The Timbre Quality of Road Traffic Noise: A Case of Tambourine-Like Jingles from Nairobi City, Kenya

OCHUNGO AKECH ELISHA[1,2,*], OSANO SIMPSON NYAMBANE[1],
GICHAGA FRANCIS JOHN[1]
[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

Abstract: - Road traffic noise presents a significant environmental challenge as it affects human health. This study examined the timbre attributes of road traffic sounds captured in Nairobi city to highlight tambourine-like zills (jingles). This was done by employing a unique perceptual metaphor, likening them to the blended resonant and shimmering elements of a tambourine. Utilizing audio recordings from 42 sampled sites across Nairobi's road network, the study processed about 2,185 segments to compute essential acoustic parameters such as spectral centroid, spectral roll-off, Mel-frequency cepstral coefficients (MFCCs), and zero-crossing rate (ZCR). These metrics were linked to equivalent sound levels (Leq) from on-site assessments and passenger car unit (PCU) indicators to reveal timbre differences influenced by vehicle categories and flow patterns. Findings demonstrate that heavy vehicles like trucks and buses feature modest spectral centroids (800–1100 Hz), limited spectral roll-off (1200–1600 Hz), and lower ZCR (0.04–0.07), resembling the sustained, foundational tone of a tambourine's drum surface, especially in elevated Leq scenarios (e.g., 84.2 dBA along Thika Road). On the other hand, motorcycles and low-capacity passenger service vehicles (PSVs) present higher spectral centroids (approximately 1800 Hz), expanded spectral roll-off (2200–3200 Hz), and increased ZCR (0.09–0.14), akin to the crisp, metallic accents of tambourine cymbals, more common in moderate Leq contexts (e.g., 68.7 dBA near Uhuru Park). The MFCC evaluations highlighted unique spectral shapes, with reduced averages for heavy vehicles denoting subdued timbres and greater deviations for motorcycles, thereby suggesting abrupt changes. A graphical representation of spectral roll-off across vehicle groups illustrated a gradient from deep resonance to sharp percussion. In conclusion, findings provide practical guidance for noise reduction strategies. This is particularly important for guiding the modulation of the acoustic soundscape for inclusive, human-centered noise abatement planning. This resultant "tambourine model" is a tool for noise perception.

Key-Words: - Road Traffic Noise, acoustic parameters, timbre, tambourine, jingles, modelling, acoustic descriptors.

Received: January 9, 2026. Revised: March 16, 2026. Accepted: July 11, 2026. Published: September 8, 2026.

1 Introduction

The growing urban expansion and increase in vehicle numbers have elevated the issue of road traffic noise as a growing concern in modern societies, as they contribute to health problems, including elevated stress levels, hearing loss, and disrupted quiet rest. While characteristics such as volume or loudness are often the main focus of noise mitigation efforts, the qualitative aspects of noise, including timbre, that is, the distinctive sonic character that separates equivalent sounds in frequency and amplitude, significantly influence how disruptive the noise feels. Timbre involves a complex mix of frequency distributions and time-based variations, making certain types of noise more disruptive than others despite having similar overall sound pressure levels.

Assessing road traffic noise levels has primarily relied on the use of energy-based metrics such as the equivalent continuous sound level (Leq). Such metrics typically provide important information concerning the noise exposure duration and its intensity, though they fail to capture the nuanced perceptual attributes that determine annoyance and health impacts.

This gap is more evident in developing cities, such as Nairobi, Kenya. Recent progress in psychoacoustics and environmental sound has shown that timbre-related features are important in shaping human reactions to urban soundscapes, with significant effects on cardiovascular health, cognitive function, and general psychological well-being.

The study thus proposes an innovative interpretive method to bridge this gap by examining road traffic noise timbre, drawing an analogy to a tambourine, an instrument that uniquely combines deep vibrations from its drumhead with sparkling overtones from its metallic jingles (zils). This comparison helps in understanding the layered nature present in urban sounds, that is, the steady low-end rumble from heavy goods vehicles against the fleeting high-frequency components from agile, lightweight vehicles and/or passenger cars.

The tambourine metaphor offers several advantages, including an intuitive understanding, due to the dual sonic character of the instrument, thus allowing for the provision of an accessible framework so that non-specialists can comprehend complex acoustic phenomena, the perceptual suitability that is evident in the contrast between the resonant and the percussive elements which reflect the subjective experience of different traffic noise sources, the analytical utility since it guides feature selection and interpretation, focusing attention on parameters that distinguish these timbral qualities, and policy communication as it aids the dialogue between the public, acousticians, and urban planners..

An intriguing case study for road traffic noise research is Nairobi city. This is because it faces a diverse vehicle fleet, rapid urbanization, and limited noise mitigation infrastructure. The city's soundscape reflects the problems that are faced by many developing urban centers, where a complex acoustic environment is created by mixed traffic patterns, including heavy trucks, matatus (PSVs), motorcycles, and non-motorized vehicles. Unlike cities in developed nations with relatively homogeneous vehicle types and stricter emission standards, Nairobi's traffic noise exhibits extreme variability in both noise level and timbre.

Based on the field data from Nairobi city's road network, the study team applied digital audio signal processing techniques to derive timbre indicators—spectral centroid, spectral roll-off, MFCCs, and ZCR—from 2,185 audio recordings, associating them with Leq measurements and PCU data. This study attempts to present a novel way of understanding road traffic noise timbre, which has received few prior interpretive models for traffic noise characterization. Specifically, the study aimed to: (a) to quantify timbre characteristics across different vehicle categories from the 42 sampled sites in Nairobi, (b) to establish relationships between timbre descriptors and conventional metrics (Leq, PCU), (c) to illustrate and apply the tambourine metaphor for improved noise mitigation approaches, and (d) to provide actionable recommendations for urban acoustic planning applicable in the context of developing nations.

The value is in supporting evidence-based guidelines, such as targeted sound controls and urban spatial planning layout optimizations, to lessen road traffic noise impacts on public health. This research adds to the growing body of knowledge on urban soundscapes, especially in emerging economies, where mixed traffic compositions and infrastructure limitations that are present generate discernible acoustic challenges that are not adequately addressed by existing frameworks from other cities that have been developed in the world. The subsequent sections in this paper provide the materials and methods, the results, the discussion of the findings, and the conclusion.

2 Materials and Methods

2.1 Materials

Continuing exposure to Road Traffic Noise, documented and discussed extensively in especially health-related literature, is associated with serious health concerns. The World Health Organization's Environmental Noise Guidelines have identified road traffic noise as a major environmental health risk, with supported evidence associating long-term exposure to cardiovascular disease, metabolic disorders, and adverse mental health problems. Scholars, including those in [1], have shown mechanistic studies that link noise exposure to vascular dysfunction through oxidative stress and inflammatory responses. Nighttime noise has been particularly emphasized as a source of disturbances that trigger sympathetic nervous system activation.

Other researchers, such as [2], conducted a comprehensive review of the literature on both auditory and non-auditory health effects, [3]. They noted that the negative impact of noise on human health is manifested in multiple areas, including hearing loss, elevated blood pressure, ischemic heart disease events, mental strain, cognitive deficits in children, and sleep disturbance. These maladies occur due to road traffic noise exposure levels commonly encountered in urban environments, with dose-response relationships suggesting no safe threshold below which impacts are absent. The economic burden of road traffic noise-related health effects is substantial, with estimates suggesting billions of dollars in annual healthcare costs and lost productivity across many European cities, for example, [4].

Timbre, loosely taken as a distinct pitch and loudness of sound, allows us to differentiate between sound sources. In ordinary contemporary understanding, it is a sound quality, independent of pitch, intensity, and loudness. When discussing the sound timbre phenomenon, we are essentially talking about perception through the human hearing lens, [5]. The human auditory system and cognitive abilities play an important role in moulding how emotions and reactions are conveyed through timbre. How we perceive and interpret sound ultimately discerns the timbre's ability to differentiate instruments, create textures, and evoke moods, [6].

Sound psychology normally highlights the role of timbre in determining the level of irritation, often going beyond basic loudness measures in predicting annoyance responses. Timbre arises from the spectral envelope, that is, the distribution of energy across frequency bands, and the temporal envelope changes in amplitude over time. Both features are vital for authentic noise evaluation. Work done by [7] provides an in-depth review of the perceptual characteristics of environmental sounds. It identifies brightness, roughness, and sharpness as key dimensions that influence subjective responses independent of the overall level.

