PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

Environmental noise event classification based on self-organizing map using psychoacoustic features and spatial filtering

Jure MUROVEC[1]; Luka ČUROVIĆ[1]; Tadej NOVAKOVIĆ[1], Jurij PREZELJ[1]

[1] Faculty of Mechanical Engineering, University of Ljubljana, Slovenia

ABSTRACT

One of the main challenges of environmental noise monitoring is the classification of noise events. This is critical for the assessment of individual noise source contribution to the overall noise level. Numerous attempts have been made to provide algorithms for such classification, however with low certainty. The main reason for poor results is the lack of proper noise features. For this purpose, a new feature "noise source direction" has been applied, based on a small horizontal microphone array. It uses a never seen before algorithm which is computationally inexpensive and provides exceptional beam-pattern directivity. The array is implemented as a part of an autonomous environmental noise source classification system. Some of the features are provided directly from the array itself and are combined with psychoacoustic metrics. Artificial Neural Network (Self Organizing Map) performs the classification based on extracted spatial and acoustical features of environmental noise sources.

Keywords: Environmental noise, Classification, Self-Organizing maps, Psychoacoustics, Localization

1. INTRODUCTION

Environmental noise is defined as unwanted outdoor sound, created by human, animal or environmental activities and it has become a growing world problem. Obtaining satisfactory measurements both in the time domain and spatially is one of the main challenges in environmental noise monitoring. The noise level at the point of measurement depends on a combined contribution from all surrounding noise sources. Also, Determining the contribution of a specific noise source is a frequent task acoustic laboratories have to perform. A typical example is depicted in Figure 1.

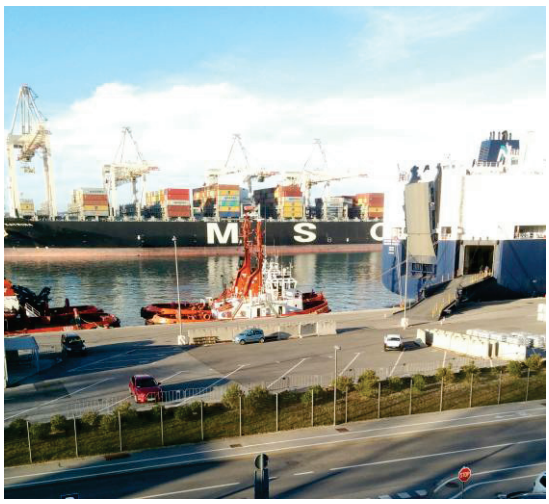

Figure 1 – Environmental noise source measurement point where multiple noise sources are present

[1] jure.murovec@fs.uni-lj.si

The problem arises when residual noise levels are comparable or even higher than the observed noise source. The contribution of main noise sources to the total noise level can be determined by extensive measurements under different operating conditions of individual sources. Different noise sources can operate simultaneously or at different time intervals while the duration of their operation is rarely deterministic. Therefore, the presence of personnel is always required to sample and record the operating conditions and the associated sound pressure levels (SPL). Another argument against extended periods of monitoring time are the changes in weather conditions which have a significant effect on monitored noise levels, (1-3). Continuous noise measurements are a credible source of information and a prerequisite for extrapolation when modeling, for example, traffic noise, (3,4). Expensive equipment, long measurement time and the need for educated staff are what makes such measurements expensive. Those around shipyards, ship ports, airports, factories, mines, and other large industrial areas, as well as in the city centers, usually require permanent installation of sound level measuring equipment. In such cases, the measurement equipment is left unattended and the recording of the environmental noise levels, including residual noise, is performed automatically. If operators are present during the noise measurements, residual noise events can be excluded from the measurements and noise properties, including the classification of noise sources and their directions, can be recorded in situ. However, it is impractical and costly to assure the presence of certified personnel at the noise monitoring location over long periods of time.

