Beyond noise levels : vehicle classification using psychoacoustic indicators from pass-by road traffic noise and their correlations with speed and temperature

Beyond noise levels: vehicle classification using psychoacoustic indicators from pass-by road traffic noise and their correlations with speed and temperature

Ablenya Barros [1*], Michiel Geluykens [2], Frederico Pereira [3], Elisabete Freitas [4], Susana Faria [5], Luc Goubert [6], Cedric Vuye [7]

[1] Sustainable Pavements and Asphalt Research, University of Antwerp, Groenenborgerlaan 171, 2020 Antwerp, Belgium. ablenya.barros@uantwerpen.be *Corresponding author

[2] University of Antwerp, Groenenborgerlaan 171, 2020 Antwerp, Belgium, michiel.geluykens@weit.co.at

[3] University of Minho, Center of Computer Graphics, Campus de Azurém, 4800-058 Guimarães, Portugal. frederico.pereira@ccg.pt

[4] University of Minho, ISISE, Campus de Azurém, 4800-058 Guimarães, Portugal. efreitas@civil.uminho.pt

[5] University of Minho, Centre of Mathematics, Campus de Azurém, 4800-058 Guimarães, Portugal. sfaria@math.uminho.pt

[6] Belgian Road Research Center, Woluwedal 42, 1200 Brussels, Belgium. l.goubert@brrc.be

[7] Sustainable Pavements and Asphalt Research, University of Antwerp, Groenenborgerlaan 171, 2020 Antwerp, Belgium. cedric.vuye@uantwerpen.be

Abstract

Environmental noise control regulations typically employ noise level descriptors to set limits for noise exposure. However, other noise characteristics, such as frequency content, temporal patterns and masking, have been proven to influence the perception of acoustic environments. Psychoacoustic indicators offer an objective means of establishing relationships between physical characteristics of noise and the human auditory sensation phenomena. This study explored psychoacoustic indicators of pass-by vehicle noise across different vehicle categories, driving speeds, and temperatures. Moreover, the indicators were exploited as features to train a classification algorithm to predict vehicle category. Over 2000 vehicle noise samples were collected using the Statistical Pass-By (SPB) method, categorized into three classes according to ISO 11819-1, besides an additional class for delivery vans. Correction coefficients were obtained for temperature and speed to noise levels, loudness, roughness, sharpness and fluctuation strength. Then, the differences in these indicators based on vehicle category were then discussed. A vehicle-category predictive model using the three vehicle categories defined in ISO 11819-1 yielded 84% accuracy. Including vans as an extra vehicle category dropped accuracy to 72% due to their misclassification with passenger cars. Combining these two categories increased overall accuracy to 86%. These findings could enable a less visual-dependent vehicle categorization so that vehicle fleets worldwide are more consistently classified in terms of noise. Additionally, psychoacoustic indicators appear to be valuable features for vehicle classification systems aimed to resemble the human auditory experience.

Keywords: Machine Learning, Logistic Regression, Audio Signal Classification, Statistical Pass-By

1. Introduction

Physical and psychological health outcomes have been correlated with long-term exposure to road traffic noise (Stansfeld et al., 2021). Awareness of this hazard resulted in the development of guidelines for assessing and managing environmental noise, including the Environmental Noise Guideline for the European Region (WHO, 2018) and the Environmental Noise Directive (END) (European Union, 2002). The END requires European Union member states to perform strategic noise mapping for major roads, railways, airports, and agglomerations every five years. Large-scale urban traffic noise maps represent noise exposure via noise level descriptors as the energy-based index L_{den} (the day–evening–night level), and are drawn based on characteristics of the traffic flow microstructure, such as vehicle type, driving speed and road surface type (Morel et al., 2016). In the end, strategies concerning noise abatement rarely account for acoustical factors other than exposure levels, even though it is recognized that noise level reductions alone do not necessarily result in more positive impressions of an acoustic environment (Fiebig et al., 2020).

Psychoacoustic indicators are metrics that establish functional relationships between complex spectral content, temporal envelopes and periodicities of acoustic signals with the human auditory sensation phenomena. The application of psychoacoustic indicators has received increasing attention owing to the recent paradigm shift from focusing on physical noise exposure to a more holistic approach based on the human perception of urban sound environments (Engel et al., 2021).

Psychoacoustics has been used to assess road traffic noise. As stated by Moreno et al. (2023), noise intensity metrics such as noise levels are insufficient to provide all relevant information about a vehicle passage and, thus, are not suitable for evaluating the total emission energy. Methods used to collect audio samples to calculate psychoacoustic indicators include the Close-ProXimity (CPX) method (Soares et al., 2017; Guo et al., 2018; Freitas et al., 2018), which mainly captures noise from the near field tyre/road interaction. Although this procedure enables assessing rolling noise from distinct road surfaces, it does not account for the actual traffic flow in a receiver-oriented approach (Ascari et al., 2022), where other noise generation mechanisms play a role in total vehicle noise. Other researchers have conducted long roadside noise measurements and retrieved several pass-by vehicles without distinguishing among them (Lo Castro et al., 2018); studies such as Raggam et al. (2007) and Gille et al. (2016), although captured individual pass-by vehicles, aimed at synthetically creating and exploring a vehicle ensemble.

Few publications are available on the acoustic and psychoacoustic descriptors of noise produced by isolated pass-by vehicles. The approach typically employed in these studies is to link the indicators to noise-induced annoyance, and only a limited number of (controlled) instances are presented. Morel et al. (2016) identified the perception of spectral and temporal features of pass-by vehicle noise in reported annoyance levels; Paviotti and Vogiatzis (2012) linked annoyance levels to the roughness of powered two-wheelers; Altinsoy (2021) showed that loudness, tonality, roughness and fluctuation strength are related to the perceived annoyance induced by pass-by noise of passenger cars.

Recording individual vehicle noise from traffic flow can be labour-intensive. George et al. (2013) describe challenges encountered in data collection from a two-lane undivided road, such as the simultaneous passage of different vehicles, horn sounds and random but identifiable background noises. The Statistical Pass-By (SPB) is a standardized methodology (ISO 11819-1:2023) to measure and compare the acoustic quality of road surfaces by leveraging the vehicle fleet. Strict requirements of free-field conditions, low background noise, and pass-by vehicles driving at a constant speed and sufficiently spaced from other vehicles guarantee that recordings taken from these measurements comprise neat audio samples of individual vehicles. Another aspect to account for in roadside measurements is that environmental and driving conditions affect the noise levels and spectra of pass-by vehicles (Yang et al., 2020). Standardized procedures have been established to normalize noise levels to a common scale, with ISO/TS 13471-2 (2022) addressing temperature normalization and ISO 11819-1 (2023) incorporating a speed-correction method. Thus, these effects must also be explored so that psychoacoustic indicators of pass-by noise in uncontrolled driving conditions, such as those encountered in typical traffic flow scenarios, can be grouped and compared.

According to the SPB method, pass-bys are classified into three categories: passenger cars, dual and multiple-axle heavy vehicles. The classification is performed visually by the measurement operator based on aspects such as number of seats, vehicle size (related to gross vehicle mass), and number of axles. Although this classification method seems easy to implement, an issue mentioned in Annex A of ISO 11819-1:2023 is that the vehicle fleet worldwide is very diverse, limiting the applicability of a standardized visual-based vehicle classification system.

Audio signal classification involves extracting suitable features from sound/noise signals to determine classes to which the signals most closely fit. The choice of feature extraction and classification/clustering algorithms depends on the application's domain. Research on auditory classification has resulted in extensive libraries of computed features that can be divided into physical and perceptual characteristics. Physical characteristics are captured directly from the magnitudes of the audio waveform or short-time spectral values, while perceptual characteristics are estimated by auditory models and thus are linked to the human perception of noise. Psychoacoustic indicators belong to the realm of time and frequency-domain perceptual characteristics of noise (Chaki, 2020).

In audio classification problems, a set of relevant features can be called the "acoustic signature" (Kandpal et al., 2013). Various studies have used supervised learning techniques to detect and classify vehicle noise using their acoustic signatures. Examples include Dawton et al. (2020), who extracted features from the frequency-domain representation of vehicle noise using successive short-time Fourier transforms and trained a support vector machine (SVM) to classify pass-bys into cars, scooters/motorbikes or buses. Kandpal et al. (2013) used features only in the time domain to train an artificial neural network (ANN) to distinguish between cars, trucks or motorbikes. George et al. (2013) used smoothed log energy for detecting noise peaks from pass-bys and extracted mel frequency cepstrum coefficients from fixed regions around the

detected peaks to train an ANN; the authors performed a classification task into heavy, medium and light vehicles. All these studies yielded accuracies above 65%.

There is, however, a limited amount of studies available on psychoacoustic-based audio signal classifications. One example of its application in environmental acoustics is a clustering algorithm for soundscapes based on psychoacoustic indicators, among other acoustic features (Rychtáriková and Vermeir, 2013). Still, employing psychoacoustics particularly in vehicle classification is an unexplored attempt. The appeal of this approach is evidenced by Morel et al. (2012), who observed a congruence between the psychoacoustic indicators and a perceptual vehicle typology created in a free categorization task of pass-by vehicle noise via listening tests with subjects.

The present study aims to investigate psychoacoustic indicators from pass-by vehicle noise from a large sample. The initial research question is twofold: to what extent are the indicators impacted by environmental and driving conditions, meaning air temperature and speed? Secondly, are the indicators sensitive enough to capture differences in pass-by vehicle noise among vehicle categories? The answers to these questions led to a third query: do psychoacoustic indicators enable enough information to derive patterns and create a relationship that allows predicting vehicle category at a reasonable accuracy? The motivation behind performing a classification task is to overcome the limitation presented in ISO 11819-1:2023 regarding a visual-based classification system for vehicle fleets worldwide. From a broader perspective, these results enable evaluating whether a vehicle classification based on psychoacoustic indicators as objective indicators of the human auditory sensation is feasible.

2. Materials and methods

2.1 Research methodology overview

An extensive SPB measurement campaign was performed to build a dataset of pass-by noise of different vehicle categories at a wide air temperature and speed range. The acoustic similarity between the test locations was evaluated via CPX and texture measurements, which allowed data polling from the SPB measurements, resulting in a dataset of 2199 observations. The SPB results were analysed via the average maximum A-weighted sound level ($L_{A,max}$) spectra. Next, the psychoacoustic indicators (loudness, sharpness, roughness, and fluctuation strength) were calculated for each pass-by. To investigate the effects of air temperature and speed on the indicators, linear regression models were employed and their statistical significance was evaluated, allowing correction coefficients to be determined. The indicators could then be normalized to a reference temperature (20 °C) and speed (50km/h) and the pass-bys of each vehicle category were combined so that differences between psychoacoustic indicators among categories could be explored. Given the large sample size and the observed sensitivity to vehicle category, the psychoacoustic indicators were used as features to create a vehicle classifier.

