URBAN NOISE MONITORING IN THE STADTLÄRM PROJECT - A FIELD REPORT

Jakob Abeßer[1], *Marco Götze*[2], *Tobias Clauß*[1], *Dominik Zapf*[1]
Christian Kühn[1], *Hanna Lukashevich*[1], *Stephanie Kühnlenz*[3], *Stylianos Mimilakis*[1]

[1] Fraunhofer Institute for Digital Media Technology (IDMT), Ilmenau, Germany
[2] IMMS Institut für Mikroelektronik- und Mechatronik-Systeme
gemeinnützige GmbH, Ilmenau, Germany
[3] Software-Service John GmbH, Ilmenau, Germany
jakob.abesser@idmt.fraunhofer.de

ABSTRACT

As noise pollution in urban environments is constantly rising, novel smart city applications are required for acoustic monitoring and municipal decision making. This paper summarizes the experiences made during the field test of the Stadtlärm system for distributed noise measurement in summer/fall of 2018 in Jena, Germany.

Index Terms— internet of things (IoT), smart city, acoustic scene classification, sensor network, noise level measurement

1. INTRODUCTION

Urban dwellers are often exposed to high levels of noise from a variety of sources such as road traffic, construction sites and public sport and music events. Ideally, the city administration needs to systematically investigate any complaint. However, this task is hardly feasible due to the high personnel and cost involved.

As previously introduced in [1], the main goal of the Stadtlärm research project was to develop a distributed noise monitoring system, which supports city management by continuously measuring noise levels and sources. The developed system of 12 distributed sensors was subjected to a field test lasting several months after one and a half years of development. This paper presents a field report and discusses various experiences that have been made. We hope that this experience report will also be useful for other research projects in the field of IoT and Smart Cities. Related research projects propose similar solutions for noise monitoring in urban environments (see for instance [2, 3, 4]). Among others, the special focus during the Stadtlärm project was on checking legally prescribed noise limits.

The paper will be structured as follows. Section 2 briefly describes the individual components of the Stadtlärm noise monitoring system. Then, Section 3 discusses practical hardware considerations such as the microphone selection, weather resistance, moisture protection, as well as the long-term operating status of the sensor devices. Different measurements towards on-site sound propagation will be described in Section 4. Finally we will illustrate how the set of acoustic classes was refined and the algorithm for acoustic scene classification was improved during the project duration in Section 5.

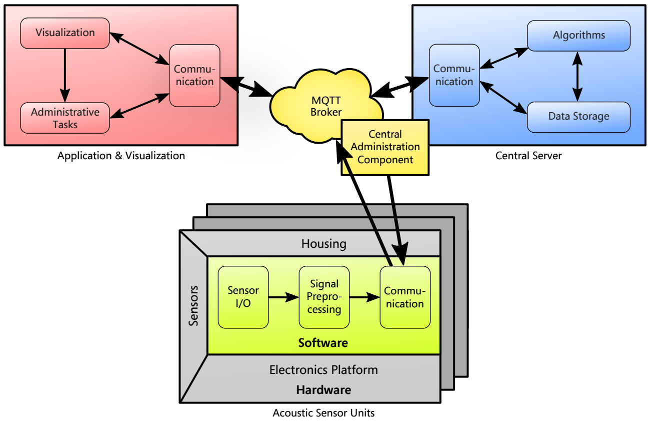

Figure 1: Stadtlärm system overview. The block diagram illustrates the data measurement and processing on the acoustic sensor units, communication via an MQTT broker, as well as data storage and visualization [1].

2. SYSTEM OVERVIEW

The Stadtlärm noise monitoring system consists of multiple distributed sensors, i. e., embedded systems that record and process audio data, and server-side services for further audio processing, data storage and access, as well as user applications. The overall system's backbone is a broker-based communications architecture. All communications among components and services utilize MQTT via a central broker, which also handles authentication and authorization. We used MQTT as the underlying communication protocol. Therefore, we defined a topics hierarchy and extended the publish/ subscribe paradigm by a convention enabling a request/response mechanism. The communications architecture is discussed in more detail in [5].

