Towards Measuring and Forecasting Noise Exposure at the VELTINS-Arena in Gelsenkirchen, Germany

Pitchapa Ngamthipwatthana*, Marco Götze+, András Kátai*, and Jakob Abeßer*, *Member, IEEE*
*Fraunhofer Institute for Digital Media Technology IDMT, Ilmenau
+IMMS Institut für Mikroelektronik- und Mechatronik-Systeme gemeinnützige GmbH, Ilmenau

***Abstract*—Urban environments often subject residents to noise levels that pose significant health risks, including hearing loss, sleep disturbances, and decreased cognitive function. To address these concerns, we deployed five sensor devices for acoustic monitoring to measure and analyze noise around the VELTINS-Arena in Gelsenkirchen, Germany, a major venue for sports and concert events. This paper presents a detailed analysis of long-term noise data collected via installed acoustic sensors, focusing on identifying noise patterns and predicting future noise levels, as well as investigating the dependence of noise exposure on audible sound sources. Our research evaluates various time series modeling techniques, both traditional and deep learning-based, to forecast noise levels effectively. The study identified repetitive patterns in urban noise levels influenced by daily rush hour traffic and spectator movements, while local weather conditions, especially wind, significantly impacted noise measurements. Event-related noise, particularly from speech, cheering, and mechanical noises, varied significantly with the type of event and sensor placement, highlighting the need for strategic sensor positioning.**

***Index Terms*—Noise Measurement, Noise Forecasting, Sound Event Detection, Internet of Sounds.**

I. Introduction

Excessive exposure to noise in urban environments poses significant health risks to individuals. Prolonged exposure to high noise levels can lead to hearing loss, sleep disturbances, stress, and reduced cognitive performance. To study these dangers, acoustic monitoring technologies offer a valuable solution. These technologies enable the real-time measurement and analysis of noise levels in urban areas, providing data that can inform noise control measures and urban planning strategies aimed at reducing noise exposure and improving overall well-being in urban environments. From a high-level perspective, the development of distributed acoustic sensors for noise monitoring is highly relevant to the emerging Internet of Sounds (IoS) research area.

The VELTINS-Arena[1] in Gelsenkirchen, Germany, is a regular venue for sports events and concerts (see Figure 1). It is directly connected to several adjacent highways and a tram line, which are used for the arrival and departure of tens of thousands of spectators before and after the events. Especially during concerts, there are occasional complaints from residents about excessive noise. As part of the open innovation lab (OIL) Gelsenkirchen[2], which provides a testbed for new smart city-related technologies and sensor systems, five acoustic sensor units were installed near the arena (see Figure 2) in April 2023.

[1]https://en.veltins-arena.de

[2]https://openinnovationlab.gelsenkirchen.de

Fig. 1: VELTINS-Arena (source: FC Schalke 04)

In this paper, we study in detail the recorded measurement data, which capture both noise levels as well as predominant sound sources. As one focus of our research, we evaluate different traditional and deep learning-based time series modeling methods for noise level forecasting. More concretely, our aim is to answer the following research questions:

(RQ1) Which long-term repetition patterns can be observed in the sound analysis results (e. g., daily rush hour commuter traffic on nearby highways, arrival and departure of spectators at major events)

(RQ2) Can we observe a correlation between local weather conditions and the sound event detection (SED) results?

(RQ3) Which sound sources contribute most to high noise levels, in particular during events at the VELTINS arena?

(RQ4) To what extent is it possible to predict the noise exposure around the ARENA during major events based on the measurements taken during previous (comparable) events?

(RQ5) Is the noise exposure for sports and concert events consistent and how does it depend on the number of spectators?

II. Related Work

In this section, we give an overview of related research projects focusing on acoustic monitoring in urban environments. We first discuss three selected projects in detail before drawing more general conclusions on aspects such as privacy considerations and modeling approaches.

During the "Stadtlärm" project [1], [2], an acoustic sensor was developed to measure noise level and detect predominant

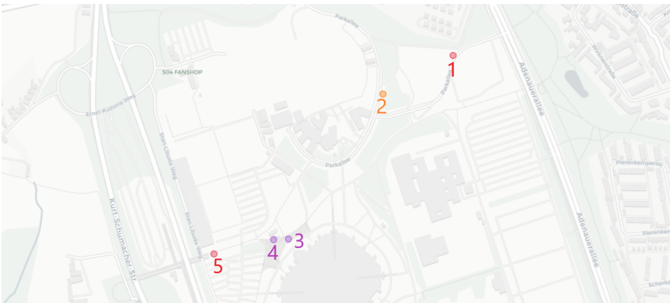

Fig. 2: Sensor location at the Veltins ARENA in Gelsenkirchen, Germany

sound events. Between 2017 and 2018, thirteen sensors were installed in Jena, Germany, at different locations in a central park area surrounded by tram tracks and noisy streets. Noise level measurements followed the German “Technical Instructions on Noise Abatement” (TA Lärm) [3]. A central server instance processed the distributed measurements to identify the most likely local noise sources such as construction sites, sports events, and concerts. In addition, the system includes a web-based application for real-time visualization of noise exposure throughout the city.