Noise perception, especially in urban areas [8], is controlled by contextual factors such as the visual environment, expectations, and activity. Timbre influences the subjective annoyance more than anticipated in traditional dose-response models. For example, psycho-acoustic sharpness, which is related to high-frequency content, quantifies the perceived "acuteness" of a sound. It can be expressed using Equation 1:

$$S = 0.11.\frac{\int 24z'.g(z).N'(z)dz}{\int 24N'(z)dz} \quad (1)$$

z' in the equation represents the critical band rate on the Bark scale, which is an important tool in psychoacoustics. It provides a way that we can use to analyze and understand how humans perceive sound. $g(z)$ is a weighting function that underscores higher frequencies. $N'(z)$ represents a specific loudness that is present at each critical band. This metric shows how high-frequency components present in traffic noise, such as tire squeal or engine hum, can amplify the perceived harshness of sound beyond what is predicted by A-weighted sound pressure level alone.

A few years back, [9] conducted a systematic review to enhance the WHO's guidelines, an exercise that confirmed that spectral characteristics significantly modify annoyance responses, with sounds containing prominent high-frequency components or temporal irregularity being rated as more disturbing than low-frequency rumble at equivalent Leq values.

Modern techniques, such as machine learning (ML) approaches, have been used to advance the identification of noise sources and classify them using neural networks trained on acoustic features. The spectral centroid and spectral roll-off play an important role in differentiating vehicle types, [10]. A vehicle with its engine running and in motion generates a complex sound from its parts, such as the engine, exhaust, air conditioner, and other mechanical components. The analysis of this sound for the purpose of vehicle identification is an interesting practice. This is because heavy vehicles typically exhibit lower values due to bass-heavy engine combustion and exhaust noise, while lighter vehicles and motorcycles reveal higher values from high-frequency tire-road interaction and aerodynamic noise, [11]. The spectral centroid, on the other hand, is a measure of spectral brightness or center of mass, and is particularly useful in distinguishing tonal (low SC) from noisy (high SC) traffic sounds. It is calculated as in Equation 2.

$$SC = \frac{\sum k\, fk|X(fk)|}{\sum k|X(fk)|} \quad (2)$$

Where fk represents frequency bins and |X(fk)| represents the magnitude spectrum from the Fast Fourier Transform (FFT). Recent models that have incorporated temporal variations of spectral centroid have achieved improved accuracy in dynamic urban environments where multiple vehicle types overlap.

Spectral roll-off, commonly defined at the 85% cumulative energy threshold, helps identify the effective bandwidth of noise sources calculated as in Equation 3:

$$SRO = f_r \; where \; \sum_{k=0}^{r} |X(k)|^2 = 0.85. \sum_{k=0}^{N} |X(k)|^2 \quad (3)$$

This parameter is especially informative for differentiating low-bandwidth sources (heavy vehicles) from broadband sources (motorcycles, tire noise), with recent refinements accounting for urban reverberation effects that broaden spectral content through multipath reflections.

The feature extraction is the process of extracting and tackling hidden information in the raw data signal. Cepstrum, such as Mel Frequency Cepstrum Coefficient (MFCC), can extract harmonics and sidebands of the spectrum of the signal. MFCC is a framework feature extraction technique that has been adopted in many areas and frequently reported to be useful for various applications, including in acoustics. Here, it is used as a standard in speech recognition and audio classification because it reflects the perceptually relevant frequency response curve modeled by the mel scale, which approximates human auditory perception, [12]. The mel scale transformation is defined in Equation 4 as:

$$MEL\,(f) = 2595.\log_{10} 1 + \frac{f}{700} \quad (4)$$

Road traffic noise comes from different motor vehicle parts, including the engines [13], horns [14], etc. In the analysis of road traffic noise, MFCC is able to capture perceptual differences by emphasizing frequencies most relevant to human hearing, with lower-order coefficients representing the spectral envelope shape and higher-order coefficients detailing fine spectral structure. The calculation, as shown in Equation 5, involves: (i) computing the power spectrum via FFT, (ii) applying a triangular mel-scale filter bank, (iii) taking the logarithm of filter bank energies, and (iv) applying the discrete cosine transform (DCT).

$$MFCC_n = \sum_{m=1}^{m} \log(E_m) cos \frac{\pi n(m-0.5)}{M} \quad (5)$$

where Em represents energy in the m-th mel-filter, and M is the number of filters (typically 20-40). Recent enhancements made include gammatone cepstral coefficients (GTCC), which use gammatone functions that approximate the impulse response of the basilar membrane to model auditory filters, calculated as shown in Equation 6.

$$GTCC_k = \sum_{m=1}^{m} \log(G_m) cos \frac{\pi k(m-0.5)}{M} \quad (6)$$

Where G_m represents the gammatone filter bank energies. Studies done from 2023-2025, to determine real-time traffic noise classification, have incorporated the use of MFCCs with convolutional neural networks (CNNs), thus achieving accuracies exceeding 95% in mixed urban scenarios with overlapping sources, [15].

Noise monitoring methods implemented in urban environments have conventionally applied Leq for exposure assessment and distress forecasting, though they increasingly incorporate timbre descriptors for refined perceptual evaluation. While Leq models average the acoustic energy over time, they fail to capture the perceptual nuances that determine subjective responses. Psychoacoustic models like loudness (ISO 532-1) incorporate spectral weighting as in Equation 7:

$$N = 0.08. \int_0^{24} N'(z) dz \quad (7)$$

Where N is loudness in sones, and N′(z) represents specific loudness distributed across the Bark scale. To improve annoyance prediction in urban parks, recent work by [16] combines Leq with timbre descriptors, finding that models that incorporate sharpness and roughness are able to explain 30-40% more variance in subjective ratings than Leq alone.

The research effort by [17] has outlined ten fundamental questions in soundscape research, underscoring the need for holistic approaches that are able to integrate physical measurements with perceptual evaluations. Following this, researchers in [18] also introduced soundscape diversity metrics that are able to quantify temporal and spectral variability, which is analogous to biodiversity indices that are present in ecology. This further inspired the effort by [19], who demonstrated that individual sensitivity to soundscape quality varies substantially, necessitating person-centered assessment approaches, [20].

Humans have the natural capability to effortlessly identify various sounds [21]; for example, many people can easily distinguish between speech and music, car and truck sounds, baby and adult speech quality [22], various speakers, noise, and enjoyable sounds. Figure 1 shows a general classification of audible sound. In the modern day, however, machines are also able to classify between various sounds as humans, but with faster computing power. This is what is called machine hearing whose knowledge literature is domiciled in acoustic scene classification, [23]. The concept of acoustic scene classification (ASC) was pioneered in the literature by [24]. The machine is trained through a process known as machine learning (ML) to classify various sounds present in an acoustic scene, as shown in the framework diagram by [25].

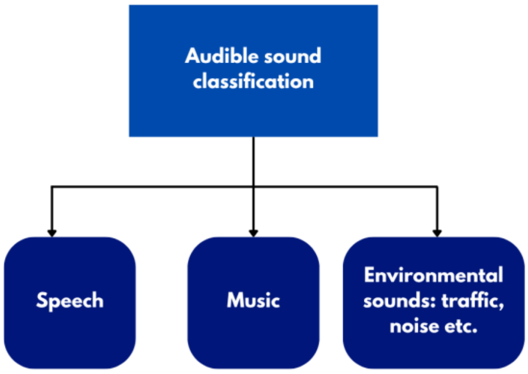

Fig. 1: Sound classification
Source: Adapted from [26]

Sound types are classically expressed along four perceptual dimensions, namely duration, loudness, pitch, and timbre. Using the lens of psychoacoustics, an archetypal type of psychophysics focused on analyzing the relationship between human auditory perception and sound properties. It is worth noting that timbral qualities require indirect and multidimensional methods to be fully explored because they cannot be easily and unequivocally specified to a listener; hence the use of machine learning tools, [21]. This is based on semantic listening [27], which is the listener's interpretation of a sound as a code, such as the horn of a car or the noise from the road traffic flow.

Usually, an ML process requires robust and discriminatory features to enable the machine to learn accurately and quickly. The computing intelligence of a machine is dictated by the number of training layers and data fed into it. In normal cases, the whole dataset is not fed into the computer or machine during its training to learn its properties; instead, a reduced-sized representation of the signals is used to train the machine. This compact representation of a signal is called a feature. The challenge is to extract the right features to ensure the success of an ML algorithm. The features must be compact in size but must highlight the characteristics of the signal. The Features are so selected that they reduce the size of a signal substantially while still describing the signal as a whole accurately. The reduced version of the signal improves the computational complexity and time complexity of the ML algorithms, which in turn makes it more suitable for real-time applications. This means that the feature extraction is a process of dimension reduction of a signal, which makes the signal more suitable for ML algorithms. This paper focuses only on the features extracted from road traffic noise signals.

