

Automatic Detection of Long-Term Audible Noise Indices from Corona Phenomena on UHV AC Power Lines

T. WSZOLEK* AND M. KŁACZYŃSKI

AGH University of Science and Technology, Faculty of Mechanical Engineering and Robotics,
Department of Mechanics and Vibroacoustics, Al. A. Mickiewicza 30, 30-059 Krakow, Poland

One of the most important tasks in outdoor acoustic monitoring stations is automatic extraction of the measured signal parameters. In case of corona discharge noise from ultra-high voltage alternating current (UHV AC) power lines it is necessary to select properly the corona audible noise (CAN) parameters to be monitored for noise indicators calculation, as the monitored signal and the background noise strongly fluctuate. A combined selection of distinctive features of CAN is necessary in order to distinguish the actual signal from the external interference. The vast amount of recorded data is difficult to store and process. Therefore, particular attention was devoted to define of the collected parameters used for automatic calculation of the CAN long-term noise indicators. In addition, several new CAN parameters were introduced, including spectral moments, spectral coefficients of tonal components contribution, and power coefficients in selected frequency bands; as it allowed more efficient selection of samples with non-zero contribution from CAN. The artificial neural network was applied for final classification of the measured samples. Selected and properly filtered samples provided the basis for calculations of long-term noise indicators. Efficiency of the said method was tested for the measurement sections with the recorded sound signal and aural qualification of the CAN intensity.

DOI: [10.12693/APhysPolA.125.A-93](https://doi.org/10.12693/APhysPolA.125.A-93)

PACS: 43.60.Lq, 07.05.Mh

1. Introduction

The acoustic signal of corona discharge, as well as the corona discharge process itself, is a random phenomenon by its nature and it strongly depends on weather conditions and the actual condition of the conductor's surface which are also affected by many random factors [1]. It leads to a number of problems regarding estimation of basic noise indicators — L_{AeqT} and L_{den} defined in ISO 1996-2:2007 — used in corona audible noise evaluation [2–5] when they are based on short-term measurements and one needs to expect the risk of time samples not being representative enough, as well as the long-term measurement (continuous monitoring) when the risk of the indicators overestimation because of insufficient filtration of external interference occurs, especially in the periods when the S/N ratio is not very high. For the case of power transmission lines, such a situation is often encountered in fair weather conditions, when the corona audible noise (CAN) is rather weak, while the environmental interference is above normal, e.g. due to intensive agricultural activity. Extraction of CAN becomes particularly important for continuous monitoring systems when the risk of the long-term indicator being biased due to environmental interference occurs, such risk being difficult to eliminate. In order to optimize interference filtration, it is necessary to register the noise spectra in the acoustic frequency band with such registration step adequate

for local prevailing at the monitoring site, so that on one hand the information is not lost (if the registration step is too long) and on the other, the problems with processing of excessive data files for too short registration step can be avoided. However, it is always recommended to adopt a system able to automatically determine some noise parameters in real time and classify the registered data with an extra possibility to record time-marks (time-stamps) when essential disturbances occur. It is necessary not only for determination of the long-term indicators but also for the estimation of partial uncertainties of the indicators extraction at various levels of external interference.

In authors' research focused on CAN detection, the statistical levels measurements in 1/3 octave bands [6] have been used as a preliminary method for external interference filtering, as said method was extensively tested in continuous monitoring systems. In the course of further processing, the distinctive features of CAN are determined which are useful in automatic detection of such a signal. In order to increase the efficiency of such an approach, in addition to the spectral moments implemented in the previous papers [3, 6, 7], some new parameters of the acoustic signal have been introduced [4], namely the coefficients of tonal component contributions and the coefficients of signal power registered in the tonal and noise bands. Verification of usability of individual parameters in the feature vector and the degrees of their correlation were carried out using the cluster analysis.