The Sounds of New York City (SONYC) project [4] is a large-scale initiative that aims to monitor, analyze, and mitigate urban noise pollution in New York City. The project uses Raspberry Pi devices equipped with microphone modules to collect noise data from various locations across the city. Noise level measurements are supplemented by sound event detection algorithms in order to better identify the prevailing noise sources. Although manual review of noise complaints from local residents often leads to a delayed response from the city administration, the proposed automatic noise monitoring system provides real-time noise measurements and can therefore improve the efficiency and effectiveness of noise regulation and enforcement.

In the CENSE project [5], a dense network of low-cost acoustic sensors was deployed in Lorient, France, to enhance numerical noise modeling with real-time in situ data. Additional perceptual assessments were performed using questionnaires to measure perceived noise annoyance and sound quality from the perspective of pedestrians (outdoor) and residents (indoor). The acoustic data recorded and transmitted from the sensors include both noise level statistics and third-octave spectral data, which, due to their low-frequency resolution, do not allow the reconstruction of personal and speech-related information, but are nevertheless a useful feature for sound event detection (SED).

In general, we observe two opposing paradigms in the distribution of tasks between edge sensors and centralized servers in urban monitoring scenarios. In the first paradigm, the analysis of recorded audio data is conducted entirely on the sensor side [1] or on the server side based on spectrogram-like data with very coarse frequency resolution [5]. These privacy-by-design approaches prevent information of data protection concern, such as randomly recorded spoken language, from being reconstructed. In the second paradigm, (compressed) audio data is transmitted for analysis from the sensors to the server [6]. On the one hand, this makes it possible to perform SED with significantly higher computing power, but at the same time raises data protection concerns. As potential countermeasures, the recorded audio data can be anonymized, for example, by removing or scrambling segments with detected voice or encrypting the transmitted data stream end-to-end [7].

Although most monitoring approaches implement separated sound classifications on each sensor nodes, Vidaña-Vila et al. [8] exploit the acoustic redundancy at nearby sensors and propose a two-stage architecture, which further aggregates classification results from individual sensor nodes to increase classification reliability. In addition to sound event detection, other potential tasks that can be implemented using acoustic monitoring include anomaly detection, crowd counting, and localization of sound events [7]. A multimodal analysis that also includes visual sensor data has the potential to lead to a more robust classification. Such an approach can combine audio-visual data as done in the MARVEL project [7] or the StreetAware dataset [9] or combine audio data with spatiotemporal textual information [10] to integrate location and time context in sound analysis.

III. Methodology

A. Acoustic Sensor

The acoustic sensors used for this project include custom electronics and can be powered either permanently or by buffering the nightly power of streetlamps using batteries. For the purpose of this study, all sensors were mounted on lampposts around the VELTINS arena as shown in Figure 3. Wi-Fi is used for communication. Depending on the specific hardware configuration, the electronic design employs custom PCBs for energy supply and the integration of both peripheral microphones. The hardware design builds on an earlier version developed in the “StadtLärm” project (see [11], [1], [2] for an in-depth discussion of the original architecture). In the current configuration, the custom base PCB and the originally used Raspberry Pi Compute Module 3 (CM3) has been replaced by a regular Raspberry 4B with either 4 or 8 GB of RAM with a separate power supply. The memory depends on the use cases to be fulfilled by the sensor unit in question and thus on the number of processes to be run in parallel for the audio analysis.

The earlier single-microphone setup has been extended to a binaural recording setup with two MEMS microphones (TDK InvenSense ICS-4343) in ear-like distance to allow detection of moving sound sources such as vehicles [12]. To reduce wind noise and issues with rain, the drill holes behind which the MEMS microphones are placed face towards the bottom (and thus still towards ground-level sound sources) and have been covered by acoustically transparent membranes. Figure 4 shows the system electronics of the audio sensor inside its weather-proof housing. The communication approach has been simplified such that all audio analysis results are

Fig. 3: Acoustic sensor (grey box) mounted on a lamppost in the Open Innovation Lab area at Gelsenkirchen. Photo: comNET GmbH

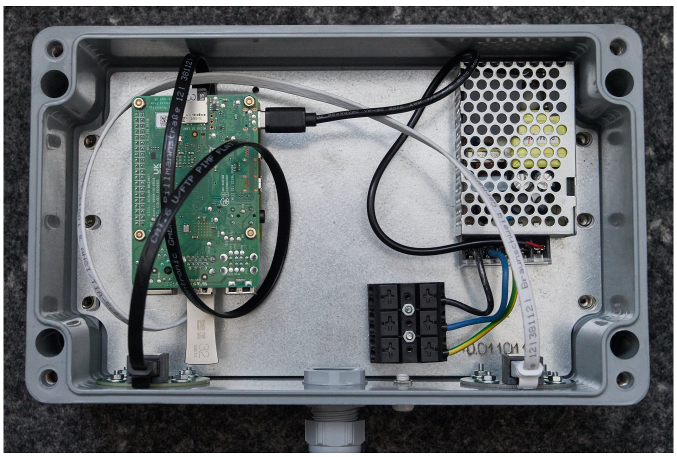

Fig. 4: Interior of a sensor box (PCB stack on the left, power supply on the right, microphones on the bottom edge).

decoded by a local process on the sensor and written directly (InfluxDB line protocol over HTTPS) into a remote time series database instance (InfluxDB v1). Several Grafana dashboards are implemented to visualize all sensor data for municipal end users.