From the perspective of developing a classification system, this research methodology consists of four stages: sensing, segmentation, feature extraction, and classification (Kandpal et al. 2013). For audio data, sensing handles the practical issues related to physical sensors, such as

microphone setup and positioning. The segmentation stage deals with retrieving relevant information from the input data stream provided by the sensors, meaning, in this case, obtaining separate audio input from individual pass-by vehicles. Here, these two steps are covered by following the strict requirements of ISO 11819-1 as a data collection method: after each measurement, an already classified and timestamped dataset was obtained (Figure 1). Feature extraction aims at finding characteristics that aid in distinguishing vehicles among different classes. The psychoacoustic indicators were employed to build the feature vector as they proved to differ among vehicle categories. Lastly, the classification stage comprises training a classifier and performing the prediction task. For that, multinomial logistic regression was used – one of the classic supervised machine learning algorithms capable of doing multi-class classification.

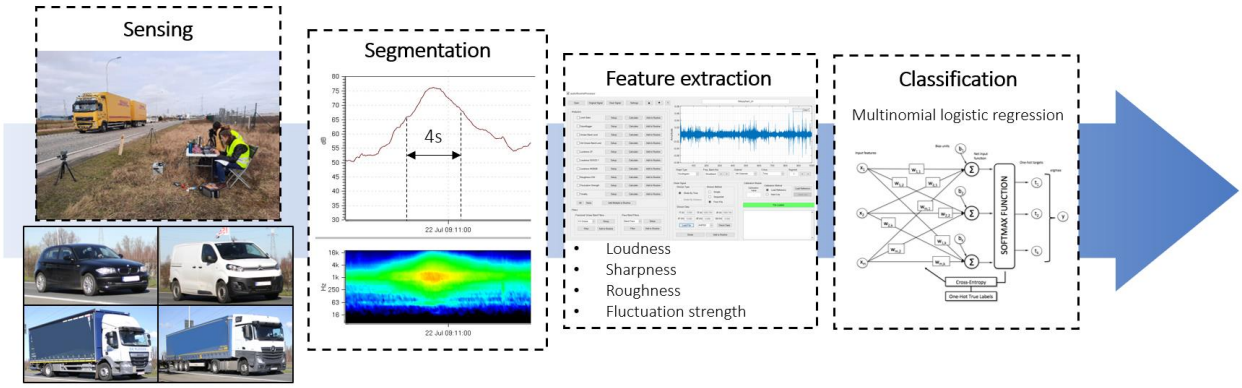

Figure 1: Development stages of a vehicle classification system based on psychoacoustic indicators of vehicle noise from SPB measurements

2.2 SPB, CPX and texture measurements

2.2.1 SPB method and measurement locations

SPB measurements were performed according to ISO 11819-1 (2023), a recently published revision of the 1997's last version. One microphone was placed at 1.2 m height and 7.5 m from the center of the measured lane. A sonometer model NTi XL2 was used to continuously record the environmental noise levels as audio files and capture the $L_{A,max}$ of each pass-by in broadband and one-third octave bands. The speed radar KR-10 SP was used for speed measurements.

Each pass-by vehicle is assigned by the operators to pre-defined categories having common features that are easy to visually identify in the traffic stream, such as the number of axles and body size. These characteristics are presumed to result in similarities in sound emission. The categories defined in the standard are passenger cars (P), dual-axle heavy vehicles (HD) and multiple-axle heavy vehicles (HM). Vehicles that do not correspond to these categories' descriptions are not taken into account.

Among the changes proposed in the new version of the ISO SPB standard, HD and HM vehicles, although flagged separately on site, shall be combined into a single group (heavy

vehicles) for analysis purposes. To do so, 2.7 dB is added to the $L_{A,max}$ of HD vehicles. As the two resulting categories (light and heavy vehicles) present very distinguishing features, they are assumed to be sufficient to describe the acoustic performance of pavements, which is the primary goal of the SPB method.

"Medium" vehicles, such as delivery vans, do not fit either category. The presence of vans in urban environments is significant due to their broad use for activities such as postal services and the construction industry. Vans are visually distinguishable from passenger cars and HD in their dimensions, cargo space, number of seats and tyre size. Given the number of vans spotted by the operators over this SPB measurement campaign, they were recorded as an extra vehicle category.

The SPB measurements were performed in three locations in the Antwerp region (Belgium) throughout a two-month measurement campaign. Besides compliance with ISO 11819-1 requirements for an acoustic free field around the microphone, these locations were chosen for having pavement surfaces in hot asphalt mixtures in moderate to good condition and a traffic composition with a large share of heavy vehicles. Table 1 presents a description of the locations. The mean profile depth (MPD) was retrieved from texture measurements as described in Section 2.2.2.

Table 1: Test locations description

	Location 1	**Location 2**	**Location 3**
Street name, municipality	Keetberglaan, Zwijndrecht	Krijgsbaan, Zwijndrecht	Stuivenbergvaart, Mechelen
Asphalt type	Stone Mastic Asphalt	Stone Mastic Asphalt	Chipped dense asphalt concrete type 2C*
Max. aggregate size (mm)	10	10	10
Mean profile depth (mm)	1.2	1.0	1.6
Surface between lane and microphone	Asphalt parking lane	Cobblestone parking lane and concrete bike lane	Concrete block pave bike lane and sidewalk
N◦ of measurement days	4	2	3
N◦ of vehicles measured	1403	406	390

*a bituminous surface formerly common in highways in Belgium. It comprises a layer of dense asphalt concrete with 10 mm aggregates scattered over the hot surface and pressed again with a steel roller (BRRC, 2020).

After outlier removal, 2199 vehicles were recorded: 823 passenger cars, 188 vans, 85 HD and 1103 HM. This dataset is an extract from a larger measurement campaign presented at Geluykens et al. (2022); see this reference for a more extensive description of the test locations and their selection procedure.

2.2.2 CPX and texture

Close-ProXimity (CPX) measurements were carried out according to ISO 11819-2 (2017) by the Belgian Road Research Centre (BRRC) at the three sites in sections of 200 m, which mid-

point corresponded to the microphone position of the SPB measurements. The tyre/pavement noise was measured with the CPX trailer. CPX differs from the SPB method as it only considers tyre/road noise and not engine noise or propagation effects in the path to the roadside. The CPX measurements were conducted at a reference speed of 50 km/h, using the Standard Reference Test Tire (SRTT, P1), a representative of passenger cars tyres, and the CPX level ($L_{CPX:P}$) was reported for 20-m segments. The air temperature was 20 ± 2 ^{o}C, conveniently close to the reference temperature prescribed in ISO 13471-1 (2017) not to require temperature correction to the noise levels.

Texture measurements were also conducted by the BRRC in the same 200-m sections, with two repetitions per location, according to ISO 13473 series. The laser beam mounted on a vehicle had a 0.2 mm diameter, high sampling frequency (78 kHz), vertical measuring range of 64 mm and a vertical resolution of 1 µm. The profile was measured with equidistant steps of 1 mm.

2.3 Road traffic pass-by noise recordings

The noise levels were continuously recorded throughout the SPB measurements as .wav audio files. A pass-by duration was defined as 4 s: from 2 s before to 2 s after a pass-by produced the peak noise level, which occurs approximately when the vehicle midpoint passes closest to the microphone. Independently of speed, this time window resulted in a good trade-off between signals long enough to capture information on the vehicle passage and short enough to exclude unwanted noises such as those from other vehicles. Still, an analysis of the impact of speed on the psychoacoustic indicators is presented in Section 3.3.

Among previous studies that also used pass-by methods to measure road traffic noise and calculate psychoacoustic indicators from short fragments, Morel et al. (2012) selected a pass-by duration ranging between 3 s to 9 s. Kandpal et al. (2013) have taken 5-s signals around the passing moment to use in an ANN classifier for ground vehicles, although the authors extracted acoustic features other than psychoacoustic indicators. In a similar context, Dawton et al. (2020) chose a 2-s signal as input for an SVM vehicle classifier.

The audio files recorded in this measurement campaign, timestamps, air temperature and vehicle category for each relevant vehicle passage can be retrieved in an open-source online data repository (Grangeiro de Barros and Vuye, 2023).

2.4 Acoustic and psychoacoustic indicators

From the 4-s audio excerpts, psychoacoustic indicators were calculated in a MATLAB-based environment employing algorithms from PsySound3 (Cabrera et al., 2014). Loudness (N) calculation uses Fastl & Zwicker's model for time-varying signals and assumes free field frontal incidence, as detailed in ISO 532-1 method B (2017). Sharpness (S) calculation is also described in Fastl and Zwicker (2007). Roughness (R) is given by Daniel and Weber (1997) and fluctuation strength (FS) is based on the Zhou et al. (2015) model. Besides these parameters, we also determined the difference between the peak $L_{A,max}$ and the $L_{A,max}$ at 2 s before the peak was

reached, named ΔL. The final dataset is also available in the repository (Grangeiro de Barros and Vuye, 2023).

The calculation outputs are given in percentile levels, which describe the statistical evolution of each parameter within the 4-s time window. A percentile *n* means that this value was exceeded in *n*% of the interval. Normally, the 5% percentile represents exceptional events in the signal, 95% characterizes a quasi-continuous situation, and the 50% percentile resembles a probable situation (Rychtáriková and Vermeir, 2013). Considering these noise excerpts are short and consistently variable, the 50^{th} percentile may be the best representative.

The following indicators are discussed: $L_{A,max}$ and N_{50} are explored as sound intensity measures. Loudness describes perceived noise intensity better than sound pressure level, as it accounts for spectral and temporal masking effects caused by the frequency selectivity characteristic of the human hearing system (Fastl and Zwicker, 2007). R_{50} and FS_{50} are both measures of noise amplitude modulation. Roughness relates to the perception of rapid periodic temporal changes (15-300 Hz), while the sensation measured by fluctuation strength is produced by temporal variations slowly enough (< 20 Hz) to be perceived by the human ear (Camacho et al., 2008). Information about the spectral envelope given through the spectra's centre of gravity is retrieved by S_{50}. Sharpness of narrow-band noises increases sharply at high center frequencies; for this reason, it is considered a noise high-frequency content descriptor (Fastl and Zwicker, 2007).

2.5 Impacts of speed and temperature on psychoacoustic indicators

Air temperature and speed affect vehicle noise levels and spectra. Consequently, it is expected that these driving and environmental conditions also impact psychoacoustic indicators derived from pass-by noise. Considering that, it is necessary to normalize these indicators to a common scale, enabling the grouping of vehicles by category and facilitating further comparative analysis.

