Urban sound monitoring beyond dB(A):

Insights from a city-wide sensor deployment

L. Cassens et al.

Abstract

Conventional urban noise monitoring relies on aggregated A-weighted sound pressure levels, which provide limited insights into the present sound sources and temporal structure of the sound environment. This study presents results from a city-wide deployment of 38 low-cost, source-aware sound sensors in Delft, the Netherlands, accompanied by a survey on perceived sound prevalence and annoyance. The sensors measured loudness levels, intermittency ratios (IR), and the presence of 11 sound event classes, as predicted by an on-device machine learning model.

The results reveal substantial spatial and temporal heterogeneity in the urban sound environment. Median loudness and intermittency varied strongly across locations, indicating that they represent complementary acoustic dimensions. Sound event presence ratios and causes of intermittency often differed, highlighting the importance of distinguishing between overall source presence and their presence in form of disruptive, intermittent events. Correlation analyses show that louder and event-like detections of vehicles and sirens are more strongly associated with reported annoyance than overall presence ratios. While source-specific intermittent detections relate to annoyance, the global IR did not correlate with perceived short-burst annoyance, suggesting that aggregated intermittency metrics may not directly translate into subjective experience.

The findings demonstrate that source-aware, edge-based sensor networks provide insights beyond conventional dB(A)-based monitoring and offer a scalable and privacy-friendly approach for urban sound monitoring. The deployment and analysis provided valuable details about Delft's soundscape. Future research should validate perceptual associations in larger samples and evaluate potential inaccuracies of machine learning-based sound event detections.

1. Introduction

The urban sound environment profoundly influences the physical health [1], psychological well-being [2], and everyday behaviour [3] of urban residents. It shapes how people experience public spaces, how they interact with one another, and how liveable urban areas are perceived to be [4,5]. Sound environments emerge from complex interactions among multiple sound sources, including road traffic, construction sites, social human activity, and natural sounds. A deep understanding of both the temporal and spatial composition of these sound environments, and of how they are perceived by citizens, is therefore crucial for developing effective policies to mitigate noise pollution and promote healthier and more pleasant cities [6–8].

The type and acoustic characteristics of a sound strongly influence how individuals interpret and respond to the sound [7]. For example, a lively public square characterised by conversation and occasional bird calls may be experienced as vibrant and attractive, whereas a street dominated by traffic noise may be perceived as annoying, even if the overall sound energy is higher in the former. These differences reflect not only contextual meaning but also distinct spectral, temporal, and source-related properties of the acoustic environment. Such variations underline the importance of detailed, fine-grained information on the urban sound environment and its relationship to human perception and annoyance.

Noise sensors are a central tool for monitoring urban sound environments. These devices are widely used to assess the acoustic impacts of major noise sources such as traffic and aviation [9]. Conventional sensors primarily measure sound energy in A-weighted decibel levels (dB(A)), a metric that approximates the frequency sensitivity of human hearing [10]. For the purpose of policy evaluation and planning, such measurements are commonly aggregated over time (e.g., yearly) to construct noise maps. Within the European Union, environmental noise is routinely assessed using the indicator [11], an averaged day-evening-night level that applies penalties to evening and night-time noise to reflect their greater potential for annoyance and health [12].

Recent advances in sensor technology have substantially expanded the capabilities of urban sound monitoring. Acoustic sensors are no longer limited to measuring loudness levels; they can also detect and classify the specific sound sources that contribute to the sound environment [9–11]. This additional information is valuable for several reasons. First, effective policy-making requires an understanding of which sources are responsible for noise problems, not only how loud an area is. Interventions to reduce traffic noise, for example, differ fundamentally from those aimed at managing nightlife. Second, people's response to sound depend partly on its source and context [7,12]. Research shows that individuals may be more sensitive to certain sources than to others [13,14], and that noise annoyance is influenced by factors such as perceived necessity, controllability, and personal agency [15–17]. Being able to identify and quantify the presence of specific sound sources, therefore, provides a richer basis for assessing and designing targeted mitigation policies.

Despite these technological advances, a deep understanding of the temporal and spatial composition of urban sound environments, and of how these environments are experienced by citizens, remains limited. To date, large-scale deployments of source-aware noise sensors in real urban settings have been rare [11]. Notable exceptions include the CENSE project [18], which deployed low-cost sensors to differentiate traffic, bird song, and speech, and the SONYC project in New York City [19], which also utilized machine-learning for sound classification. While these initiatives demonstrated the technical feasibility of city-wide monitoring, their analyses have focused primarily on methodological development and operational performance. As a result, in-depth empirical insights into the fine-grained spatial and temporal composition of urban sound environments, and their implications for human perception, are still largely lacking.

This study aims to address this gap by providing detailed empirical insights into the urban sound environment and its perception by residents. We examine the spatial and temporal distribution of sound events, loudness patterns, and their relationship with noise annoyance in the city of Delft, a medium-sized city in the Netherlands. To this end, we deployed 38 low-cost, machine-learning-based noise sensors capable of detecting eleven distinct sound sources, as well as measuring intermittency ratios, acoustic sharpness, and A-weighted sound pressure levels. The sensors were installed using a citizen science approach with broad coverage of the city. To capture perceptual responses, we conducted a noise annoyance survey among local residents. In this paper, we present the results of this data collection effort, focusing in particular on the temporal and spatial patterns of urban sound and their relationship to human perceptions of noise annoyance.

This study extends the current literature in three ways. First, it provides detailed empirical insights into the sound environment of a typical medium-sized Dutch city. Second, it demonstrates how sensor data on sound events can be used to analyse urban sound environments beyond aggregated dB(A) levels, thereby offering a richer and more perceptually meaningful representation of urban noise. Third, by linking objective sensor measurements to reported noise annoyance, the study contributes to a better understanding of how specific sound sources and acoustic characteristics relate to human experience in real-world settings.

The remainder of this paper is organised as follows. Section 2 describes the study area, sensor deployment, survey, and data collection. Section 3 presents the combined results and discussion, including intermittency patterns, dominant sound events and other spatial and temporal sound event pattern. The paper concludes with recommendations for sound-event-aware monitoring and future research.

2. Methods

We deployed 38 low-cost open hardware sound sensors in and around Delft, The Netherlands. In this section, we first discuss the research area and deployment. After that, we describe the technical details of the deployed sensor. Finally, we discuss the survey that was administered to the residents who deployed the sensors at their premises.

2.1 Research area and participants

Delft is a mid-sized Dutch city, located between The Hague and Rotterdam. The city has a population of around 104,000 people spread over 24 km². A significant portion of its population are students due to the presence of Delft University of Technology. The university campus is located south of the old town and is the largest campus in the Netherlands. The old town itself is largely pedestrianised, comprising multiple small canals and historic buildings. Many cafes and restaurants are located in this historic centre. Delft is adjacent to the highway A13 in the northeast and the A4 in the southwest. The presence of a lively student population, a pedestrianised city centre, and major roads leads to various distinct sound environments. This diversity makes Delft an interesting case study for a sensor deployment focused on sound event detection. For comparison, we included the town of Pijnacker in our research area. Pijnacker is situated 5 km east of Delft and has 27,000 residents. The sensors in both places are depicted in Figure 1.