B. Audio Processing

The audio sensor is equipped with three capabilities: measuring noise levels, detecting sound events, and monitoring traffic, including vehicle detection and classification. However, since traffic monitoring is beyond the scope of this paper, the focus of this section is on noise level measurements, sound event detection, and noise forecasting methods.

1) Noise Level Measurement: Noise level measurements are implemented in accordance to the specifications in the German "TA Lärm" [3]. A general technique in noise level measurement is to apply additional frequency weightings to better model the frequency-dependent sensitivity of human auditory perception [13]. We measure the A-weighted sound level (LAF), the C-weighted sound level (LAC), as well as the time-weighted sound level (LF), each with a temporal resolution of 8 measurements per second and finally store the temporal average over periods of one seconds. In this paper, we focus solely on the LAF measurements.

2) Sound Event Detection: In addition to noise level measurements, the sensor code implements sound event detection (SED) functionality, which aims to identify the most prominent sound sources at specific times [14]. We use the "VGG-like" SED model from [15] with 222k parameters and train it using the USM (V2) dataset proposed in [16], which includes synthetically generated five-second long soundscapes composed of between 2 and 6 sounds with random sound level and stereo placement. The audio clips used to generate the soundscapes have been selected from the FSD50k dataset [15] under the restriction that their licenses allow commercial use. Given these constraints, we make the commercial usability of the SED model as transparent as possible. The sensor code triggers the SED model every 2.5 seconds and analyzes the past 5 seconds of recorded audio. Here, we currently only consider one microphone channel and do not take stereo information into account. The SED outputs individual probability values of each of the 26 sound classes *airplane*, *alarm*, *birds*, *bus*, *car*, *cheering*, *church bell*, *dogs*, *drilling*, *glass break*, *gunshot*, *hammer*, *helicopter*, *jackhammer*, *lawn mower*, *motorcycle*, *music*, *rain*, *sawing*, *scream*, *siren*, *speech*, *thunderstorm*, *train*, *truck*, *wind*, which cover different sound class categories such as for instance human-made sounds, vehicle sounds, construction site sounds, as well as security-related sounds.

C. Time Series Modeling

Both research questions **(RQ1)** and **(RQ4)** introduced in Section I are related to the temporal progression of the noise level captured by the audio sensors. We therefore treat the noise level measurements as time series and study different time series modeling and forecasting techniques. In general, these techniques can be used to solve various tasks, including classification, anomaly detection, regression, forecasting, and clustering [17]. In this study, we compare five modeling approaches for noise level forecasting, which will be described in detail in the following subsections. For the training procedure, we train all models using the mean squared error (MSE) loss and the Adam optimizer for 100 epochs. Moreover, we incorporated early stopping with a patience of 10 epochs to prevent overfitting.

1) Naïve Forecaster: The Naïve Forecaster [18] simply uses the last observed value in the training set as the forecast for all future time points. It does not take into account any patterns or trends in the data. Due to this naive modeling approach, its main purpose is to provide a lower baseline performance for a given dataset. Given the value of the time series $\mathbf{y}_T$ at the last observation time T, the prediction of the next value is simply $\hat{y}_{T+1} = y_T$.

2) Linear Model: The "linear model" implements a linear regression to map an explanatory variable to a response variable. In the context of time series forecasting, this can be represented as $\hat{y}_t \approx \beta_0 + \beta_1 y_{t-1}$ where $\hat{y}_t$ is the predicted value at time t, y_{t-1} is the value at time t, β_1 is the coefficient (slope) for the predictor variable, β_0 is the intercept [18], [19]. We implemented the linear model using a dense layer without an activation function to map a single input to a single output similar to the Naïve Forecaster model.

3) Dense Model: We implement another forecasting model ("dense model") by stacking two dense layers with 64 neurons, each with a rectified linear unit (ReLU) activation function, to predict future values based on the previous values step by step as before in the linear model.

4) Multi-step Dense Model: All models discussed previously predict each time step individually and therefore cannot model any temporal changes in the input data. To overcome this limitation, we implement another forecasting model ("multi-step dense model"), which allows to process multiple time steps as input to predict a single output. Similarly to the dense model, the multi-step dense model includes two dense layers of 32 neurons and ReLU activation functions and a final linear output layer.

5) Long Short-Term Memory (LSTM): The Long Short-Term Memory (LSTM) model [20] was designed to better mitigate the vanishing and exploding gradients that often occur during the training of recurrent neural networks (RNN). The model implements an internal memory unit and an input, forget, and output gate to control which new information should be used to update the memory cell, be forgotten, or be output by the memory cell, respectively. LSTM models have been shown to effectively capture long-term dependencies in sequential data. This makes them particularly suitable for tasks such as natural language processing, time series forecasting, and speech recognition [17]. In our study, we evaluated two LSTM models. The "Single-LSTM" model has one LSTM layer with 50 neurons, and the more complex "Multi-LSTM" model includes three LSTM layers with 128 neurons each. In "Multi-LSTM" models, the last LSTM layer is followed by a dense layer with 64 neurons and the ReLU activation function.

IV. Experiments

In this section, we describe several experiments to answer the research questions posed in Section I. Understanding long-term repetition patterns in noise levels is not merely a technical challenge; it plays a key role in enhancing urban living conditions and public well-being. These patterns provide essential information for effective noise mitigation in urban planning.