Sound metaphors and comparisons to musical instruments enrich technical descriptions and facilitate interdisciplinary communication, yet tambourine-inspired frameworks are novel in road traffic noise literature. Perceptual studies employ multidimensional scaling to map timbre space, consistently identifying dimensions corresponding to brightness (correlated with spectral centroid) and roughness (related to fluctuation strength), as shown in Equation 8:

$$FS = 4 \cdot \sum_i \left(\frac{m_i - \bar{m}}{\bar{m}}\right)^2 \quad (8)$$

Where m_i represents modulation amplitudes and $\bar{m}$ their mean.

Following these advancements, researchers in [28] applied such frameworks in the urban context, which revealed cultural differences in timbre perception [6] and the importance of audio-visual interactions [29] in shaping the overall assessment. Subsequently, researchers in [30] validated soundscape descriptors across eighteen languages through listening experiments, confirming that timbre-related terms, bright, harsh, and dull, exhibit cross-cultural consistency while context-dependent attributes, pleasant and chaotic, show more variation. This supports the use of universal acoustic features such as the spectral centroid while recognizing the need for culturally informed interpretation.

The notion of timbre originates from Western musical tradition. Timbre analysis is used to identify a sound source by deploying perceptual and acoustic dimensions, mainly the spectral centroid and the attack time of a sound. Timbre analysis of urban soundscapes has been discussed by [31] and [32]. They have emphasized smart city applications of acoustic monitoring, stressing the importance of intelligent feature extraction and perceptually grounded assessment frameworks. Ecologists have used the same approach to identify animal calls by detecting their sounds, [33]. Recent developments in transformer architectures for processing MFCC sequences have improved urban noise event detection [34] and segmentation, with potential applications in real-time monitoring systems.

It is efforts like these that have laid the foundational motivation of the current research, in which the timbre analysis of Nairobi city road traffic noise data is the test bed. The work relied on the fact that vehicle type and driving speed have a big influence on the vehicle noise spectrum, [35]. The current research stood on the pillar of work done by [36], whose work evaluated the sound insulation performance of building components that are affected by traffic noise. This work contributes to knowledge in urban soundscape studies in the African context, with a view to providing insights on how to mask low-frequency noise from road traffic. The low-frequency noise (LFN) from road traffic is a typical urban environmental noise pollution. It is difficult to eliminate using traditional noise control techniques. Sound masking has been proven to be an effective way to improve noise perception, [37]. Granted, African urban areas continue to produce road traffic noise research work, especially the application of GIS to generate urban noise maps. What is lacking, however, is a detailed timbre characterization of the noise sources.

The heterogeneity of African urban traffic, including informal public transport (matatus), high motorcycle prevalence, and mixed traffic volumes on the roads, creates a unique acoustic environment distinct from those in Western cities where most applied acoustic research has concentrated over time.

Despite substantial progress in Africa's soundscape research, noticeable gaps remain in applying perceptual metaphors like the tambourine concept to road traffic noise and, by extension, timbre analysis. Existing studies have emphasized objective metrics but tend to neglect intuitive frameworks that are able to bridge technical analysis with policy communication and public engagement. While MFCC and spectral analysis are common in sound classification research, their systematic application to diverse traffic environments, such as in the case of Nairobi's mix of trucks, matatus, motorcycles, and non-motorized vehicles, still remains limited, particularly regarding correlations with Leq and PCU measures.

Furthermore, cultural and contextual differences in perceptual responses remain underexplored in African urban contexts, where soundscape characteristics differ from Western world models. The assumption that dose-response relationships derived from European or North American populations apply universally requires empirical validation in diverse settings. This study addresses these gaps by introducing the tambourine metaphor as an accessible interpretive framework, conducting comprehensive timbre analysis on Nairobi city's road traffic noise recordings, establishing relationships between acoustic features, Leq, and traffic composition, and providing a methodological template that is applicable to similar urban environments in developing regions.

2.2 Methods

From 42 strategically selected locations across Nairobi city, Kenya, Audio recordings were collected, with locations encompassing major roads, residential neighborhoods, and commercial hubs. The sites were also chosen to represent the diversity that exists in urban acoustic environments, including high-traffic corridors, such as Thika Road, Mombasa Road, and Uhuru Highway, mixed-use areas, such as the Central Business District (CBD), Westlands, and Parklands, residential zones such as Kilimani and Eastleigh, and near institutional facilities such as schools, Arya School in Ngara, and parks such as Uhuru Park. This sampling strategy captured different traffic densities, vehicle compositions, and built environment characteristics that affect sound propagation and reflection.

Each recording done in a certain location yielded multiple 6-second audio segments. The audio was captured at a sampling rate of 44,100 Hz using calibrated measurement-grade equipment. The dataset finally comprised a total of 2,185 segments. This provided robust statistical power for analysis.

The recording sessions took place during peak traffic periods, that is, 6:00 AM to 6:00 PM on weekdays, to capture the representative flow conditions and vehicle mix present. Concurrent with audio recordings, other data, such as the vehicle type, that is, heavy trucks, buses, passenger service vehicles (PSVs/matatus), SUVs, motorcycles, bicycles, traffic flow characteristics, comprising of the volume, composition, speed estimates, and environmental conditions temperature, humidity, wind speed, were systematically collected. The measurement locations were georeferenced and characterized by their proximity to noise-sensitive areas such as schools, hospitals, and residential buildings.

Data from two sources complemented acoustic measurements: the Noise Summary Field Data (compiled in the FINAL DATA.xlsx document), which contained on-site equivalent continuous sound level (Leq) measurements recorded using a Type 1 Sound Level Meter (SLM) calibrated according to IEC 61672-1 standards. The measurements employed A-weighting to capture how the human ear perceives sound, and followed ISO 1996-2 protocols for environmental noise assessment, and the Passenger Car Unit (PCU) Conversion Data, which entailed the passenger car unit calculations converting heterogeneous vehicle counts into standardized traffic load estimates, accounting for vehicle size and impact on traffic flow. The PCU factors applied were: Heavy trucks = 3.0, Buses = 2.5, SUVs = 1.0, Motorcycles = 0.5, Bicycles = 0.2. The naming of sites was consistent by matching location identifiers across data sources (for example, "AROUND ARYA SCHOOL" to "ARYA(NGARA)"). This enabled the reliable integration of acoustic features with Leq and traffic composition data.

Processing of the audio signals was done using Python 3.13.7 and the torchaudio library. The torchaudio library was selected for its computational efficiency and integration with deep learning frameworks. The preprocessing pipeline included:

2.2.1 Resampling

This involved resampling all the audio segments to 22,050 Hz. This was done to balance temporal resolution with computational requirements and align with standard practice in environmental sound analysis, while preserving the frequency range relevant to road traffic noise analysis.

2.2.2 Channel Reduction

This involved converting stereo recordings to mono by averaging the channels, reducing dimensionality while preserving important spectral information.

2.2.3 Duration Standardization

This involved the trimming or zero-padding of the 6-second audio segments to 4-second segments during preprocessing, thus ensuring uniform temporal windows for feature extraction while minimizing edge effects. This also ensured consistent comparison across all samples.

2.2.4 Normalization

This involved peak amplitude normalization to account for the differences found in recording gain. This was followed by RMS energy normalization to standardize the loudness across segments.

To characterize traffic noise timbre, four complementary acoustic features were extracted. Each feature provides distinct insights into spectral and temporal structure:

2.2.5 The Zero-Crossing Rate (ZCR)

This quantifies the frequency of signal polarity changes, which serve as a time-domain indicator of high-frequency content and noisiness, calculated as in Equation 9:

$$ZCR = \frac{1}{2(N-1} \sum_{i=1}^{N-1} |\mathrm{sgn}(\mathrm{x[i]} - \mathrm{sgn}(\mathrm{x[i-1]})| \quad (9)$$

The main purpose is to capture the percussive, shimmering quality similar to tambourine jingles. High ZCR values indicate rapid signal oscillations associated with high- frequency components characteristic of tire noise, motorcycle engines, and wind turbulence. Low ZCR suggests smoother, more tonal content from low-frequency engine combustion and exhaust rumble. ZCR was computed frame-wise using 2048- sample windows with 50% overlap, then averaged across the segment.

2.2.6 Spectral Centroid (SC)

Spectral Centroid (SC), which represents the center of mass of the magnitude spectrum, thus providing a measure of spectral brightness, as presented in Equation 10:

$$SC = \frac{\sum_k f_k |X(f_k)|}{\sum_k |X(f_k')|} \quad (10)$$

Its purpose is to differentiate "dark" (low-frequency dominated) from "bright" (high-frequency rich) timbres. Heavy vehicles with large-displacement engines produce low-frequency exhaust noise, thus yielding small spectral centroid values between 800 and 1100 Hz, analogous to the resonant drumhead of a tambourine. Light vehicles and motorcycles generate higher centroid values of 1800+ Hz, from tire-road interaction and high-RPM engines, resembling the bright zils. Spectral centroid was computed per frame using the Hamming-windowed FFT (4096 points) with 75% overlap, then temporally averaged.

