A Distributed Sensor Network for Monitoring Noise Level and Noise Sources in Urban Environments

Jakob Abeßer, Robert Gräfe, Christian Kühn,
Tobias Clauß, Hanna Lukashevich
Fraunhofer Institute for
Digital Media Technology (IDMT)
Ilmenau, Germany
Email: jakob.abesser@idmt.fraunhofer.de

Marco Götze
IMMS Institut für Mikroelektronik-
und Mechatronik-Systeme gemeinnützige GmbH
Ilmenau, Germany

Stephanie Kühnlenz
Software-Service John GmbH
Ilmenau, Germany

***Abstract*—Smart city applications for acoustic monitoring become essential to cope with the overall increasing noise pollution in urban environments. This paper gives an overview over a distributed sensor network for noise monitoring in the German city of Jena. Several acoustic sensor units allow for classifying among various acoustic scenes and events using Convolutional Neural Networks (CNN) and measuring different noise level parameters. Connected by a communication system based on MQTT (Message Queue Telemetry Transport), these sensors communicate measurement data to a central server for data post-processing and storage. Finally, a web-based application allows for various real-time visualizations of noise exposure distributed over the city.**

***Keywords*-Smart city; acoustic scene classification; event detection; sensor network; noise level measurement; TA Lärm**

I. Introduction

The overall noise exposure in urban environments can be affected by various acoustic sources such as car and railway traffic, construction sites, as well as public sports and music events. Especially in adjacent residential areas, local residents often file complaint to the city administration about too high noise levels. While urban noise is known to be a source of health-related issues, the problem of noise mitigation is not systematically addressed in most cities [1]. The type of noise source has an influence on which noise protection regulations are used. We developed a permanent acoustic monitoring system that is distributed among the most relevant places of both sound emission and immission. In addition to a localized sound level measurement, we identify active noise events and scenes in order to attribute the overall noise pollution to particular sound sources. The system can provide valuable input data for systematic municipal planning in order to improve the city's residents' quality of living.

The "StadtLärm" (German: "city noise")[1] research project aims to install a smart city system for systematic noise level measurement combined with noise source classification in the city of Jena in the state of Thuringia, Germany. The project focuses on acoustic measurements within and around a large park area along the Saale river ("Obere Aue", "Seidelpark", "Rasenmühleninsel") in a central position in the city as shown in Figure 3. The area is surrounded by two main streets that connect the inner city to the adjacent highway A4 as well as multiple tram and train tracks. There are various sound emission sources that affect the noise impact in the surrounding residential areas. For instance, several restaurants and open air venues organize music events on a regular basis. These events need to be approved beforehand by the city administration. Furthermore soccer games frequently take place in the sport stadium, which attract large groups of singing and cheering fans.

Complaints about regularly appearing noise disturbances need to be followed up by administration employees, who perform manual noise level measurements at the local places of interest. This procedure is both ineffective and time-consuming as it does not allow for a systematic noise level control of all officially authorized events. For one-time complaints, it is impossible to reconstruct past noise levels via measurements. In this paper, we propose a distributed network of sensor units that allows for a continuous monitoring of noise exposure in an urban environment. Among others, these measurement results shall facilitate in-advance planning of cultural events that require noise level predictions based on knowledge about past events.

We face two main challenges due to the geographic location of the target area. Due to the valley-like elevation profile in Jena, most residential areas (sound immission locations) have higher elevation than the event locations monitored in this project (sound emission locations). Furthermore, the sound propagation is heavily influenced by the prevailing westerly wind direction.

II. Related Research

Computational environmental sound analysis aims to extract semantic information from the acoustic surrounding, be it at home or in an urban environment [2]. In the past years,

[1] www.stadtlaerm.de, www.idmt.fraunhofer.de/StadtLaerm

various publications addressed the task of acoustic scene classification and event detection. Deep learning algorithms, particularly convolutional and recurrent neural networks (CNN, RNN), showed state-of-the-art performance in other research fields such as image processing and speech processing [3]. As a consequence, these methods were successfully adopted to environmental sound classification [4]. Here, learning algorithms are applied to discover spectro-temporal patterns within the analyzed audio signal that are distinctive for certain sound classes.