A. (RQ1): Temporal Loudness Patterns

In this experiment, we study the LAF measurements recorded between April and November 2023 at sensor 3, which is located near the entrance to the VELTINS arena, and sensor 5, which mainly captures traffic noise at the nearby tram station (see Figure 2). We resample the measurements to a 15-minute time resolution and group them by workdays (Monday to Friday) and weekend days (Saturday and Sunday). Furthermore, we derive two additional groups from five days in which concert events (two days) or soccer games took place at the VELTINS Arena (see Table I). Although the concert events took place mainly at night, the soccer games started at different times (15:30, 18:30, and 20:30).

TABLE I: Investigated events at the VELTINS Arena between April to November 2023 grouped by soccer games (S) and concerts (C).

Events	Date	Start	End	# Spectators
FC Schalke 04 - Hertha BSC (S)	14.04.2023	20:30	22:30	62,721
FC Schalke 04 - SV Werder Bremen (S)	29.04.2023	18:30	20:30	62,721
FC Schalke 04 - Eintracht Frankfurt (S)	20.05.2023	15:30	17:30	62,721
Herbert Grönemeyer (C)	09.06.2023	20:00	23:00	50,000
Olé auf Schalke Festival (C)	14.10.2023	13:00	22:00	45,000

The average LAF values are shown for each of the four groups in Figure 5 (sensor 3) and Figure 6 (sensor 5). In general, the noise level values recorded closer to the stadium (sensor 3) show a greater variance as indicated by the larger error bars. As expected, LAF values are higher during the day than during the night and higher during the week than on weekends. When focusing on public events, we generally observe an increase in noise, especially in the afternoon and evening. Sensor 5 captures the noise of arriving and departing fans before and after concerts in the stadium.

Using the "Olé auf Schalke" concert as an example, one can recognize in Figure 6 corresponding loudness peaks between 10:00 and 13:00 as well as after 20:00 (blue dashed line). In particular, we observe high LAF values during 00:00-01:00, resulting in a high standard deviation for the concert curves in both sensors.

For soccer matches, we expect a reduced noise level during half-time breaks, which can be seen in Figure 5 around 16:15, 19:15, and 21:15, as expected from the three soccer matches that started at different times. Although we anticipated that soccer events would have higher noise levels than concert events due to a larger number of spectators (62,721 versus 50,000), sensor 3 indicates that concert events have higher noise levels than soccer events. In contrast, on sensor 5, soccer events exhibit higher noise levels between 13:30 and 19:00, which we assume is due to the transport between the tram station and the arena. We will discuss this in more detail in Section IV-E.

B. (RQ2): Influence of Local Weather Conditions

Local weather conditions can significantly affect noise measurements and sound event detection (SED). Wind, in particular, can create substantial background noise captured in sensor microphones. By studying the correlation between wind speed and the probability of wind noise detected by the SED model, we aim to filter out wind-related noise, ensuring that other sound sources are accurately identified.

For this experiment, we used wind speed measurements recorded at the "ERLE" weather monitoring station (IGELSE58) located at 51.56°N, 7.09°E (elevation 42 m) [3] every 4 minutes. We resampled the measurements to a 15-minute time resolution. We compared the wind speed data (in

[3]https://www.wunderground.com/weather/de/gelsenkirchen

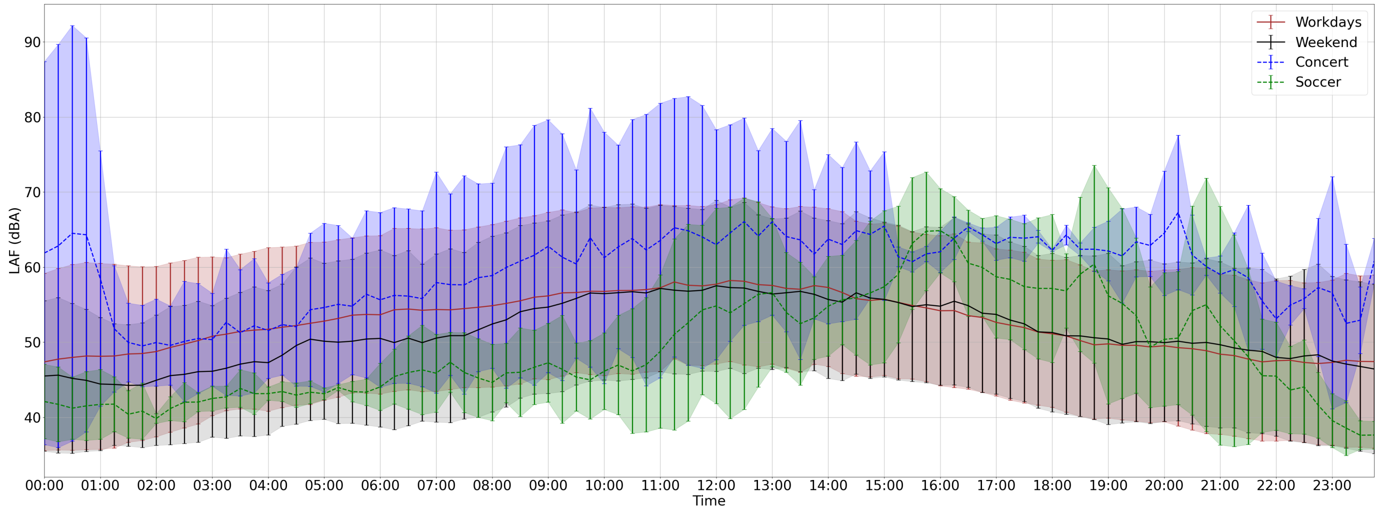

Fig. 5: Noise monitoring near the arena (Sensor 3) on weekdays, weekends and major events