The known effect of air temperature is an increase in noise levels with a decrease in temperature. The relationship between noise level and temperature presented in ISO/TS 13471-2 (2022) was determined from a compilation of several investigations; it considers the slope of the linear relationship between $L_{A,max}$ and temperature, named temperature coefficient (γ_t), given in dB/°C. γ_t is defined for tyre class and road surface type. Although the $L_{A,max}$ – temperature relationship is known, currently, there is no indication in the literature of the expected correlation between temperature and the psychoacoustic indicators retrieved from road traffic noise, if any. Therefore, this study analysed the correlation between the indicators and air temperature or speed using Pearson's correlation coefficient for linear relationships and Spearman's correlation coefficient for monotonic relationships. A temperature coefficient for $L_{A,max}$ was also calculated instead of using those specified in ISO 13471-2 (2022). The indicators presenting a linear correlation with temperature were normalized to a 20 °C reference temperature using the linear regression slope for each vehicle category (Equation 1).

$$I_{norm,i} = I_{meas,i} - \alpha_{air} \times (T_{air,i} - 20) \quad (1)$$

Where:

- $I_{norm,i}$: acoustic/psychoacoustic indicator normalized to 20 °C for vehicle i
- I_{meas} : acoustic/psychoacoustic indicator as initially measured for vehicle i
- $T_{air,i}$: air temperature measured during the passage of vehicle i
- α_{air}: slope from the linear regression model between the I_{meas} and T_{air}

Similarly to temperature, ISO 11819-1 (2023) describes that $L_{A,max}$ can be normalized to a reference speed using the slope of the linear regression between this parameter and the logarithm of speed. Thus, correlations between the psychoacoustic indicators and the log of speed were explored. The values corrected for temperature were then normalized to 50 km/h for the vehicle categories that displayed a significant linear correlation (Equation 2).

$$I_{norm1,i} = I_{norm,i} - \alpha_{speed} \times log_{10}(\frac{v_i}{50}) \quad (2)$$

Where:

- $I_{norm1,i}$: acoustic/psychoacoustic indicator normalized to 20 °C and 50 km/h for vehicle i
- v_i : speed of the pass-by vehicle *i*
- α_{speed}: retrieved from the linear regression between the I_{norm} and log(v_i)

2.6 Multinomial logistic regression for vehicle classification

Logistic regression is an analytic technique for multivariate modelling of categorical outcome variables. Multinomial logistic regression (MLR) is the generalization of binary logistic regression for categorical outcomes with more than two classes.

The *LogisticRegression* function from the *Scikitlearn* library in Python was used to implement the logistic regression classifier. The algorithm uses the cross-entropy loss for multinomial logistic regression, 'lbfgs' solver, L2 regularization, C=1.0, and the softmax function to find the predicted probability of each class. The dataset was stratified by the vehicle categories and split randomly into a training set of 70% and a testing set with the 30% remaining data. The models' performance was measured via recall for each class: the ratio of true positives to the total number of observations in that class. The overall model's accuracy was obtained by dividing the total number of correct predictions by the total number of predictions.

Three models were developed: one considering the three ISO categories, another considering the four categories in the dataset, and a third in which passenger cars and vans were combined into a single class.

2.7 Data preparation

2.7.1 Multicollinearity

Logistic regression (LR) analysis is a useful complement to ordinary linear regression (OLS) when dealing with categorical response variables. However, since LR calculates changes in the

log odds of the response variable instead of changes in the variable itself, the underlying assumptions for OLS regression, such as linearity, normal distribution and homoscedasticity, are not required for LR. Yet, strong multicollinearity can negatively impact prediction accuracy. Multicollinearity among predictors can lead to an inflation of the coefficient estimates. This inflation is captured by the Variance Inflation Factor (VIF), which is, therefore, a multicollinearity indicator. A predictor with a VIF score higher than 5 can be predicted by other independent variables in the dataset (Midi et al., 2013). VIF scores were calculated using the *variance_inflation_factor* function from the *Statsmodels* library in Python.

2.7.2 Dataset balancing: under sampling and over sampling

While this dataset has many observations for passenger cars and HM, it falls short in Vans and HD (n = 823, 1103, 188 and 85, respectively). Imbalanced datasets introduce bias into the predictive tasks as there are insufficient examples of the minority classes for a model to learn the decision boundary effectively. Class imbalance can be addressed by re-sampling the original dataset, by over-sampling the minority classes and/or under-sampling the majority classes. Synthetic Minority Oversampling Technique (SMOTE) is a data augmentation technique that uses a k-nearest neighbour algorithm to interpolate new synthetic instances in feature space (Nitesh Chawla, et al. 2002). *SMOTE* and *RandomUnderSampler* functions from the *imblearn* library in Python were used. The sampling strategy comprised undersampling the two largest categories to 500 observations each and oversampling the two smallest categories to the same sample size, achieving a final dataset with 2000 observations in total.

3. Results and discussions

3.1 Texture and CPX

Figure 2 shows the one-third octave band texture spectra at a reference value of 10^{-6} m in the megatexture and most of the macrotexture range for two measurement repetitions at each location. Good repeatability is observed between repetitions for the texture range below 250 mm. Some variability is visible for larger wavelengths, probably due to transversal inhomogeneities of the pavement texture.

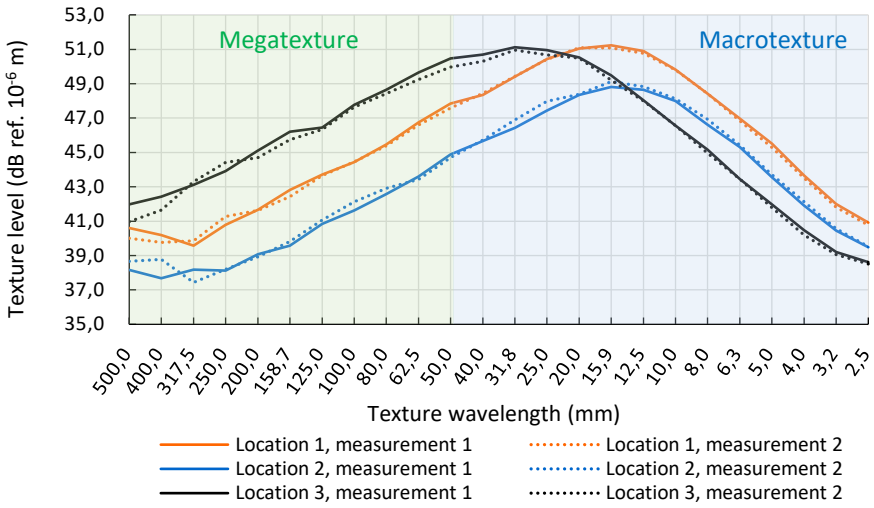

Figure 2: Texture spectra (average of 1-m segments)

The texture wavelength at which the texture level peaks is about 1.2 times the maximum aggregate size (10 mm) for the SMA surfaces at locations 1 and 2. The same is, however, not observed for the chipped dense asphalt concrete surface of Location 3.

The spectra of Location 1 and 2 have a similar shape, although the first presents 2 to 3 dB higher texture levels. ISO 13473-5 (2009) indicates that the texture level in the octave band centered at 63 mm (L_{tx63}) plays a direct role in the tyre/road noise generation as these wavelengths correspond to the tyre/pavement contact patch length. L_{tx63} is 51.6, 49.5 and 54.4 dB for Locations 1, 2 and 3, respectively.

Location 3 presents the highest megatexture (wavelengths between 50 mm and 500 mm). This texture level drops to values below the two other locations for shorter wavelengths (<25 mm) in the macrotexture range. An increased megatexture combined with a smaller macrotexture may result in a higher tyre/road noise (Sandberg and Ejsmont, 2002).

Figure 3 displays the three locations' one-third octave band noise spectra determined from the CPX measurements with the P1 tyre. The standard deviations are displayed as error bars around the data points. Table 2 presents the broadband $L_{CPX:P}$.

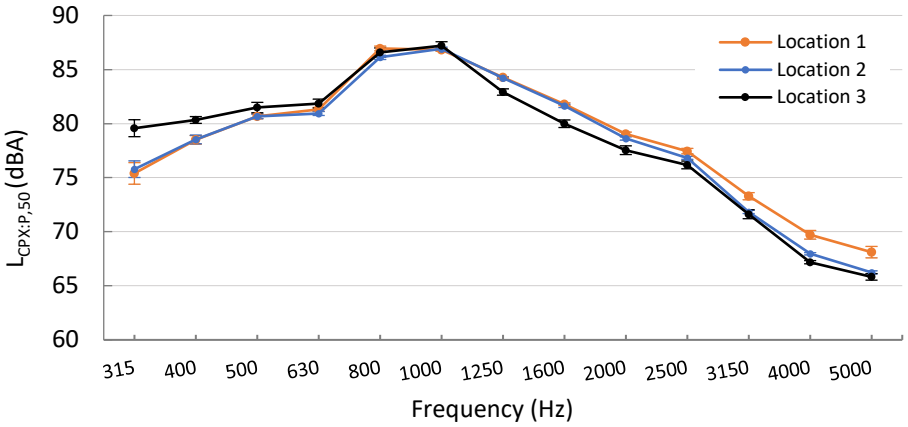

Figure 3: $L_{CPX:P}$ in 1/3 octave bands (average from 20-m segments)

Table 2: Broadband CPX levels ($L_{CPX:P,50}$ in dBA – mean ± standard deviation)

Location 1	Location 2	Location 3
92.8 ± 0.1	92.6 ± 0.1	92.8 ± 0.3

The relatively low standard deviation (up to 0.8 dBA) indicates homogeneity of the $L_{CPX:P}$ across the 20 m segments in all frequency bands. This implies that the noise levels recorded in front of the SPB microphone position accurately represent the whole 200 m sections.

Locations 1 and 2 present similar noise levels in most frequency bands. Between 315 and 630 Hz, the $L_{CPX:P}$ of Location 3 is slightly higher than the other two, with a maximum difference of 3.8 dBA observed at 315 Hz. These values are expected as, at these noise frequencies, the tyre/road noise of passenger cars increases with texture levels in the megatexture range (Sandberg and Ejsmont, 2002). These differences are reduced and the three spectra merge at 800 and 1000 Hz. Above 1250 Hz, the noise levels at Location 3 are lower, differing from Location 1 by a maximum of 2.3 dBA at 5000 Hz.

Despite the differences in the noise spectra, Table 2 shows that the broadband CPX levels are very similar across locations. Consequently, it can be inferred that the road surfaces have comparable acoustic performance.