2.2.7 Spectral Roll-Off (SRO)

Spectral Roll-Off (SRO) that is, identifying the frequency below which a specified percentage (typically 85%) of the total spectral energy is contained. It is calculated using Equation 3.

Its purpose is to quantify the bandwidth and high-frequency extent, directly supporting the tambourine metaphor. Roll-off values, considered low, of between 1200 and 1600 Hz indicate the present energy concentration in bass frequencies characteristic of heavy vehicle rumble, that is, the drumhead quality, while high values of between 2200 and 3200 Hz suggest broadband spectra with substantial high-frequency content that is the jingle quality, which is typical of motorcycles and tire noise. The 85% threshold was selected as standard in music information retrieval and audio classification.

Ochungo Akech Elisha,
Osano Simpson Nyambane,
Gichaga Francis John

2.2.8 Mel-Frequency Cepstral

Mel-Frequency Cepstral Coefficients (MFCCs). This involves providing a compact, perceptually motivated representation of the spectral envelope shape. It is calculated using Equation 11.

$$MFCC_n = \text{DCT}\ (\log(MelSpectrum) \quad (11)$$

The extraction process involves the computation of the power spectrum via FFT, the application of a triangular mel-scale filter bank (40 filters, 0-11025 Hz), taking the logarithm of filter outputs, approximating perceived loudness, and through the application of a discrete cosine transformation for decorrelation. The first 13 coefficients were retained, excluding the 0th coefficient representing overall energy. They were then converted to the decibel scale calculated using Equation 12:

$$dB = 20\log_{10}(amplitude) \quad (12)$$

And the z-score is normalized using Equation 13:

$$X' = \frac{X-\mu}{\delta} \quad (13)$$

μ and σ in this case represent the mean and standard deviation across the dataset. The purpose is to capture the human-like perception of spectral shape. Lower coefficients represent broad envelope characteristics, whilst the higher coefficients encode fine spectral details. For each segment, the mean and standard deviation were computed across the 13 coefficients.

Through careful site and timestamp matching, the extracted acoustic features were systematically linked to the corresponding Leq measurements and PCU values. This enabled analysis of relationships between timbre characteristics, overall sound pressure level, and the traffic composition.

The analysis also employed standard descriptive and inferential statistics that were implemented with the pandas and SciPy libraries: they involved the grouping of features aggregated by vehicle type and measurement site [38], the summary statistics such as the mean, median, standard deviation, and interquartile range computed for each feature, a correlation analysis, using the Pearson correlation coefficients calculated between acoustic features and Leq/PCU measures, and the visualization of relationships between the data relationships shown through scatter plots, bar charts, and box plots and implemented using Matplotlib. For the bar chart data visualizations of spectral centroid and roll-off, vehicle types were ordered by ascending mean value, creating an intuitive gradient from low-frequency, resonant sources to high-frequency, percussive sources, which reinforces the tambourine metaphor.

The entire analysis was conducted on standard computational hardware running Python 3.13.7 with key dependencies such as torchaudio 2.1.0 for Audio I/O and feature extraction, pandas 2.1.0 for data manipulation and aggregation, matplotlib 3.8.0 for visualization, and SciPy 1.11.0 for statistical analysis. The analysis code is documented in the timbre quality of noise.ipynb, a Jupyter notebook enabling reproducible workflows and interactive exploration.

3 Results

3.1 Overview

A comprehensive analysis of the 2,185 audio segments revealed systematic variations in timbre characteristics across vehicle categories and measurement sites. This strongly supports the tambourine metaphor. The results are organized by acoustic features, followed by their integration with Leq and PCU data.

The results showed that heavy trucks exhibited the lowest mean spectral centroid values (950 Hz; SD = 180 Hz). This reflects the energy concentration in low-frequency bands, which are mostly associated with large-displacement diesel engines and exhaust rumble. Heavy trucks are thus positioned at the "drumhead" end of the tambourine spectrum, with sustained, resonant timbres. Buses revealed slightly higher values of 1050 Hz and SD = 210 Hz; though distinguishable from trucks, they are still firmly placed in the low-frequency category. Passenger service vehicles (PSVs/matatus) occupied the middle position with 1400 Hz and SD = 280 Hz, with their higher centroids resulting from smaller engines that operate at higher revolutions per minute (RPMs) and the increased tire-road noise contribution. Sports Utility Vehicles (SUVs) showed moderate values of 1350 Hz and SD = 240 Hz. This is similar to PSVs but with greater variability, reflecting a mix of vehicle conditions and maintenance states that is common in Nairobi’s mixed fleet. Motorcycles presented the highest spectral centroids, of 1900 Hz and SD = 340 Hz, approaching the "jingle" end of the tambourine spectrum. Their bright, sharp timbres are derived from high-frequency engine harmonics, chain noise, and pronounced tire noise at smaller contact patches.

Ochungo Akech Elisha,
Osano Simpson Nyambane,
Gichaga Francis John

3.1.1 Site-based Patterns
This revealed that high-traffic corridors such as Thika Road showed lower mean spectral centroids of 1100 Hz, since they are dominated by heavy vehicle rumble, while areas near parks and residential zones exhibited higher values of 1500 to 1700 Hz, where lighter vehicles and motorcycles constitute larger proportions of the mix.

3.2 Spectral Roll-Off Analysis

3.2.1 Bandwidth Characteristics
The distribution of the spectral centroid values across vehicle categories is represented in Figure 2. The figure illustrates an increase in the centroid frequency from trucks and buses to motorcycles and PSVs. This indicates a transition from darker, low-frequency timbres to brighter, high-frequency timbres.

Spectral roll-off measurements strongly correlated with spectral centroid ($r = 0.87$, $p < 0.001$), though they provided additional information about high-frequency extent. Heavy trucks showed limited spectral roll-off (mean = 1400 Hz, SD = 260 Hz), with 85% of energy concentrated below 1600 Hz, illustrating narrow-bandwidth, bass-dominated spectra. Buses exhibited a slightly broader bandwidth (mean = 1550 Hz, SD = 290 Hz). PSVs and SUVs demonstrated intermediate roll-off values (2000-2200 Hz), indicating moderately high-frequency content. Motorcycles showcased the highest roll-off frequencies (mean = 2800 Hz, SD = 480 Hz), with a great part of the energy extending to 3200 Hz in some cases.

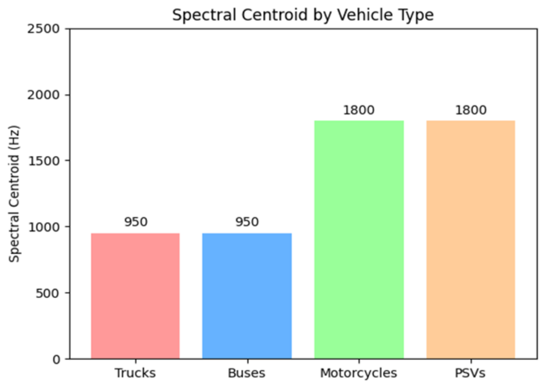

Fig. 2: The Spectral Centroid by Vehicle Type characteristic of broadband noise sources
Source: Created by the authors

3.2.2 Gradient visualization
The bar chart visualization, sorted by ordering the vehicles by their ascending mean roll-off value, created a clear gradient, as shown in Figure 3, from heavy trucks, with deep resonance through buses and PSVs, with mixed character to motorcycles, with sharp percussion. This visual representation effectively represents the tambourine metaphor's descriptive power, showing the continuous variation rather than discrete categories.

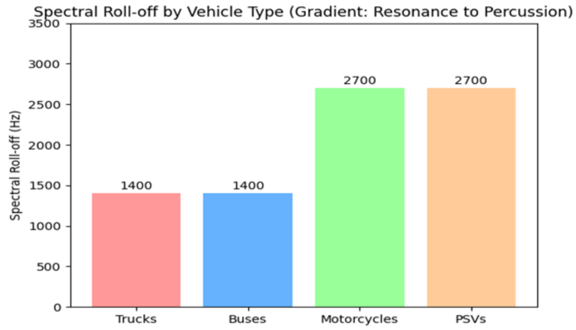

Fig. 3: Spectral Roll-off by Vehicle Type showing the gradient, from resonance to percussion
Source: Created by the authors

3.3 Analysis of the Zero-Crossing Rate (ZCR)

3.3.1 Temporal Characteristics
The analysis of ZCR, as represented in Figure 4, revealed the systematic differences present in signal noisiness and high-frequency temporal structure. Heavy trucks: Mean ZCR = 0.06 (SD = 0.015), indicating smooth, quasi-periodic waveforms. Buses: Mean ZCR = 0.07 (SD = 0.018). PSVs: Mean ZCR = 0.10 (SD = 0.024). SUVs: Mean ZCR = 0.09 (SD = 0.022). Motorcycles: Mean ZCR = 0.12 (SD = 0.031). This shows the rapid signal fluctuations. The correlation between ZCR and spectral centroid was found to be moderate ($r = 0.64$, $p < 0.001$). This suggests that while these features are related, they partially capture the independent aspects of timbre, with centroid representing the spectral balance, and ZCR capturing the temporal texture.