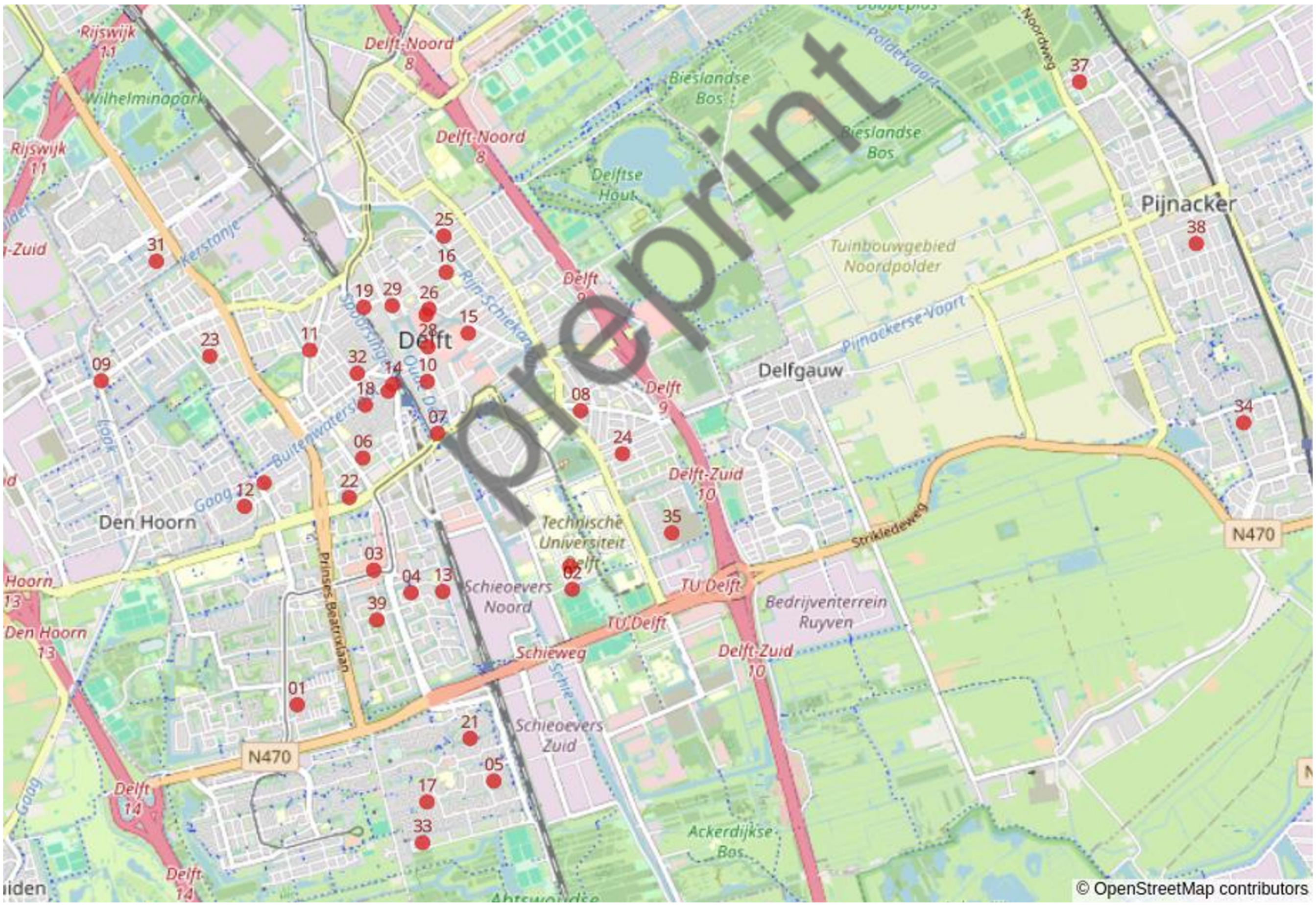

Figure 1: Sensor locations in the research area consisting of Delft and Pijnacker

We invited residents of both cities to volunteer to host sound sensors provided and installed by the authors. In return, each volunteer got a brief report about their measurements. The project was advertised through the university's communication channels and the authors' private networks. This led to 39 participants, one of whom was excluded from the analysis due to a hardware failure. While this citizen science approach provides less control over the sensor locations, no clear skew in the volunteers' locations was observed (see Figure 1 for the sensor locations). Five were located in the pedestrianised old town, 30 in other residential areas of Delft, and 3 in the north, centre, and south of Pijnacker. The sensors vary in their installation heights, with most deployed on the ground or first floor. Most sensors faced the street, while some faced a backyard. In the backyard facing cases, the residents indicated that they spend a significant amount of time at those locations.

The sensors were deployed between May and October 2025. This period granted favourable weather conditions for the solar-powered sensors. Furthermore, more human activity was expected during the Summer, resulting in a more diverse soundscape. The exact deployment duration varies slightly between locations, depending on the availability of the citizens hosting the sensors. Due to a software bug, the sound event prediction was only operational for an average of six weeks, while loudness was monitored during the whole deployment duration. Furthermore, each sensor may reduce its sampling ratio based on the available solar power and battery charge. The sampling ratio is defined as the duration during which the sensor is monitoring the sound environment as a fraction of the total time passed. The ratio is adjusted once per day. The average sampling ratio across all sensors was 95%, meaning the sensors monitored, on average, 57 minutes per hour. The lowest average sampling ratio for a sensor was 77 %.

2.2 Sensor and measurements

We utilised low-cost open hardware sensors [20] for the deployment (see Figure 2). These sensors have two important advantages over traditional sound level meters: Firstly, they can detect sound events at the edge (meaning no audio or spectral information needs to be permanently recorded). The open-source nature of the sensors provides full transparency and control over data collection. Secondly, deployment costs are very low, with a unit cost of 55 EUR. Thus, such low-cost sensors can be deployed at a much larger scale. This specific deployment incurred hardware costs below 2,500 EUR. While traditional, industrial sound level meters provide more accurate loudness measurements, no more than one or two locations could have been covered within a similar budget.

Figure 2: Image showing one of the deployed sound sensors on a balcony facing an intersection.

The sensors measured the following sound metrics relevant to the human perception of sound: A-weighted sound levels, presence of eleven different sound events, and acoustic sharpness. Below, each of these metrics is explained individually.

A-weighted sound levels were measured eight times per second. Humans perceive very low and very high frequencies as quieter than a 1 kHz tone with equivalent sound energy [21]. A-weighting corrects for the loudness perception through frequency weighting. Various definitions for loudness exist in the literature, but for the remainder of this article, loudness refers to the A-weighted sound levels.

The high temporal resolution for loudness allows the calculation of intermittency ratios (IR). The IR reflects the fraction of the total sound energy attributable to loud, distinct events [9]. Thus, a low IR indicates that a sound environment is dominated by constant background noise. Contrary, a high IR indicates that most sound energy originates from events well above the median noise level. The IR was calculated after the deployment. The exact calculation method will be explained further below.

The presence of sound events was predicted in 3-second intervals on the sensor. Edge-processing is necessary to preserve citizens' privacy, as audio may not be permanently recorded in public. For each 3-second interval, the sensors' machine learning model may deem none, one, or multiple sound events present. The sensor applies short-term fast Fourier transforms followed by a conversion from uniform frequency bins to the non-linear Mel frequency scale. The Mel scale is more representative of human perception [22] and reduces the frequency resolution, thereby improving the sensor's efficiency. The resulting 3-second Mel spectrograms are then processed by a convolutional neural network [23], which classifies the sound fragment. Specifically, the model distinguishes between 11 sound events: vehicle noise (such as cars and scooters), honking, aviation noise (from airplanes and helicopters), sirens (such as those from emergency vehicles and air raid alarms), speech and crowd noise (e.g., individual or groups talking or screaming), dog barking, bird song, ringing of church bell, music, wind induced noise (as an artifact from wind-microphone interaction and wind interacting with the environment), and lastly, rain induced noise (e.g. rain hitting surfaces such as the street or the sensor itself). The classification is not mutually exclusive, meaning the model is capable of detecting multiple simultaneous sound events, such as vehicle noise combined with honking.