As shown in Figure 1, the system's components can be partitioned into three main functional groups. The acoustic sensor units are embedded field devices for acoustic data acquisition (see Section 3). On the software side, the sensors run embedded Linux and a custom application (implemented in Go), which handles communication (data, metadata, and management aspects). This application is also responsible for integration with the actual audio retrieval and processing software (implemented in Python). This software performs level measurements according to requirements of the German

https://doi.org/10.33682/s9w3-5341

TA Lärm regulations (see Section 4) as well as acoustic scene classification based on deep neural networks (see Section 5). By only transmitting the noise level and classification analysis results, privacy by design is implemented (as no speech recognition / surveillance is possible) and the amount of data for mobile communications is reduced.

The Stadtlärm audio service runs on a central server. After receiving measurement data from the acoustic sensor units, the service complements their data processing by computing long-term noise level parameters as defined in the German TA Lärm regulations. Data is stored and made available by request through an MQTT API. The Stadtlärm web application runs on a separate server and offers a web frontend to the data accumulated by the system tailored for administrative employees in departments concerned with urban noise. As such, it represents sensors on a city map, indicating status and current noise levels per location. On a per-sensor basis, historical level and classification data can be viewed, and reports can be created based on this.

Central to the system, an MQTT broker (Mosquitto) runs on another server and acts as a communications hub with central authentication/authorization. The broker is complemented by a central administration component serving as a registry of acoustic sensor units, a monitor of the overall system's status and load, and providing facilities to manage the sensors.

3. SENSOR HARDWARE

The acoustic sensor units are custom embedded devices developed in the course of the project. Each unit consists of a computation platform (Raspberry Pi 3 Compute Module Lite) and a microphone integrated on a custom PCB that also addresses aspects of robustness (robust power supply, hardware watchdog, etc.) and communications (M.2 slot for a wireless modem). The required computing power was estimated based on an initial implementation of the audio processing software (implemented in Python and utilizing the *keras* deep learning framework). Thanks to further optimizations made during the algorithmic development in the course of the project, the final CPU load is about 50% (across its 4 cores, i.e., equivalent to 200% load on a single core) and thus more than adequate, with ample reserves for communications and management as well as potential future adaptations.

For communications, both mobile wireless communications and utilizing public WiFi networks were considered. Each sensor unit is outfitted with a wireless modem and a SIM card. The total data volume used per month and unit is less than 1 GB (even including a remote firmware update or two), with the net amount of audio data being communicated by a single sensor being about 9 MB per day. As the sensors were meant to be mounted on lighting posts, assuming a permanent mains supply was an unrealistic assumption. The devices were therefore fitted with an additional power management PCB and a rechargeable battery (150 Wh), which enable off-grid operation with a periodic recharge (of at least 5 hours per day, typically overnight). Both computations and always-on mobile communications (with data being sent in near-real-time) result in a continuous power consumption of about 4 W. For this, the battery's capacity is ample; the situation is further relaxed by the fact that, in the field trial, lighting posts are powered for significantly more than 5 hours in the annual average, topping 16 hours in winter.

The sensor systems are fitted in weatherproof housings (for cost reasons, off-the-shelf) meant to be mounted on lighting posts of varying heights and diameters, at heights of 3 m or more. For this, the housings were fitted with steel strap clamps, which are flexible with respect to a post's diameter. The housing's external dimensions are 25 x 35 x 15 cm^3, so it is large enough to fit all the components and even leaves some room for optional equipment. It features a lockable hinged door, enabling easy but restricted access to its interior, necessary for wiring the device up to mains power during installation. For the pilot trial, the devices were otherwise assembled and configured manually prior to their being rolled out.

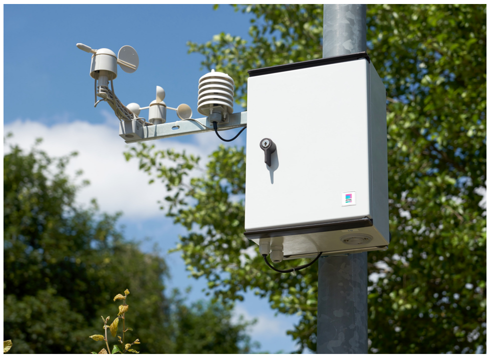

Figure 2: A Stadtlärm sensor equipped with the optional weather station components.

