Available online at www.sciencedirect.com

ScienceDirect

Energy Reports 6 (2020) 223–230

www.elsevier.com/locate/egyr

Tmrees, EURACA, 25 to 27 June 2020, Athens, Greece

Monitoring, profiling and classification of urban environmental noise using sound characteristics and the KNN algorithm

Eleni Tsalera[a], Andreas Papadakis[b], Maria Samarakou[a,*]

[a] *Department of Informatics and Computer Engineering, University of West Attica, Agiou Spyridonos, Egaleo, 12243, Greece*
[b] *Department of Electrical and Electronics Engineering Educators, School of Pedagogical and Technological Education, Athens, 14122, Greece*

Received 1 August 2020; accepted 30 August 2020

Abstract

Environmental noise is a key factor affecting the quality of life in modern societies as they influence an extended set of human activities. Unwanted sounds, typically characterized as noise, can be of many types and vary in their impact and the ways to be confronted on behalf of competent public authorities. In this work, we describe environmental noise in a qualitative manner using sound-specific features from the time and spectral domains. These features consist of 8 temporal (including RMS, standard deviation, Zero Crossing Rate), 11 spectral (including spectral envelope slope, skewness, spectrum mass center, peak amplitude crest, spread and skewness) and 4 perceptual (including Mel Frequency Cepstral Coefficients) descriptors. Based upon a set of 8 discriminant types of unwanted sounds, typically met in urban environments (car horn, children playing, dog barking, drilling, engine idling, jack hammer, siren and street music), we specify a methodology of matching environmental noise into these categories. Using training and test data from the UrbanSound8K public dataset, we use the K-Nearest Neighbors (KNN) algorithm for classification. The algorithm has been configured to allow from 1 to 3 neighbors, while three distance metrics (Euclidean, Chebyshev and cosine) have been employed to create 9 models that achieve performance between 70% and 85%.

Keywords: Urban noise; Sound classification; Supervised machine learning; KNN algorithm

1. Introduction

Noise is an unwanted or objectionable sound and environmental noise has negative impact upon human activities. While such impact varies among people, sounds of increased intensity and duration can damage hearing capabilities and lead to hearing loss. Noise effects have been studied in detail in the past [1,2] identifying a series of adverse effects including hearing impairment (i.e. the increase in the threshold of hearing), interference with speech communication, sleep disturbance, physiological (including stress, arousal response and cardiovascular effects) and

[*] Corresponding author.
E-mail address: marsam@uniwa.gr (M. Samarakou).

https://doi.org/10.1016/j.egyr.2020.08.045

mental health and behavioral effects. Noise, even of low intensity, has negative impact in the cases where mental work is being performed or tranquility is pursued such as educational institutes and hospitals.

In Europe, environmental noise is recognized as one of the main local environmental problems [3]. It is also recognized that the problem of environmental noise is difficult to confront on behalf of public authorities due to the limited elaboration (or even lack) of fine-grained quantitative metrics. Noise measurements include the intensity (or loudness, measured in decibel) and the duration (measured in milliseconds), which can determine the impact on the human hearing system. According to WHO in outdoor living area, the moderate annoyance is defined at 50 dB while the serious annoyance is at 55 dB. The respective figures for indoors and dwelling are at 35 dB and 30 dB [1]. The types and the characteristics of the noise pollution vary, and noises of similar intensity have different impact as well as cause and origination. They can be confronted with different ways. In this respect the capability of better understanding the characteristics of the noise are important for better confrontation and decision making. While the existing sound-related frameworks mainly consider the intensity level of the noise, further processing and understanding of the noise types is still under investigation. Noise categorization can allow for more efficient management and decision making.

The ***objectives*** of our work include:

(a) The investigation of the characteristics of selected urban environmental noises and their statistical noise profiling based on a set of sound-related properties (features).
(b) The selection of sound features to extract for eight noise types (car horn, children playing, dog bark, drilling, engine idling, jack hammer, siren and street music), in order to perform noise classification using the KNN (K-Nearest Neighbors) machine learning algorithm.
(c) The evaluation of the results employing noise samples from a publicly available sound database.