Acoustic sharpness indicates how much of the sound is caused by high-frequency components as opposed to low-frequency components. Higher sharpness is associated with an increased annoyance response [24]. Due to computational constraints, the sensor calculates a simplified version of acoustic sharpness, which is not fully compliant with common standards such as Zwicker's sharpness method. Thus, acoustic sharpness should not be interpreted in absolute values, but rather comparatively between different sensors or over time. For further information about the sensor, we refer to the dedicated papers [10,23] .

2.3 Noise annoyance questionnaire

The noise questionnaire, given to the citizens who hosted the sensors, was distributed after deployment and asked about people's perceptions of the sound environment throughout the deployment period. Specifically, participants were asked how common certain sounds were, using a 5-point Likert scale. They were also asked which sound sources annoy them, following the ICBEN standard, which also uses a 5-point Likert scale [25]. On a similar scale, participants were then asked which sound characteristics contributed most to the annoyance. These characteristics included the overall volume of noise, short but loud sound bursts, low-frequency noise, high-frequency noise, and noise during the day/night.

2.4 Analysis

To determine how common certain sound events are across the city, we quantify spatial patterns in sound event detections obtained from the sensor network. Specifically, sound event presence ratios are calculated and interpreted in the context of Delft. The machine learning model for sound event detection runs directly on the sensors and outputs, for each 3-second sound fragment i at sensor s, a probability $p_{s,i,e} \in [0,1]$ for each sound event class e. A sound event is considered detected if its probability exceeds 0.5:

$$d_{s,i,e} = 1[p_{s,i,e} > 0.5] \quad (1)$$

The sound event presence ratio $\mathrm{PR}_{s,e}$ is defined as the proportion of sound fragments at sensor s in which sound event e is detected, relative to the total number of fragments N_s:

$$\mathrm{PR}_{\mathrm{s,e}} = \frac{1}{N_s}\sum_{i=1}^{N_s} d_{s,i,e} \tag{2}$$

Thus, this metric captures how frequently a given sound event occurs at each location over the observation period.

To analyse typical daily activity patterns of sound events, we calculate diurnal sound-event presence ratio curves. For each minute of the day, sound fragments recorded during that minute are aggregated across all sensors and days. The diurnal presence ratio curve can be defined as:

$$\mathrm{DPR}_{\mathrm{e}}(m) = \frac{1}{|I_m|}\sum_{(s,i)\in I_m} d_{s,i,e} \tag{3}$$

Where $m \in \{1, \dots, 1440\}$ denotes the minute of the day, I_m the set of all sound fragments recorded during minute m, and $|I_m|$ the number of fragments in that set.

The resulting diurnal curves represent the average probability of detecting each sound event at a given time of day and can be interpreted as typical daily activity patterns. The curves, therefore, provide insights into when specific sound sources are most common at the city level.

To characterise loudness patterns at each sensor location, we consider three dimensions: median loudness, day–night cycle, and the intermittency ratio (IR). Median loudness is represented by aggregated A-weighted sound pressure levels and is used instead of the mean to reduce sensitivity to outliers. The day–night cycle captures systematic differences in sound levels between day, evening, and night-time. It represents the median loudness at a location for each hour of the day.

The intermittency ratio (IR) reflects the extent to which sound levels are dominated by distinct events rather than relatively constant background noise [26]. High IR values indicate environments where short-term peaks contribute substantially to overall sound levels, whereas low IR values suggest more continuous sound conditions. The IR has been calculated in accordance with [26]. To account for systematic differences in sound levels over the course of the day, the background level was estimated hourly. This avoids underestimating intermittency during periods with lower sound levels. A sound event was classified as intermittent if its sound energy surpasses the background noise level by 10 dB(A).

The final part of the analysis investigates how the sensor-based measurements relate to people's perception. For this, sensor metrics are linked to responses from the noise annoyance survey. First, we assess the relationship between the perceived commonness of sound events and their corresponding presence ratios from the sensors. Second, we analyse how noise annoyance from various sound sources corresponds to measured presence ratios of these sources. Finally, we examine how median loudness and intermittency ratios are associated with perceived noise annoyance.

3. Results and discussion

This section presents the empirical findings and discusses their implications. We first provide an overview of the main acoustic metrics and associated findings. We then examine additional spatial and temporal patterns in sound event detections. Next, we assess intermittency and the strength of the day–night cycle as dimensions complementary to overall loudness. Finally, we relate the measured metrics to people's perception as reported in the survey.

3.1 Overview of important metrics per sensor

Table 1 contains the median dB(A), day-night pattern, IR, causes of intermittency, and presence ratios for four sound events. The day-night pattern shows the median loudness per hour over all days, split by day (07:00 - 19:00, yellow), evening (19:00 - 23:00, orange), and night-time (23:00 - 07:00, purple). The causes of intermittency are presented as a bar chart showing the ratio of sound events that cause intermittency, as defined by the IR (i.e., 10 dB above background levels). Thus, the bar chart shows which sources produce the most intermittent sounds at a given location. The first bar represents traffic noise (blue), followed by human noise (orange), bird song (green), and lastly all remaining sound events (grey). The sound event presence ratios represent how much of the time a sound event is present and include all events, not just the intermittent ones. For traffic, only events above 60 dB(A) are considered. For the geographic location of each sensor, please refer to Figure 1.

Median loudness ranges from 37.7 to 59.0 dB(A). Three of the five loudest places are directly adjacent to major roads (3, 7, 14), while the remaining two are in the pedestrianised old town (28) and a student housing complex (22). The three loud roadside locations have the highest presence ratio for traffic noise above 60 dB(A), while sensors in the pedestrianised old town express and the student housing complex show high ratios of crowd noise. Traffic is typically understood to be a major noise source [27], but the sound levels observed in the pedestrianized old town indicate that human activity can produce sound levels comparable to traffic noise. The quietest locations exhibit low presence ratios of both traffic and crowd noise.

The intermittency of the sound environment varies substantially between locations. The lowest IR (5%) was measured at sensor 2, installed on the 16th floor of a residential building. The elevated position increases the distance to most sound sources, which reduces their relative contribution of short-term peaks compared to background levels. Sensors 28, 34, and 38 have the highest IRs (all above 85%), indicating that a large share of the sound energy at these locations comes from intermittent events at least 10 dB above the background level.

The dominant causes for the most intermittent locations vary, between traffic (28, 32) and speech and other sources (38). In general, substantial variations exist in causes of intermittency, with many sites being dominated either by traffic or bird song, while other sound environments are interrupted by human speech or other sources. Notably, of the eleven locations where bird song was the dominant cause of intermittency, only one sensor faced a street. The other sensors were facing gardens or backyards. Consequently, gardens and backyards appear to have a substantially different sound environment than residential streets, potentially allowing biological sounds to dominate the acoustic environment. The distinction between traffic-dominated and bird-dominated intermittency suggests that similar IR values may reflect fundamentally different acoustic qualities, which may have different perceptual implications.

Table 1: Overview of important metrics for each measurement location.