Another challenge was the integration of the microphone. Industry-grade microphones in outdoor-capable housings cost easily as much as the rest of the sensor system, which is why the goal with respect to the microphone was to use low-cost hardware in conjunction with state-of-the-art processing to nevertheless deliver high-quality audio recognition results. To this end, a MEMS microphone (ICS 43434, with an I^{2}S interface) was selected based on an systematic measurement and comparison of various microphone models. The easiest way of integrating that with the housing was to place the microphone directly behind a 4 mm drill hole at the bottom. Frequency response measurements and field recordings confirmed this to be adequate for the purpose of noise measurement and acoustic scene classification. However, in the course of the pilot trial, wind noise manifested itself as an issue of underestimated impact, requiring to make adaptions to the underlying acoustic scene classification model (see Section 5).

As weather may have profound effects on audio propagation and reception, a subset of sensor units in the pilot trial were extended by an inexpensive weather station measuring wind speed and direction as well as temperature and ambient humidity (Fig. 2). Internally, these additional components are connected via USB, proving the extensibility of the hardware platform. The weather data (measured by 3 devices placed in representative locations during the field test) correlate closely with weather data available from public providers, with plausible local deviations nevertheless being evident.

4. NOISE LEVEL MEASUREMENT

4.1. Sound Propagation in the Target Area

We performed an experiment to investigate the sound propagation in the target area of the field test. Therefore, given different configurations of source-to-sensor distances, we measured the drop in sound level. As a rule of thumb, we expect the sound level to drop by 6dB for a doubling of the distance. The test setup included an om-

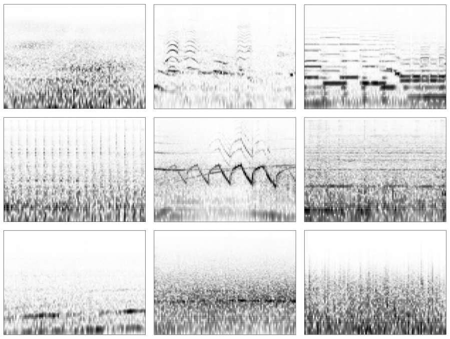

Figure 3: Examples of two-second long spectrogram patches for the classes car, conversation, music, roadworks, siren, train, tram, truck, and wind (from top left to right bottom), which are processed by the neural network.

nidirectional loudspeaker (GlobeSource Radiator by Outline) with a mobile power supply unit, which was powered by a truck battery. The measurement device was a NTi Audio XL2 with a calibrated Earthworks M30 as microphone. The Earthworks microphone was phantom-powered by the internal batteries of the NTi XL2 audio. We initially set up the SPL at a distance of 1 meter to 90 dB and for the second play-through to 100 dB.

As test signals, we used a simple sinusoidal signal with frequency of 1 kHz as well as a pink noise signal. A third test signal was a compilation of audio recordings covering different acoustic scenes. We repeated the test procedure at six different locations within the target area. From the results, we found that the measured SPL levels were slightly higher than expected. Possible reasons could be the rustling of the leaves or the wind, sound reflections by trees and buildings, as well as noises caused by pedestrians and cyclists nearby.

4.2. Case Study: Noise Pollution caused by Soccer Games

Residents from the nearby residential areas around the target area often complain about noise from soccer games, which happen typically on the weekends. As a show case, we analyzed the loudness curves, which have a temporal resolution of 0.125 s, recorded at the sound emission location i. e., the soccer stadium as well as a higher residential area around 600 m away. By computing the cross-correlation between overlapping 30 second long segments of the loudness curve, we derive the local maximum cross-correlation. Based on an additional test-recording nearby the second sensor, we were able to associate peaks in the cross-correlation curve with typical acoustic events during a soccer game such as drumming, fan cheering, and announcements from the stadium speaker. This provides clear evidence that noise from the soccer game is audible even in surrounding residential areas.