As basis signal representation, different spectrogram types are used ranging from the constant-Q transform [5] to the perceptually motivated mel-frequency spectrogam [6]. In addition to a single spectrogram layer, Takahashi et al. [7] as well as Piszak [4] propose to provide additional information about the temporal gradient of the spectrogram by adding first and second order derivatives as additional depth channels to the CNN. The proposed network architectures range from parallel convolutional layers as proposed by [5] to successive convolutional and recurrent layers [8], [9]. In our previous work [10], we compared two hybrid models proposed by Takahashi et al. [7] as well as Salamon and Bello [6] that combine a sequence of convolutional layers for feature learning and a couple of fully connected layers for classification. In particular, we investigated the robustness of the model against additional environmental background noise (wind, rain, thunderstorm), a reduced number of network parameters, as well as a preceding autoencoder-based signal compression. We refer the reader to [11] as well as [2] for an extensive literature review on methods for acoustic scene and event detection.

Several related research projects exist that similarly try to establish sensory networks for sound analysis in smart city application scenarios. The EAR-IT project proposed strategies for large-scale indoor and outdoor sensor networks [12]. Noise level measurement close to traffic infrastructure was investigated in the LIFE+ project DYNAMAP [13]. Finally, the SONYC research project (Sounds of New York City) has similar goals as the StadtLärm project as it aims at large-scale noise source classification and monitoring in an urban environment [6], [14]. Maijala et al. proposed a similar distributed sensor system for sound pressure level measurement and sound source classification [15]. In contrast to the system presented in this paper, the authors propose to perform an initial sound source classification on the sensor device and then transmitting compressed audio recordings in active time segments to a cloud service for further analysis. In the StadtLärm project, transmission of compressed audio is avoided due to the privacy-by-design criterion (compare Section III-B2).

III. Proposed Noise Monitoring Framework

Figure 5 gives an overview of the proposed framework. The individual components will be explained in the following sections.

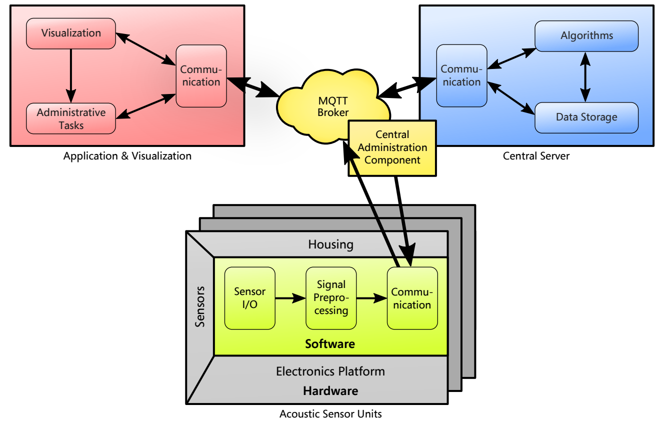

Figure 1. Framework flowchart illustrating the visualization, processing, and sensor platforms, which communicate via an MQTT broker.

A. System Architecture & MQTT Broker

An MQTT broker forms the center of the overall system architecture. MQTT is a lightweight M2M communication protocol popular in the context of the Internet of Things (IoT) in which multiple clients pass messages via a central broker. Messages are addressed to topics arranged in a user-defined hierarchy (e. g., “devices/1234/state”). Clients can publish on or subscribe to certain topics. The broker may restrict access on a per-topic basis. In the StadtLärm system, the open-source Mosquitto broker[2] and MQTT v3.1.1. are used. This way, all communications go through a centralized broker, which offers several advantages: First, physical communication relationships grow linearly with the number of participants even if logical communications are m:n. This facilitates adding clients/parties. Secondly, authentication, authorization, and logging can be handled centrally on the broker. Finally, MQTT provides a standardized protocol with readily available protocol implementations. At the same time, more complex interaction patterns can be realized.

While event-driven data processing maps naturally on the publisher/subscriber paradigm, the StadtLärm system also requires a request/response mechanism between logical peers. To this end, a convention has been elaborated that defines a fixed sub-structure of request/response topics. This structure enables all requests to be handled using a single abstract implementation. Sub-topics “req” and “res” are used for requests and responses, respectively, and client IDs are used on another subordinate level. The convention also makes it obvious which sub-topics the requester and responder have to subscribe to and where the response is published. Both requests and responses are transmitted in JSON (JavaScript Object Notation) format (see [16] for more

[2]https://mosquitto.org/

details). Since this approach is built upon MQTT primitives, it works transparently with any MQTT broker.