In order to control and reduce environmental noise, we must find new approaches of how to effectively measure it. People can very skillfully perceive and judge the general characteristics of the surrounding sound field. The computer does not have enough capacity to match this, so developing computational methods to automatically extract this information also has great potential in a variety of applications. Designing an environmental noise classification system that mimics the operation of educated personnel as much as possible should be the next logical step. A considerable amount of manual work can be saved by automatically determining the contribution of each individual noise source to the total noise level.

As almost all environmental noise sources are placed far from the measurement location, we can consider that the incoming sound rays propagate in a two-dimensional plane parallel to the ground. This leads to the assumption that one-dimensional microphone array can be applied for the purpose of spatial filtering and classification of environmental noise sources.

2. AUTOMATION OF ENVIRONMENTAL NOISE MONITORING

In the last decades, automatic noise identification and classification has become a very active subject of research. It can be directly or indirectly implemented into a very wide area of topics, including speech recognition, pattern identification, and context-aware applications. In order to achieve automation of environmental noise monitoring, the equipment should mimic the operator's activities during measurements. Personnel classifies noise events depending on the noise level threshold, which can be similar to the background noise level, the direction of the noise source and the subjective noise recognition capabilities. The classification is therefore based on the experience of the operator.

A pattern classification algorithm typically consists of a feature extractor and a classifier as seen in Figure 2. Regarding classifiers, different systems and algorithms for automatic classification of environmental noise sources were already proposed including Hidden Markov Model, Neural Networks, k-NN, Support Vector Machine, and Fuzzy logic, dynamic time warping, and Gaussian Mixture Models (GMM) (5, 6, 7).

Figure 2 – Block diagram of a typical noise classification system

Deep convolutional neural networks (CNN) have the ability to learn discriminative spectro-temporal patterns and that makes them well suited to environmental sound classification. First, they are capable of capturing energy modulation patterns across time and frequency when applied to spectrogram-like inputs, which has been shown to be an important trait for distinguishing between different, often noise-like, sounds such as engines and jackhammers (7). Neural networks tend to be extremely precise and more often than not, outperform other classification algorithms, (8-11). The problem with CNN is the vast amount of training data and time that is needed to fine-tune the classifier. Additionally, the algorithms are designed or learned for a specific kind of noise source to acquire the best possible performance, (12). Generally, artificial neural networks are divided into two main categories: supervised and unsupervised. Self-Organizing Maps (SOMs) are competitive and unsupervised training networks of feed-forward type. There is no training data and no expected output for the learning process. SOMs are able to discover statistical similarities in the input space which they exploit and automatically cluster the data into different input classes. A SOM is built from nodes or neurons and can be visualized as a grid-like neural network array. The radius of surroundings is given by the number of neurons in the surroundings of the best matching unit (BMU) /winning neuron. Weight vectors of the same dimension as the input and at a particular position in the map space are associated with each node. Kohonen self-organizing neural networks have two layers, an input layer and an output layer, known as the competitive layer. Each input neuron is connected to every neuron on the output layer, the latter being organized as a two-dimensional grid, (13).

While one part of the research deals with the classification itself, the other one concentrates on feature extraction. The commonly used features including the well-known mel-frequency cepstral coefficients or more specialized features such as histograms of sound events, or histogram of gradients learned from time-frequency representations. Although there has been some use of psychoacoustic metrics outside of their initial purpose, they have not yet been implemented in any kind of system for classification of environmental noise sources.

Psychoacoustics is the science of the human perception of sound. It studies the relationship between sensory perception (psychology) and physical variables (physics). Equations were developed to calculate a set of metrics to objectively describe the complex human perception of sound quality, (14). Loudness N (DIN 45631/A1), tonality T (DIN 45681) and sharpness S (DIN 45692) have already been standardized. Psychoacoustics is already widely accepted as an important approach in the design and manufacturing of products to attract and retain customers. This is especially true in the automotive industry, as car manufacturers have realized that the produced sound could also serve as a desired acoustic signal, giving the driver feedback about the functioning of the car. Gradually, the psychoacoustic technique has been spreading to other industrial areas, from the car door closing sound quality, design of centrifugal fans for vacuum cleaners, end-of-line inspection and gear fault diagnosis to a concept for the general detection of machine faults. If we want to mimic the operator's activities during measurements, the use of features that describe the human perception of sound is quite obvious.