The *structure* of the paper is the following: Section 2 considers the contribution of sound in noise pollution and briefly discusses similar work in the field. Section 3 describes our methodology including the selected types of urban environmental noise, the features used for their description and their profiling based upon key descriptors. The KNN algorithm is briefly discussed as well. Section 4 discusses the usage of the classification algorithm with different parameters values, forming nine scenarios and the results that have been achieved when applied upon the testing set. The paper includes the conclusions and the framework for future work.

2. Similar work on environmental noise

To assess and subsequently manage environmental noise, the European Commission issued the Environmental Noise Directive (END)2002/49/EC. According to this framework, main cities are expected to measure noise exposure and provide accurate mappings of noise pollution levels to support decision making and trigger further actions [4,5]. The strategic noise mapping has been endorsed by the World Health Organization. The determination of noise levels has been based upon the maximum intensity to assess (a) annoyance and (b) sleep disturbance. The intensity has been calculated from the sound pressure level and has been harmonized with reference values. Noise mapping encompasses noise contour mapping and estimation of exposure.

The intensity of the noise is extracted and represented in a straightforward manner, but it is considered only a basic feature in terms of Psycho-Acoustic Annoyance (PA), as frequency characteristics are not included. Noise of similar intensity can lead to different perception failing to provide information related to the subjective annoyance and their psychoacoustic properties [6]. Towards this direction, Zwicker annoyance model considers additional factors including loudness, sharpness, fluctuation strength and roughness [7,8]. In this view, an elaborate analysis of the sound characteristics is needed, including an extensive and well-organized set of sound-related features that allow further insight in sound properties.

2.1. Similar work

In the context of *smart city* efforts, infrastructure for retrieving noise measurements (based on Wireless Acoustic Sensor Network, WASN) and further acoustic environment description have been pursued under various contexts. The processing of the measurements is increasingly extending beyond intensity monitoring into more elaborate analysis and identification of noise types. Alsouda et al. [6] have retrieved sounds using Raspberry Pi Zero W and performed classification using the MFCC features as fed into the SVM and KNN supervised learning algorithms.

Mitilineos et al. [9] has developed a two-level sound classification (based on neural networks) for sound monitoring in the context of monitoring and managing cultural heritage.

Kong et al. [10] has performed time–frequency (T–F) segmentation and classification framework trained upon weakly labeled audio data using convolutional neural network and global weighted rank pooling. Piczak [7] has employed convolutional neural networks of two fully connected layers for classifying short audio clips of environmental sounds, trained upon a low-level representation of audio data (segmented spectrograms) with deltas.

Detection of audio events in noisy environments with the intensity of background noise comparable to audio events has been performed using the bag of words approach [8]. In such efforts, quantitative evaluation of the detection accuracy of sound event analysis systems is performed through comparison of system output with reference test data [11,12].

3. Methodology and implementation

3.1. Methodology

Our methodology includes the following steps: Initially, we select the noise types (classes) of interest, i.e. sounds met in urban environments. Labeled sound excerpts of different types of sound has been preprocessed (including normalization and segmentation) resulting into the training and the testing sets. Then we select the sound characteristics (features) are extracted from the training files and after standardizing their values, they are fed into the KNN classification algorithm. The classification algorithm creates a set of classification models, enabling fine-tuning of the algorithm (in terms of distance type used and number of neighbors considered).

Using the model, classification is performed upon the testing data. The evaluation takes place through comparison of the estimation with the ground truth. (See Fig. 1.)