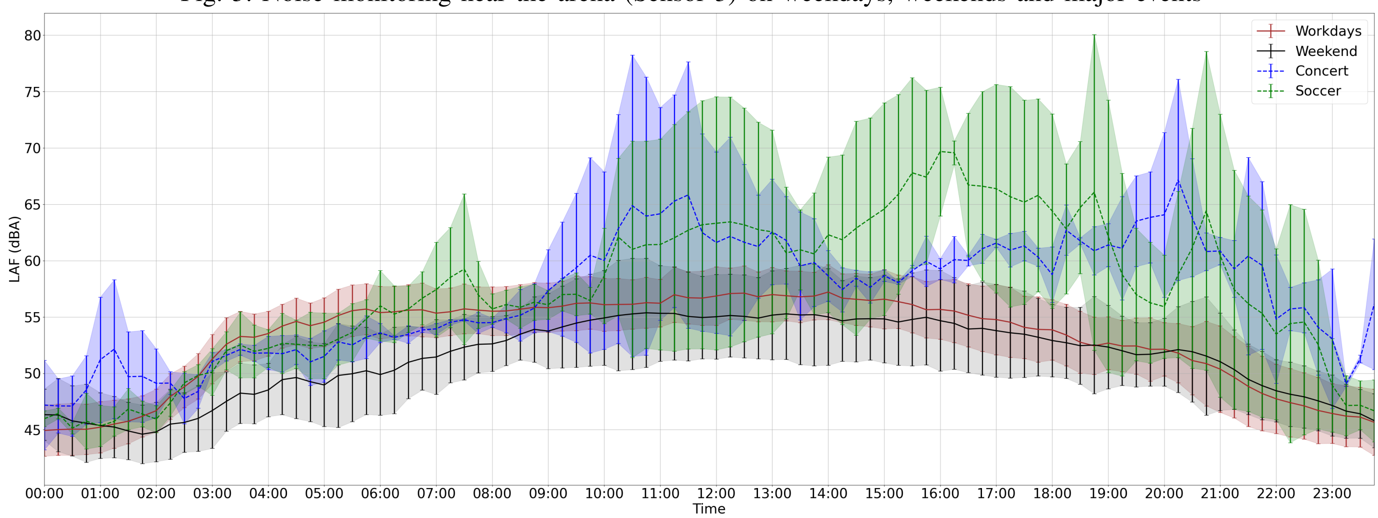

Fig. 6: Noise monitoring near the tram station (Sensor 5) on weekdays, weekends and major events

km/h) with the probability of the wind noise class detected by the SED model on days with high wind speeds, which are listed in Table II. Figures 7 and 8 illustrate the wind speed and wind probability values over the course of two selected days in November 2023 with a high and a low average wind speed, respectively.

In general, we observe that wind noise probability detected by the SED model does not show a significant correlation with the wind speed data. On days with high wind speeds, the probability of detecting wind noise is consistently high, as expected, except for the last day (28.11.2023), where the probability decreases when the wind speed falls below 2 km/h. However, it seems that the SED model is not able to distinguish between wind noise from different levels of wind speed. Although the characteristic sounds of strong and weak winds differ, we assume that the wind noise examples used to train the SED model capture various wind speeds. Our initial objective was to determine whether wind had any influence on our recordings. We focused primarily on comparing our existing recording data with available online wind speed data. Therefore, we excluded wind direction and other factors from our analysis. These were outside our main focus but could be explored in future research for more insights.

TABLE II: Correlation between wind speed and wind class (SED) from Sensor 3.

Date	Correlation (Sensor3)	Avereage wind speed (km/h)
16.04.2023	0.12	4.3
19.04.2023	-0.14	4.1
02.11.2023	0.38	8.5
04.11.2023	0.5	5.1
24.11.2023	0.16	3.1
28.11.2023	0.65	3.1

C. (RQ3): Temporal Correlation between Noise Level and Active Sound Sources

This experiment aims to identify the sound sources that contribute the most to the overall noise level. Such findings can lead to more effective noise control strategies at specific sensor