Practically all classification algorithms are implemented in combination with many different types of signal attributes. Earlier attempts were focused on the development of a universal procedure to classify all possible noise events. However, this proved to be impossible. More recent algorithms are specialized for a specific type of environmental noise. Studies have shown that the use of ANN for the purpose of environmental noise monitoring is the most promising technique going forward, (9-11). They do have their shortcomings that need to be addressed and worked on. As said before, it is impossible to create a universal database for recognition of all noise sources/events. Therefore, the best performing ANN are those who are trained and designed for a specific type of environmental noise. Also, the training requires a vast amount of data and a human expert until the classification is fully autonomous. This is time-consuming and is in contradiction with lowering the costs and making the environmental noise monitoring cheaper and easier to use and implement.

Motivated by the observations that humans are highly effective in distinguishing different noise sources, we must pay close attention to how environmental noise measurements are done. Consciously or not, we spatially filter our acoustic space in which we are present. All algorithms used in systems for automatic classification of environmental noise are based on a single microphone channel sound pressure signal. Therefore, a microphone array in combination with beamforming techniques should be implemented for the purpose of automatic noise source classification. This opens up the possibility of filtering noise events and sources before the extraction of features for the purpose of classification. Additionally, the synthesized signal provides an improved Signal-to-noise (SNR) ratio which naturally improves the recognition of noise sources.

An autonomous environmental noise source classification system is presented in this paper. It consists of a small microphone array for the purpose of spatial and noise filtering. To mimic the operations carried by personnel, psychoacoustic features are extracted and then fed into an artificial neural network. SOM performs the classification and thus significantly reduces the time required for post-processing identification of each noise source. The system acts as a human expert, using sound source localization and psychoacoustic metrics in its decision making. Reducing the need for human interaction lowers the cost and complexity of long term environmental noise monitoring.

3. PROPOSED SYSTEM

3.1 System's characteristics

The proposed system consists of 4 channel horizontal microphone array, A/D converter and a computer to process the information. All the main parameters are presented in Table 1. The four microphones measure the sound field simultaneously and their signals are processed for the purpose of noise source localization in a two-dimensional plane. The use of horizontal microphone array is justified by the fact that almost all environmental noise sources are placed far from the measurement location. The incoming sound rays, therefore, propagate in a two-dimensional plane parallel to the ground.

Table 1 – Main parameters of the proposed system

Parameter	Value
Sample rate	192 kHz per channel
Bits per sample	24
Feature extraction time constant	125 ms
Array dimension	fi 7 cm x 30 cm
Spatial resolution	6°
Number of features	8
Number of possible noise source classes	Infinite

The direction of the most dominant noise source at a particular time is calculated with a highly advanced, never seen before algorithm with exceptional beam-pattern directivity. The four acquired signals are then delayed and added up to provide a synthesized signal with a higher SNR for the purpose of feature extraction as all microphone array based systems do. What makes this system remarkable is the additional use of sound source localization directly for pre-classification and as a standalone feature as seen in Figure 3. This brings us one step closer to designing a human-like environmental noise measurement system that has great potential in a variety of applications.