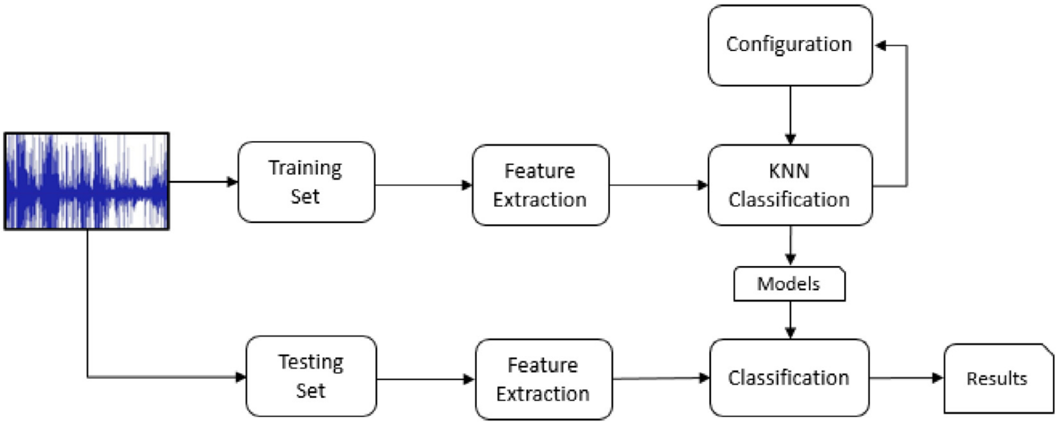


Fig. 1. The audio features of the training files are extracted and fed into the KNN model for training the classification model. Classification is then performed upon the testing files. The evaluation is performed with comparison of the classification results with the ground truth.

In the following subsections we describe in more detail the steps of our methodology.

3.2. Noise classes and preparation of training and testing sets

High-level classification of environmental noise into discrete depends on multiple criteria. The rate of intensity changes allows discrimination into *continuous*, i.e. of relatively stable intensity, *intermittent* where the intensity increases and decreases rapidly, *impulsive* characterized by a sudden change of intensity and the *low-frequency* noise. Another criterion is the origination, i.e. whether the sound comes from anthropogenic or other activities.

In our work, we consider 8 different types (classes) of noise, as depicted in Table 1, that are frequently appearing in urban areas and burden urban noise pollution.

We have used the *UrbanSound8K* [13] audio dataset, which is a subset of Freesound collaborative database with more than 400,000 sounds [14]. The selected sounds are audio recordings of street scenes with various levels of

Table 1. The noise types we consider include car horn, children playing, dog barking, drilling, engine idling, jack hammer, siren and street music. They are characterized based upon the intensity change rate and their origination.

Classes	Class name	Intensity change rate	Activity origin
Car horn	1	Impulsive	Transportation
Children playing	2	Intermittent	Human activities
Dog barking	3	Intermittent	Animal
Drilling	4	Continuous	Industrial
Engine idling	5	Continuous	Transportation
Jack hammer	6	Intermittent	Industrial
Siren	7	Intermittent	Transportation
Street music	8	Intermittent	Human activities

traffic and human activity. The audio excerpts have been recorded in different outdoor locations in residential areas and city center.

Based on the training audio sets, a 64-minute audio file has been created with each audio class lasting 8 min. The sampling rate has been 44.1 kHz. The recordings of each sound type have been made under different condition. 80% of the audio file has been selected for training file and the remaining 20% for testing. The audio classes have been shuffled per second so that the same class would not be displayed in successive seconds.

3.3. Sound event profiling

Audio content analysis [15] is based upon the extraction of descriptive sound characteristics to perform automatic identification. Descriptors are spectral, temporal, and perceptual (correlated with the human auditory system). For the audio analysis we have extracted 8 temporal descriptors, 11 spectral descriptors and 4 perceptual descriptors. Temporal descriptors include time RMS, time standard deviation, time Zero Crossing Rate [16]. Spectral descriptors include slope of the decrease of the spectral envelope, skewness [15,17,18], centroid calculating the mass center of the spectrum [19], crest as the peak amplitude of the spectrum [20], spread and skewness.