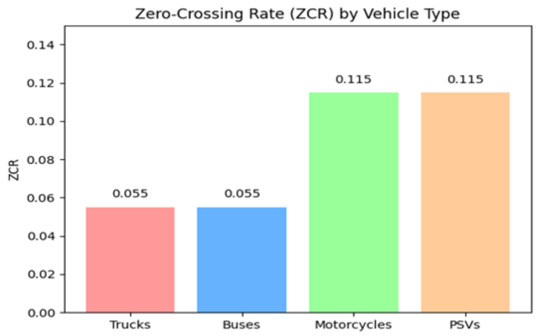

Fig. 4: The Zero-Crossing Rate (ZCR) by Vehicle Type
Source: Created by the authors

3.3.2 Perceptual Implications

The high ZCR obtained from the motorcycle recordings corresponds to the crisp, "sizzling" quality similar to tambourine jingles. The low ZCR in heavy vehicles produces the sustained, "droning" character like that of the drumhead. This difference has important implications for the subjective annoyance, since high-ZCR sounds are often perceived as more intrusive even when their overall sound levels are lower.

3.4 MFCC Analysis

3.4.1 Spectral Envelope Characterization

The analysis of the 13-coefficient MFCC vectors revealed distinct patterns: for heavy trucks, their mean MFCC values ranged from -0.8 to -1.2 (after normalization), with low standard deviation (0.4-0.6), thus indicating stable, homogeneous spectral envelopes dominated by fundamental frequency and low harmonics.

From Figure 4 the following can be discerned: Buses: similar to trucks but with slightly greater variability (SD = 0.6-0.8), reflecting more diverse operational states, PSVs and SUVs: Intermediate mean values (-0.3 to -0.6) with moderate variability (SD = 0.8-1.0), consistent with mixed spectral content, and Motorcycles: Higher mean values (-0.1 to +0.3) with substantial variance (SD = 1.3-1.6), indicating rapidly changing spectral shapes associated with acceleration events, gear changes, and modulated high-frequency content.

3.4.2 Coefficient-Specific Patterns

Lower-order MFCCs (coefficients 1-4) showed the greatest discrimination between the vehicle types, capturing broad spectral envelope differences. Higher-order coefficients (9-13) exhibited more overlap but greater variance for motorcycles, suggesting fine spectral structure variability.

3.5 Integration with Leq Measurements

3.5.1 Timbre-Level Relationships

The analysis of acoustic features relative to that of the equivalent sound level revealed systematic patterns: Sites with high Leq (>80 dBA) predominantly exhibited resonant timbre characteristics, that is, low spectral centroid (850-1100 Hz), limited roll-off (1300-1600 Hz), and low ZCR (0.05-0.07). Such sites, such as Thika Road (Leq = 84.2 dBA), were dominated by heavy vehicles producing sustained bass rumble, sites with moderate Leq (70-80 dBA), displayed mixed timbre profiles with intermediate feature values, reflecting diverse vehicle compositions, and sites with lower Leq (<70 dBA), showed tendencies toward percussive characteristics, that is, a higher spectral centroid (1500-1800 Hz), extended roll-off (2200-2800 Hz), and elevated ZCR (0.10-0.14). Such sites, like those around the vicinity of Uhuru Park (Leq = 68.7 dBA), featured lighter traffic with higher motorcycle proportions.

This inverse relationship between Leq and spectral brightness reflects the fact that whilst heavy vehicles produce lower-frequency sounds, they generate greater overall energy due to larger engines and mass.

3.5.2 Correlation Statistics

The correlational values obtained are: Spectral centroid vs. Leq: $r = -0.52$ ($p < 0.001$), Spectral roll-off vs. Leq: $r = -0.48$ ($p < 0.001$), ZCR vs. Leq: $r = -0.41$ ($p < 0.001$), and MFCC mean vs. Leq: $r = -0.38$ ($p < 0.001$). The moderate negative correlations obtained suggest that the high noise levels found in the traffic environment of Nairobi city are primarily associated with darker and more resonant timbres, while lower levels, although they are not necessarily less annoying, tend to exhibit brighter and more percussive acoustic qualities.

3.6 Passenger Car Unit (PCU) Analysis

3.6.1 Traffic Composition Effects

The PCU analysis showed how the vehicle mix shapes the overall timbre characteristics. High PCU sites (>1000 PCU/hour), which are dominated by heavy vehicles, exhibited mean spectral centroids of 900-1050 Hz and roll-off values of 1300-1500 Hz, firmly in the resonant category. Moderate-PCU sites (500-1000 PCU/hour) are mixed-composition sites that showed intermediate values with high variance and reflected temporal fluctuations as different vehicle types passed the measurement position. Low-PCU sites (<500 PCU/hour) are sites that often feature higher motorcycle proportions as compared to heavy vehicles. These sites displayed brighter timbres with centroids of 1400-1700 Hz.

3.6.2 Vehicle Type Proportions

Statistical analysis controlling for total vehicle count revealed that PCU-weighted composition predicts timbre features more effectively than simple vehicle counts. A 10% increase in heavy vehicle proportion, by PCU, corresponds to: a spectral centroid decrease of 80 Hz, a spectral roll-off decrease of 120 Hz, a ZCR decrease of 0.015, and a MFCC mean decrease of 0.2 (normalized units). These relationships allow for predictive modeling of the expected timbre characteristics based on traffic survey data, thus forming a valuable tool for transport planning and traffic impact assessment.

3.7 Temporal Patterns

3.7.1 Variation of Noise Levels

Figure 5 shows the average equivalent continuous sound level (Leq) observed at different times of the day. Peak noise levels coincided with periods of increased traffic activity, while lower levels were recorded during off-peak periods.

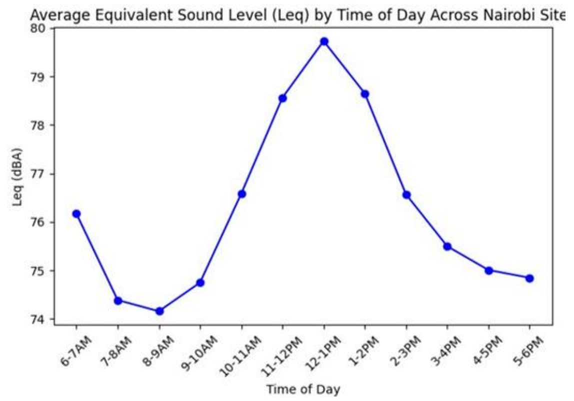

Fig. 5: Variation of the Average Equivalent Sound Level (Leq) in dBA with time
Source: Created by the authors

To examine the relationship between traffic composition and noise levels throughout the day, Figure 6 compares Leq and PCU variations over the monitoring period. Similar temporal variations suggest that traffic composition predominantly contributes to the observed noise levels.

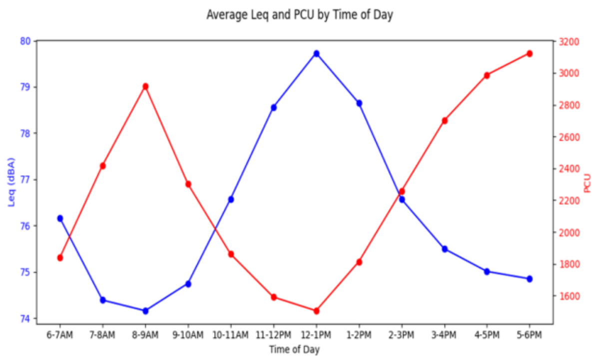

Fig. 6: The Leq and PCU Variation observed at different times of the day
Source: Created by the authors

Figure 7 reveals the relationship between the average Leq and passenger car units (PCU) per time slot. The negative trend shows that locations that experience higher traffic volumes generally experience lower noise levels.