Semantically, the topic hierarchy is structured based on the associated system entities on the top level with the paths “admin”, “devices”, and “proc” for the processing service (central server). The central administration component, which is implemented in Kotlin[3] and Spring[4], runs on a server and monitors the overall system and allows for the available field/sensor devices’ metadata to be queried via corresponding topics. It is also responsible for remotely configuring the sensor devices. These devices each have a sub-tree of their own named after their respective serial numbers, with topics for their online state, metadata, as well as audio and (optionally) environmental sensor data. On the sensor devices, this interface is implemented by an application that wraps audio processing software (compare Section III-B) and takes care of data aggregation, buffering, and communications with the administration component.

The processing component (central server) publishes measurement results, which include noise level measurements as well as acoustic scene and classification results, in its sub-hierarchy. Furthermore, it offers request/response interfaces for retrieving historical data including noise measurements as well as potentially noise-inducing public events (concerts, parties, etc.) that were officially registered with the public administration. This data allows correlating noise level measurements with the presence/activity of potential noise sources.

The application/visualization component has no representation of its own as it does not offer any request/response interfaces of its own. Instead, it consumes live data from the sensor devices and the processing component as well as request-based data from the latter in order to realize the user interface for the official in charge (compare Section III-D).

B. Acoustic Sensor Units

As discussed in [2], the use of microphones for urban sound sensing is motivated by their relatively low installation cost and their robustness towards possible occlusion and environmental weather conditions such as fog, wind, or rain.

Hardware-wise, the sensor units (as shown in Figure 2) build upon a custom hardware platform with a Raspberry Pi 3 compute module (CM3) at its core, i.e., a quad-core ARM BCM2837 SoC with 1 GiB of RAM and 1.5 GiB of flash. The hardware has been optimized for robustness and versatility, utilizing electronic components made for industrial applications (-40..85 °C) and featuring an M.2-type modem slot and optional battery management for a 12 V, 13 Ah battery. This enables operation on lamp posts, where the units can recharge overnight. As a testament to the platform’s extensibility, an optional weather station component is available, recording, in particular, wind speed and direction for use in conjunction with audio algorithms. A MEMS microphone at the bottom is used for sound recording.

Figure 2. Sensor unit (open, door removed).

The current positioning of the acoustic sensors in the target area is illustrated in Figure 3. We followed three decision criteria. First, the two main streets surrounding the target area as well as the main tram and train tracks need to be monitored since traffic noise is one of the main sources of noise pollution. Secondly, we aim to install a small grid of three sensors in near free field conditions in the northern part of the target area in order to be able to perform experiments towards noise level interpolation as well as sound source localization. Finally, we selected three particular example areas where noise pollution, which is perceived in a residential areas, can be traced to close-by sound immission locations. In all three example areas, a street and a tram track run in between the sound immission and emission locations. Hence, we place sensors at both locations as well as at the intermediate traffic sound source in order to analyze to what extent the traffic noise overshadows the targeted sound source’s noise level.

1) Noise Level Measurement: The noise level measurements implement several noise level parameters mentioned in “Technische Anleitung zum Schutz gegen Lärm - TA Lärm” (German: “Technical Instructions on Noise Protection”) and defined in German DIN norms [17]. The norm DIN EN 61672-1 (electro-acoustics—sound level meter) defines the sound level. The calculation of the rating level L_r is

[3]https://kotlinlang.org/
[4]https://spring.io/

Table I
NOISE PARAMETERS AS DEFINED IN THE GERMAN "TA-LÄRM" [17] WITH BRIEF EXPLANATION AND LOCATION OF COMPUTATION.