We have also considered the sound spectrum in the range of [133, 6854] Hz as 40 Mel Frequency Cepstral Coefficients (MFCCs). Mel frequency scale models the human auditory system by mapping the actual frequency to the perceived pitch [2,17,21]. For example, Mel-3 which is included in Table 2 is the coefficient of Mel-scale for the frequency range of [267, 400] Hz. In total 62 features' values have been extracted for non-overlapping windows of 1 second length. Table 2 presents the average values and the Coefficients of Variation (CoV) for 10 features per class. The Coefficient of Variation (also known as Relative Standard Deviation, RSD) is defined as the ratio of the standard deviation to the mean. The mean value of each feature has been calculated before standardization. The CoVs are calculated to compare the distributions of the values [22].

$$CV = \frac{s}{|\overline{x}|}, \text{ where } s \text{ is the standard variation and } |\overline{x}| \text{ is the absolute value of the mean} \tag{1}$$

Large value of CoV indicate dispersion around the mean value, and it is due to the choice of a wide variety of sounds per class (i.e. recordings from many different car horns or children's voices).

A basic discrimination among sound can be identified through different mean values for specific classes as depicted in Table 2. For example, classes 3 and 4 can be separated, due to distinct average values in the majority of the parameters. The CoV indicate the uniformity or dispersion of values around the mean. The class with the highest uniformity is the 6th due to the fact that jack hammers produce a similar sound. The selection of features in terms of discriminative importance is still an open issue in classification.

3.4. Machine learning for classification

KNN machine learning algorithm is a widely used classification method based on distance metrics. A testing point is represented as an n-dimensional vector, where n is the number of parameters that characterize it. Its class

Table 2. The average value and the Coefficients of Variation (CoV) of 10 features per class. The values of CoV indicate the dispersion around the mean value per class. The concentration of values around the mean per sound class for a specific feature is an indication of the discriminative power of the specific feature.

Classes		Class 1: Car horn	Class 2: Children playing	Class 3: Dog barking	Class 4: Drilling	Class 5: Engine idling	Class 6: Jack hammer	Class 7: Siren	Class 8: Street music
Features									
Skewness	Mean	4.84	5.12	6.25	3.76	7.30	2.76	5.27	5.77
	CoV	21.5	16.5	19.6	19.4	24.4	17.7	25.8	18.4
Centroid	Mean	1081	965	825	1668	303	3985	1169	543
	CoV	59.5	50.4	54.8	36.9	91.2	6.5	30.9	60.7
Time ZCR	Mean	0.06	0.07	0.05	0.09	0.04	0.20	0.07	0.04
	CoV	40.7	24.9	43.0	20.6	57.3	8.2	31.4	53.6
Crest	Mean	0.15	0.15	0.22	0.10	0.26	0.06	0.16	0.19
	CoV	34.3	24.2	30.3	21.3	37.8	16.5	48.5	33.8
Kurtosis	Mean	29	32	45	17	63	10	34	40
	CoV	48.7	34.1	39.4	50.1	40.4	40.9	49.8	32.9
Spread	Mean	767	817	473	861	598	2370	801	598
	CoV	30.7	23.4	66.1	17.4	57.5	8.5	58.7	59.6
Decrease	Mean	−0.31	−0.29	−0.31	−0.09	−1.78	−0.01	−0.06	−0.58
	CoV	199	150	297	267	62	111	246	109
Time std	Mean	0.15	0.12	0.14	0.14	0.19	0.07	0.15	0.15
	CoV	28.6	27.3	31.1	10.1	31.8	13.3	29.3	29.9
Time RMS	Mean	−16.9	−19.5	−19.3	−17.2	−15.1	−23.6	−17.4	−17.3
	CoV	16.2	17.4	16.1	10.1	21.6	5.0	18.9	15.2
MEL 3	Mean	2.06	1.84	3.15	0.99.1	2.11	−0.59	2.27	3.13
	CoV	46.5	40.8	38.9	115	10.9	32.7	29.5	41.2

is decided upon the distance of its neighbors. To configure the algorithm, the following parameters have to be decided:

- The distance metric to be applied (indicatively Euclidean, Chebyshev and cosine).
- The number of neighbors to be considered.