					Sound event presence ratios			
ID	Median dB(A)	Day-night cycle	IR (%)	Causes of Intermittency	Traffic >= 60 dB(A)	Speech and crowd	Bird song	Church bells
01	48.27		21.16		13.08	7.61	14.40	0.03
02	52.74		5.30		38.73	8.66	12.73	0.05
03	58.96		22.49		82.93	1.65	5.08	0.10
04	44.35		15.94		19.09	10.01	15.33	0.08
05	39.86		25.77		22.23	5.88	17.19	0.03
06	45.65		37.98		48.22	10.00	12.91	0.10
07	58.38		23.51		81.35	5.63	5.41	0.02
08	45.14		22.63		7.77	8.20	22.14	0.35
09	48.62		21.22		38.65	5.15	18.16	0.07

10	40.89		21.06		6.99	10.07	13.00	1.76
11	49.42		30.11		56.67	7.31	22.84	0.01
12	43.03		22.55		11.61	11.54	9.40	0.10
13	44.26		14.70		13.17	5.63	37.39	0.12
14	56.19		29.69		87.94	8.52	4.40	0.10
15	50.82		26.37		16.68	48.74	3.36	1.32
16	44.21		27.77		22.81	14.82	8.90	1.55
17	44.69		20.09		19.98	5.80	18.69	0.03
18	38.33		26.69		11.30	12.06	6.40	0.41
19	41.32		30.72		32.50	25.61	6.90	0.07
20	41.25		22.16		8.01	11.43	8.65	0.16
21	42.51		23.25		33.49	6.18	27.15	0.04
22	56.24		18.57		39.70	32.06	8.13	0.10
23	43.73		19.52		20.50	6.65	18.08	0.03
24	42.95		19.32		15.13	9.24	31.26	0.07
25	43.26		20.47		9.19	9.05	17.37	0.11
26	52.86		34.28		22.74	56.39	7.84	1.31
27	44.68		19.39		2.34	10.03	3.29	3.34
28	53.57		49.45		27.54	49.49	7.14	2.93
29	37.70		30.35		1.60	17.12	28.13	1.08
30	48.58		17.49		13.82	8.09	3.45	0.09
31	45.42		25.41		41.72	6.70	55.67	0.07
32	40.94		20.67		14.41	19.28	6.26	0.89
33				ignored due to malfunction				
34	40.05		47.08		32.63	18.66	20.76	0.09
35	46.00		25.92		16.52	3.98	5.35	0.04
36	51.64		19.93		31.93	1.96	28.48	0.02
37	44.02		33.84		32.69	7.93	28.00	0.02
38	39.81		41.72		8.06	14.26	9.74	0.24
39	49.67		14.25		3.60	7.48	42.80	0.13
All	44.69		24.97		26.51	13.39	16.11	0.45

Note: The day night cycle shows the median dBA per hour over all days, color-coded based on day (yellow), evening (orange), and night time (purple), as defined by L_{den}. The causes of intermittency show the ratio of intermittent events for traffic (blue), human noise (orange), bird song (green), and the sum of other sources (grey).

3.2 Sound event presence ratios

Figure 3 shows the average sound event presence ratios over all sensors. Traffic noise is present 49% of the time, followed by bird song (16.1%) and human speech (13%). Although traffic is the most prevalent sound source, it was not always the dominant driver of intermittency as seen above. It is therefore important to distinguish between the mere presence of a source and the intermittent disruption of the environment by a source. Given the overall low presence ratio for bird song, the number of locations where it caused most intermittency is remarkable.

Aviation was detected 6.5 % of the time. This rate is likely inflated by acoustic aviation and road traffic sounds, leading to potential false positives in the classification model. Therefore, aircraft detections were excluded from the results unless mentioned otherwise. Wind and rain detections are also excluded, as they primarily reflect weather-related microphone interactions rather than persistent urban sound sources. The remaining sound event types were rare on average but can occur more frequently at specific locations, such as church bell sounds near churches.

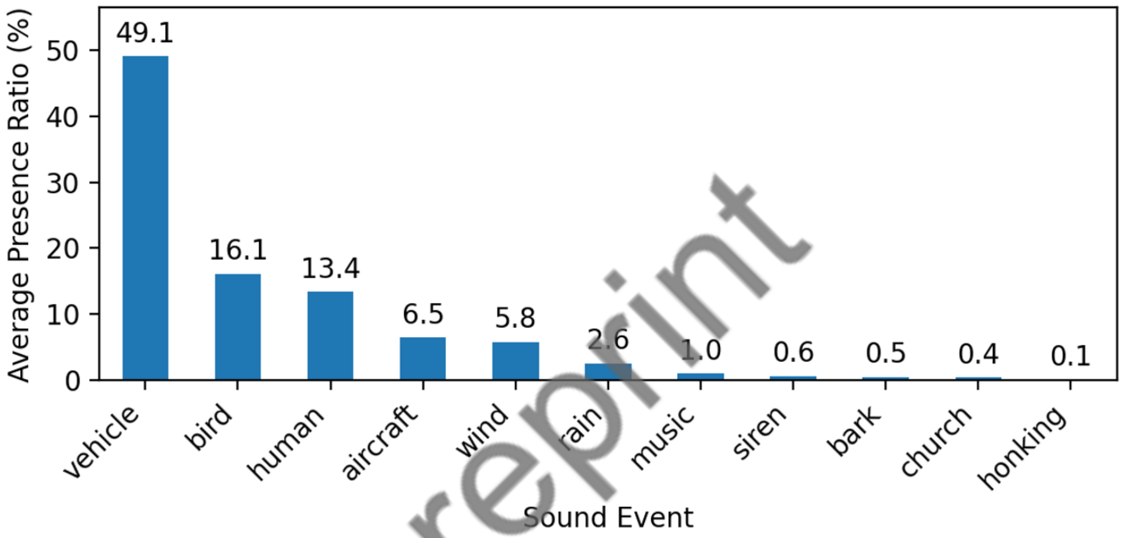

Figure 3: Average sound event presence ratios (%) aggregated across all sensors. The values represent the proportion of 3-second sound fragments classified as the respective event.

3.2.1 Spatial sound event patterns

The presence of traffic noise along major roads and the concentration of speech noise in the city centre were already discussed in relation to the overview of important metrics. Next, we discuss additional spatial observations of sound event presence ratios.

Church bells are predominantly detected in the historic city centre of Delft (Figure 4), where several churches are located. Although church buildings are also present in other neighbourhoods, they did not result in substantial detections by the sensors' machine learning model. Thus, church bell ringing emerges as a distinctive and spatially concentrated feature of Delft's centre.

Dog barking does not exhibit a clear spatial clustering pattern and appears to be highly localised. This indicates that such sounds are largely dependent on hyperlocal conditions rather than broader neighbourhood characteristics (e.g., residents owning a dog in the immediate vicinity). Bird song also shows a strong dependence on the local environment. The sound energy of bird calls is considerably lower than that of other sources, such as traffic or church bell ringing. Thus, bird song is likely to be detected only in close proximity to the source. As already mentioned, nearby gardens or the absence of other sound sources may increase the amount of audible bird song.

Sirens are present along certain major streets. Music was primarily present in the city centre and the student housing complex. Honking followed no clear pattern in our data, which may be partly due to the very low overall presence ratio of the sound.

Overall, these spatial patterns show that, beyond the dominance of traffic noise, several sources are strongly shaped by the urban environment and micro-environmental conditions. While some sound events follow urban infrastructure (e.g., roads), others seem to correlate with human activities, such as crowd noise, which may be caused by shopping or nightlife. These spatial differences highlight the heterogeneous nature of the urban acoustic environment.