	Parameter	Explanation	Sensor / Server
L_{AF}	Sound event level	A-Frequency and F-Time-weighted sound pressure level	Sensor
L_{Aeq}	Average sound level	Mean sound level over a given time period	Server
L_{AFT}	Takt maximal sound level	Peak sound pressure level within five seconds. Necessary for impulsiveness surcharge	Server
K_{i}	Impulsiveness	Surcharge for short-term sound pressure levels	Server
L_{r}	Rating level	Average sound pressure level with surcharges and rating time over 16h/8h (day/night)	Server

defined in the norms DIN 45681 (Acoustics—Determination of tonal components of noise and determination of a tone adjustment for the assessment of noise immission) as well as DIN 45645-1 (Determination of rating levels from measurement data) in combination with DIN 45641 (Averaging of sound levels). Both the rating level and other long-term noise level aggregation methods are computed on the server-side instead of on the sensors. As both require long-term accumulation, a potential sensor malfunction would lead to data loss. To avoid this, each sensor sends a package each second with eight sound level measurements to the server every second for storage and further processing. The real-time processing on the sensor units includes filtering and weighting of the audio signal. These dense sound level measurements allow for a visualization of the noise level live map (compare Section III-D). Table I summarizes all noise parameters that are computed in the StadtLärm system on the sensor and server sides.

Table II
TAXONOMY OF SOUND EVENTS AND SCENES.

Acoustic Scene	Acoustic Event
Music Event	Applause, Busking, Club Concert, Open Air Concert
Public Place	Conversation, Shouting
Roadworks	Jackhammer
Sports Event	Applause, Chants
Traffic	Car, Horn, Siren, Train, Tram

2) Acoustic Scene and Event Classification: On the acoustic sensor units, the audio stream is recorded at a sample rate of 32 kHz and processed using a Short-Time Fourier Transform (STFT) with a hop size of 882 (20 ms) and a window size of 1024 (23.2 ms). The magnitude spectrogram is furthermore logarithmically compressed to be less sensitive to the sound level dynamics of the microphone. For the purpose of data compression, the magnitude spectrogram is mapped to a logarithmically spaced frequency axis of 49 bins using a triangular-shaped filter bank between 50 Hz and 15 kHz with 6 bins per octave resolution.

Table II gives a taxonomy of the five acoustic scenes of interest in the StadtLärm project and the corresponding 14 sound events. We compiled audio recordings for training classification models from various publicly avaible datasets such as the Urban Sound Dataset [14], the TUT Sound Events (real audio) 2016 development set [18], and the IEEE AASP public & private datasets [11]. We apply different data augmentation methods such as time stretching, pitch shifting, dynamic range compression, and adding of environmental background noise to enlarge our training dataset.

As shown in Figure 4, our model for automatic acoustic scene and event classification is based on the hybrid deep neural network proposed by Takahashi et al. in [7]. The model has three pairs of convolutional layers with ReLU activation layers [3] for feature learning, two fully connected layers for classification followed by a final sigmoid layer that provides class-likelihood values for all scene and event classes. The network architecture is inspired by the VGG CNN architecture [19] since larger convolution kernels (e.g. 5 x 5) are replaced by pairs of stacked layers with smaller 3 x 3 kernels without intermediate pooling. This reduces the number of trainable model parameters and allows for more expressive features due to three additional non-linearities. In contrast to [7], we only use two-second-long magnitude spectrogram patches and only one depth dimension, whereas Takahashi et al. proposed to use four-second-long patches and to add the first two time derivatives as additional depth dimensions of the spectrogram patch.

Audio monitoring technologies need to first solve several privacy requirements before being successfully integrated in smart city environment [20]. By running the acoustic scene and event classification algorithm directly on the sensor units, only classification results and noise level measurements are transmitted to the server. This procedure ensures the privacy-by-design property.

C. Central Server

The central server hosts a database to store the measurement data collected by the sensor units and distributes data on demand as needed by the other clients. Therefore, the server subscribes to two MQTT topics: measurement data, i. e. , noise level measurements and classification results, from the sensor units and requests from other clients such as the application/visualization client. Using a request/response mechanism, the server passes the measurement data combined with corresponding timestamps and geo-locations to the application/visualization component for real-time visualization of the distributed measurements.