While the Euclidean distance is calculated by the sum of the distances of the individual coordinates, the Chebyshev distance is the maximum of these individual distances. On the other hand, cosine distance is calculated from the cosine of the angle between the individual vectors through the inner product and it is a measure of similarity when the vectors are parallel to each other or conversely a measure of disparity when the vectors are perpendicular to each other [23].

$$Euclidean\ distance = \sqrt{\sum_{i=1}^{n}(r_i - t_i)^2} \tag{2}$$

$$Chebyshev\ distance = \max_{i}\{|r_i - t_i|\} \tag{3}$$

$$Cosine\ distance = \left(1 - \frac{\vec{r_i}\,\vec{t_i}}{\left|\vec{r_i}\right|\left|\vec{t_i}\right|}\right) \tag{4}$$

where r_i and t_i are the training vector and the testing vector respectively in each time window. Each training and testing point have n-coordinates$(x_1, x_2, x_3, \ldots, x_n)$, resulting from the values of the extracted features, which in our case$n = 23$.

The predicted class is determined according to the most frequently encountered class of these neighbors when these are more than one. We configured the algorithm to use the distance metrics Euclidean, Chebyshev and cosine (Eqs. (2)–(4)) with k = 1, 2 and 3 neighbors, resulting in 9 models.

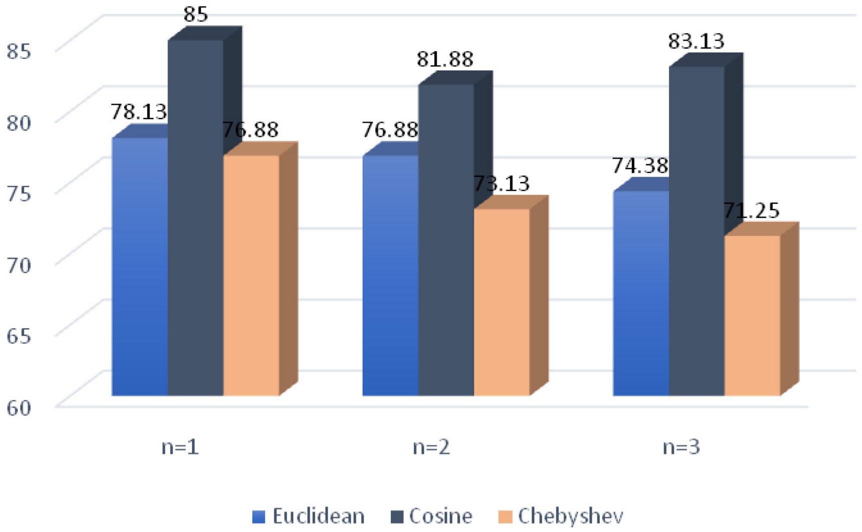
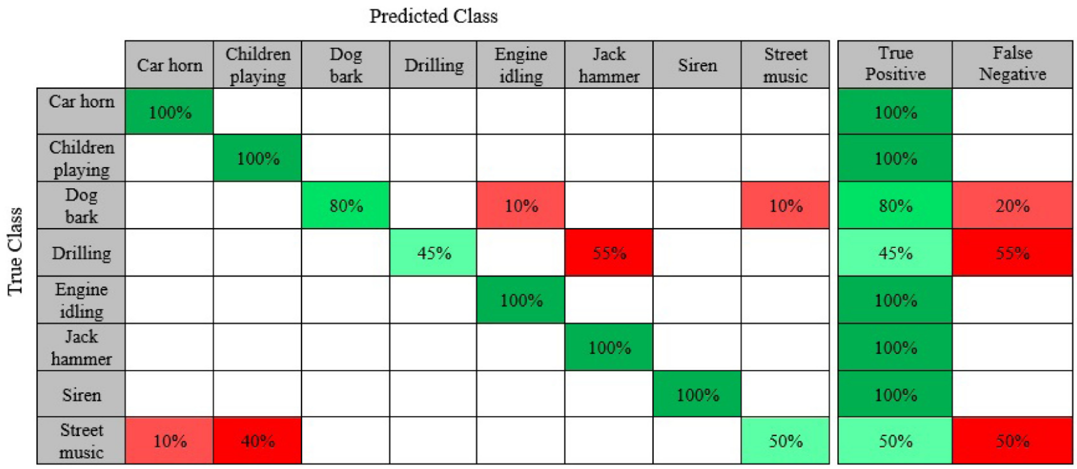


Fig. 2. Comparative illustration of the performance (vertical axis) of the scenarios under consideration (horizontal axis). Each distance metric has been applied with n = 1, 2 or 3 nearest neighbors and each combination is considered as a different scenario resulting in 9 different models.