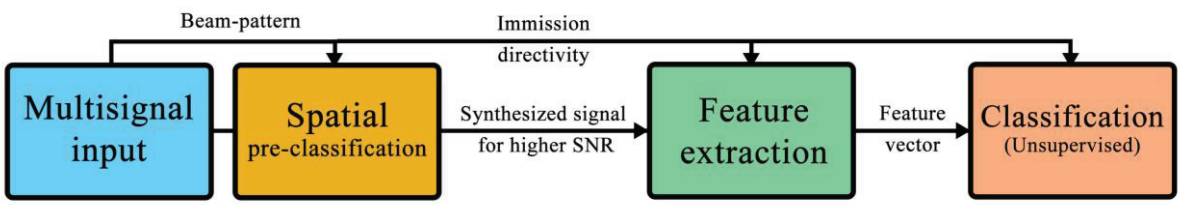

Figure 3 - Block diagram of our environmental noise measurement system. Spatial pre-classification is performed prior to feature extraction

3.2 Noise source localization

Array processing relies on multiple microphones arranged in a certain geometry. Among all the established methods in signal processing, those based on beamforming are probably the most common. They are simple and have a low computational cost that makes them a priori well suited for long

measurement times. However, their performances strongly depend on the array geometry and number of microphones, (6). The localization of conventional beam-formers is limited to low frequencies, and thus alternative developing a new algorithm was necessary. To keep the system compact and inexpensive, the array needed to be as small as possible with the number of microphones as little as possible. The frequency range of environmental noise sources goes as low as 20 Hz which translates to a wavelength of 17 m. Known beamforming techniques would have to be performed on huge arrays, making them permanently stationary.

Due to the innovative nature of the algorithm, the details of mathematical operations are not present in this paper. However, the simulation of the array's frequency response is presented in Figure 4 and compared to the delay-and-sum beamforming. The algorithm is able to localize the noise source in the lowest of frequencies while retaining small distance between the microphones.

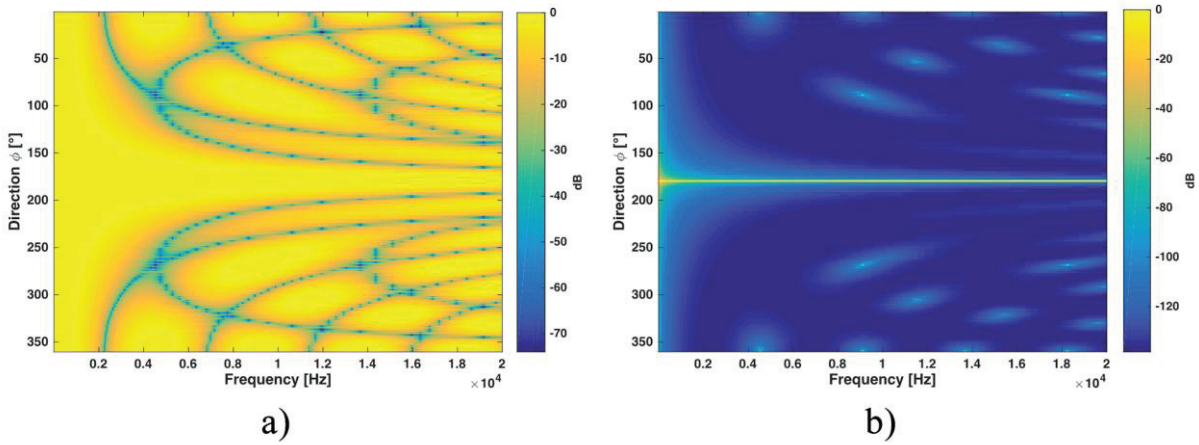

Figure 4 – Simulation of array's frequency response for: a) Classical delay-and-sum beamforming, b) Our algorithm

The detected direction of the most dominant noise source at a given time, as seen in Figure 5, determines which parts of the whole measurement will be put into the classification algorithm. It is pointless to extract features when the detected direction is not stationary as this means that there is no clear-cut dominant noise source. This way we avoid large quantities of unnecessary data, reducing the computational time for classification with SOM.

3.3 Feature extraction

Seven features are extracted from the synthesized signal. Joining psychoacoustic metrics of loudness, sharpness, tonality and roughness are crest factor, zero crossings and one third octave spectrum. The calculation is performed every 125 ms or 24000 samples. Another additional feature provided by the beam pattern is immission directivity, (6), which indicates how dominant the detected noise source is. It can be used for either spatial classification (by filtering out only the most dominant sources) or as one of the inputs for the final classification. Extracted feature vector serves as an input for SOM.

