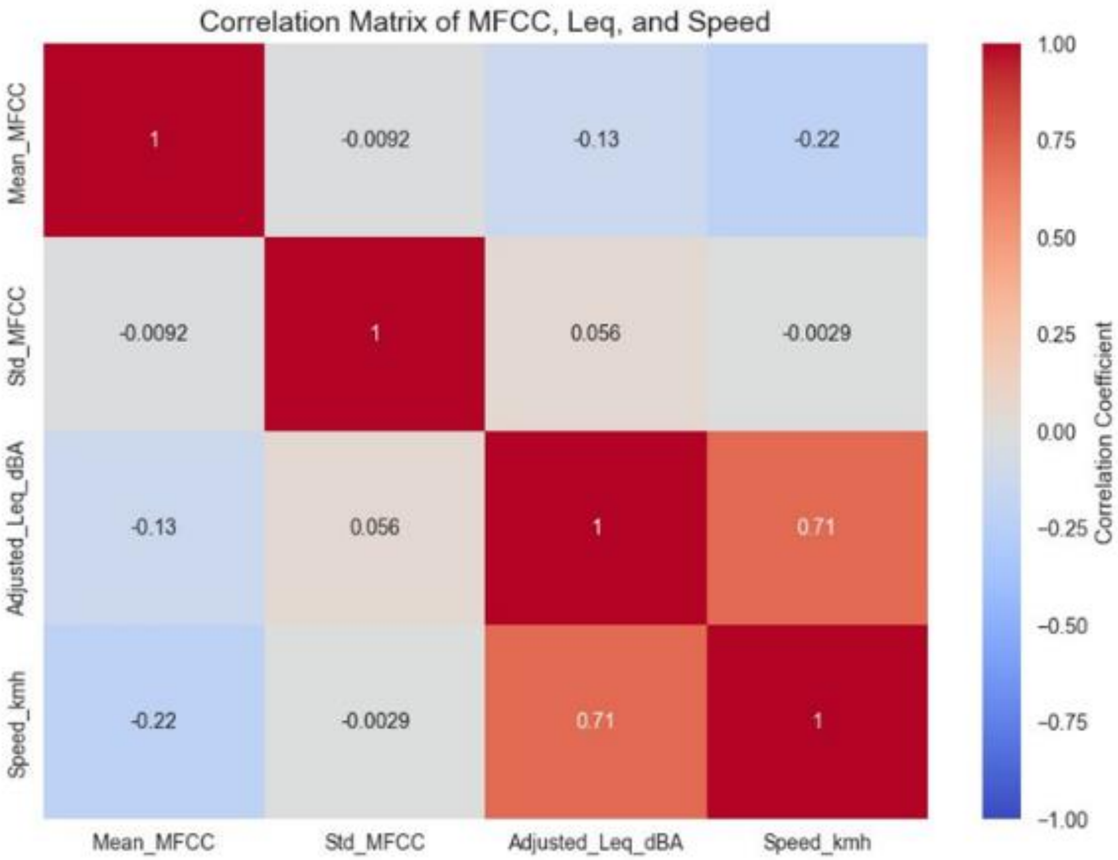

Figure 7: Correlation Matrix of MFCC, Leq, and Speed. Visualizes relationships between variables, confirming moderate correlations. (Source: Created by the authors.)