In the paper [4], categorization of registered samples into one of two classes related to "1" (corona effect) and "0" (no corona effect) was carried out using the Support Vector Machines technique. In this study, an artificial

*corresponding author; e-mail: tadeusz.wszolek@agh.edu.pl

neural network (Multi-Layer Perceptron) has been used for automated classification of CAN. Selection of the network architecture, as well as the neuron activation function and the network training method, were carried out experimentally. Two variants of the training vector were used, composed of either four or two CAN categories. Depending on environmental conditions (occurring disturbances), different number of categories of CAN may prove more useful. Also two feature vectors consisting of 19 or 10 acoustic and non-acoustic parameters were used, representing distinctive CAN features.

2. Research methodology

In this paper and in the present authors' previous papers [3, 4, 11], recordings from a mobile continuous acoustic monitoring station were used for detection and identification of CAN. The monitoring measurements were carried out in several locations near UHV AC transmission lines, e.g. near a 400 kV double circuit line with double ($2 \times 525 \text{ mm}^2$) and a triple ($3 \times 350 \text{ mm}^2$) bundle of subconductors. The monitoring station registered noise parameters in the mode of sound level continuous spectrum measurement in 1/3 octave bands in the frequency range between 20 Hz and 20 kHz. Additionally, A-, C-, and Z-weighted sound levels were registered. The averaged spectrum data was recorded to a buffer-type file every 10 s, while the other data including L_{Aeq} values and statistical levels (L_{A50} , L_{A90}) of the 1/3 octave spectrum and the A-, C-, and Z-weighted sound levels have been recorded every 15 minutes. The registered data were used for extraction of distinctive features of CAN, and further for determination of the short- (L_{AeqT}) and long-term (L_{den}) noise indicators used for assessment of the corona noise nuisance. Large number of details, as carried by results in spectral analysis, complicates interpretation and recognition of the essential information contained in the registered signal. Therefore, from the spectrum (frequency domain) or multi-spectrum (time and frequency domain), specific (distinctive) features are determined that are useful for a specific problem of interest. In the case of CAN spectrum, the following parameters related to its density distribution are determined which can be considered distinctive features:

- normalized (to M_0) spectral moments of the 1st and 2nd order (M_1 , M_2) for the whole frequency band (20 Hz to 20 kHz),
- normalized (to M_{w0}) spectral moments of the 1st and 2nd order (M_{w1} , M_{w2}) for the frequency band characteristics for the wind effects (20 Hz to 400 Hz),
- normalized (to M_{c0}) spectral moments of the 1st and 2nd order (M_{c1} , M_{c2}) for the frequency band characteristic to the broadband part of CAN (1 kHz to 10 kHz).

Normalized spectral moment of the m -th order is described by a general relation is defined as:

$$M_m^*(t) = \frac{\sum_{i=0}^N |G_n(f_i)| [f_i]^m}{\sum_{i=0}^N |G_n(f_i)|}, \quad (1)$$

where $G_n(f)$ is the frequency spectrum of the n -th data record, f_i is the central frequency for the i -th frequency band defined for the spectral analysis, f_0 is the lower band frequency, f_N is the upper band frequency (N depends on the bandwidth), and m is the order number.

The effectiveness of such an approach may be increased by narrowing the analyzed frequency bands separately to the "noisy" spectrum part and the tonal spectrum part. Therefore the following parameters have been used:

- PC_1 coefficient of the signal power calculated for the 20 Hz to 400 Hz frequency band where rainfall noise effects and the tonal components of corona noise are encountered;
- PC_2 coefficient of the signal power calculated for the 1 kHz to 10 kHz frequency band where the noise components of corona process signal are located;
- TC_{100} and TC_{200} coefficients useful in detection of 100 Hz and 200 Hz tonal components in the spectra of registered samples [9].