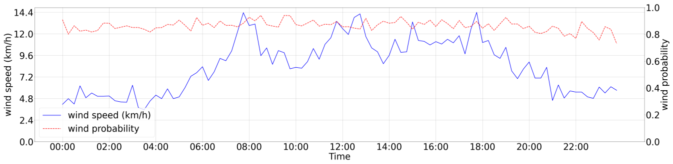

Fig. 7: Wind Speed and wind probability on 02.11.2023

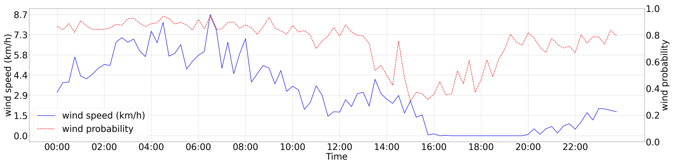

Fig. 8: Wind Speed and wind probability on 28.11.2023

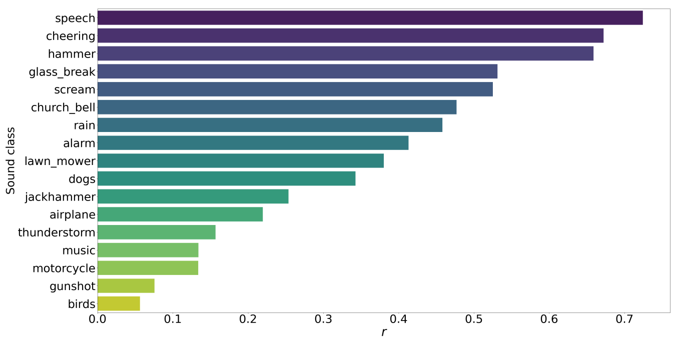

Fig. 9: A temporal correlation during major events (soccer matched and concerts) between noise level and sound class from sensor 5

locations. We computed the Pearson correlation coefficient between LAF measurements and SED predictions of sensor 5 within a 15-minute time resolution during both shorter sampling periods (major events, see Section IV-A) and a longer sampling period (April to November 2023). In Figure 9 and Figure 10, respectively, the sound classes detected by the SED model are shown in descending order, ranked according to the correlation coefficients r for both scenarios. The five main sound events that correlate with high noise levels are *speech*, *cheering*, *hammer*, *glass break*, and *scream*. In contrast, during long periods from April to November (Figure 10), the top five sound events are *scream*, *speech*, *lawn mower*, *cheering* and *birds*. Considering the time of the season in the data is also important, as certain sounds such as *lawn mower* (or other sounds of outdoor activity) are not expected during the winter season.

D. (RQ4): Noise Level Forecasting

In this experiment, our objective was to develop forecasting models to predict noise levels based on past measurements. This predictive ability will help implement noise control strategies and support noise pollution management at event locations.

We evaluated six different time series models introduced in Section III-C based on LAF measurements from sensor 3 with a 15-minute time resolution. For single-step input models (Baseline, Linear, and Dense), we predict the value at a given

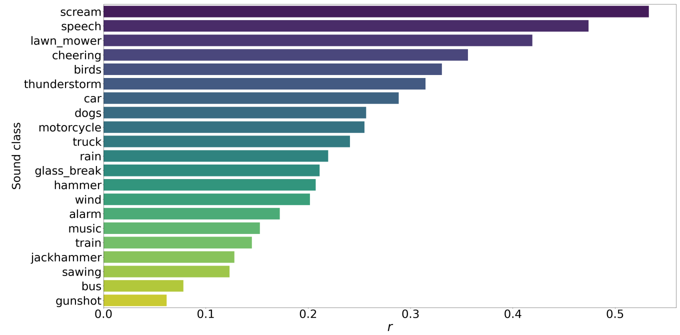

Fig. 10: A temporal correlation during April-November 2023 between noise level and sound class from sensor 5

time step based on the value at the previous time step. In contrast, for multi-step input models (Multi-step dense, LSTM, and Multi-LSTM), we use a longer context (such as 96 time steps, i. e., 24 hours) to predict the next value. As listed in Table III, we designed four experiments to systematically investigate how predicting LAF values depends on the temporal context covered by the training data. The duration covered by the training data ranges from half a month (Experiment 1) to three months (Experiment 4).

The model was trained and evaluated using the Mean Absolute Error (MAE) metric defined as

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{1}$$

with N denoting the length of the test sequence and y_i and $\hat{y}_i$ denoting the true and predicted value at the i-th timestep, respectively.

Table IV presents the evaluation results grouped by forecast model and experiment. We draw the following conclusions regarding the noise forecasting around the ARENA. First, the Linear, Dense, and Multi-step dense models can not surpass the baseline model due to their limited modeling capacity. Secondly, the Multi-LSTM model is the most effective model as it demonstrates the lowest and most consistent MAE values across all experiments. Interestingly, adding more training data does not necessarily improve performance. Instead, we observe the same performance in predicting LAF values for half, one, or two months of training data (Experiments 1-3) and worse performance for three months of training data (Experiment 4). Moreover, simple LSTM models do not capture sudden changes in noise patterns effectively. This limitation highlights the need for more sophisticated models such as Multi-LSTM to handle the dynamic nature of noise levels during major events.

TABLE III: Time coverage of training and test data used for evaluating loudness forecasting models.

Experiment	Training data	Testing data
Experiment 1	Second half of May 2023	June 2023
Experiment 2	May 2023	June 2023
Experiment 3	May, June 2023	July 2023
Experiment 4	May, June, July 2023	August 2023

TABLE IV: Evaluation results of all experiments.