3.4 SOM Classification

Input for the SOM are all the feature vectors when the direction is stationary and/or the immission directivity is high enough. The neuron lattice is a two-dimensional rectangular grid and its size is chosen beforehand. Usually, the number of neurons should be slightly larger than the expected number of noise sources. The size of the lattice can be fine-tuned with the help of different characteristics of the generated output SOM. We can observe the number of times each neuron won while the SOM was competitively learning or the statistical properties of neuron and input vector distances such as mean value and standard deviation. SOM was chosen because of its unsupervised nature which isn't time-consuming nor does it need any training database. Classification based only on the direction of noise sources and immission directivity would be difficult for moving targets, such as trains, automobiles,

etc. SOM also helps to recognize similar or even the exact same noise sources regardless if their location varies with time.

The output of the system is the direction plot, immission directivity plot and noise source class plot as seen in Figure 5. Combining the output information with a simultaneous sound level meter (SLM) measurements gives us a calibrated SPL and the contribution of each individual noise source to the total noise level can be determined.

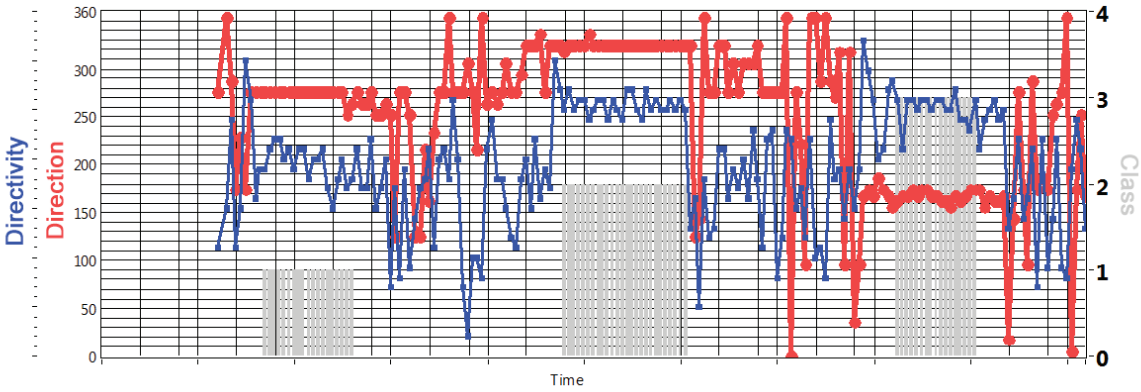

Figure 5 - Example of the output of the proposed system: Noise is classified only when the direction is stable and/or immission directivity is high enough. Three different noise sources are recognized based on extracted features

4. EXPERIMENTAL RESULTS

To test the prototype system, we performed a typical measurement of environmental noise near the Port of Koper, Slovenia. Multiple noise sources at the measurement point (MP) were expected as seen in Figure 6. The system was left unattended for approx 5 hours.

Figure 6 – Measurement point near the Port of Koper, (15).

Classification with SOM was performed once the recording has been finished. We chose a 3 by 2 neuron lattice as we expected at least three dominant sources around the MP (Ship MSC Katrina, ship NEPTUNE and the local traffic of a nearby road). The classification of the 5-hour data was carried

out instantly because of the spatial filtering. Only data points where the noise source direction is stable and the immission directivity is high enough are run through the SOM. The performance of the localization algorithm was excellent. Frequencies observed in the measurements had far larger wavelengths than the distance between four microphones. The detected direction of the sound emission proved to be stable and accurate, which is clearly seen in Figure 8.