The signal power coefficients of the m -th order (also called the 'relative power coefficients') are determined by means of the formula

$$PC_m = \frac{\sum_{f=f_{\text{lower}}}^{f_{\text{upper}}} G_n(f)}{\sum_{f=20}^{20000} G_n(f)}, \quad (2)$$

where f_{lower} and f_{upper} are the limits of the selected frequency band, and tonal coefficients are calculated according to

$$TC_x = \left. \frac{\partial G_n(f)}{\partial x} \right|_{x=100,200}, \quad (3)$$

where x is the frequency of a given tonal component at 100 Hz and 200 Hz.

In the authors' previous papers [3, 6–8, 10] several versions of the feature vector were presented and proper selection of the feature space was extensively discussed. The space of the feature vector present at the origin of the considerations on its final dimension (19 parameters) is based on the spectral coefficients statistical data of the A-weighted sound level and the meteorological data described as

$$X_n = \langle M_0, M_{c0}, M_{w0}, M_1, M_{c1}, M_{w1}, M_2, M_{c2}, M_{w2}, PC_1, PC_2, TC_{100}, TC_{200}, RH, RF, WS, L_{A90}, L_{A50}, L_{A50(1-10) \text{ kHz}} \rangle, \quad (4)$$

where the parameters not listed earlier are as follows: L_{A90} is the L_{90} statistical level for A-weighted sound level (carrying the information concerning the interference level), L_{A50} is the L_{50} statistical level for A-weighted sound level, $L_{A50(1-10) \text{ kHz}}$ is the L_{50} statistical

level of *A*-weighted sound level determined for the 1 kHz to 10 kHz frequency band, *RF* is the rain fall level (in [mm/h]), *WS* is the wind speed (in [m/s]), *RH* is the relative humidity (in [%]).

3. Research material

The database used in the present study contained 17151 records sampled every 15 minutes, with the imposed recognition qualification thresholds “0” — no corona effect; “0.25” — low corona effect; “0.5” — average corona effect; “1” — high corona effect. The categories of “0.25” and “0.5” can also mean different interference masking distinctive features of CAN, whereas category “1” means intensive and not interfered CAN. This classification was proposed by a subjective decision based on listening to sample audio recordings of the data and was necessary to use artificial intelligence methods (Multi-Layer Perceptron training process). The database included 417 records registered during low rainfall (0–0.5 mm/h) average rainfall (0.5–2 mm/h) and intense rainfall (> 2 mm/h) conditions among which 361 records were marked with wind speed value 0–1.1 m/s. The records with the rainfall accompanied by a low value of wind speed are particularly useful, because of the high intensity of corona effect and low level of external interference (caused by wind).

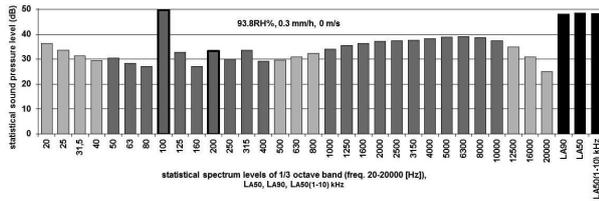


Fig. 1. Spectrum of the high corona effect.

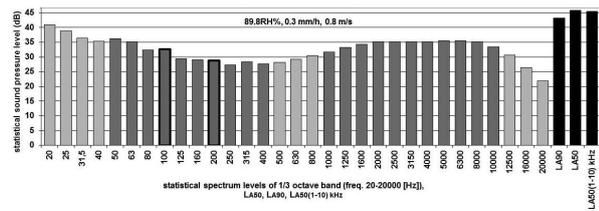


Fig. 2. Spectrum of the average corona effect.

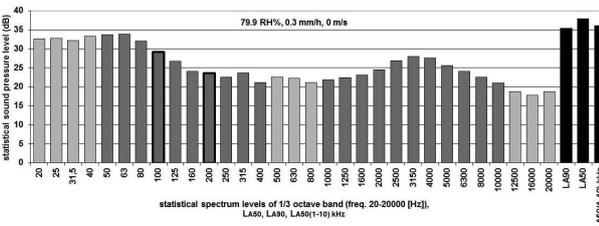


Fig. 3. Spectrum of the low corona effect.