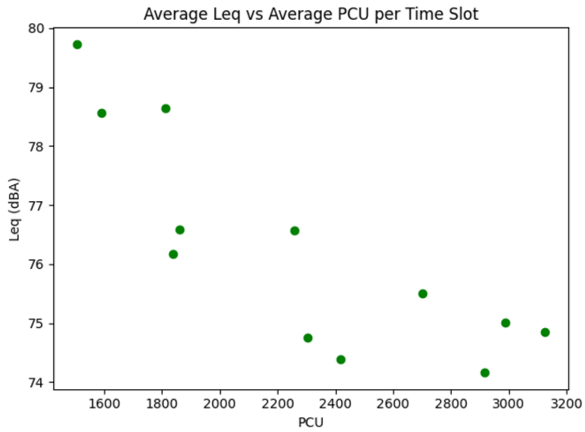

Fig. 7: The relationship between the average Leq in dBA and Passenger Car Units (PCU) per time slot
Source: Created by the authors

Figure 8 shows the relationship between the average traffic speed and the average Leq. The observed trend suggests that vehicle speed is a contributor to the traffic noise generation experienced, generated through mechanisms such as tire-road interaction and aerodynamic effects.

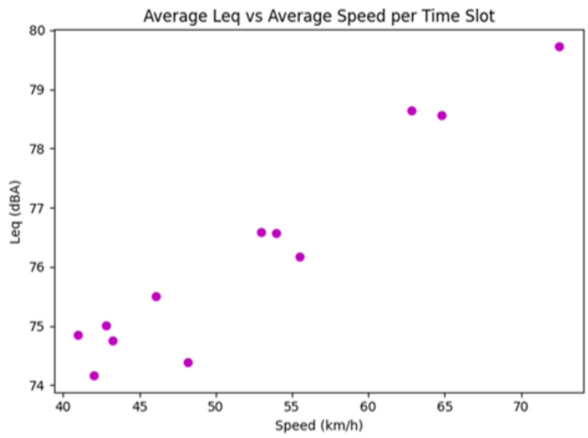

Fig. 8: The relationship between the average Leq in dBA and the average Speed in km/h
Source: Created by the authors

3.7.2 Within-site Variability

A substantial temporal variation in the timbre was found, even at single locations during the recording periods. Sites with mixed traffic exhibited a coefficient of variation (CV) for spectral centroid that ranged from 15-25%, while those with heavy traffic exhibited a lower variation (CV = 8-12%), indicating an even more stable acoustic character.

The variability influences the development of assessment protocols, in that the range of timbral qualities experienced over extended periods of time may not be adequately captured in single short-term measurements.

3.7.3 Event Detection
Characteristic feature trajectories could be seen from the individual recordings, which contained discrete vehicle pass-by events. Sharp peaks in the spectral centroid and ZCR during the closest approach were from the motorcycle pass-by events, while heavy vehicle events showed gradual increases in the overall level with stable timbre characteristics.

4 Discussion

4.1 Interpretation through the Tambourine Metaphor
From the results, it can be seen that timbre from road traffic noise in Nairobi city has a multifaceted nature. This strongly confirms the tambourine metaphor as an interpretive framework.

4.1.1 Resonant Component: The Drumhead
Sustained low-frequency rumbles, which are produced by heavy vehicles, have a spectral centroid of 800-1100 Hz. They permeate the urban environment like the fundamental tone of a tambourine membrane that has been struck.

This resonant quality stems from resonant bass frequencies produced by the large diameter pipes that are used in exhaust systems, strong fundamental frequencies and lower order harmonics produced during diesel combustion at low RPM by the large-displacement engines, sustained low frequency radiation contributed by structural vibration from vehicle mass and frame resonances, and propagation advantages, such as an omnipresent background rumble produced when low frequencies suffer less atmospheric absorption and diffract around obstacles more effectively.

This resonant component dominates especially in high-Leq places (>80 dBA), that is, along major arterials such as Thika Road, where a continuous low-frequency wash is created by the heavy vehicle density, which constitutes the foundational layer of Nairobi’s soundscape.

4.1.2 Percussive component: The Jingles
Motorcycles and light vehicles generate bright, transient sounds, that is, a spectral centroid of 1500-1900 Hz, analogous to tambourine jingles, which are sharp, attention-capturing, and spatially localized. This percussive quality derives from: the high-frequency engine harmonics since small engines operating at high RPM (>6000 RPM) produce energy-rich upper harmonics, the interaction between the tire and road surface which produce high frequency noise, aerodynamic turbulence from exposed mechanical components and minimal acoustic shielding, which allow wind noise radiation, and transient events, such as rapid acceleration, braking, and gear changes that create temporal discontinuities.

These percussive elements, which often occur at moderate Leq values (65-75 dBA), may be disproportionately annoying due to their attention-grabbing character and temporal unpredictability. Sites like those around the vicinity of Uhuru Park, with a Leq of 68.7 dBA but high motorcycle activity, are an example of scenarios where subjective disturbance may exceed predictions from obtained noise levels alone.

4.1.3 The Blended Soundscape
Most urban locations exhibit a temporal mixture of resonant and percussive components, creating a complex "polyphonic" soundscape. The metaphor works because, as a tambourine is played dynamically, road traffic noise is also made up of different sound components, and the balance continuously shifts. The rumbling sounds produced by heavy vehicles provide a persistent foundation, punctuated by the frequent passing of motorcycles and light vehicles. The understanding of this brings about a practical value whereby, in the same way that percussionists can modulate tambourine timbre by adjusting the way they strike it, pointing to the drumhead or jingles, urban planners are also able to design the soundscape character of cities through the employment of traffic management strategies that influence the vehicle mix.

4.2 Health and Annoyance Implications

4.2.1 Resonant Timbres and Physiological Stress
The relationship between resonant timbres (those with low centroid and roll-off) and elevated Leq levels reveals intensified health risks in high-traffic zones. Noise that is of low frequency is particularly problematic because it causes: (i) Physiological penetration from bass frequencies. They induce vibrations throughout the whole body, potentially triggering stress responses beyond the auditory pathways. (ii) Sleep disturbance: Low frequencies penetrate easily, thus disrupting sleep even with the windows closed. (iii) Warning sounds are masked. (iv) Cardiovascular activation: Previous research [1] suggests that prolonged exposure to low-frequency noise may contribute to cardiovascular dysfunction.

Sites such as Thika Road (Leq of 84.2 dBA) and a dominant resonant sound character represent environments where both energy-based and timbre-based health concerns converge, warranting prioritized noise mitigation.

Previous research suggests that prolonged exposure to low-frequency traffic noise may contribute to cardiovascular health effects. Therefore, sites such as Thika Road, with a high noise level (Leq = 84.2 dBA) and a dominant resonant sound character, should be prioritized for noise mitigation because they pose both high noise-level and sound-quality concerns.

4.2.2 Percussive Timbres and Cognitive Disruption

Percussive noise that is characterized by high ZCR and spectral variance may increase annoyance and disrupt cognitive functions in many ways. These ways include: startle responses, in which high-frequency transients trigger involuntary orienting responses and cause fragmented attention, which prevents habituation, maintaining vigilance and psychological arousal, interfering with task interference, and interfering with speech intelligibility. Thus, even sites with moderate Leq levels with percussive character may still require attention thus require attention. WHO Environmental Noise Guidelines emphasise event characteristics alongside energy metrics, aligning with this understanding.

4.2.3 Vulnerable Populations

Children in schools near mixed-traffic areas face particular risks from percussive timbres, as cognitive development and learning are sensitive to acoustic disruption. Studies documenting learning deficits from traffic noise should consider timbre alongside level, as classrooms exposed to high-ZCR motorcycle noise may experience greater impacts than those with equivalent Leq from steady traffic flow. Elderly populations may be affected more by low-frequency resonant components because age-related changes can increase sensitivity to vibrations, thus making them more vulnerable to health effects caused by noise.

4.3 Mathematical Modeling of Timbre-Based Annoyance

4.3.1 Conceptual Timbre Annoyance Index

The observed timbre gradient can be formalized into a predictive model for subjective responses. A framework for a timbre annoyance index A_T is proposed, as shown in Equation 14, which combines normalized acoustic features.

$$TA_T = W_1 \frac{SC}{SC_{max}} + W_2 \frac{SRO}{SRO_{max}} + W_3 ZCR + W_4 \cdot \delta_{MFFCC} \quad (14)$$

where:

- *SC*, *SRO*, and *ZCR* are as previously defined
- δ_{MFCC} represents the standard deviation across MFCC coefficients, capturing spectral variability
- W_i are empirically derived weights from perceptual studies

Based on psychoacoustic literature and our observations, we suggest initial weights: $W_1 = 0.35$ (centroid, reflecting brightness), $W_2 = 0.25$ (roll-off, bandwidth), $W_3 = 0.25$ (ZCR, temporal texture), $W_4 = 0.15$ (MFCC variance, spectral dynamics). This index predicts that both very resonant (low values, dominated by low-frequency terms) and very percussive (high values, dominated by high-frequency terms) sounds can be annoying through different mechanisms, yielding a non-monotonic relationship with overall annoyance when combined with Leq. The proposed index is conceptual and has not yet been calibrated or validated using subjective response data.