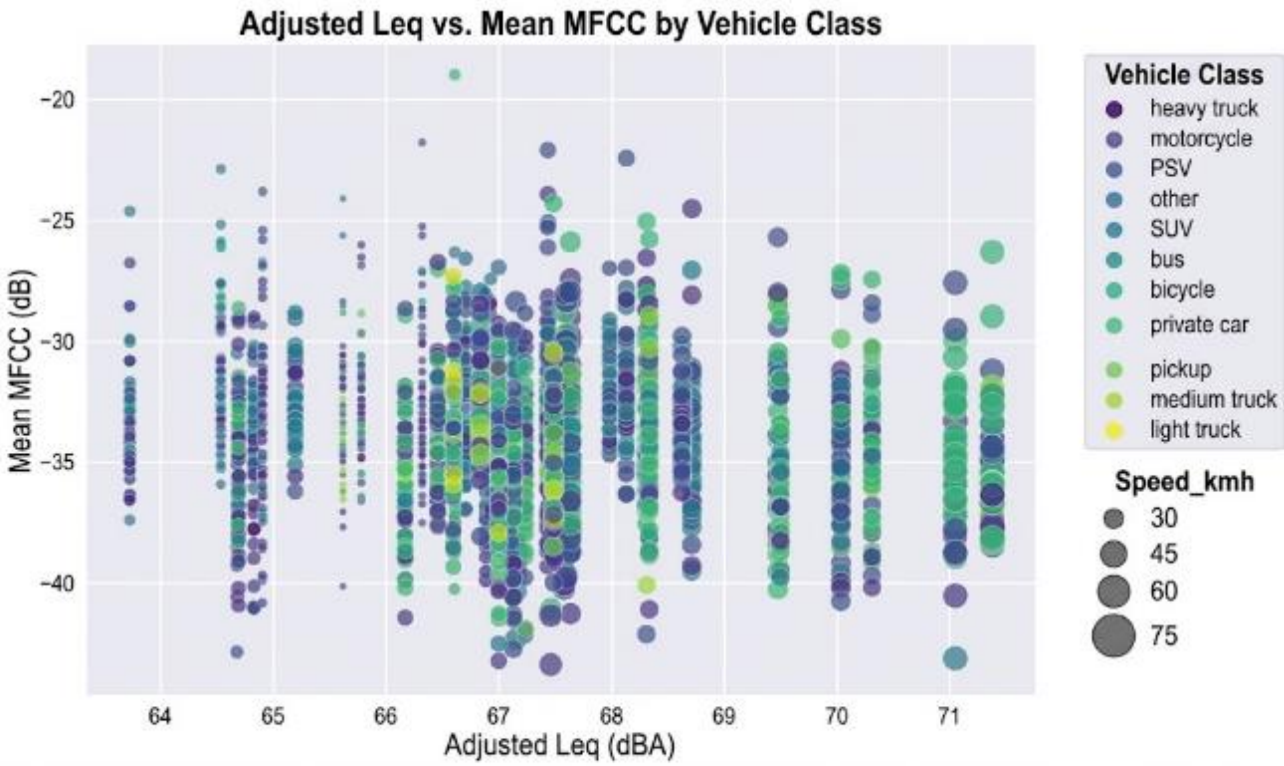

Figure 8: Adjusted Leq vs. Mean MFCC by Vehicle Class. Illustrates a negative correlation, with PSVs showing higher noise levels and lower MFCCs. (Source: Created by the authors.)