3.2 SPB noise level spectra

Considering that the ISO 11819-1 (2023) requirements for free-field conditions and background noise are met, the differences between testing sites lie mainly in the road surface characteristics that impact tyre/road noise generation. From the CPX results, it is acceptable to consider the three locations as similar in these terms and pool the SPB measurements to obtain a more robust dataset. Figure 4 depicts the average $L_{A,max}$ spectra for each vehicle category from all locations. These averages were obtained from the $L_{A,max}$ corrected for temperature according to ISO/TS 13471-2 (2022) and then normalized to 50 km/h for each one-third octave band. The error bars represent the standard deviation.

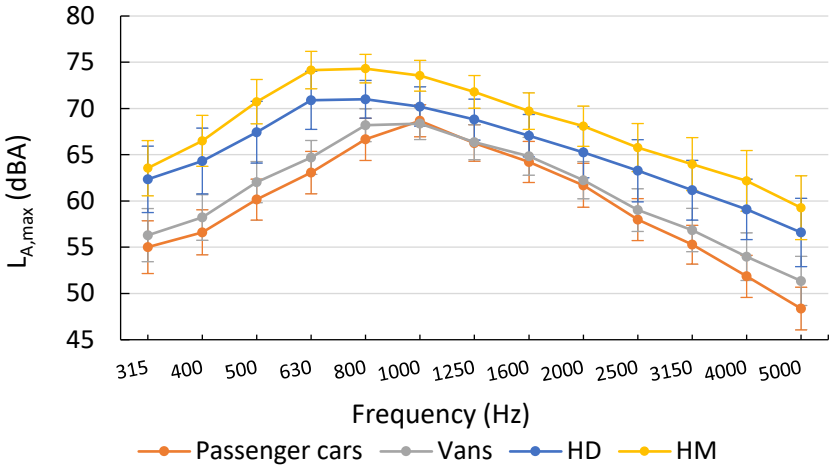

Figure 4: Average $L_{A,max}$ spectra (at reference 50km/h and 20 °C)

Passenger cars and vans present a similar $L_{A,max}$ peak around the 800-1000 Hz third-octave bands and their spectra almost coincide between 1000 and 2500 Hz. For frequency bands below 1000 and above 2500 Hz, the average noise levels of vans are slightly higher than those of passenger cars, but mostly by no more than 2 dBA, probably due to contributions from engine and car body radiation rather than from tyre/road interaction. The noise spectra of HD and HM exhibit comparable shapes, with the second shifted upwards by 2.1 ± 0.7 dBA on average. This difference suggests that these two groups are similarly affected by the different tyre/road noise-generating mechanisms, resulting in a consistent vehicle group once the noise levels for HD are incremented, as proposed in ISO 11819-1 (2023).

It should be recalled that SPB measurements capture vehicle noise, a combination of both tyre/road noise and power-unit noise. Power-unit noise composes the largest share of vehicle noise at low speeds, but tyre/road noise dominates vehicle noise generation as speeds increase. This transition occurs at a lower speed for light vehicles; at 50 km/h, tyre/road noise is already the primary source of vehicle noise for passenger cars, while for heavy vehicles, the share of power-unit noise is still considerable (Li, 2018). Therefore, the vehicle noise samples of heavy vehicles in this study are expected to have a significant share of engine noise. Still, when investigating tyre/road noise, tyre characteristics play a major role in differentiating tyre/road noise emission among vehicle categories. Regardless of the type of pavement, the range between the loudest and quietest tyre can reach up to 10 dB, as observed by Sandberg and Ejsmont (2002). Therefore, the similarities between passenger cars to vans and HD to HM observed in Figure 4 can be attributed to passenger cars having type C1 tyres, while on vans mostly C1, but also C2 tyres are mounted. HD are equipped with C2 and, in some cases, with C3 tyres, while HM are equipped with C3 only.

The spectral energy contribution rate (P(*k*) in Equation 3) was calculated, as proposed by Yang et al. (2020). This approach rules out magnitude differences, facilitating the visualization of the noise frequency components' contribution and the identification of mechanisms affecting vehicle noise generation among the vehicle categories. P(*k*) is the percentage of the $L_{A,max}$ at

each k-th one-third octave band (L_k) in relation to the total $L_{A,max}$ (L). *k* ranges from 6.3 Hz to 20 kHz. Only the spectral energy contribution rates in the 315 to 5000 Hz frequency bands are displayed in Figure 5, as it contains almost the entire energy share.

$$P(k) = \frac{10^{0.1*L_k}}{10^{0.1*L}} \times 100, k = 6.3, 8, \dots, 20000 \quad (3)$$

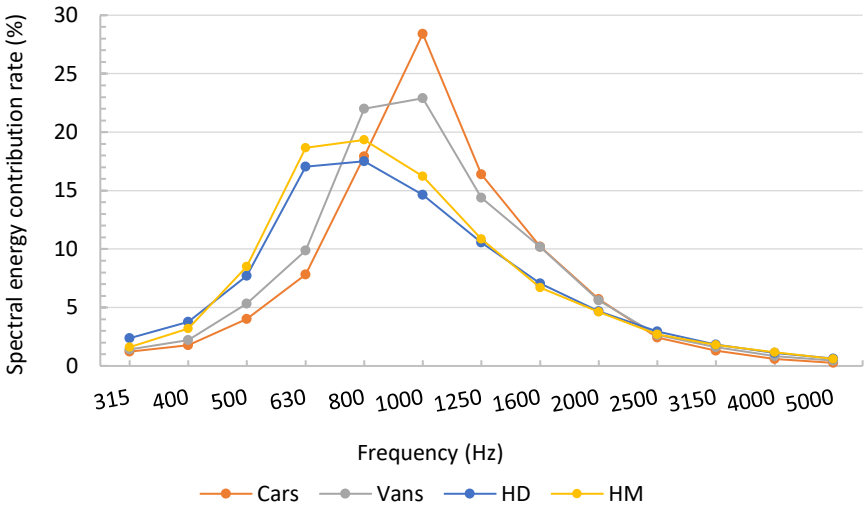

Figure 5: Spectral energy contribution rate

More than half of the spectral energy contribution rate for HD and HM is present at frequency bands up to 800 Hz. For passenger cars, the contribution rate becomes more significant at the 1000 and 1250 Hz bands.

Although the noise levels were normalized to a reference speed, the speed impact on the energy contribution rates is not entirely ruled out, especially as heavy vehicles travel slower than light vehicles (Table 3). Since power-unit noise is linked to low-frequency components of vehicle noise (Roberts, 2010), the more significant energy contribution rates in the low frequencies confirm that HD and HM vehicle noise is affected by power-unit noise.

Regarding tyre/road noise, low-frequency components primarily result from tyre vibrations induced by the shocks of irregularities in the megatexture range, whereas energy in the high frequencies is linked to air pumping caused by noisy air displacement in/out of the tyre tread upon contact with the pavement. The crossover frequency between the two phenomena occurs at 500-630 Hz and 800-1200 Hz for trucks and car tyres, respectively (Sandberg and Ejsmont, 2002). Figure 5 supports the notion that impact mechanisms are more significant in generating tyre/road noise for heavy vehicles, while air pumping is more prominent for light vehicles.

From 2000 Hz, Figure 5 displays similar spectral energy contribution rates for all vehicle categories. Given that the microphone is placed in the far field in the SPB method, the received signal is impacted by propagation effects such as sound absorption in the travelling medium and interaction with the ground. Typically, the higher noise frequencies are more easily attenuated

in the propagation path than low frequencies (Lercher, 2019), explaining the reduced contribution rate in the higher bands.

3.3 Dataset overview and psychoacoustic indicators correlations with temperature and speed

Table 3 shows the average of a set of features that describe the 4 s vehicle passages for the four categories. Apart from the parameters described in Section 2.4, Table 3 presents the average speed, the A-weighted equivalent continuous sound level ($L_{A,eq}$) and the 10 and 90% percentile statistical noise levels.

Table 3: Acoustic and psychoacoustic indicators averages and standard deviation with neither temperature nor speed correction

Feature	P	Vans	HD	HM
Amount	823	188	85	1103
Speed (km/h)	57.4 ± 10.0	56.1 ± 10.0	51.0 ± 7.0	51.6 ± 7.2
$L_{A,max}$ (dBA)	75.1 ± 2.6	75.6 ± 2.4	78.8 ± 2.7	81.5 ± 2.0
ΔL (dBA)	10.9 ± 4.2	10.4 ± 4.1	13.5 ± 4.2	14.5 ± 4.2
$L_{A,eq}$ (dBA)	71.6 ± 2.3	72.0 ± 2.2	75.0 ± 2.1	77.7 ± 2.4
L_{10} (dBA)	74.4 ± 2.8	74.9 ± 2.7	78.1 ± 2.5	80.7 ± 2.9
L_{90} (dBA)	66.2 ± 2.2	66.8 ± 2.0	68.6 ± 2.1	70.4 ± 2.4
N_{50} (sone)	22.52 ± 2.63	23.99 ± 2.63	30.77 ± 3.39	36.83 ± 4.79
S_{50} (acum)	1.254 ± 0.060	1.293 ± 0.065	1.263 ± 0.055	1.302 ± 0.058
R_{50} (asper)	0.058 ± 0.010	0.060 ± 0.009	0.065 ± 0.009	0.072 ± 0.011
FS_{50} (vacil)	0.479 ± 0.070	0.479 ± 0.052	0.467 ± 0.052	0.493 ± 0.076

These averages are, however, calculated from vehicles that passed at different speeds and air temperatures. As described in Section 2.5, these environmental and driving conditions impact tyre/road noise and may consequently affect the psychoacoustic indicators. A normalization to reference values is necessary to allow direct comparison among categories.

Henceforth, the calculations and discussions will be limited to the chosen relevant indicators $L_{A,max}$, ΔL, N_{50}, S_{50}, R_{50}, and FS_{50}. Table 4 shows Pearson's correlation coefficient and simple linear regression coefficients between these indicators and air temperature for each vehicle category. The air temperature in the measurements ranged from 16.0 to 36.4 °C, averaging 24.3 ± 4.2 °C.