5. ACOUSTIC SCENE AND EVENT CLASSIFICATION

5.1. Acoustic Classes

In the final project stage, the initial number of acoustic scene classes (as presented in [1]) was reduced from 18 to 9 by focusing on the most relevant noise sources in the target area. In order to improve the applicability of the Stadtlärm system for traffic monitoring applications, the “truck” class was added to distinguish between the four most relevant vehicle types in the target site—cars, trains, trams, and trucks. During test recordings, we found that wind noises often overshadow other present noise sources, hence we added “wind” as an additional sound class. The intuition was to get a better sense of the classification confidence of other recognized sound sources. The initial sound classes “busking”, “music event”, and “open-air” were merged to a unified class “music”, which includes all music-related events. For the class “siren”, we improved the training data by adding recordings of local police cars, fire trucks, and ambulances. At the same time, we discarded the sound event classes “applause, “chants”, “horn”, and “shouting” as well as the (more ambiguous) sound scenes “public place” and “sports events” for now.

5.2. Acoustic Scene Classification

The classification model, which processes the recorded audio stream on the sensor units, was improved following the convolutional neural network (CNN) architecture proposed in [6]. The network is based on the VGG model paradigm that includes pairs of convolutional layers with small filter size and no intermediate pooling operation. In the applied architecture, four such layer pairs are concatenated with increasing number of filters (32, 64, 128, 256). After each layer pair, we apply a (3,3) max pooling in order to gradually decrease the spatial resolution and to encourage learning more abstract spectrogram patterns in higher layers. In order to improve the model’s generalization towards unseen input data, we apply batch normalization between the convolutional layers, global average pooling between the convolutional and dense layers, dropout (0.25) between the final dense layers, as well as L2 regularization (0.01) on the penultimate dense layer.

Given the project-specific set of acoustic classes discussed in the previous section, we assembled training data from various publically available datasets such as the Urban Sound dataset[1], the TUT acoustic scenes 2016 dataset[2], as well as the freefield1010 dataset[3].

5.3. Results & Observations

Figure 3 shows examples of two-second long spectrogram patches for all 9 classes as they are processed by the neural network. These plot give a better intuition of the difficulty of the classification task. The classes “conversation”, “music”, “roadworks”, and “siren” show distinctive patterns of either harmonic components (horizontal lines) or percussive components (vertical lines). For the vehicle-related classes “car”, “train”, “tram”, and “truck”, the sounds are more complex and noisy. The choice of an analysis window of two seconds (and a processing hop-size of one second) allows for a pseudo real-time processing of the recorded audio streams at the sensor units. However, such short analysis windows cannot capture slowly changing sounds such as passing vehicles. The noisy nature of wind also becomes apparent from the plot.

Figure 4 shows the receiver operation characteristic (ROC) curves as well as the precision-recall curves, which we obtained in a classification experiment on a separate test set, which was not used in the model training process. For each class, we obtained an