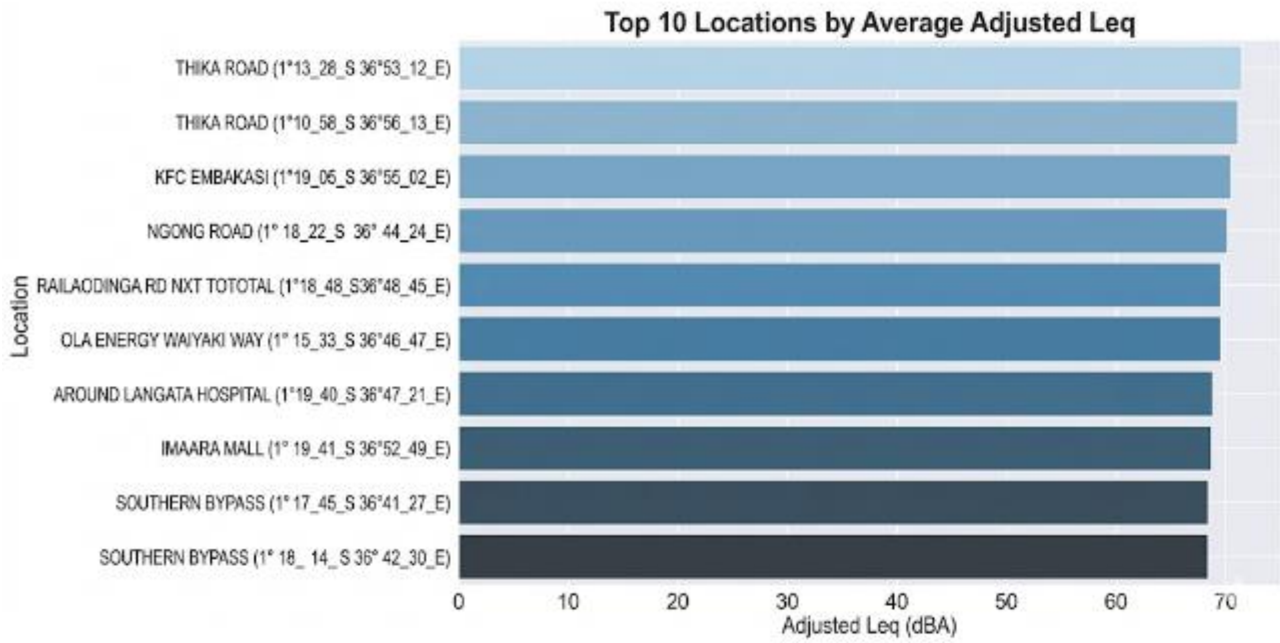

Figure 9: Top 10 Locations by Average Adjusted Leq. Highlights locations like Total Energies Outering with elevated noise levels. (Source: Created by the authors.)

International Journal of Neural Networks and Advanced Applications
DOI: 10.46300/91016.2026.10.1 Volume 10, 2026

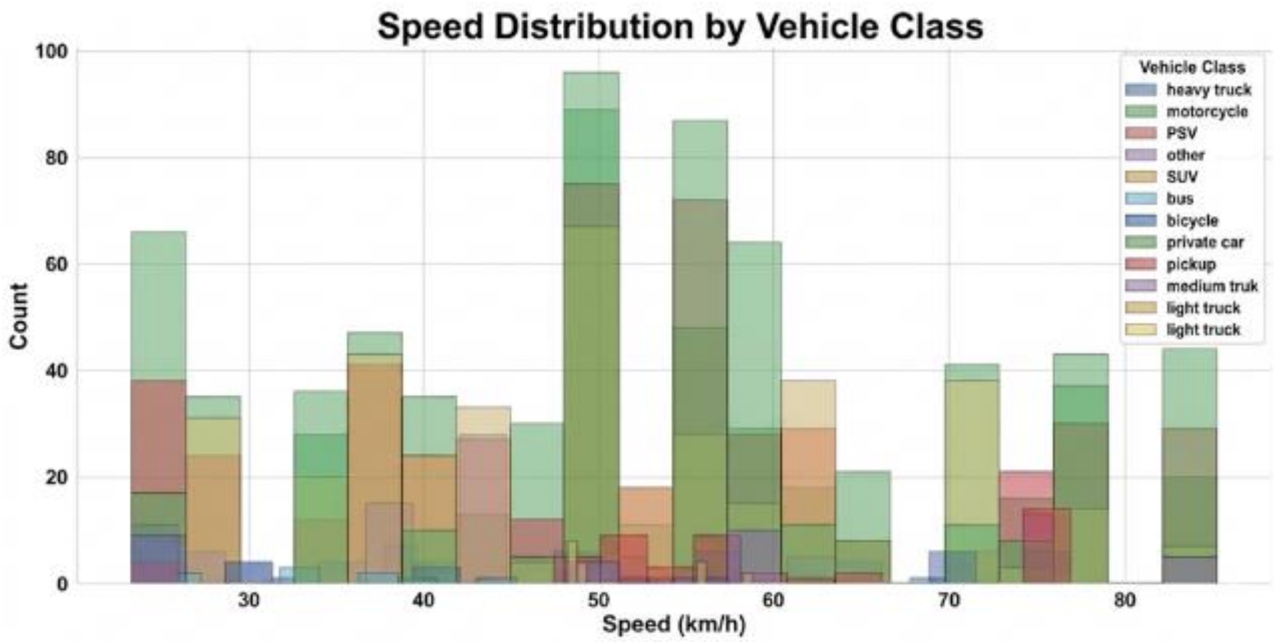

Figure 10; Speed Distribution by Vehicle Class. Shows speed distribution, with PSVs reaching higher speeds (up to 76.33 km/h). (Source: Created by the authors.)