Furthermore, the data can be accessed by the city administration for processing noise complaints and predicting

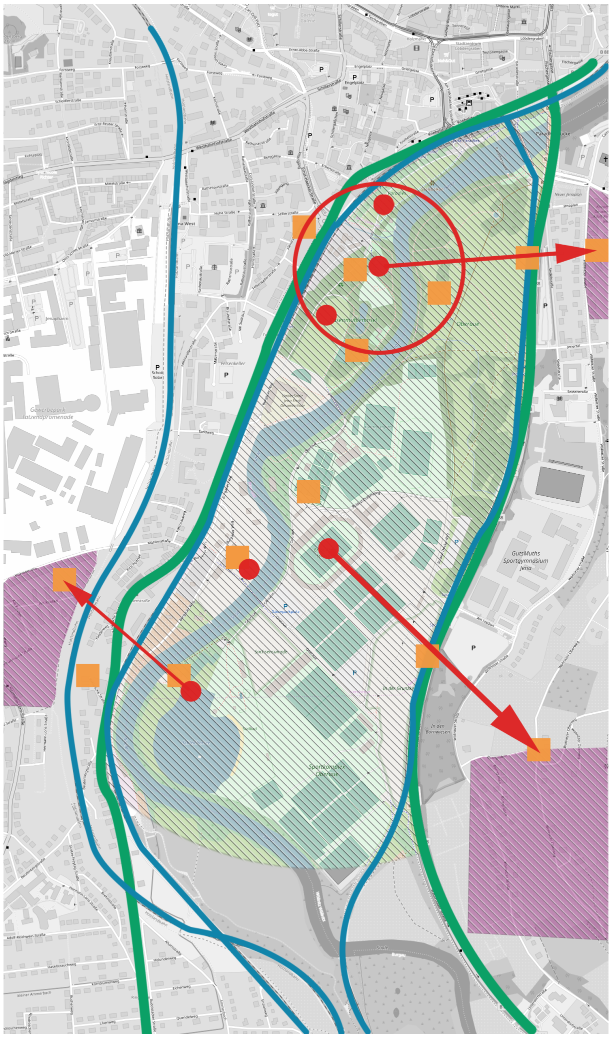

Figure 3. Targeted area in the city of Jena. Orange squares represent acoustic sensor units. Residential areas are illustrated in purple. Exemplary sound emission locations (e.g. open air venues, sport stadium) are shown as red circles. Small sensor grid is illustrated as group of sensors in the northern part of the target area. Most relevant traffic elements are streets (green) as well as train and tram tracks (blue). Red arrows connect sound emission locations and close-by residential areas.

noise levels in order to decide whether to approve certain events, such as open air concerts. Depending on the request parameters, the server furthermore performs averaging of noise pressure levels over larger time periods. Also, the activity of acoustic scenes and events is post-filtered using heuristics based on estimated minimum durations of certain sound types in real-life scenarios.

D. Application & Visualization

The StadtLärm web application is an interactive map, based on an open-source framework[5]. It visualizes various noise measurements obtained from the central server

[5]https://mapbender3.org/

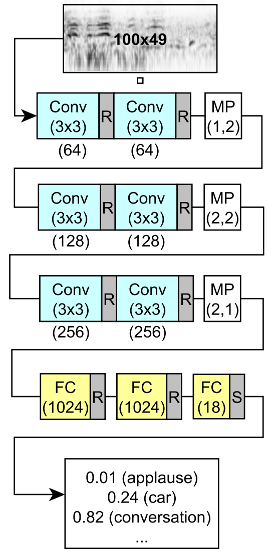

Figure 4. A hybrid deep neural network provides a mapping from two-second-long spectrogram patches to likelihood measures for the relevant acoustic scene and event classes. The network components in the figure are convolutional layers (Conv), rectified linear unit activation functions (R), max pooling layers (MP), and fully connected layers (FC). The number of filters for each convolutional layer is given in parantheses below.

(compare Section III-C) and supports the decisions of the city administration. In addition to the map, the application provides an interface to a desktop application that displays the three-dimensional noise propagation in an interactive 3D-landscape model. The base map of the StadtLärm web application for the city administration of Jena is extracted from the city-owned map portal “Kartenportal Jena”[6]. This portal also provides various additional layers with information about possible noise sources, such as road construction sites, locations of bottle banks, and playgrounds. Furthermore, it allows to overlay areas of traffic noise mapping.

The noise level values measured at the sensor units can be displayed in the map application in real-time or aggregated over a selectable time period. For example, the user gets a condensed overview of the past noise situation by eliminating the interferences caused by rush-hour traffic. The measured noise level values can be matched with publicly approved events in the map that are automatically retrieved from city-owned open data event calendar[7].