Model	MAE1	MAE2	MAE3	MAE4	Model Params
Baseline	0.846	0.847	0.867	0.896	-
Linear	0.844	0.843	0.864	0.891	2
Dense	0.844	0.844	0.864	0.890	4,353
Multi-step dense	1.274	1.211	1.052	0.952	4,193
Single LSTM	0.866	0.828	0.833	0.845	10,451
Multi-LSTM	**0.809**	**0.809**	**0.809**	**0.834**	338,049

TABLE V: ANOVA results between each group of events for noise levels at different Sensors

Sensor	Event Type	Avg Loudness (dB)	Std Dev (dB)	F-statistic	p-value
Sensor 3	Concerts	59.53	9.56	141.6	8.85e-29
	Soccer Games	49.05	9.39		
Sensor 5	Concerts	56.98	6.06	0.27	0.61
	Soccer Games	57.35	8.53		

E. (RQ5): Influence of Event Type and Number of Spectators

Following RQ4, where we predicted the average loudness curve over a month without specifically considering events with higher loudness, we observed that the predictions were generally good, but failed to accurately capture sudden changes in noise levels. Therefore, it is crucial to account for events and other factors such as the number of participants to improve the accuracy of noise forecasting algorithms in the future.

A one-way Analysis of Variance (ANOVA) was performed to compare the average loudness calculated over days of events between the two event groups shown in Table I (3 soccer games and 2 concerts). For sensor 3, the analysis revealed a statistically significant difference, with concerts being significantly louder ($F = 141.61, p < 0.001$) than sport events. In contrast, sensor 5 did not show a significant difference in noise levels between concerts and soccer games ($F = 0.27, p = 0.606$). Detailed results of ANOVA are shown in Table V. In addition, the scatter plot in Figure 11 illustrates the relationship between average loudness (in dBA) and audience size for concerts and soccer games, measured by both sensors.

From these results, we conclude that concerts are generally louder than soccer games at Sensor 3, but this difference is not observed at Sensor 5. This indicates that the location of the noise measurement sensors relative to the venue of the event clearly affects the noise levels observed (compare Figure 2). Moreover, the type of event and its specific characteristics (e.g., music vs. crowd cheering) have a stronger influence on noise pollution than the number of spectators.

V. Conclusion

In this paper, we analyzed long-term measurement data to capture both noise levels and predominant sound sources around the VELTINS arena in Gelsenkirchen, Germany. Our study aimed to provide information on patterns, influences, and sources of noise, and to evaluate traditional methods based on deep learning for noise level forecasting. We addressed five key research questions to achieve these objectives.

Our findings revealed several long-term repetition patterns, such as daily rush hour traffic and spectator movements during events, which help to understand regular fluctuations in noise levels. We also observed a significant influence of local

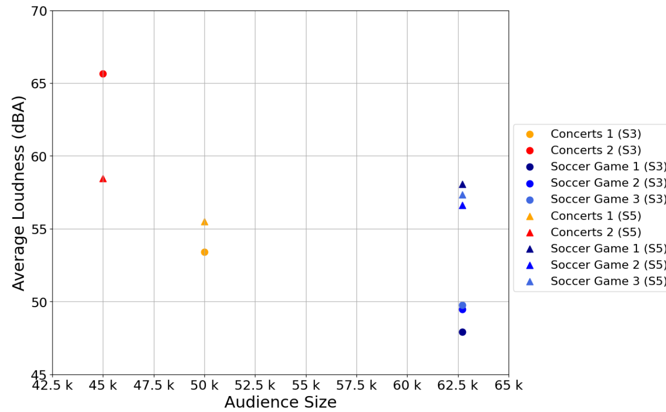

Fig. 11: Comparison of average loudness (from sensor 3 and 5) in each events and audience sizes for concerts and soccer games

weather conditions, particularly wind, on noise measurements. This correlation allows for filtering out wind-related noise to ensure accurate identification of other sound sources in the future. During events, specific sound sources such as speech, cheering, and mechanical noises were identified as the main contributors to high noise levels. In terms of noise forecasting, the Multi-LSTM model proved to be the most effective, showing the lowest and most consistent MAE values in different training scenarios. However, more training data did not necessarily improve performance, highlighting the importance of data quality and relevance. Our ANOVA analysis and scatter plots demonstrated that concerts generally produce higher average loudness compared to soccer games, with significant differences observed at one sensor location but not the other. This indicates that the placement of the acoustic sensors relative to the event venue is crucial for accurate noise measurement. Furthermore, the number of participants alone is not a straightforward predictor of noise levels, as the type of event and its characteristics contribute significantly to the generated noise.

For future work, we propose enhancing noise forecasting by developing multivariate models that incorporate multiple input data sources. Factors such as work day versus weekend, day versus night, and other contextual variables could be included in deep learning models for time series analysis. More complex models that consider these additional factors could lead to better noise forecasting before, during, and after larger events at the arena.

Acknowledgments

This research was funded by internal funding from the Fraunhofer-Gesellschaft. The authors would like to thank the members of the Open Innovation Lab (OIL) and the city administration of Gelsenkirchen, Germany.

References

[1] J. Abeßer, M. Götze, S. Kühnlenz, R. Gräfe, C. Kühn, T. Clauß, and H. Lukashevich, “A Distributed Sensor Network for Monitoring Noise

Level and Noise Sources in Urban Environments," in Proceedings of the 6th IEEE International Conference on Future Internet of Things and Cloud (FiCloud), Barcelona, Spain, 2018, pp. 318–324.

[2] J. Abeßer, M. Götze, T. Clauß, D. Zapf, C. Kühn, H. M. Lukashevich, S. Kühnlenz, and S. I. Mimilakis, "Urban Noise Monitoring in the Stadtlärm Project - A Field Report," in Proceedings of the Detection and Classification of Acoustic Scenes and Events 2019 Workshop (DCASE2019), New York University, NY, USA, 2019, pp. 1–4.