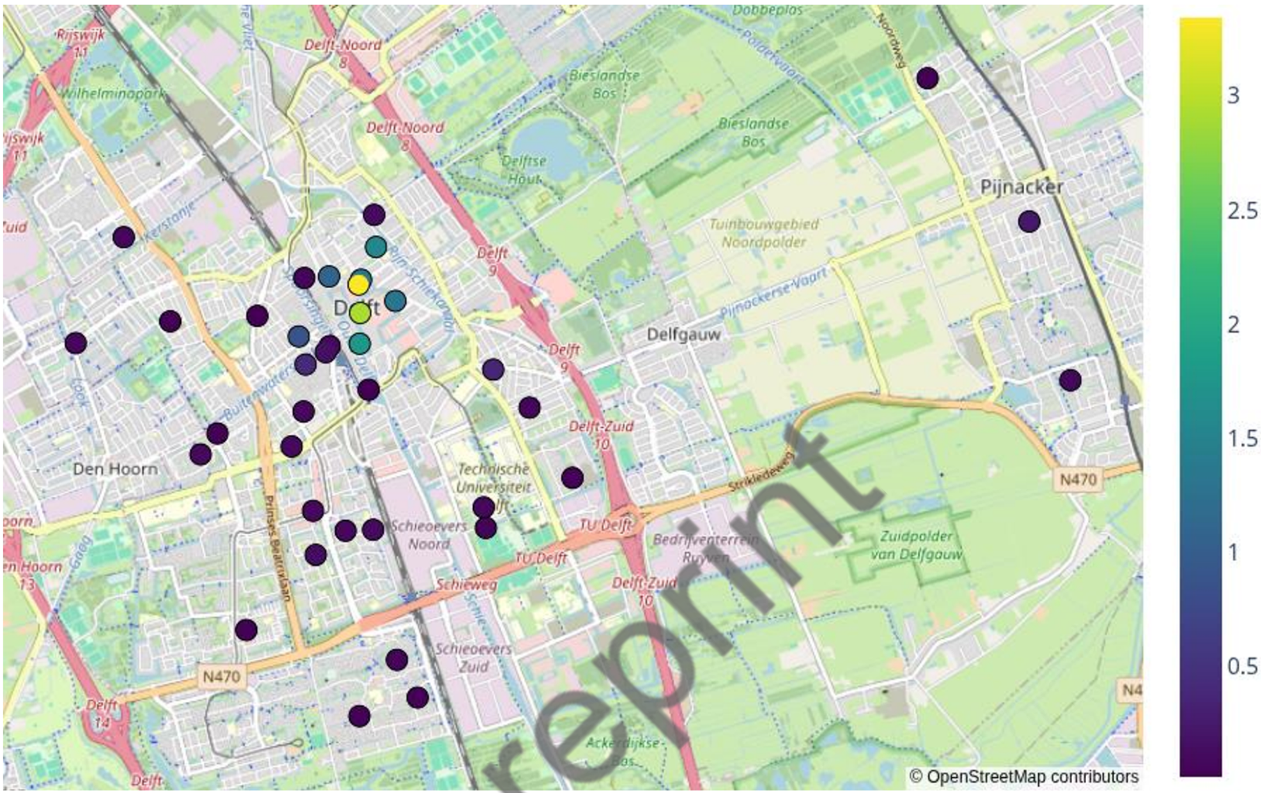

Figure 4: Spatial distribution of church bell presence ratios (%) across all sensors.

3.2.2 Temporal sound event patterns

Moving from spatial to temporal patterns (Figure 5), clear, highly structured dynamics emerge in church bell activity. Pronounced peaks occur at each full hour. During the day, smaller but consistent peaks occur additionally each half hour. The magnitude of the hourly peaks increases throughout the morning, reaching a maximum at 12:00. Afterwards, the intensity drops and the pattern repeats until midnight. In Delft, the number of bell strikes corresponds to the hour of the day, resulting in the highest number of strikes at 12:00 (midday and midnight). The regularity and precise timing of these detections also confirm the sensor's ability to correctly classify church bell events, as the observed temporal pattern matches the known ringing practice in the city.

The diurnal traffic noise pattern shows a higher detection frequency during the night than during the day. At first glance, this appears to contradict established traffic intensity patterns, which typically exhibit morning and evening rush-hour peaks. However, the presence ratio reflects the proportion of sound fragments classified as traffic, rather than the amount of traffic or loudness of traffic. Instead, the results suggest that traffic constitutes an omnipresent urban background sound: During daytime and evening hours, distant traffic noise may be overshadowed by louder, nearby sources. Thus, traffic noise presence ratios drop as they only capture traffic noise not masked by other sounds. At night, when competing sources such as human speech diminish, more of the background traffic noise may be picked up, leading to higher traffic noise presence ratios. This omnipresence of traffic complicates its quantification, as it remains unclear whether faint background traffic should trigger a sound event detection and how this could be implemented.

Human crowd noise peaks around 8 p.m., coinciding with lower traffic noise detections and thereby supporting the previously discussed masking or dominance effect. More generally, human speech is predominantly present during daytime and evening hours and least prevalent during the late night and early morning. Bird song detections match expected bird activity patterns, with a strong peak around 5 a.m. and a second, less pronounced peak around 8 p.m. This is consistent with expected foraging behaviour. Diurnal patterns for honking, sirens, and rain (not shown) exhibit less clear patterns. A notable exception is the monthly national air-raid siren test, which was clearly detected as a siren event, with a median loudness of 74 dB(A) and a maximum of 102 dB(A).