The results showed three main noise sources as predicted: the two ships and the nearby road. The classification can be seen in Figure 7. Classes 2, 4 and 6 suggested residual noise sources with random direction and immission directivity. We can see that immission directivity varies from source to source as it is directly linked to the noise source's frequency range. In Figure 8a we can observe the detection of a vehicle as the direction goes from 100° down to 0° and then to around 300°. The vehicle, in this case, passed the MP from right to left according to Figure 7. In cases of ships MSC Katrina and Neptune (Figure 8b and 8c), the direction was stable as one would expect. Even though they were anchored almost one behind the other in respect to the MP, the system was able to differentiate between them. The features used for classification performed well and expand the use of spatial filtering: even though the dominant noise sources may share the same direction, they can be detected individually. Figure 8d depicts no particular dominant noise source as the immission directivity is low and the direction is not stationary. These data points were not included in the SOM classification, thus no class is assigned.

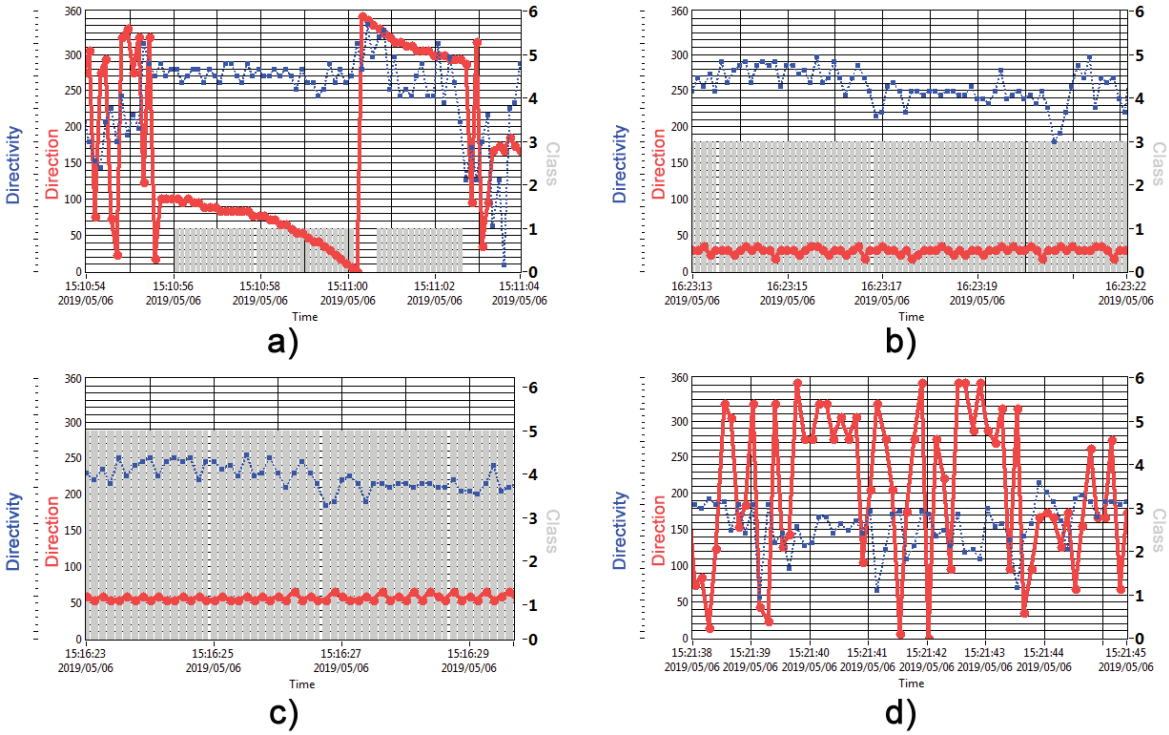

Figure 8 – Automatic classification od different noise sources using SOM: a) Traffic on a nearby road, b) MSC Katrina, c) Neptune and d) residual noise

The results clearly show the advantages of the proposed system. It can classify periods of time where there is a dominant noise source. These ques can be paired with simultaneous SLM measurements, therefore the contribution of each individual noise source to the total noise level can be calculated. The use of SOM proved that it can classify different noise sources quicker as there is no need for any kind of training process. Only a couple of minutes of was needed to tune the parameters and perform the classification, instead of hours of manual work by either pressing the trigger on SLM or post-processing with "pause" and "back erase".