4. Evaluation

The results are summarized in Fig. 2. The best prediction rate (85%) has been achieved using neighbor with the distance metric to be the cosine. The cosine metric performs better in all neighbor-cases while comparing the number of the neighbors the usage of one neighbor is the best.

For the evaluation we consider the *true positive* percentage, where the classifier correctly identifies the event and the *true negative* where the classifier correctly rejects the sample as the reference does not indicate an event either. In the *false positive* the classifier erroneously identifies the sample as a relevant event, while in *false negative* the classifier rejects the sample, although the reference indicates an event. The confusion matrix for the classification model with 1 neighbor and cosine distance metric is shown in Fig. 3.



Predicted Class

True Class

	Car horn	Children playing	Dog bark	Drilling	Engine idling	Jack hammer	Siren	Street music	True Positive	False Negative
Car horn	100%								100%	
Children playing		100%							100%	
Dog bark			80%		10%			10%	80%	20%
Drilling				45%		55%			45%	55%
Engine idling					100%				100%	
Jack hammer						100%			100%	
Siren							100%		100%	
Street music	10%	40%						50%	50%	50%

Fig. 3. Confusion matrix for 1-NN classification model with cosine distance. With green are marked the correct prediction percentages, while with red are marked the mistaken prediction percentages, e.g. class 4 (drilling) has been identified correctly in 45% while it has been incorrectly classified as class 5 (jack hammer) in 55%. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

For the classes of car horn, children playing, engine idling, jack hammer and siren, we have 100% true positive, while drilling and street music are often misclassified with jack hammer and children-playing respectively.

5. Conclusions — Future work

In this work we have investigated the aspect of urban noise as factor of environmental pollution and its negative impact upon human activities. Starting from the typical measurement related to noise, i.e. related to the intensity as expressed in dB, we have considered a set of features (descriptors) based on the time, frequency and perceptual domains. A set of eight discrete noise types have been analyzed based on these descriptors and trained a KNN-based model using a set of test excerpts. The classification system has been used and achieved ratio of 70% to 85%.

In terms of future work, we consider the functionality of Sound Event Detection in which a continuous sound stream is separated into specific parts of sound events. This functionality has been performed in our case in manual way, while efforts are taking place to fully automate it. Further investigation of the impact of the features in terms of number and selection is also an interesting point for our future work considering that IoT sensors and equipment are typically of low complexity and processing capabilities and the number of features used should be rationalized.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

Part of this work has been performed and supported in the MS “Master in Energy” program, which has been executed in the University of West Attica (former TEI of Athens).

References

[1] Berglund B, Lindvall T, Schwela DH. Guidelines for community noise. World Health Organization Occupational and Environmental Health Team; 1999, http://dx.doi.org/10.1260/0957456001497535.

[2] Goyal N, Purwar RK. Analyzing mel frequency cepstral coefficient for recognition of isolated english word using DTW matching. Int J Res Comput Commun Technol 2014;3(4):436–43.

[3] European Commission. Future noise policy. Green paper, Commission of the European communities; 1996. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:51996DC0540&from=EN.

[4] Murphy E, King EA. Strategic environmental noise mapping: Methodological issues concerning the implementation of the EU Environmental Noise Directive and their policy implications. Environ Int 2010;36(3):290–8. http://dx.doi.org/10.1016/j.envint.2009.11.006.

[5] Kephalopoulos S, Paviotti M, Anfosso-Lédée F, Van Maercke D, Shilton S, Jones N. Advances in the development of common noise assessment methods in europe: The CNOSSOS-EU framework for strategic environmental noise mapping. Sci Total Environ 2014;482:400–10. http://dx.doi.org/10.1016/j.scitotenv.2014.02.031.