Table 4: Simple linear regression analysis results between (psycho)acoustic indicators and air temperature (p-values < 0.05 in red)

	Category	$L_{A,max}$	ΔL	N_{50}	S_{50}	R_{50}	FS_{50}
Pearson's correlation coefficient	P	-0.39	0.04	-0.34	<0.01	-0.26	0.26
	Vans	-0.41	0.05	-0.31	0.08	-0.32	0.09
	HD	-0.28	-0.06	-0.28	-0.03	-0.03	0.10
	HM	-0.16	0.03	-0.12	-0.06	-0.11	0.05
p-value	P	<.001	.304	<.001	.984	<.001	<.001
	Vans	<.001	.530	<.001	.282	<.001	.229
	HD	.009	.605	.008	.771	.761	.343

	HM	<.001	.395	<.001	.030	<.001	.073
Slope	P	-0.28	0.04	-0.22	1.06×10^{-5}	-6.42×10^{-4}	4.58×10^{-3}
	Vans	-0.28	0.05	-0.20	1.29×10^{-3}	-7.47×10^{-4}	1.17×10^{-3}
	HD	-0.16	-0.05	-0.21	-3.88×10^{-4}	-6.98×10^{-5}	1.18×10^{-3}
	HM	-0.11	0.03	-0.13	-8.97×10^{-4}	-2.89×10^{-4}	9.61×10^{-4}
Standard error	P	0.02	0.04	0.02	5.2×10^{-4}	8.1×10^{-5}	5.8×10^{-4}
	Vans	0.05	0.08	0.05	1.2×10^{-3}	1.6×10^{-4}	9.6×10^{-4}
	HD	0.06	0.10	0.08	1.3×10^{-3}	2.2×10^{-4}	1.2×10^{-3}
	HM	0.02	0.03	0.03	4.1×10^{-4}	7.6×10^{-5}	5.3×10^{-4}

A statistically significant negative linear correlation ($p<0.05$) exists between air temperature and $L_{A,max}$ and N_{50} for all vehicle categories. R_{50} exhibited a significant negative linear correlation across all vehicle categories, except for HD, which may be attributed to the relatively smaller sample size of this group. Moreover, a significant negative linear correlation ($p=0.03$) was observed between air temperature and S_{50} for HM only. Conversely, for P only, a significant negative linear correlation with FS_{50} was found ($p<0.001$). Spearman's correlation coefficient was calculated to ensure that monotonic, although non-linear, correlations between the indicators and temperature were not disregarded. The only extra significant correlation not presented in Table 4 retrieved from this non-parametric test is given by FS_{50}, for HM ($p = 0.002$).

For dense asphalt surfaces, the temperature coefficients in ISO/TS 13471-2 (2022) for tyres C1, C2 and C3 are, respectively, -0.10, -0.07 and -0.06 in dB/°C. Although smaller, these values are in the same order of magnitude as the slopes for $L_{A,max}$ in Table 4. The differences between the ISO coefficients and those retrieved in this study can be attributed to the broader temperature range covered in the standard (5°C and 35 °C), while the air temperatures of our measurement campaign cover only the higher end of this range.

Similarly to the A-weighted noise level, N_{50} also shows significant negative linear relations, as expected, considering that both parameters are related to sound energy content. The slopes for both indicators versus temperature are quite similar for all vehicle categories, even though loudness, in sones, is spread over a wider range of values than $L_{A,max}$ in dBA (Figure 6 a and b).

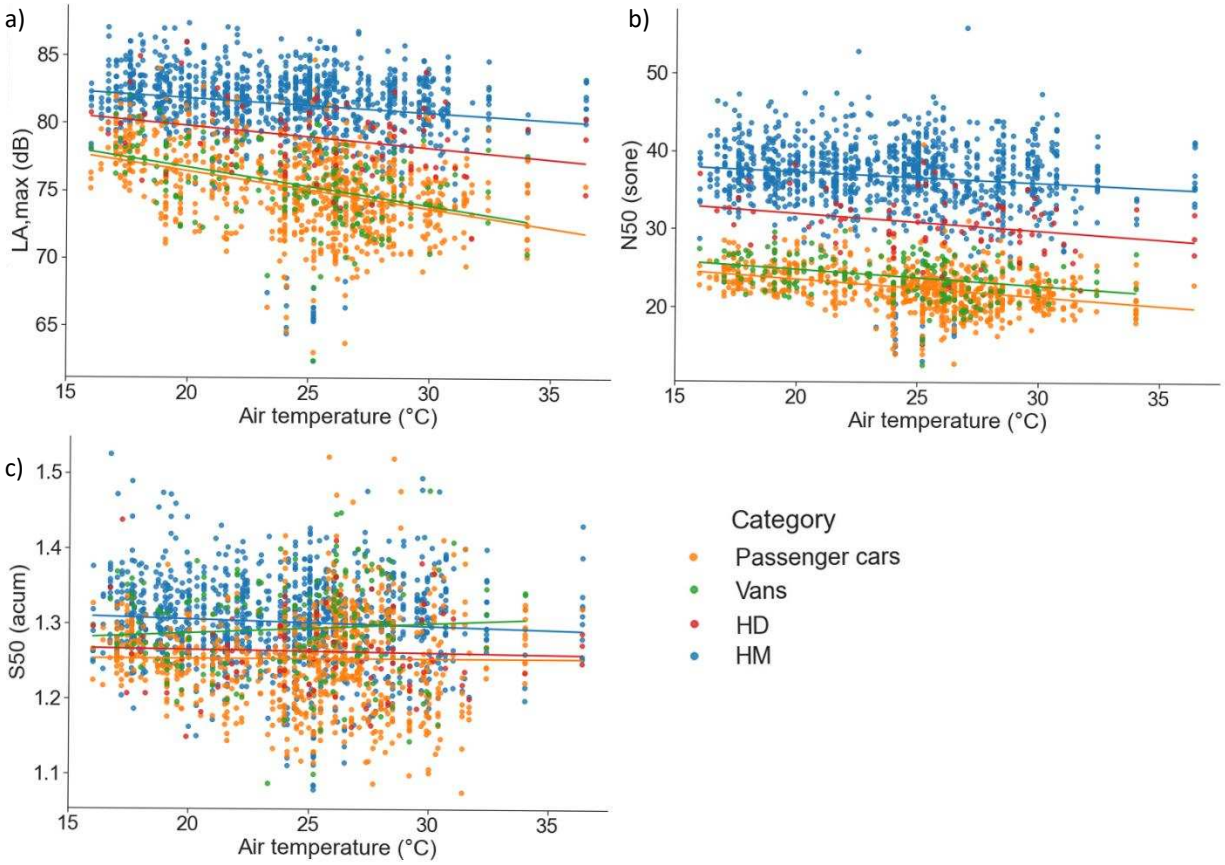

Figure 6: Scatter plot and regression line of a) $L_{A,max}$, b) N_{50} and c) S_{50} versus air temperature

The same temperature coefficients are applied to all one-third octave bands in ISO/TS 13471-2 (2022). According to the standard, although some studies indicate that temperature influence is more pronounced at low (below 800 Hz) and high frequencies (above 1250 Hz), insufficient consistency exists to consider a frequency-dependent effect of temperature on road traffic noise. The lack of correlation between S_{50} and temperature (see also flat linear trendlines in Figure 6 c) supports the premise that temperature changes have no significant impact on high-frequency components.

The non-significant correlation for ΔL seems logical as it is derived from subtracting two noise levels equally impacted by temperature; possibly for the same reason, the temperature does not affect low-modulated noise fluctuations, given by FS_{50}. On the other hand, there is no clear physical explanation for fast-modulated sounds (R_{50}) to show a significant negative correlation with temperature; in this case, the correlation might not indicate a causal relationship.

The significant correlations found for only a single vehicle category for a certain indicator were neglected and $L_{A,max}$, N_{50} and R_{50} were normalized to 20 °C for all vehicle categories, according to Equation 1. Subsequently, the linear correlations between the indicators (already corrected for temperature) and the speed log (Table 5) were checked.

Pearson's correlation coefficients are statistically significant for most indicators, except for S_{50} for Vans, HD and HM, and FS_{50} for HD. For HD, Spearman's correlation coefficient was

significant in FS_{50} ($p = 0.022$). Therefore, except for S_{50}, all parameters were normalized to 50 km/h for each vehicle category, as per Equation 2.

The significant positive linear correlations between $L_{A,max}$ and log of speed agree with the behaviour described in ISO/TS 11819-1. This standard proposes a fixed speed coefficient of 25 for heavy vehicles and dense asphalt surfaces. The slope retrieved for HD is very close to this value (24.92), while for HM, although smaller, it is still relatively similar. A reason for the smaller slope may be more significant power-unit noise contributions in total vehicle noise for HM than HD, given that power-unit noise is less sensitive to speed changes than tyre/road noise (Sandberg and Ejsmont, 2002).

Table 5: Simple linear regression analysis results between (psycho)acoustic indicators (corrected for temperature) and log of speed (p-values < 0.05 in red)

	Category	$L_{A,max}$	ΔL	N_{50}	S_{50}	R_{50}	FS_{50}
Pearson's correlation coefficient	P	0.67	0.26	0.29	0.19	0.31	-0.48
	Vans	0.60	0.16	0.22	-0.04	0.25	-0.35
	HD	0.58	0.25	0.33	-0.05	0.22	-0.21
	HM	0.40	0.26	0.18	0.02	0.10	-0.13
p-value	P	<.001	<.001	<.001	<.001	<.001	<.001
	Vans	<.001	.025	.003	.540	.001	<.001
	HD	<.001	.021	.002	.666	.044	.051
	HM	<.001	<.001	<.001	.447	.001	<.001
Slope	P	23.87	14.69	9.63	0.15	0.04	-0.45
	Vans	19.67	8.78	6.70	-0.04	0.03	-0.24
	HD	24.92	18.10	18.00	-0.04	0.03	-0.18
	HM	18.60	17.69	14.06	0.02	0.02	-0.16
Standard error	P	0.94	1.91	1.11	0.03	4.2×10^{-3}	0.03
	Vans	1.94	3.88	2.33	0.06	0.01	0.05
	HD	3.81	7.68	5.66	0.10	0.02	0.09
	HM	1.28	1.99	2.29	0.03	0.01	0.04

N_{50} presents similar behaviour as $L_{A,max}$, but smaller Pearson's correlation coefficients and slopes. Peak intensity descriptors are expected to be more sensitive to speed variations than noise intensity descriptors that cover the whole signal sensation.

ΔL presents a positive linear correlation with the log of speed. Naturally, the faster the vehicle, the farther away it is from the microphone at 2 s before it produces the peak noise level, resulting in a smaller $L_{A,max}$ at that moment. In addition, faster vehicles provoke a higher $L_{A,max}$ peak; thus, subtracting these two noise levels results in larger ΔL. Overall, the slopes for this indicator mean that a 10 km/h increase in vehicle speed in the range of 30-50 km/h increases ΔL by 0.9 - 1.8 dBA.

For slowly modulated noise, the negative Pearson's correlation coefficient for FS_{50} implies that speed reduces the sensation of sound fluctuation. Going further through the modulation frequency range, roughness, on the other hand, increases with speed. In Morel et al. (2016), the pass-by noise of two-wheeled vehicles in acceleration also produced higher roughness values than those driving at a constant speed, attributed partially to the presence of engine noise but also to the speed increase.