Basic measurement results will be displayed and communicated to the citizens via free and anonymous access to the StadtLärm application. In contrast, the city council is provided an extended account with further internal administrative information like conditions of the events’ approvals and local residents’ complaints. During the ongoing event,

[6]http://www.jena.de/kartenportal
[7]https://www.jena.de/de/kultur/veranstaltungskalender/246470

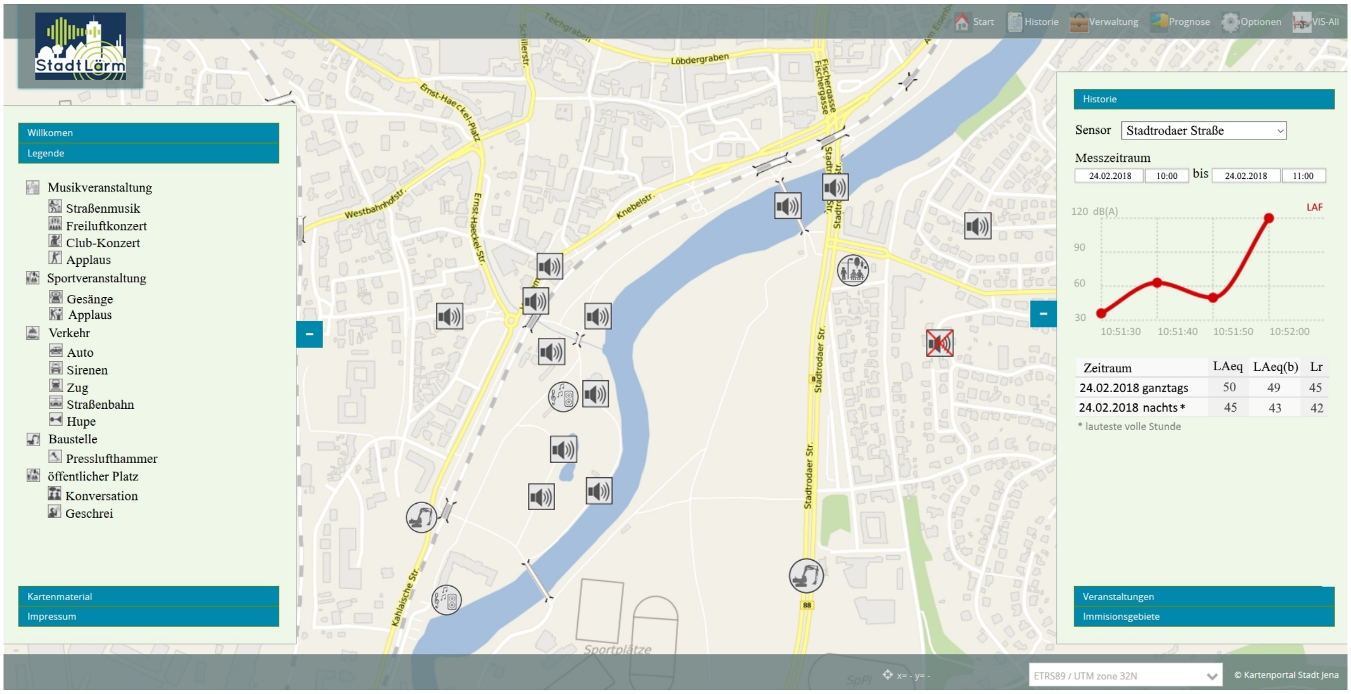

Figure 5. Screenshot of StadtLärm web application (in German). Sensor locations (loudspeaker symbols) as well as recognized sound events and scenes are shown at their respective geo-location on top of the city map. On the right, real-time noise level measurements at particular sensor locations are illustrated.

the regulatory agency of the city administration can be informed automatically via email if noise level limits are exceeded. A long-term goal is to facilitate the administrative decision procedure even more by predicting future noise situations of planned events based on previously measured data.

IV. Conclusion & Outlook

This paper gives an overview of a distributed sensor network for noise level and noise source monitoring as developed in the StadtLärm research project. After reviewing related scientific work and research projects on smart city acoustic monitoring, we introduced the overall system architecture containing an MQTT broker for communication purposes, distributed acoustic sensor units that allow for noise level measurement and classification of acoustic scenes and events, as well as a web-based end user application and visualization.

In the future, the proposed framework will be extended for further application scenarios in an urban environment. For instance, the sensor units along the main streets allow for a monitoring of traffic flow as well as for identifying active vehicle types. Also, long-term noise level analysis will allow to identify quiet regions within the city, which can be included in touristic recommendations.