4.3.2 Integration with Conventional Metrics

A comprehensive annoyance model that may combine energy and timbre components is proposed and shown in Equation 15:

$$A_{total} = \alpha f(L_{eq} + \beta A_T + \gamma f\big(L_{eq}\big) A_T \quad (15)$$

$f\ (L_{eq})$ describes the dose-response function, for example, one obtained from the WHO guidelines, while the interaction term accounts for the synergistic effects. This formulation suggests that sound quality timbre modulates the relationship between sound level and perceived response. For example, percussive sounds may be more annoying at moderate noise levels compared to resonant sounds, which may become increasingly disturbing at high levels.

This framework should be validated through subjective surveys, which is essential. However, the framework does provide a hypothesis-driven starting point for field studies within Nairobi and other similar cities.

4.4 Urban Planning and Noise Mitigation Strategies

4.4.1 Vehicle-specific Regulations

The clear difference in timbre characteristics among vehicle types enables targeted noise management approaches. These include, (i) Heavy vehicle management, that is, route optimization through the restriction of truck access to sensitive areas, especially during vulnerable periods such as night and school hours and use of low-frequency absorbers: in road surface treatments and noise barriers especially in residential zones adjacent to truck routes, (ii) Enforce stricter regulations on low-frequency emissions, (iii) Regulate the use of motorcycles through high-frequency mufflers, enforcing speed limits, (iv) Encouraging the use of electric vehicle to eliminate combustion noise.

4.4.2 Traffic Composition Optimization
The PCU analysis revealed that small shifts in the vehicle mix, to a large extent, alter the soundscape. Urban planning approaches could consider time-based access by separating heavy and light vehicle flows temporally, dedicated vehicle lanes to reduce the acoustic "clutter" of mixed timbres, and public transport improvements, such as incentivizing matatus and buses over private vehicles.

4.4.3 Acoustic Barrier Design
Traditionally, acoustic barriers are designed to cater for sound attenuation. A timber-aware design could be employed to improve their performance by using resonant absorbers, which target the 800-1100 Hz range in noise; diffusive surfaces, which scatter high-frequency sounds; and hybrid systems that address both resonant and percussive components.

4.4.4 Soundscape-informed Urban Spatial Planning
Beyond noise abatement, cities can proactively improve their acoustic environments by establishing quiet zones: designated areas with vehicle restrictions, creating green spaces strategically located to mask undesirable timbre components, and arranging land uses that are compatible with the surrounding environment's timbre compatibility. The tambourine metaphor provides a simple way to facilitate stakeholder communication in city planning processes.

4.5 Methodological Considerations

4.5.1 Feature Selection and Interpretation
Selected acoustic features, that is, the spectral centroid, roll-off, ZCR, and MFCCs, were found to be effective, but alternative methods may also be considered. Alternative methods include: (i) Roughness, which measures the rapid changes that occur in sound, thus quantifying amplitude modulation in the 15- 300 Hz range, (ii) Sharpness, which directly addresses high-frequency annoyance, (iii) Tonality, which distinguishes tonal sound from the broadband components, and (iv) Fluctuation strength, which capture the slow variations in sound over a period of time.

4.5.2 Temporal Dynamics
Six-second segments provide snapshots but may miss pass-by dynamics, traffic flow phases, and cumulative exposure.

4.5.3 Propagation and Environmental Effects
Propagation modifies timbre through distance, atmospheric absorption, ground effects, and urban canyon effects.

4.6 Limitations

4.6.1 Sampling limitations
Peak-hour focus, spatial coverage, and seasonal effects limit generalizability.

4.6.2 Validation Gap
The present study focused on descriptive analysis and correlation assessment. More advanced statistical and predictive modelling approaches were beyond the scope of the study and are recommended for future research.

4.6.3 Subjective Annoyance
No listening experiments, psychoacoustic surveys, or subjective annoyance assessments were conducted.

5 Conclusion

5.1 Summary of Findings
The findings of this study demonstrate the value of the tambourine metaphor as an interpretive framework for understanding timbre in road traffic noise in Nairobi city, Kenya. The analysis of the 2,185 audio segments across 42 sampling locations revealed that:
(i) There is a clear timbre differentiation. Different vehicle categories produce noise that varies across a spectrum from resonant sounds (produced by heavy trucks and buses) to percussive sounds (produced by motorcycles) at the opposite end, (ii) Heavy vehicles exhibit low spectral centroids (800- 1100 Hz), limited roll-off (1200-1600 Hz), and low ZCR (0.04-0.07), while motorcycles produce higher

values across all features, (iii). High Leq environments (>80 dBA) are mostly dominated by resonant timbres. In contrast, sites with moderate Leq are often dominated by percussive character, and (iv) the composition of traffic, expressed as a PCU-weighted vehicle mix, predicts timbre characteristics of road traffic noise, thus making it possible to forecast from traffic survey data.

The contribution of this study lies mainly in the application of timbre-based acoustic descriptors and a perceptually grounded interpretive framework to characterize road traffic noise in a developing-city context. Future work may extend this framework through predictive modelling, uncertainty analysis, and subjective validation studies.

Acknowledgments:
The authors are grateful to the university administration for providing a conducive work environment.

References:

[1] Münzel, T., Daiber, A., Steven, S., Tran, L. P., Ullmann, E., Kossmann, S., Schmidt, F. P., Oelze, M., Xia, N., Li, H., Pinto, A., Wild, P., Pies, K., Schmidt, E. R., Rapp, S., & Kröller-Schön, S. (2017). Effects of noise on vascular function, oxidative stress, and inflammation: Mechanistic insight from studies in mice. European Heart Journal, 38(37),2838–2849. https://doi.org/10.1093/eurheartj/ehx081.

[2] Basner, M., Brink, M., Davis, A., van Kamp, I., Kujawa, S. G., Münzel, T., & Smith, M. G. (2026). Auditory and non-auditory effects of noise on health: Updated evidence and future directions. Environmental Research, 286, Article 124648. https://doi.org/10.1016/j.envres.2026.124648.

[3] Jeon, J. Y., Lee, P. J., Hong, J. Y., & Cabrera, D. (2011). Non-auditory factors affecting urban soundscape evaluation. The Journal of the Acoustical Society of America, 130(6), 3761–3770. https://doi.org/10.1121/1.3652902.

[4] Khomenko, S., Cirach, M., Barrera-Gómez, J., Pereira Barboza, E., Iungman, T., Mueller, N., Foraster, M., Tonne, C., Thondoo, M., Jephcote, C., Gulliver, J., Woodcock, J., & Nieuwenhuijsen, M. (2023). Road traffic noise in European cities: Learnings from a European-wide study on the health impacts of road traffic noise. Proceedings of INTER-NOISE 2022, Glasgow, Scotland, United Kingdom. INTER-NOISE and NOISE-CON Congress and Conference Proceedings, 265(7), 678–685. https://doi.org/10.3397/IN_2022_0093.

[5] Xu, X., & Wu, H. (2021). Audio-visual interactions enhance soundscape perception in China's national parks. Urban Forestry & Urban Greening, 61, 127090. https://doi.org/10.1016/j.ufug.2021.127090.

[6] Zhang, H., Lin, J., & Chen, S. (2024). Timbre perception, representation, and its neuroscientific exploration: A comprehensive review. arXiv:2405.13661. https://doi.org/10.48550/arXiv.2405.13661.

[7] Jo, H. I., & Jeon, J. Y. (2021). Urban soundscape categorization based on individual recognition, perception, and assessment of sound environments. Landscape and Urban Planning, 216, 104241. https://doi.org/10.1016/j.landurbplan.2021.104241.

[8] Ozturk, Z. S., Kang, J., & Aletta, F. (2025). Soundscape Research in Streets: A Scoping Review. Sustainability, 17(8), 3329. https://doi.org/10.3390/su17083329.

[9] Guski, R., Schreckenberg, D., & Schuemer, R. (2017). WHO Environmental Noise Guidelines for the European Region: A Systematic Review on Environmental Noise and Annoyance. International Journal of Environmental Research and Public Health, 14(12),1539. https://doi.org/10.3390/ijerph14121539.

[10] Göksu, H. (2018). Engine speed–independent acoustic signature for vehicles. Measurement and Control, 51(10), 598–606. https://doi.org/10.1177/0020294018769080.

[11] Mayvan, D., Beheshti, A., & Masoom, 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), 53–64. https://doi.org/10.5121/ijsc.2015.6105.

[12] Abdul, Z., & Al-Talabani, A. (2022). Mel frequency cepstral coefficient and its applications: A review. IEEE Access, 10, 124308–124324. https://doi.org/10.1109/ACCESS.2022.3223444.