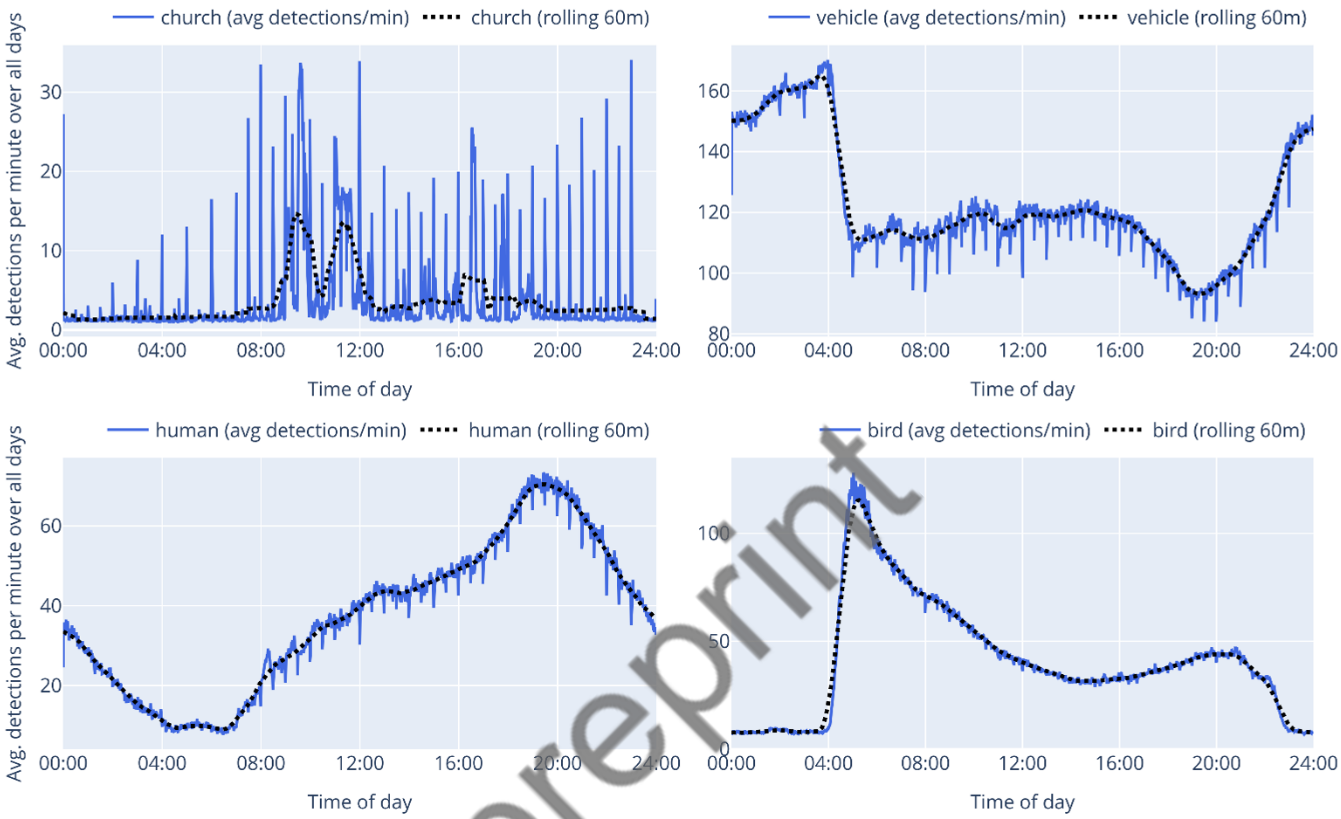

Figure 5: Diurnal presence ratios for church bells, vehicles, human speech and crowd noise, and bird song. Solid lines represent the average number of detections per minute over all sensors and days, while dashed lines show a 60-minute rolling average.

3.3 Intermittency, loudness, and strength of the day-night cycle

As outlined in the overview of key metrics, intermittency ratios vary substantially across locations and are driven by different dominant sound sources. We now examine how the IR relates to loudness levels. Figure 6 plots the average loudness against the average IR per sensor. No systematic relationship is apparent; high or low loudness does not necessarily coincide with high or low intermittency. Little correlation between the two concepts was already assumed by the authors of the IR based on simulations, which is now confirmed by measurements.

Another dimension of the sound environment related to loudness is the strength of the day-night cycle. Figure 7 illustrates examples of strong and weak day-night contrasts for both loud and quiet environments. This dimension is perceptually relevant, as humans are more vulnerable to noise during the evening and night-time [28], a principle reflected in established noise indicators such as Lden. Moreover, residents may not be present at home during daytime hours, meaning that similar average loudness levels can have different experiential implications depending on their temporal distribution. The strength of the day–night cycle, therefore, provides an additional explanatory dimension of the sound environment besides overall loudness and intermittency.

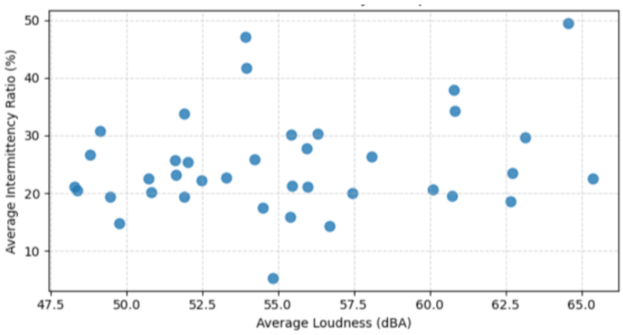

Figure 6: Scatter plot of the average loudness (dB(A)) and average intermittency ratio (IR, %) per sensor location. Each point represents one sensor.

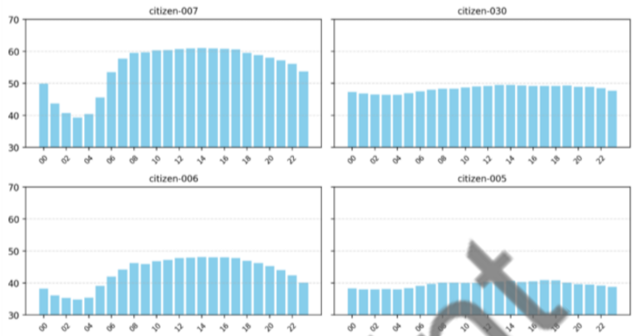

Figure 7 Median hourly loudness (dB(A)) for selected sensors, illustrating strong and weak day–night differences in both loud and quiet environments.

3.4 Measurements in relation to people's perception

Table 2 presents the Spearman correlation coefficients between sensor detections and the surveyed prevalence of perceived sounds. Correlations are reported for detected events regardless of their loudness, those with an average sound level exceeding 60 dB(A), and detections restricted to intermittent events. Given the small sample size, levels of statistical significance are of limited value; instead, we focus on the direction and magnitude of the correlations.

Most sound sources exhibit weak to moderate correlations, with the exception of aviation noise and siren sounds. Sensor detections and perceived aviation noise are only correlated when considering only loud events or intermittent events. In contrast, the correlation for siren sounds becomes weaker when the analysis is restricted to louder events and disappears entirely for intermittent even detections. For most other sources, correlations increase when restricted to louder or intermittent events. This pattern is intuitive, as louder and event-like sounds are more likely to be perceptually salient and to penetrate building insulation than quieter background noise.

Table 2: Spearman correlations between perceived sound source prevalence and sensor-based detections (all, loud, and intermittent events).

Sound source	**Spearman's ρ**	**Spearman's ρ (dBA > 60)**	**Spearman's ρ (intermittent events)**
Vehicles	0.21	0.42*	0.48**
Aviation	0.00	0.26	0.34
Speech and crowd noise	0.23	0.32	0.24

Bird song	0.58***	0.59***	0.6***
Sirens	0.18	0.08	0

*Note: *0.05 significance level. **0.05 significance level. ***0.001 significance level.*

Table 3 presents the Spearman correlation coefficients between reported annoyance and (i) surveyed sound source prevalence, (ii) all sensor detections, (iii) loud events, and (iv) intermittent events. The strongest correlations with annoyance are observed for perceived source prevalence. However, a substantial portion of the variance in annoyance remains unexplained. We consider this relationship as a de facto upper limit for sensor-based correlations with annoyance, as it is unlikely that individuals report annoyance from a source they do not perceive to be present. Consequently, correlations between sensor detections and annoyance are expected to be below these limits, which is reflected in the results.

Comparing correlations across detection types indicates that louder events are generally more strongly associated with reported annoyance than overall presence ratios. For vehicles, the negative correlation observed for overall presence ratios becomes positive when restricting the analysis to louder or intermittent events, suggesting that annoyance is primarily driven by salient traffic events rather than background traffic presence. A similar pattern is observed for sirens, where intermittent detections show the strongest association with annoyance. In contrast, aviation and speech-related detections show only modest associations across detection types. Overall, these results suggest that event-like and acoustically prominent occurrences are more relevant for annoyance than the mere frequency of detected sound fragments.

Table 3: Spearman correlations between reported annoyance and perceived sound source prevalence and sensor-based detections (all, loud, and intermittent events).

Sound source	**Spearman's ρ for annoyance and perceived presence**	**Spearman's ρ for annoyance and presence ratios**	**Spearman's ρ for annoyance and presence ratios above 60 dB(A)**	**Spearman's ρ for annoyance and intermittent events**
Vehicles	0.55**	-0.10	0.32	0.27
Aviation	0.49**	0.15	0.15	0.05
Speech and crowd noise	0.66***	0.26	0.25	0.17
Sirens	0.63***	0.05	0.15	0.35

*Note: *0.05 significance level. **0.05 significance level. ***0.001 significance level.*

The survey further asked respondents which acoustic characteristics they perceived as contributing to their annoyance, including overall sound levels and short, distinct sound events. Examining the relationship between the measured intermittency ratio (IR) and reported annoyance from short, distinct sounds reveals a negligible and statistically insignificant correlation ($\rho = 0.04$), indicating no meaningful association in the present sample. In contrast, perceived overall volume shows a moderate positive and statistically significant correlation with the measured median loudness ($\rho = 0.37$).