Overall, the linear correlations between S_{50} and log of speed were not statistically significant. By calculating the sharpness from tyre/road noise recorded in CPX measurements at different speeds, Guo et al. (2018) observed no trends or considerable changes in sharpness. The sharpness values of the constant speed and accelerating vehicles presented by Morel et al. (2016) also do not change significantly between the two driving conditions.

3.4 Comparison of acoustic and psychoacoustic indicators among vehicle categories

With values normalized to a reference temperature and speed, measures of central tendency can be compared across vehicle categories. Figure 7 to 9 display the indicator's distributions as boxplots. The averages and standard deviations are also presented in Table 6.

In general, passenger cars and HM present a more significant number of outliers due to their larger sample size and, for HM, possibly increased by variations in tyre load and number of axles. For each indicator, the interquartile range is similar across vehicle categories, showing that the distribution behaviour of these quantities is not sensible to vehicle type. A one-way ANOVA at a 5% significance level revealed overall statistically significant differences in mean across vehicle categories for all indicators. Posthoc Tukey test results regarding individual differences between pairs are discussed in the following paragraphs.

Table 6: (Psycho)acoustic indicators averages and standard deviations, corrected for temperature and speed

Feature	P	Vans	HD	HM
$L_{A,max}$ (dBA)	75.2 ± 2.0	75.9 ± 2.0	79.7 ± 2.1	81.7 ± 2.6
ΔL (dBA)	10.1 ± 4.1	10.1 ± 4.1	13.5 ± 4.2	14.4 ± 4.1
N_{50} (sone)	23.07 ± 2.37	24.54 ± 2.45	31.94 ± 3.07	37.21 ± 4.67
S_{50} (acum)	1.254 ± 0.060	1.293 ± 0.065	1.263 ± 0.055	1.302 ± 0.058
R_{50} (asper)	0.059 ± 0.009	0.062 ± 0.009	0.065 ± 0.009	0.072 ± 0.011
FS_{50} (vacil)	0.502 ± 0.061	0.489 ± 0.049	0.468 ± 0.051	0.494 ± 0.075

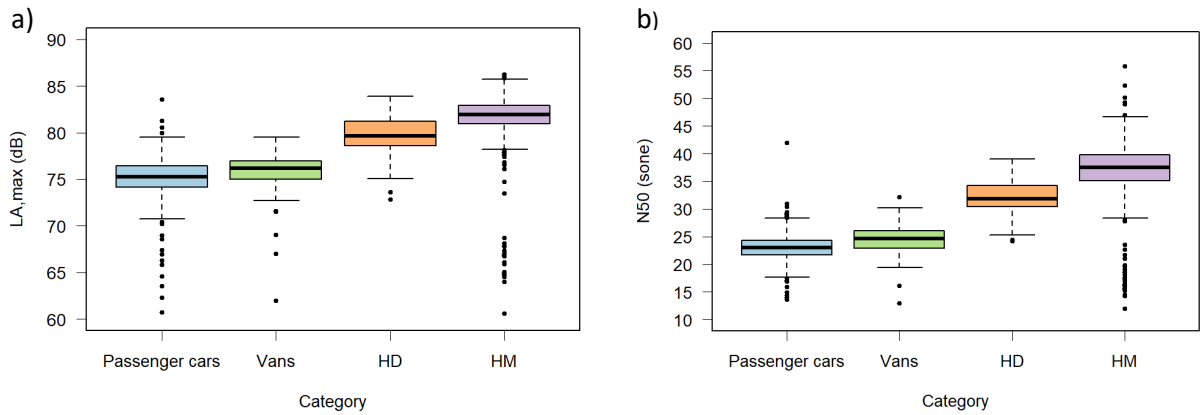

Figure 7: Boxplots of a) $L_{A,max}$ and b) N_{50}, both corrected for temperature and speed

From Table 6, the means of $L_{A,max}$ in passenger cars and vans are virtually identical, while HD is 4 and HM 6 dBA larger than the first two. The detailed discussion on these differences, also considering the $L_{A,max}$ spectra, was presented in Section 3.2.

The magnitude differences in loudness are more discrepant than $L_{A,max}$, with vans being 1.5 sones louder than passenger cars, while this difference is 8.9 and 14.1 sones for HD and HM, respectively. Posthoc Tukey tests revealed that the differences in average are statistically significant between all categories for both $L_{A,max}$ and N_{50}.

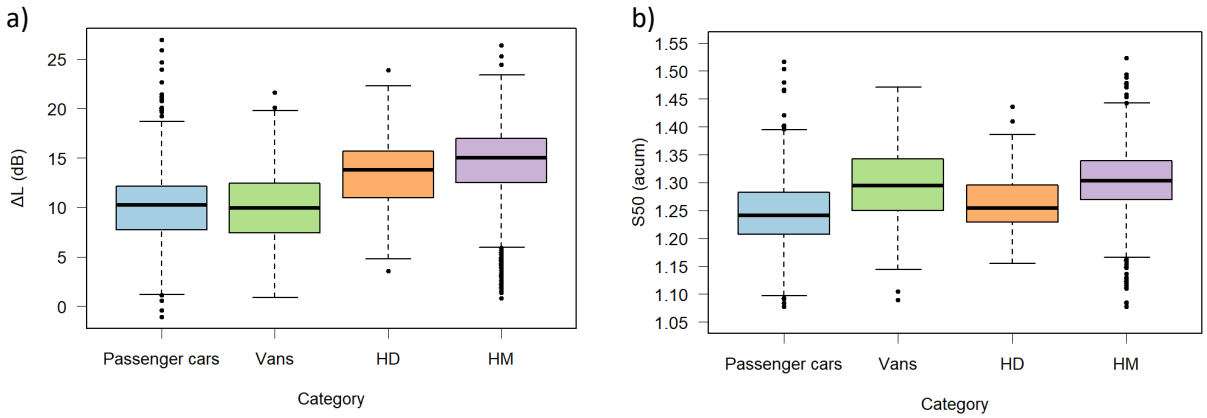

Figure 8: Boxplots of a) ΔL and b) S_{50}, both corrected for temperature and speed

ΔL of heavy vehicles is around 4 dBA larger than for light vehicles (Figure 10 a). This value is smaller than the difference in peak $L_{A,max}$, as a heavy vehicle approaching produces higher noise levels than a light vehicle at 2 s before it crosses the microphone. Tukey's test found no statistically significant differences between passenger cars and vans (p-value = 0.900) and between HD and HM (p-value = 0.222).

The magnitude variation in sharpness among vehicle categories (Figure 10 b) is not wide. HM exhibited the highest average sharpness of 1.30 acum, while passenger cars had the lowest average sharpness of 1.25 acum. This can be attributed to the differences in high-frequency noise content of these samples not being so pronounced to impact the sharpness values substantially. From the posthoc Tukey test, passenger cars and HD showed no statistically significant differences (p-value = 0.491), or vans and HM (p-value = 0.172). Morel et al. (2012) reported 1.5, 1.3 and 1.2 acum for a two-wheeled, a light and a heavy pass-by vehicle, respectively, also calculated by Zwicker's model. Fu and Murphy (2017) recorded pass-by noise of a motorcycle and a passenger car driving at certain acceleration and speed conditions to obtain samples dominated by engine noise and tyre/road noise, respectively; sharpness values yielded 1.51 acum for tyre/road noise and 1.13 acum for engine noise (employing Zwicker's model). Paviotti and Vogiatzis (2012) observed that sharpness did not allow for distinguishing between cars and two-wheeled pass-bys. Considering these results and confirmed by the literature, no clear relationship between vehicle characteristics and sharpness is observed.

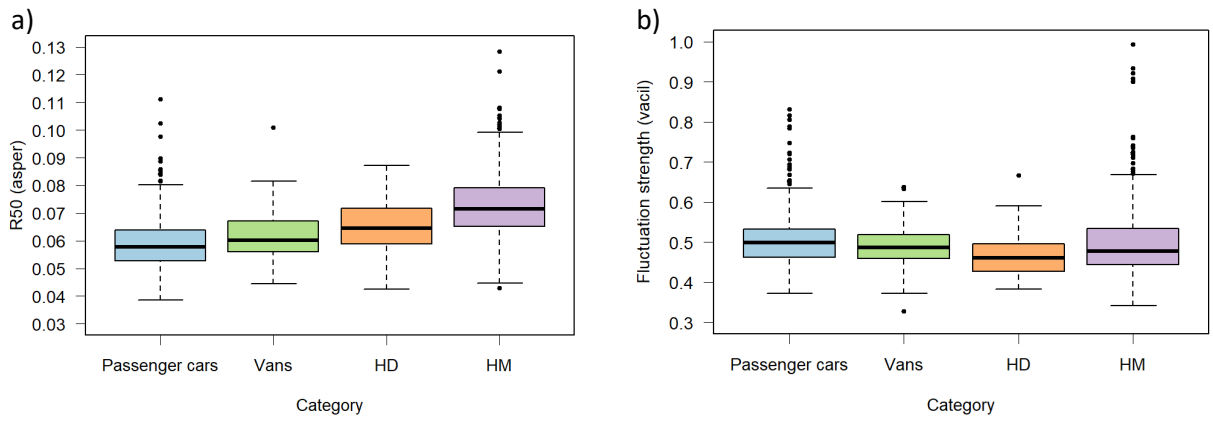

Figure 9: Boxplots of a) R_{50} and b) FS_{50} both corrected for temperature and speed

Concerning temporal aspects, Altinsoy (2022) stated that quickly modulated noise components tend to arise from engine orders or diesel knocking in internal combustion engine cars. Also, according to Morel et al. (2012), roughness in road traffic noise can be attributed to engine noise being predominant for limited-speed vehicles in urban areas. The tyre/road noise sample obtained by Fu and Murphy (2017), calculated by Zwicker's model in this study, presented R_{50} of approximately 0.036 asper, while this value was 1.7 asper for the engine noise sample. Figure 9 a) shows a consistent increase in roughness from light to heavy vehicles. The value of 0.072 yielded by HM is double the roughness of Fu and Murphy's (2017) tyre/road noise sample. This discrepancy suggests that an engine noise share in the heavy vehicles noise samples, especially HM, may have increased the average roughness compared to passenger cars and vans. Additionally, posthoc Tukey tests displayed statistically significant differences between all categories. Paviotti and Vogiatzis (2012) describe roughness as an important differentiating parameter among cars, motorbikes and scooters' pass-by noise, given its independence of loudness level and large sensitivity.