[13] Yan, H., Bai, H., Zhan, X., Wu, Z., Wen, L., & Jia, X. (2022). Combination of VMD Mapping MFCC and LSTM: A New Acoustic Fault Diagnosis Method of Diesel Engine. Sensors, 22(21), 8325. https://doi.org/10.3390/s22218325.

[14] Ma, C., Dai, Z., & Zhang, W. (2024). Recognition of car horns based on principal component analysis of Mel-frequency cepstral coefficients and support vector machine. Computers and Electrical Engineering, 120, 109666. https://doi.org/10.1016/j.compeleceng.2024.109666.

[15] Dhanalakshmi, P., Palanivel, S., & Ramalingam, V. (2009). Classification of audio signals using SVM and RBFNN. Expert Systems with Applications, 36(5), 6069–6075. https://doi.org/10.1016/j.eswa.2008.06.126.

[16] Cao, J., & Kang, J. (2023). A Perceptual Structure of Soundscapes in Urban Public Spaces Using Semantic Coding Based on the Grounded Theory. International Journal of Environmental Research and Public Health, 20(4), 2932. https://doi.org/10.3390/ijerph20042932.

[17] Kang, J., Aletta, F., Gjestland, T., Brown, L. M., Botteldooren, D., Schulte-Fortkamp, B., Lercher, P., van Kamp, I., Genuit, K., Fiebig, A., Coelho, J., Maffei, L., & Lavia, L. (2016). Ten questions on the soundscapes of the built environment. Building and Environment, 108, 284–294. https://doi.org/10.1016/j.buildenv.2016.08.011.

[18] Xiang, Y., Meng, Q., Li, M., Yang, D., & Wu, Y. (2024). Using soundscape diversity to interpret soundscape evaluation in urban green spaces—Relationship, integration, and application. Urban Forestry & Urban Greening, 97, 128371. https://doi.org/10.1016/j.ufug.2024.128371.

[19] Filipan, K., Boes, M., De Coensel, B., Lavandier, C., Delaitre, P., Domitrović, H., & Botteldooren, D. (2017). The Personal Viewpoint on the Meaning of Tranquility Affects the Appraisal of the Urban Park Soundscape. Applied Sciences, 7(1), 91. https://doi.org/10.3390/app7010091.

[20] Rudge, L., Holland, A., Schmit, J., & Shriver, W. (2025). In urban national parks acoustic indices represent avian habitat quality better than avian diversity. SSRN Electronic Journal. SSRN 5714799. https://doi.org/10.2139/ssrn.5714799.

[21] Rosi, V. (2022). The metaphors of sound: From semantics to acoustics—A study of brightness, warmth, roundness, and roughness (PhD thesis, Sorbonne Université). Cognitive Science. https://doi.org/10.13140/RG.2.2.23054.43843.

[22] Baquero Larriva, M. T., & Higueras García, E. (2023). Differences in perceptions of the urban acoustic environment in older adults: A systematic review. Population Ageing, 16, 781–813. https://doi.org/10.1007/s12062-021-09325-7.

[23] Ding, B., Zhang, T., Wang, C., Liu, G., Liang, J., Hu, R., Wu, Y., & Guo, D. (2023). Acoustic scene classification: A comprehensive survey. Expert Systems with Applications, 238, 121902. https://doi.org/10.1016/j.eswa.2023.121902.

[24] Tzanetakis, G., & Cook, P. (2002). Musical genre classification of audio signals. IEEE Transactions on Speech and Audio Processing, 10(5), 293–302. https://doi.org/10.1109/TSA.2002.800560.

[25] Gerhard, D. (2003). Audio signal classification: History and current techniques (Technical Report TR-CS 2003-07). Department of Computer Science, University of Regina. ISSN 0828-3494. ISBN: 0-7731-0456-9, [Online]. https://www2.cs.uregina.ca/~gerhard/publications/TRdbg-Audio.pdf (Accessed Date: January 24, 2026).

[26] Sharma, G., Umapathy, K., & Krishnan, S. S. (2020). Trends in audio signal feature extraction methods. Applied Acoustics, 158, 107020. https://doi.org/10.1016/j.apacoust.2019.107020.

[27] Chion, M. (2012). The three listening modes. In J. Sterne (Ed.), The Sound Studies Reader (pp. 48–53). Routledge, New York. ISBN: 9780415771313.

[28] Kang, J. (2023). Soundscape in city and built environment: Current developments and design potentials. City and Built Environment, 1. https://doi.org/10.1007/s44213-022-00005-6.

[29] Zhang, S., Zhao, X., Zeng, Z., & Qiu, X. (2019). The influence of audio-visual interactions on psychological responses of young people in urban green areas: A case study in two parks in China. International Journal of Environmental Research and Public Health, 16(18), 3465. https://doi.org/10.3390/ijerph16183465.

[30] Aletta, F., Mitchell, A., Oberman, T., Kang, J., Khelil, S., Bouzir, T. A. K., Berkouk, D., Xie, H., Zhang, Y., Zhang, R., Yang, X., Li, M., Jambrošić, K., Zaninović, T., van den Bosch, K., Lühr, T., Orlik, N., Fitzpatrick, D., Sarampalis, A., Aumond, P., Lavandier, C.,

Moshona, C. C., Lepa, S., Fiebig, A., Papadakis, N. M., Stavroulakis, G. E., Sudarsono, A. S., Sarwono, S. J., Puglisi, G. E., Jafari, F., Astolfi, A., Shtrepi, L., Nagahata, K., Jo, H. I., Jeon, J. Y., Lam, B., Chieng, J., Ooi, K., Hong, J. Y., Antunes, S. M., Alves, S., Carvalho, M. L. U., Michalski, R. L. X. N., Kogan, P., Vida Manzano, J., García Quesada, R., Suárez Silva, E., Almagro Pastor, J. A., Nilsson, M. E., Axelsson, Ö., Gan, W.-S., Watcharasupat, K. N., Jaratjarungkiat, S., Ong, Z.-T., Dökmeci Yörükoğlu, P. N., Erçakmak Osma, U. B., & Nguyen, T. L. (2024). Soundscape descriptors in eighteen languages: Translation and validation through listening experiments. Applied Acoustics, 224, 110109. https://doi.org/10.1016/j.apacoust.2024.110109.

[31] İşler, B. (2025). Urban sound recognition in smart cities using an IoT–fog computing framework and deep learning models: A performance comparison. Applied Sciences, 15(3), 1201. https://doi.org/10.3390/app15031201.

[32] Bonet-Solà, D., Vidaña-Vila, E., & Alsina-Pagès, R. (2023). Predicting the perceptual rating of a soundscape using artificial intelligence. Frontiers in Artificial Intelligence and Applications, 375, 287–288. https://doi.org/10.3233/FAIA230696.

[33] Martins, F., Segurado, P., & Marques, J. (2025). Acoustic detection and occupancy models: A systematic review with insights for future monitoring programs. Ecological Indicators, 179, 114081. https://doi.org/10.1016/j.ecolind.2025.114081.

[34] Qiu, S., Guo, M., & Ma, M. (2025). A sound event localization and detection research based on inception depth-wise convolution and adaptive efficient channel attention block. Circuits, Systems, and Signal Processing, 45, 1–23. https://doi.org/10.1007/s00034-025-03325-0.

[35] Yang, W., Cai, M., & Luo, P. (2020). The calculation of road traffic noise spectrum based on the noise spectral characteristics of single vehicles. Applied Acoustics, 160, 107128. https://doi.org/10.1016/j.apacoust.2019.107128.

[36] Cai, Y., Chen, Z., Akram, A. N. A., & Yuan, M. (2024). Urban road traffic noise spectra and their effects on spectral correction in sound insulation evaluation. Journal of Vibration and Shock, 43(19), 287–297, [Online]. https://jvs.sjtu.edu.cn/EN/Y2024/V43/I19/287 (Accessed Date: January 24, 2026).

[37] Xu, X., Cai, J., Yu, N., Yang, Y., & Li, X. (2020). Effect of loudness and spectral centroid on the music masking of low frequency noise from road traffic. Applied Acoustics, 166, 107343. https://doi.org/10.1016/j.apacoust.2020.107343.

[38] Ye, J., Kobayashi, T., & Murakawa, M. (2017). Urban sound event classification based on local and global features aggregation. Applied Acoustics, 117(Part B), 246–256. https://doi.org/10.1016/j.apacoust.2016.08.002.

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.

Data Availability:

Audio recordings, extracted features, and analysis code are available at https://github.com/ElishaAkech/SOUNDAI

Conflict of Interest

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

Artificial Intelligence Generated Content:

No part of this work came from machine-assisted thinking.