[6] Alsouda Y, Pllana S, Kurti A. A machine learning driven IoT solution for noise classification in smart cities. 2018, arXiv preprint arXiv:1809.00238.

[7] Piczak KJ. Environmental sound classification with convolutional neural networks. 2015 IEEE 25th International Workshop on Machine Learning for Signal Processing, MLSP: IEEE; 2015. p. 1–6. https://doi.org/10.1109/MLSP.2015.7324337.

[8] Foggia P, Petkov N, Saggese A, Strisciuglio N, Vento M. Reliable detection of audio events in highly noisy environments. Pattern Recognit Lett 2015;65:22–8. http://dx.doi.org/10.1016/j.patrec.2015.06.026.

[9] Mitilineos SA, Potirakis SM, Tatlas NA, Rangoussi M. A two-level sound classification platform for environmental monitoring. J Sensors 2018;2018. http://dx.doi.org/10.1155/2018/5828074.

[10] Kong Q, Xu Y, Sobieraj I, Wang W, Plumbley MD. Sound event detection and time–frequency segmentation from weakly labelled data. IEEE/ACM Trans Audio Speech Lang Process 2019;27(4):777–87. http://dx.doi.org/10.1109/TASLP.2019.2895254.

[11] Mesaros A, Heittola T, Virtanen T. Metrics for polyphonic sound event detection. Appl Sci 2016;6(6):162. http://dx.doi.org/10.3390/app6060162.

[12] Papadakis A, Tsalera E, Samarakou M. Survey on sound and video analysis methods for monitoring face-to-face module delivery. Int J Emerg Technol Learn (iJET) 2019;14(08):229–40. http://dx.doi.org/10.3991/ijet.v14i08.9813.

[13] Salamon J, Jacoby C, Bello JP. A dataset and taxonomy for urban sound research. In: Proceedings of the 22nd ACM international conference on multimedia; 2014. p. 1041–44.https://doi.org/10.1145/2647868.2655045.

[14] Collaborative repository of licensed audio, http://www.freesound.org.

[15] Lerch A. An introduction to audio content analysis: applications in signal processing and music informatics. Wiley-IEEE Press; 2012.

[16] Bachu RG, Kopparthi S, Adapa B, Barkana BD. Separation of voiced and unvoiced using zero crossing rate and energy of the speech signal. In: American society for engineering education (ASEE) zone conference proceedings; 2008. p. 1–7.

[17] Dave N. Feature extraction methods LPC, PLP and MFCC in speech recognition. Int J Adv Res Eng Technol 2013;1(6):1–4.

[18] Donnelly PJ, Blanchard N, Olney AM, Kelly S, Nystrand M, D'Mello SK. Words matter: automatic detection of teacher questions in live classroom discourse using linguistics, acoustics, and context. In: Proceedings of the seventh international learning analytics & knowledge conference; 2017. p. 218–27. https://doi.org/10.1145/3027385.3027417.

[19] Serizel R, Bisot V, Essid S, Richard G. Acoustic features for environmental sound analysis. In: Computational analysis of sound scenes and events. Cham: Springer; 2018, p. 71–101. http://dx.doi.org/10.1007/978-3-319-63450-0_4.

[20] Peeters G. A Large set of audio features for sound description (similarity and classification) in the CUIDADO project. CUIDADO IST project report, 54(0), 2004, p. 1–25.

[21] Xu M, Duan LY, Cai J, Chia LT, Xu C, Tian Q. HMM-Based audio keyword generation. In: Pacificrim conference on multimedia. Berlin, Heidelberg: Springer; 2004, p. 566–74. http://dx.doi.org/10.1007/978-3-540-30543-9_71.

[22] Heumann C, Schomaker M. Introduction to statistics and data analysis. Springer International Publishing Switzerland; 2016.

[23] Rosen KH. Handbook of discrete and combinatorial mathematics. CRC press; 1999.