In Morel et al. (2012), the fluctuation strength of pass-bys shows an increasing trend in the following order: heavy, light and two-wheeled vehicles. The fluctuation strength of pass-by noise samples analyzed by Paviotti and Vogiatzis (2012) did not present a systematic variation among classes. The differences in average seen in Figure 9 b), with passenger cars presenting the higher FS value (0.502 vacil) and HD the lowest (0.468 vacil), are not clearly associated with any particular characteristic of the vehicle categories. Tukey's posthoc test revealed that differences in means between vans to passenger cars, HD and HM were not statistically significant (p-values of 0.069, 0.078, and 0.724, respectively).

The considerable number of instances and the sensible variation of the psychoacoustic indicators with vehicle category render these attributes potentially relevant to training a machine learning model to predict vehicle categories.

3.5 Classification

Table 7 presents the VIF scores used to measure multicollinearity. Initially, all parameters showed a VIF score above 5 (see column "Before scaling"), which may be caused by structural multicollinearity. Structural multicollinearity is a mathematical artefact caused by creating new

predictors from the data itself rather than actual sampled data. As the psychoacoustic indicators are calculated from the same noise signal, some degree of structural multicollinearity is to be expected. Standardization is a feature scaling technique that aids in reducing structural multicollinearity without removing explanatory variables from the model (Ríos and Simpson, 2017). It transforms the data into zero mean and standard deviation of one. The third column of Table 7 shows that feature scaling reduced the VIF scores greatly, although $L_{A,max}$ and N_{50} still show a certain degree of multicollinearity. Removing either N_{50} or $L_{A,max}$ dropped the VIF scores to values below 5 (fourth and fifth columns in Table 7, respectively). This is expected as both loudness and $L_{A,max}$ are indicators of noise intensity. In a preliminary attempt to train the classifier, the prediction accuracy yielded by using $L_{A,max}$ was lower than using loudness, therefore, we proceeded with including only N_{50} to avoid model overfitting.

Table 7: Variance Inflation Factor

Feature	VIF score			
	Before scaling	After scaling		
$L_{A,max}$	804.00	7.36	1.80	-
ΔL	58.28	1.24	1.24	1.20
N_{50}	9.96	7.35	-	1.80
S_{50}	457.08	1.11	1.10	1.10
R_{50}	58.26	1.44	1.38	1.44
FS_{50}	68.39	1.09	1.07	1.04

The dataset was balanced with 500 instances per vehicle category and split into 70% for training and 30% for testing, resulting in 150 observations per class to assess the algorithm's accuracy. Model 1 considers the vehicle categories defined in ISO 11819-1; Table 8 shows the performance metrics per class (recall) and overall model accuracy, while Figure 10 a) displays the confusion matrix of the model. Overall accuracy reaches 84.0%, indicating the psychoacoustic attributes have patterns that match the visual-based labels with considerable precision. Although HD and HM present notable similarities in psychoacoustic indicators, as presented in Section 3.4, the misclassification between these categories hit 16%.

Model 2 (Figure 10 b) yielded 71.5% accuracy considering the four vehicle categories. The main cause of accuracy losses is 27% of the passenger cars being misclassified as vans and vice-versa. Additionally, 17% of the true HDs were predicted as HM, and 19% of the HM as HD. There was almost no misclassification between the extreme ends, e.g. passenger cars and vans by HM, but a few (8%) passenger cars and vans were classified as HD.

The performance of Model 2 demonstrates that vans are considerably similar to passenger cars in terms of auditory sensation represented by psychoacoustic indicators. This compatibility is also clear by the small average differences presented in Section 3.4 and the noise spectra displayed in Figure 4. If vans were to be included in SPB measurements, this group seems too similar to passenger cars to compose a fourth vehicle category, and could be better leveraged if combined with passenger cars. To illustrate that, a third model was built with passenger cars and vans combined; prediction performance is shown in Figure 10 c) and Table 8 (Model 3). Overall accuracy yields 85.7% and recall of the P+Vans class reaches 91%. The misclassification

between HD and HM is also reduced compared to Model 2. The greater accuracy compared to Model 1 highlights the feasibility of including vans in the passenger cars category.

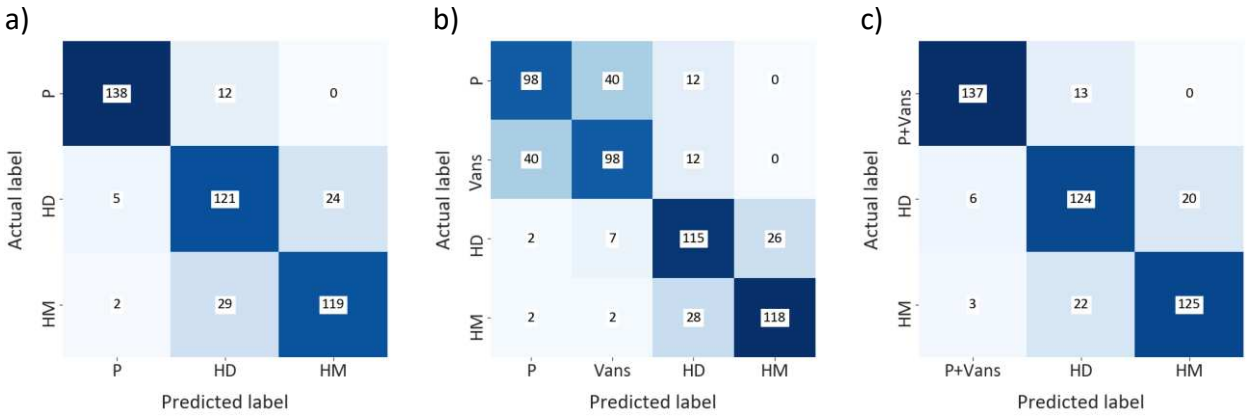

Figure 10: Confusion matrix a) ISO 11819-1 categories (Model 1), b) Four vehicle categories (Model 2), c) P and vans combined (Model 3)

Table 8: Performance metrics

Model 1	Recall	Model 2	Recall	Model 3	Recall
P	0.920	P	0.653	P + Vans	0.913
		Vans	0.653		
HD	0.807	HD	0.767	HD	0.827
HM	0.793	HM	0.787	HM	0.833
Overall accuracy	0.840		0.715		0.857

Reaching higher prediction accuracy or having a larger number of classes is a trade-off, after all, vans still yielded a recall of 65% in Model 2. The choice of one over the other depends on the model's intended use. The practicality of SPB measurements could be improved if vehicles that would otherwise not be considered were incorporated into the current vehicle classification. However, a classification system aiming to capture characteristics of the traffic flow microstructure to estimate, for example, noise-induced annoyance may be better represented by more, detailed classes even if prediction accuracy is reduced.

4. Conclusions

An SPB measurement campaign was performed in three locations on hot-mix asphalt road surfaces. CPX measurements proved their comparable acoustic performance and a pooled dataset with 2199 vehicle pass-bys was obtained, divided into four categories: passenger cars, vans and heavy vehicles with dual or multiple axles.

The noise level ($L_{A,max}$) spectra and the spectral energy contribution rate of each vehicle category presented trends that generally align with previous studies, with heavy vehicles showing higher noise levels and more energy share in the lower frequencies compared to light vehicles. The differences and similarities in the noise spectra among vehicle categories are attributed mainly to the tyre type and the more significant contribution of power-unit noise for heavy vehicles.

The psychoacoustic indicators loudness, sharpness, roughness, and fluctuation strength, besides $L_{A,max}$ and ΔL, were calculated from 4 s audio segments taken around the moment a vehicle produced the peak noise level. The effects of air temperature and speed on these indicators were checked via linear regression. $L_{A,max,}$ loudness and roughness showed a positive linear correlation with temperature. The log of speed presented positive significant linear correlations with $L_{A,max}$, ΔL, loudness, roughness and a negative correlation with fluctuation strength. To the authors' knowledge, this is the first time correction coefficients of temperature and speed for psychoacoustic indicators have been published.

The averages of the acoustic and psychoacoustic indicators per vehicle category were obtained from values normalized to 20 °C and 50 km/h. The parameters related to noise intensity ($L_{A,max}$, ΔL, and loudness) increased consistently from light to heavier categories, as expected. The same trend is observed for roughness, probably due to quickly modulated noise components from engine noise. No tendencies were observed for sharpness and fluctuation strength across vehicle categories.

A machine-learning logistic regression model to predict vehicle category was trained using ΔL and the psychoacoustic indicators N_{50}, S_{50}, R_{50}, FS_{50} as features. The psychoacoustic indicators have allowed the algorithm to derive patterns that matched the visual vehicle classification defined in ISO 11819-1 with considerable accuracy in the testing set (84%). For a model including the extra vehicle category (vans), accuracy dropped to 72%, with misclassification between passenger cars and vans being the main cause of accuracy reductions. Combining these two classes into a single group increased accuracy to 86%.

In the context of the SPB method, this outcome shows the potential of enabling a less visual and operator-dependent vehicle classification task, as it could be performed by using audio signal only. Besides, pass-by vehicles that do not fit any category defined in ISO 11819-1, and would otherwise be excluded from the measurement, could be incorporated into the existing classes due to similarities in acoustic signature, as were vans in this study. Also, new categories that fit well together in terms of auditory sensation could be explored, for instance, electric vehicles, an emerging concern for the current classification systems. That said, the issue of inconsistent vehicle classification systems described in ISO 11819-1 for the vehicle fleet worldwide could be mitigated. SPB measurements have been widely performed to assess road surfaces' acoustic quality, and the recent ISO 11819-1 update shows it remains state-of-the-art. With larger training sets and increased training repetitions, the classifier can reach higher levels of predictive accuracy.

Within the framework of environmental noise control shifting toward a human-centred approach, a vehicle classification system based on auditory noise sensation, rather than size, gross vehicle mass or number of axles, can be relevant. To expand the current approach, pass-by noise samples labelled with more and different classes than those from the SPB standard should be explored using the psychoacoustic-based classifier. Additionally, instead of a supervised classification algorithm with visual-based labels, a clustering algorithm could be employed to investigate how the cluster structure forms based on psychoacoustic indicators.