[1] https://urbansounddataset.weebly.com/urbansound.html

[2] https://zenodo.org/record/45739

[3] https://c4dm.eecs.qmul.ac.uk/rdr/handle/123456789/35

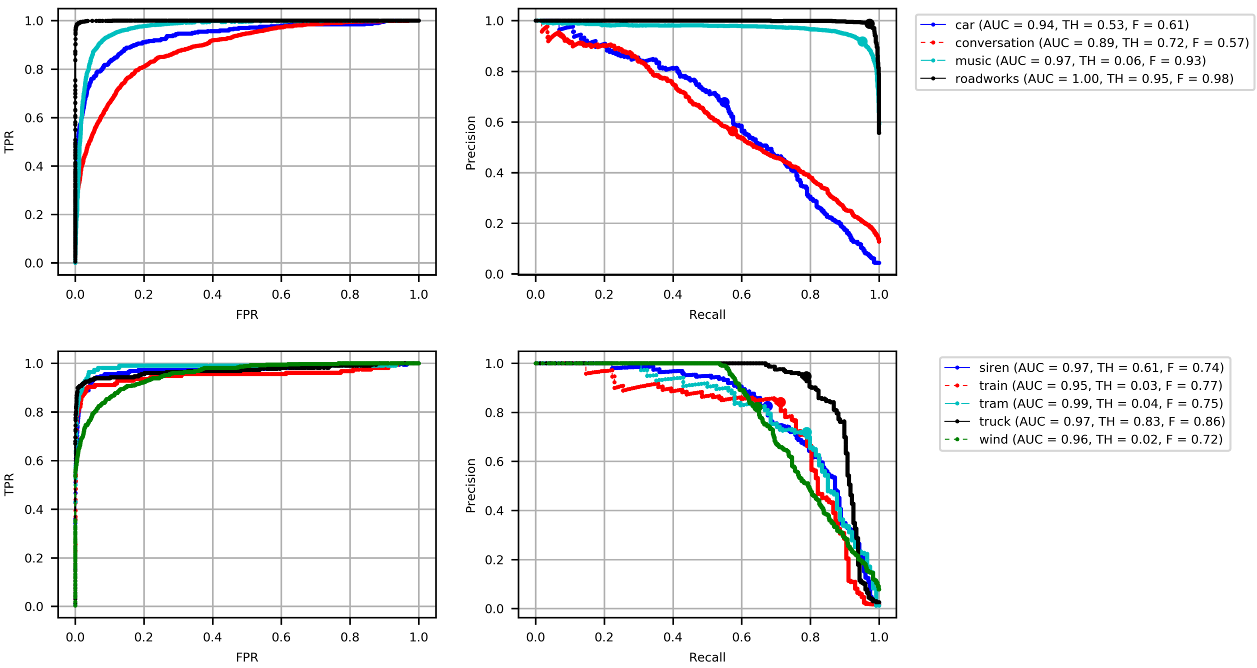

Figure 4: Receiver operation curves (ROC) and precision-recall curves for all 10 sound classes. Area-under-the-curve (AUC), optimal class-wise decision thresholds and best f-measure values are given in brackets.

optimal decision treshold on the sigmoid output values from the final network layer by maximizing the f-score. While all classes get a high AUC (area-under-the-curve) value, the precision-recall curves provide a better insight into the classifier performance.

When analyzing test recordings, which were made on the target site with the sensor, we observed a constant level of noise, either from environmental factors such as wind and rain or from the sensor hardware itself. Future steps for improving the model could be a stronger focus on data augmentation, where the high-quality audio recordings in the training data sets will be mixed with on-site background noise recordings to make the model more robust. Also we plan to test recently proposed methods for domain adaptation to achieve more robust classification results in different test environments.

6. WEB APPLICATION

We used the field test to evaluate the practical suitability of the web application. Based on continuous communication with the local city administration, we received regular feedback and optimized the main functionalities of the web application. This mainly concerned the identification of the prominent noise sources as well as the overall documentation of the noise level measurements. By providing the estimated noise source class probabilities from the model, non-expert users can better interpret the confidence of the acoustic scene classifier.

7. ACKNOWLEDGMENTS

The Stadtlärm project is funded by the German Federal Ministry for Economic Affairs and Energy (BMWi, ZF4248201LF6, ZF4248301LF6, ZF4072003LF6, ZF4085703LF6).

8. 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)*, 2018, pp. 318–324.

[2] J. P. Bello, C. Mydlarz, and J. Salamon, “Sound analysis in smart cities,” in *Computational Analysis of Sound Scenes and Events*, T. Virtanen, M. D. Plumbley, and D. P. W. Ellis, Eds. Springer International Publishing, 2018, pp. 373–397.

[3] A. J. Fairbrass, M. Firman, C. Williams, G. J. Brostow, H. Titheridge, and K. E. Jones, “Citynet—deep learning tools for urban ecoacoustic assessment,” *Methods in Ecology and Evolution*, pp. 1–12, 2018.

[4] P. Maijala, Z. Shuyang, T. Heittola, and T. Virtanen, “Environmental noise monitoring using source classification in sensors,” *Applied Acoustics*, vol. 129, pp. 258 – 267, 2018.

[5] M. Götze, R. Peukert, T. Hutschenreuther, and H. Töpfer, “An open platform for distributed noise monitoring,” in *Proceedings of the 25th Telecommunications Forum (TELFOR) 2017*, Belgrade, Serbia, 2017, pp. 1–4.

[6] Y. Sakashita and M. Aono, “Acoustic scene classification by ensemble of spectrograms based in adaptive temporal division,” in *Detection and Classification of Acoustic Scenes and Events Challenge (DCASE)*, 2018.