At first glance, the absence of a relationship between IR and reported short-burst annoyance appears to contrast with the source-specific findings above, where intermittent traffic and siren detections show meaningful associations with annoyance. This suggests that while annoyance is influenced by intermittent events, citizens may not explicitly attribute their annoyance to “short, distinct sounds” as an abstract characteristic. Alternatively, the IR may not fully capture the perceptual dimensions underlying disruptive sound events.

Overall, the survey results indicate that perceived source prevalence correlates more strongly with annoyance than sensor detections. Louder events correlate more strongly with annoyance than the overall ratio of events; further research with a larger sample size is advised to confirm these effects and to further investigate the role of intermittency in noise annoyance.

Considering the deployment results as a whole, our findings relate to previous research in various ways. Loudness patterns over the course of the day are congruent with known day-night differences, with noise levels generally lower during nighttime [29]. We further observed that the strength of this day–night cycle can vary substantially between locations, a phenomenon also reported in studies conducted in other countries [29,30]. With regard to intermittency, to our knowledge no other study investigated the causes of high IRs using sound event detection. While previous studies report spatial differences in absolute IRs[30,31], our findings suggest that the underlying causes of intermittency also differ across locations. This is of importance, as the perception of sound differs by sound type [14], suggesting that the cause of intermittency also influences how an intermittent environment is perceived. Comparing our findings to the CENSE project [18,32], one of few monitoring efforts utilizing sound event prediction, similarities in bird song activity emerge, with strong presence of bird song in early morning hours. In contrast, traffic noise patterns differ between our deployment and CENSE: While we find traffic noise to be omnipresent in the form of background noise, CENSE findings show detections to correlate stronger with rush hour traffic. This difference likely stems from differences in machine learning models and highlights an important ambiguity in sound event detection: For sounds that are continuously present in the background, it is often unclear—and somewhat arbitrary—when a detection should be triggered. The model used in this deployment seems to have a higher sensitivity to traffic noise, leading to detections of background traffic noise, while CENSE may only detect foreground traffic noise. Both approaches can be justified, but lead to significantly different outcomes: Background sounds are closely linked to overall soundscape perception, whereas foreground sound events are particularly relevant for negative perceptions such as annoyance. [33].

3.5 Implications for future sensor deployments

The results show clear advantages of source-aware sound sensors compared to the monitoring of A-weighted sound pressure levels alone. While equivalent loudness levels are important, sound event presence ratios provide information about spatial and temporal patterns. Several event detections were associated with perceived sound prevalence and noise annoyance, particularly when considering louder events. This suggests that combining loudness metrics with sound event detections allows for a more complete assessment of urban sound environments than equivalent sound pressure levels alone. For sounds which are commonly present both as background and foreground sound, such as traffic noise, results will likely differ depending on the specifics of the machine learning model. Therefore, presence ratios should only be compared if obtained from the same model.

The findings also show the importance of distinguishing between sound event presence ratios and the causes of intermittency. While both metrics are based on sound event detections, a location may show high ratios of traffic noise, and yet, loud intermittent sounds may be caused by speech or crowds instead. As sources of intermittency can vary, it is advised to consider them when interpreting IRs, as high intermittency from bird song likely has different implications for annoyance and well-being than high intermittency from sirens or honking.

With regard to the measurement locations, the use of solar-powered sensors provided flexibility in the placement of sensors and made their installation easier. However, variations in mounting height and orientation (e.g., street-facing versus courtyard-facing) introduced additional heterogeneity. More controlled sensor placement, for example, in municipal deployments where installation on public infrastructure is feasible, would allow a clearer attribution of observed patterns to environmental factors rather than specifics of the sensor placement.

Finally, on-device sound event classification proved advantageous from a public acceptance perspective, as local processing reduced privacy concerns, which led to more trust and higher citizen participation. While we acknowledge the benefits of collecting raw audio, these benefits should always be weighed against the drawbacks from its privacy implications when monitoring sound environments.

3.6 Limitations and directions for further research

Several limitations should be acknowledged. First, the survey sample size was too limited to test many associations for significance. While the observed relationships appear plausible, they should be validated in a larger sample. Second, as mentioned above, sensor orientation and mounting height varied between locations, which likely influenced measured sound levels and sound event detections. Although overarching patterns appear rather robust to this, particularly where detections align with known characteristics about the research area, interpretations for individual locations may vary depending on the sensor placement. Lastly, sound event detections are based on machine learning and therefore prone to false negatives and false positives (e.g., overestimation of aviation noise data due to its similarity to traffic noise). Should sound event detection become part of data-driven policy-making, for example, by being included in nationwide noise mapping, then these biases need to be investigated to ensure fairness in the resulting noise mitigation strategies.

4. Conclusions

This study demonstrates how sound-event aware noise monitoring can provide valuable insights beyond a conventional purely sound-energy based assessment. By deploying 38 low-cost edge-processing sensors in and around Delft, the Netherlands, we characterised spatial and temporal patterns in sound events and loudness, and analysed how these metrics relate to surveyed sound prevalence and noise annoyance.

Across locations, the urban sound environment was highly heterogeneous in both composition and temporal structure. Median loudness and intermittency vary substantially and are largely independent of each other. We find large differences in the causes of intermittency, ranging from traffic to speech and bird song, signalling the importance of considering sound sources when interpreting intermittency ratios. Moreover, sound event presence ratios and the causes of intermittency can diverge, implying that common sound sources at a location are not always the same as those disrupting the sound environment. Traffic noise exhibited interesting temporal patterns, with more detections during the night, likely because of lower overall sound levels that provide less masking of the omnipresent background traffic noise.

Linking measurements to perception, sensor detections aligned somewhat with perceived prevalence, more so when focusing on louder events, potentially due to perceptual salience and building attenuation. Annoyance correlated most strongly with perceived source commonness, while sensor-based associations were weaker and more pronounced for louder events than for overall presence ratios. Intermittency showed no meaningful association with perceived short-burst annoyance. Further research, preferably with a larger sample, on intermittency and annoyance, is advised.

Taken together, the findings support the use of sensor-based sound-event detection as a scalable and publicly acceptable approach for monitoring urban sound environments. Future work should build on this approach, with more controlled deployments and systematic evaluation of potential biases, before source-aware detections are used in data-driven policy making.

5. References

[1] Münzel T, Molitor M, Kuntic M, Hahad O, Röösli M, Engelmann N, et al. Transportation Noise Pollution and Cardiovascular Health. Circ Res 2024;134:1113–35.

https://doi.org/10.1161/CIRCRESAHA.123.323584;WEBSITE:WEBSITE:AHA-SITE;WGROUP:STRING:PUBLICATION.

[2] Montenegro AL, Leal G, Zumelzu A, Herrmann-Lunecke MG, Vergara G, Heskia C, et al. Exploring the relationship between urban acoustic environments and mental well-being. Applied Acoustics 2026;242:111092. https://doi.org/10.1016/J.APACOUST.2025.111092.

[3] Aletta F, Lepore F, Kostara-Konstantinou E, Kang J, Astolfi A. An Experimental Study on the Influence of Soundscapes on People's Behaviour in an Open Public Space. Applied Sciences 2016, Vol 6, Page 276 2016;6:276. https://doi.org/10.3390/APP6100276.