5. CONCLUSION

The proposed system proved to be effective for determening the contributions of individual noise sources to the overall noise level, especially for long term measurements. A typical environmental noise measurement was performed and three noise sources were detected and classified. Further development of the presented system should be carried on. Selection of classes sould be simpler, integration of more indicators for determination of how dominant the noise source is at a give time should be implemented and immision directivity still needs to be improved. Upgrading the direction algorithm for the purpose of detection of multiple noise should be done in the next iteration of the presented system. Full integration with the SLM is also one of the priorities. Further testing of the system is necessary, which will give us a better understanding of its shortcomings and suitability for integration to other applications, such as production line quality control, cavitation detection, acoustic imaging, etc.

REFERENCES

1. Maijala PP. A measurement-based statistical model to evaluate uncertainty in long-range noise assessments [Doctoral dissertation]. P.O. Box 1000, FI-02044 VTT, Finland: Tampere University of Technology; 2013.
2. ISO. Standard ISO 1996-2:2007. Acoustics – description, measurement and assessment of environmental noise – Part 2: Determination of environmental noise levels; 2007.
3. J. Prezelj, J. Murovec. Traffic noise modelling and measurement: inter-laboratory comparison. Appl Acoust, 127 (2017), pp. 160-168
4. P. Mioduszewski, J.A. Ejsmont, J. Grabowski, D. Karpinski. Noise map validation by continuous noise monitoring. Appl Acoust, 72 (2011), pp. 582-589.
5. M. Cowling and R. Sitte, “Comparison of techniques for environmental sound recognition,” Pattern Recognition Letters. 24, 2895–2907 (2003).
6. Murovec, J., Prezelj, J., Čurović, L., & Novaković, T. (2018). Microphone array based automated environmental noise measurement system. Applied Acoustics, 141, 106-114.
7. X. Valero and F. Alias, “Hierarchical Classification of Environmental Noise Sources Considering the Acoustic Signature of Vehicle Pass-Bys,” Archives of Acoustics. 37-4, 423–434 (2012).
8. J. Salamon, J. P. Bello, "Feature learning with deep scattering for urban sound analysis", Proc. 2015 23rd Eur. Signal Process. Conf., pp. 724-728, Aug. 2015.
9. Salamon, J., & Bello, J. P. (2017). Deep convolutional neural networks and data augmentation for environmental sound classification. IEEE Signal Processing Letters, 24(3), 279-283.
10. Sigtia, S., Stark, A. M., Krstulović, S., & Plumbley, M. D. (2016). Automatic environmental sound recognition: Performance versus computational cost. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 24(11), 2096-2107.
11. Piczak, K. J. (2015, September). Environmental sound classification with convolutional neural networks. In 2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP) (pp. 1-6). IEEE.
12. Haghmaram, R., Aroudi, A., Ghezel, M. H., & Veisi, H. (2012, September). Automatic noise recognition based on neural network using LPC and MFCC feature parameters. In 2012 Federated Conference on Computer Science and Information Systems (FedCSIS) (pp. 69-73). IEEE.
13. Pavlović, M., Zajić, G., Zajeganović, M., Ristić, D., Reljin, I., & Mijić, M. (2018, November). Classification of Room Impulse Responses using Kohonen Neural Network. In 2018 14th Symposium on Neural Networks and Applications (NEUREL) (pp. 1-5). IEEE.
14. Zwicker E, Fastl H. Psychoacoustics: Facts and models. Vol. 22. Springer Science & Business Media; 2013.ise series, EAA, Tampere, Finland; 2006, p. 1–6.
15. Slovenian Environment Agency, https:// http://gis.arso.gov.si/atlasokolja ; 2019