[3] Federal Ministry for the Environment, "Technische Anleitung zum Schutz gegen Lärm – TA Lärm," Sechste Allgemeine Verwaltungsvorschrift zum Bundes-Immissionsschutzgesetz, last accessed 3 July 2024, p. 503, 1998.

[4] J. P. Bello, C. Silva, O. Nov, R. L. Dubois, A. Arora, J. Salamon, C. Mydlarz, and H. Doraiswamy, "SONYC: a system for monitoring, analyzing, and mitigating urban noise pollution," Communications of the ACM, vol. 62, no. 2, pp. 68–77, 2019.

[5] A. Can, J. Picaut, J. Ardouin, P. Crépeaux, E. Bocher, D. Ecotière, M. Lagrange, C. Lavandier, V. Mallet, C. Mietlicki, and M. Paboeuf, "CENSE project: general overview," in Proceedings of the Euronoise conference, Madeira, Portugal, 2021, pp. 1–10.

[6] L. Luo, H. Qin, X. Song, M. Wang, H. Qiu, and Z. Zhou, "Wireless Sensor Networks for Noise Measurement and Acoustic Event Recognitions in Urban Environments," Sensors, vol. 20, no. 7, 2020.

[7] D. Bajovic, A. Bakhtiarnia, G. Bravos, A. Brutti, F. Burkhardt, D. Cauchi, A. Chazapis, C. Cianco, N. Dall'Asen, V. Delic, C. Dimou, D. Djokic, A. Escobar-Molero, L. Esterle, F. Eyben, E. Farella, T. Festi, A. Geromitsos, G. Giakoumakis, G. Hatzivasilis, S. Ioannidis, A. Iosifidis, T. Kallipolitou, G. Kalogiannis, A. Kiousi, D. Kopanaki, M. Marazakis, S. Markopoulou, A. Muscat, F. Paissan, T. P. Lobo, D. Pavlovic, T. P. Raptis, E. Ricci, B. Saez, F. Sahito, K. Scerri, B. Schuller, N. Simic, G. Spanoudakis, A. Tomasi, A. Triantafyllopoulos, L. Valerio, J. Villazán, Y. Wang, A. Xuereb, and J. Zammit, "MARVEL: Multimodal Extreme Scale Data Analytics for Smart Cities Environments," in 2021 International Balkan Conference on Communications and Networking (BalkanCom), 2021, pp. 143–147.

[8] E. Vidaña-Vila, J. Navarro, D. Stowell, and R. M. Alsina-Pagès, "Multilabel Acoustic Event Classification Using Real-World Urban Data and Physical Redundancy of Sensors," Sensors, vol. 21, no. 22, 2021.

[9] Y. Piadyk, J. Rulff, E. Brewer, M. Hosseini, K. Ozbay, M. Sankaradas, S. Chakradhar, and C. Silva, "StreetAware: A high-resolution synchronized multimodal urban scene dataset," Sensors, vol. 23, no. 7, p. 3710, 2023. [Online]. Available: https://doi.org/10.3390/s23073710

[10] J. Bai, J. Chen, and M. Wang, "Multimodal urban sound tagging with spatiotemporal context," IEEE Transactions on Cognitive and Developmental Systems, vol. 15, no. 2, pp. 555–565, 2023.

[11] M. Götze, R. Peukert, T. Hutschenreuther, and H. Toepfer, "An open platform for distributed urban noise monitoring," in 2017 25th Telecommunication Forum (TELFOR), 2017, pp. 1–4.

[12] J. Abeßer, S. Gourishetti, A. Kátai, T. Clauss, P. Sharma, and J. Liebetrau, "IDMT-Traffic: An open benchmark dataset for acoustic traffic monitoring research," Proceedings of the European Signal Processing Conference (EUSIPCO), pp. 551–555, 2021.

[13] K. Genuit, "The use of psychoacoustic parameters combined with a-weighted spl in noise description," in Proceedings of inter-noise 99: the 1999 International Congress on Noise Control Engineering, 1999, pp. 1887–1892.

[14] A. Mesaros, T. Heittola, T. Virtanen, and M. D. Plumbley, "Sound Event Detection: A tutorial," IEEE Signal Processing Magazine, vol. 38, no. 1, pp. 67–83, 2021.

[15] E. Fonseca, X. Favory, J. Pons, F. Font, and X. Serra, "FSD50K: An open dataset of human-labeled sound events," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 30, pp. 829–852, 2022.

[16] J. Abeßer, "Classifying Sounds in Polyphonic Urban Sound Scenes," in Proceedings of the 152nd Audio Engineering Society (AES) Convention, Online, 2022.

[17] S.-I. Ao and H. Fayek, "Continual Deep Learning for Time Series Modeling," Sensors, vol. 23, no. 16, 2023.

[18] R. Hyndman and G. Athanasopoulos, "Forecasting: principles and practice," OTexts.com/fpp3, last accessed 27 March 2024, 2021.

[19] C. Molnar, Interpretable Machine Learning: A Guide for Making Black Box Models Explainable, 2nd ed., 2022.

[20] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural computation, MIT Press, vol. 9, no. 8, pp. 1735–1780, 1997.