Acknowledgments

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

References

[1] M. Zappatore, A. Longo, and M. A. Bochicchio, “Crowd-sensing our smart cities: a platform for noise monitoring and acoustic urban planningd,” *Journal of Communications Software and Systems*, vol. 13, no. 2, pp. 53–67, 2017.

[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] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[4] K. J. Piczak, “Environmental sound classification with convolutional neural networks,” in *Proceedings of the 25th IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, Sept 2015, pp. 1–6.

[5] T. Lidy and A. Schindler, “CQT-based convolutional neural networks for audio scene classification and domestic audio tagging,” in *Proceedings of the Detection and Classification of Acoustic Scenes and Events Workshop (DCASE)*, Budapest, Hungary, 2016, pp. 1032–1048.

[6] J. Salamon and J. P. Bello, “Deep convolutional neural networks and data augmentation for environmental sound classification,” *IEEE Signal Processing Letters*, vol. 24, no. 3, pp. 279–283, March 2017.

[7] N. Takahashi, M. Gygli, B. Pfister, and L. V. Gool, “Deep convolutional neural networks and data augmentation for acoustic event detection,” in *Proceedings of the Interspeech Conference*, San Francisco, USA, 2016, pp. 2982–2986.

[8] S. Adavanne, P. Pertilä, and T. Virtanen, “Sound event detection using spatial features and convolutional recurrent neural network,” in *Proceedings of the 42^{nd} IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, New Orleans, USA, 2017, pp. 771 – 775.

[9] E. Cakir, S. Adavanne, G. Parascandolo, K. Drossos, and T. Virtanen, “Convolutional recurrent neural networks for bird audio detection,” in *Proceedings of the 25^{th} European Signal Processing Conference (EUSIPCO)*, Kos, Greece, 2017, pp. 1744–1748.

[10] J. Abeßer, S. I. Mimilakis, R. Gräfe, and H. Lukashevich, “Acoustic scene classification by combining autoencoder-based dimensionality reduction and convolutional neural networks,” in *Proceedings of the 2nd DCASE Workshop on Detection and Classification of Acoustic Scenes and Events*, Munich, Germany, 2017, pp. 7–11.

[11] D. Stowell, D. Giannoulis, E. Benetos, M. Lagrange, and M. D. Plumbley, “Detection and classification of acoustic scenes and events,” *IEEE Transactions on Multimedia*, vol. 17, no. 10, pp. 1733–1746, Oct 2015.

[12] D. Hollosi, G. Nagy, R. Rodigast, S. Goetze, and P. Cousin, “Enhancing wireless sensor networks with acoustic sensing technology: Use cases, applications & experiments,” in *2013 IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing*, 2013, pp. 335–342.

[13] P. Bellucci, L. Peruzzi, and G. Zambon, “LIFE DYNAMAP project: The case study of Rome,” *Applied Acoustics*, vol. 117, pp. 193–206, 2017.

[14] J. Salamon, C. Jacoby, and J. P. Bello, “A dataset and taxonomy for urban sound research,” in *Proceedings of the 22^{nd} ACM International Conference on Multimedia (ACM-MM’14)*, Orlando, FL, USA, 2014, pp. 1041–1044.

[15] 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.

[16] M. Götze, R. Peukert, T. Hutschenreuther, and H. Toepfer, “An open platform for distributed urban noise monitoring,” in *Proceedings of the 25^{th} Telecommunication Forum (TELFOR)*, Nov 2017, pp. 1–4.

[17] “German technical guidelines for noise reduction,” http://www.verwaltungsvorschriften-im-internet.de/bsvwvbund_26081998_IG19980826.htm (last visited: 02/03/2018).

[18] “Tut sound events 2016 dataset.” [Online]. Available: http://www.cs.tut.fi/sgn/arg/dcase2016/task-sound-event-detection-in-real-life-audio

[19] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in *International Conference on Learning Representations (ICLR 2015)*, 2015, pp. 1–4.

[20] A. Ståhlbröst, A. Sällström, and D. Hollosi, “Audio monitoring in smart cities - a privacy perspective,” in *Proceedings of the 12^{th} International Conference e-Society*, Madrid, Spain, 2014, pp. 35–44.