CRediT author statement

Ablenya Grangeiro de Barros: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization **Michiel Gelyukens**: Formal analysis, Investigation, Data Curation, Writing – Review, Editing **Frederico Pereira**: Software, Writing - Review & Editing **Elisabete Freitas**: Conceptualization, Supervision, Writing - Review & Editing **Suzana Faria**: Methodology, Formal analysis, Writing - Review & Editing **Luc Goubert:** Conceptualization, Methodology, Resources, Supervision, Writing - Review & Editing **Cedric Vuye**: Conceptualization, Methodology, Formal analysis, Resources, Funding acquisition, Supervision, Project administration, Writing - Review & Editing

Declarations of interest: none

References

1. Altinsoy, M. E, 2022. The Evaluation of Conventional, Electric and Hybrid Electric Passenger Car Pass-By Noise Annoyance Using Psychoacoustical Properties. Applied Sciences 12(10), 5146. https://doi.org/10.3390/app12105146

2. Ascari, E., Cerchiai, M., Fredianelli, L., Licitra, G, 2022. Statistical Pass-By for Unattended Road Traffic Noise Measurement in an Urban Environment. Sensors 22(22), 8767. https://doi.org/10.3390/s22228767

3. BRRC (Belgian Road Research Center), 2020. Handleiding voor de keuze van de asfaltverharding bij het ontwerp of onderhoud van wegconstructies. Brussels. https://brrc.be/nl/expertise/expertise-overzicht/handleiding-keuze-asfaltverharding-ontwerp-onderhoud-wegconstructies (in Dutch)

4. Cabrera, D.; Jimenez, D.; Martens, W, 2014. Audio and Acoustical Response Analysis Environment (AARAE): a tool to support education and research in acoustics. Internoise 16-19 November 2014, Melbourne, Australia

5. Camacho, A., Piñero, G., Diego, M., González A, 2008. Exploring Roughness Perception in Car Engine Noises through Complex Cepstrum Analysis. Acta Acustica United With Acustica 94(1), 130-140. https://doi.org/10.2392/AAA.918015

6. Chaki, J., 2021. Pattern analysis based acoustic signal processing: a survey of the state-of-art. International Journal of Speech Technology 24, 913–955. https://doi.org/10.1007/s10772-020-09681-3

7. Chawla, N. V., Bowyer, K. W., Hall, L. O., Kegelmeyer, W. P., 2002. SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research 16, 321–357. https://doi.org/10.1613/jair.953

8. Daniel, P., Weber, R., 1997. Psychoacoustical roughness: Implementation of an optimized model. Acustica 83, 113-123.

9. Dawton, B., Ishida, S., Hori, Y., Uchino, M., Arakawa, Y., Tagashira, S., Fukuda, A., 2020. Initial evaluation of vehicle type identification using roadside stereo microphones, Proceedings IEEE Sensors Applications Symposium, SAS, 1–6.

10. EN ISO 11819-1; EN ISO. Acoustics—Measurement of the Influence of Road Surfaces on Traffic Noise—Part 1: Statistical Pass-By Method. The International Organization for Standardization: Geneva, Switzerland, 2023.

11. EN ISO 11819-2; EN ISO. Acoustics — Measurement of the influence of road surfaces on traffic noise — Part 2: The close-proximity method. The International Organization for Standardization: Geneva, Switzerland, 2017.

12. EN ISO 13473-5; EN ISO. Characterization of pavement texture by use of surface profiles — Part 5: Determination of megatexture. The International Organization for Standardization: Geneva, Switzerland, 2009.

13. EN ISO 532-1; EN ISO. Acoustics — Methods for calculating loudness — Part 1: Zwicker method. The International Organization for Standardization: Geneva, Switzerland, 2017.

14. EN ISO/TS 13471-1; EN ISO. Acoustics — Temperature influence on tyre/road noise measurement — Part 1: Correction for temperature when testing with the CPX method. The International Organization for Standardization: Geneva, Switzerland, 2017.

15. EN ISO/TS 13471-2; EN ISO. Acoustics — Temperature influence on tyre/road noise measurement — Part 2: Correction for temperature when testing with the pass-by methods. The International Organization for Standardization: Geneva, Switzerland, 2022.

16. Engel M., Fiebig, A., Pfaffenbach, C., Fels, J. A., 2021. Review of the Use of Psychoacoustic Indicators on Soundscape Studies. Current Pollution Reports 7, 359–378. https://doi.org/10.1007/s40726-021-00197-1

17. European Union. Directive 2002/49/EC relating to the Assessment and Management of Environmental Noise. Official Journal of the European Communities; 2002. No. L 189.

18. Fastl, H., Zwicker, E., 2007. Psychoacoustics: facts and models. 3rd edition. Springer, Berlin, Germany.

19. Fiebig A., Jordan P., Moshona C.C., 2020. Assessments of Acoustic Environments by Emotions - The Application of Emotion Theory in Soundscape. Frontiers in Psychology 11, 573041. https://doi.org/10.3389/fpsyg.2020.573041

20. Freitas, E., Martins,F., Oliveira, A., Rocha Segundo, I., Torres, H., 2018. Traffic noise and pavement distresses: Modelling and assessment of input parameters influence through data mining techniques. Applied Acoustics 138. 147-155. https://doi.org/10.1016/j.apacoust.2018.03.019

21. Fu, Y., Murphy, D., 2018. Spectral Modelling Synthesis of Vehicle Pass-by Noise. Internoise 2017, 27-30 August, Hong Kong.

22. Geluykens, M., de Barros, A.G., Goubert, L., Vuye, C., 2022. Empirical Study on Temperature Influence on Noise Measurements with the Statistical Pass-By Method. Sustainability 14, 2099. https://doi.org/10.3390/su14042099

23. George, J., Cyril, A., Koshy, B., Mary, L., 2013. Exploring Sound Signature For Vehicle Detection And Classification Using ANN. International Journal on Soft Computing (IJSC) 4(2). https://doi.org/10.5121/ijsc.2013.4203

24. Gille, L., Marquis-Favre, C., Weber, R., 2016. Noise sensitivity and loudness derivative index for urban road traffic noise annoyance computation The Journal of the Acoustical Society of America 140, 4307. https://doi.org/10.1121/1.4971329

25. Grangeiro de Barros, A., Vuye, C., 2023. Psychoacoustic indicators of pass-by road traffic noise [Data set]. Zenodo. https://doi.org/10.5281/zenodo.7904680

26. Guo, Z., Yi, J., Xie, S., Chu, J., Feng, D., 2018. Study on the Influential Factors of Noise Characteristics in Dense-Graded Asphalt Mixtures and Field Asphalt Pavements. Shock and Vibration, 5742412. https://doi.org/10.1155/2018/5742412

27. Kandpal, M., Kakar, V. K., Verma, G., 2013. Classification of ground vehicles using acoustic signal processing and neural network classifier. International Conference On Signal Processing And Communication (ICSC), Noida, India, 512-518. https://doi.org/10.1109/ICSPCom.2013.6719846.

28. Lercher, P., 2019. Combined Transportation Noise Exposure in Residential Areas. Encyclopedia of Environmental Health (Second Edition), 695-712. https://doi.org/10.1016/B978-0-12-409548-9.11280-1

29. Li, T., 2018. Influencing Parameters on Tire–Pavement Interaction Noise: Review, Experiments, and Design Considerations. Designs 2(38). https://doi.org/10.3390/designs2040038

30. Lo Castro, F., Brambilla, G., Iarossi, S., Fredianelli, L., 2018. The LIFE NEREiDE project: psychoacoustic parameters and annoyance of road traffic noise in an urban area. Euronoise 2018 - Conference Proceedings.

31. Midi, H., Sarkar, S.K., Rana, S., 2010. Collinearity diagnostics of binary logistic regression model. Journal of Interdisciplinary Mathematics 13(3), 253-267. https://doi.org/10.1080/09720502.2010.10700699

32. Morel, J., Marquis-Favre, C., Dubois. D., Pierrette, M., 2012. Road Traffic in Urban Areas: A Perceptual and Cognitive Typology of Pass-By Noises. Acta Acustica United With Acustica 98, 166-178.

33. Morel, J., Marquis-Favre, C., Gille, L.A., 2016. Noise annoyance assessment of various urban road vehicle pass-by noises in isolation and combined with industrial noise. Applied Acoustics 101, 47–57. https://doi.org/10.1016/j.apacoust.2015.07.017

34. Moreno, R., Bianco, F., Carpita, S., Monticelli, A., Fredianelli, L., Licitra, G. Adjusted Controlled Pass-By (CPB) Method for Urban Road Traffic Noise Assessment. Sustainability. 2023; 15(6):5340. https://doi.org/10.3390/su15065340

35. Paviotti, M., Vogiatzis, K., 2012. On the outdoor annoyance from scooter and motorbike noise in the urban environment. Science of the Total Environment 430, 223–230. https://doi.org/10.1016/j.scitotenv.2012.05.010

36. Raggam, R., Cik, M., Holdrich, R., Fallast, K., Gallasch, E., Fend, M., Lackner, A., Marth, E., 2007. Personal noise ranking of road traffic: Subjective estimation versus physiological parameters under laboratory conditions. International Journal of Hygiene and Environmental Health 210, 97–105. https://doi.org/10.1016/j.ijheh.2006.08.007

37. Roberts, C., 2010. Low Frequency Noise from Transportation Sources. Proceedings of 20th International Congress on Acoustics, ICA 2010, Sydney, Australia.

38. Ríos, A., Simpson, J., 2017. A sequential augmentation method to eliminate multicollinearity. Quality Engineering 29(4), 588-604. https://doi.org/10.1080/08982112.2016.1258474

39. Rychtáriková, M., Vermeir, G., 2013. Soundscape categorization on the basis of objective acoustical parameters. Applied Acoustics 74, 240–247. https://doi.org/10.1016/j.apacoust.2011.01.004

40. Sandberg, U., 2003. The Multi-Coincidence Peak around 1000 Hz in Tyre/Road Noise Spectra. Euronoise 2003, 19-21 May 2003, Naples, Italy.

41. Sandberg, U.; Ejsmont, J. A., 2002. Tyre/Road Noise Reference Book. Informex.

42. Soares, F., Freitas, E., Cunha, C., Silva, C., Lamas, J., Mouta, S., Santos, J. A., 2017. Traffic noise: Annoyance assessment of real and virtual sounds based on close proximity measurements, Transportation Research Part D: Transport and Environment 52, 399-407. https://doi.org/10.1016/j.trd.2017.03.019

43. Stansfeld, S., Clark, C., Smuk, M., Gallacher, J., Babisch, W., 2021. Road traffic noise, noise sensitivity, noise annoyance, psychological and physical health and mortality. Environmental Health 20, 32. https://doi.org/10.1186/s12940-021-00720-3

44. Winroth, J., Kropp, W., Hoever, C., Beckenbauer, T., Männel, M., 2017. Investigating generation mechanisms of tyre/road noise by speed exponent analysis. Applied Acoustics 115, 101-108, https://doi.org/10.1016/j.apacoust.2016.08.027

45. WHO, 2018. Environmental Noise Guideline for the European Region. Copenhagen, Denmark.

46. 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

47. Zhou, T., Zhang, M., Li, C., 2015. A Model for Calculating Psychoacoustical Fluctuation Strength. Journal Audio Engineering Society 63 (9), 713-724. https://doi.org/10.17743/jaes.2015.0070