[4] Gilmour LRV, Bray I, Alford C, Lintott PR. Natural soundscapes enhance mood recovery amid anthropogenic noise pollution. PLoS One 2024;19. https://doi.org/10.1371/journal.pone.0311487.

[5] Tsaligopoulos A, Kyvelou SS, Karapostoli A, Bobolos N, Tsintzou T, Lekkas DF, et al. Sound complexity as a strategy for livable and sustainable cities: The case of an urban waterfront. Noise Mapping 2023;10. https://doi.org/10.1515/noise-2022-0173.

[6] Murphy Enda, King EA. Environmental noise pollution: noise mapping, public health, and policy. Elsevier; 2022.

[7] Farina A. Soundscape ecology: Principles, patterns, methods and applications. vol. 9789400773745. Springer Netherlands; 2014. https://doi.org/10.1007/978-94-007-7374-5.

[8] Raimbault M, Dubois D. Urban soundscapes: Experiences and knowledge. Cities 2005;22:339–50. https://doi.org/10.1016/j.cities.2005.05.003.

[9] Peng B, Wang KIK, Abdulla WH. Environmental noise monitoring using distributed hierarchical wireless acoustic sensor network. Internet of Things 2024;28:101373. https://doi.org/10.1016/j.iot.2024.101373.

[10] Cassens L, Kroesen M, Calvert S, van Cranenburgh S. Low-cost solar-powered urban soundscape sensor. HardwareX 2026;25:e00753. https://doi.org/10.1016/j.ohx.2026.e00753.

[11] Oliveira HFR; S;, Machado JJM;, Leone R, Moroni D, Remagnino P, Filipa A, et al. Sound Classification and Processing of Urban Environments: A Systematic Literature Review. Sensors 2022, Vol 22, Page 8608 2022;22:8608. https://doi.org/10.3390/s22228608.

[12] Raimbault M, Dubois D. Urban soundscapes: Experiences and knowledge. Cities 2005;22:339–50. https://doi.org/10.1016/j.cities.2005.05.003.

[13] Brink M, Schäffer B, Vienneau D, Foraster M, Pieren R, Eze IC, et al. A survey on exposure-response relationships for road, rail, and aircraft noise annoyance: Differences between continuous and intermittent noise. Environ Int 2019;125:277–90. https://doi.org/10.1016/j.envint.2019.01.043.

[14] Axelsson Ö, Nilsson ME, Berglund B. A principal components model of soundscape perception. J Acoust Soc Am 2010;128:2836–46. https://doi.org/10.1121/1.3493436.

[15] Kjellberg A, Landström U, Tesarz M, Söderberg L, Åkerlund E. THE EFFECTS OF NONPHYSICAL NOISE CHARACTERISTICS, ONGOING TASK AND NOISE SENSITIVITY ON ANNOYANCE AND DISTRACTION DUE TO NOISE AT WORK. J Environ Psychol 1996;16:123–36. https://doi.org/10.1006/jevp.1996.0010.

[16] Song C, Li H, Ma H, Han T, Wu J. Effects of Noise Type and Noise Sensitivity on Working Memory and Noise Annoyance. Noise Health 2022;24:173–81. https://doi.org/10.4103/nah.nah_6_22.

[17] Kim J, Lee S, Kim S, Song H, Ryu J. Quantitative study on the influence of non-acoustic factors on annoyance due to floor impact noise in apartments. Applied Acoustics 2023;202:109144. https://doi.org/10.1016/j.apacoust.2022.109144.

[18] Can A, Picaut J, Ardouin J, Crepeaux P, Bocher E, Ecotiere D, et al. CENSE Project: general overview. Focus on Microscopy 2021. https://doi.org/10.34894/VQ1DJA.

[19] Mydlarz C, Sharma M, Lockerman Y, Steers B, Silva C, Bello JP. The Life of a New York City Noise Sensor Network. Sensors 2019, Vol 19, Page 1415 2019;19:1415. https://doi.org/10.3390/S19061415.

[20] Cassens L, Kroesen M, Calvert S, van Cranenburgh S. Low-cost solar-powered urban soundscape sensor. HardwareX 2026;25:e00753. https://doi.org/10.1016/j.ohx.2026.e00753.

[21] Suzuki Y, Takeshima H, Iti Suzuki Y. Equal-loudness-level contours for pure tones. J Acoust Soc Am 2004;116:918–33. https://doi.org/10.1121/1.1763601.

[22] Stevens SS, Volkmann J, Newman EB. A Scale for the Measurement of the Psychological Magnitude Pitch. J Acoust Soc Am 1937;8:185–90. https://doi.org/10.1121/1.1915893.

[23] Cassens L, Kroesen M, Calvert S, Van Cranenburgh S. URBAN SOUND CLASSIFICATION ON THE EDGE: EXPLORING THE ACCURACY-EFFICIENCY TRADE-OFF 2025. https://doi.org/10.61782/fa.2025.0940.

[24] Schell-Majoor L, Ewert SD, Kollmeier B, Rennies J. Concurrent categorical scaling of sound quality measures (CCSM) – a database of psychoacoustic measures roughness, sharpness, tonality, loudness and annoyance of artificial and real sounds. Https://DoiOrg/103205/Zaud000073 2025;7:Doc10. https://doi.org/10.3205/ZAUD000073.

[25] Fields JM, De Jong RG, Gjestland T, Flindell IH, Job RFS, Kurra S, et al. STANDARDIZED GENERAL-PURPOSE NOISE REACTION QUESTIONS FOR COMMUNITY NOISE SURVEYS: RESEARCH AND A RECOMMENDATION. J Sound Vib 2001;242:641–79. https://doi.org/10.1006/JSVI.2000.3384.

[26] Wunderli JM, Pieren R, Habermacher M, Vienneau D, Cajochen C, Probst-Hensch N, et al. Intermittency ratio: A metric reflecting short-term temporal variations of transportation noise exposure. J Expo Sci Environ Epidemiol 2016;26:575–85. https://doi.org/10.1038/jes.2015.56.

[27] Khomenko S, Cirach M, Barrera-Gómez J, Pereira-Barboza E, Iungman T, Mueller N, et al. Impact of road traffic noise on annoyance and preventable mortality in European cities: A health impact assessment. Environ Int 2022;162:107160. https://doi.org/10.1016/j.envint.2022.107160.

[28] World Health Organization. Nigth noise guidelines for europe 2009:162.

[29] Alburquenque MF, Marzzano Ríos A, Lambarri MC, Arellano RT, Lillo DC, Orellana SA. Identification of daily environmental noise patterns in two different urban sites in Santiago, Chile. Rev Med Chile 2020;148:582–93.

[30] Rahim Hamidi A, Novo P, Remy Kubwimana J, Clark SN, Nimo J, Umutoni C, et al. City-wide space-time patterns of environmental noise pollution in Kigali, Rwanda. Environmental Research Letters 2025;20:124024. https://doi.org/10.1088/1748-9326/ae1f2c.

[31] Brambilla G, Confalonieri C, Benocci R. Application of the Intermittency Ratio Metric for the Classification of Urban Sites Based on Road Traffic Noise Events. Sensors 2019, Vol 19, 2019;19. https://doi.org/10.3390/s19235136.

[32] Lavandier C, Aumond P, Can A, Gontier F, Lagrange M, Petit G. Urban sensor network for characterizing the sound environment in Lorient (France) through an automatic assessment of traffic, voice and bird presence ratios n.d.

[33] Han Z, Kang J, Meng Q. The effect of foreground and background of soundscape sequence on emotion in urban open spaces. Applied Acoustics 2022;199:109039. https://doi.org/10.1016/j.apacoust.2022.109039.