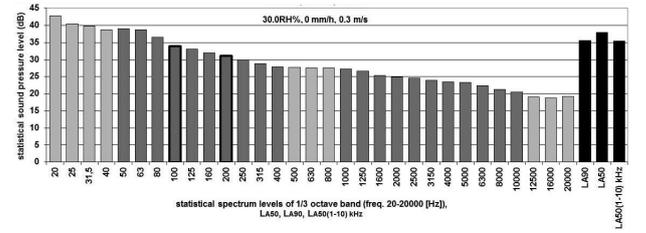


Fig. 4. Spectrum of no corona effect.

Example spectra registered by acoustic monitoring stations are shown in Figs 1–4. Respective values of the feature vector and the CAN classification into four categories: high corona effect, average corona effect, low corona effect, and no corona effect are shown in Table I.

Verification of the selected feature vector’s quality (4) was carried out using the cluster analysis [10]. In the classical cluster analysis, matrix of distances between objects is calculated in order to determine the similarity of these objects. Reversing the assumption, in order to estimate the correctness of feature space choice (the measure of features dissimilarity), the objects were created from the individual elements of the feature vector (Eq. (4)) taking into account the entire database (17151 records). To calculate distance between parameters of the presented feature vector was used the Euclidean metric and for group of the objects - average method. The results of this verification are shown in two figures: dendrogram of the features for intense corona effect situation (Fig. 5) and the remaining situations, including no corona effect conditions (Fig. 6).

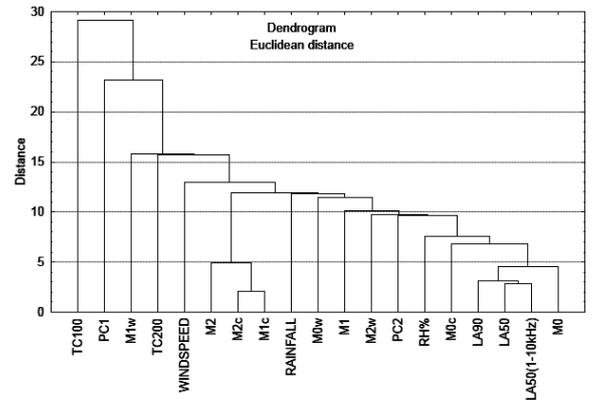


Fig. 5. The results of cluster analysis (dendrogram) for the feature vector selected for an intensive corona process.

As it can be noticed for the case presented in Figure 5, the features which correlate are the following: L_{A90} , L_{A50} and $L_{A50(1-10)}$ kHz, then the group of spectral moments of 0-th order (M_0 , M_{c0} , M_{w0}). Also in the case shown in Fig. 6, high correlation can be noticed for moments L_{A90} , L_{A50} and $L_{A50(1-10)}$ kHz together with M_0 , M_{c0} and then M_{c1} , M_{c2} and M_2 .

TABLE I

Example values of the corona effect feature vector (shown in Figs. 1–4).

Corona effect	M_0	M_{c0}	M_{w0}	M_1	M_{c1}	M_{w1}	M_2	M_{c2}	M_{w2}	PC_1	PC_2	TC_{100}	TC_{200}	RH	RF	WS	L_{A90}	L_{A50}	$L_{A50(1-10)}$ kHz
“1” – high	1045	410	321	630	3150	125	5642	3893	160	0.4	0.4	40	9	93.8	0.3	0	48.3	48.5	48
“0.5” – aver.	896	264	331	400	3150	125	4869	3515	150	0.6	0.2	–2	0	30	0	0.3	35.4	37.9	35.6
“0.25” – low	790	268	270	500	3150	125	5166	3658	144	0.57	0.32	0	0	79.9	0.3	0	35.4	37.9	36.1
“0” – no	896	264	331	400	3150	125	4869	3515	150	0.6	0.2	–2	0	30	0	0.3	35.4	37.9	35.6

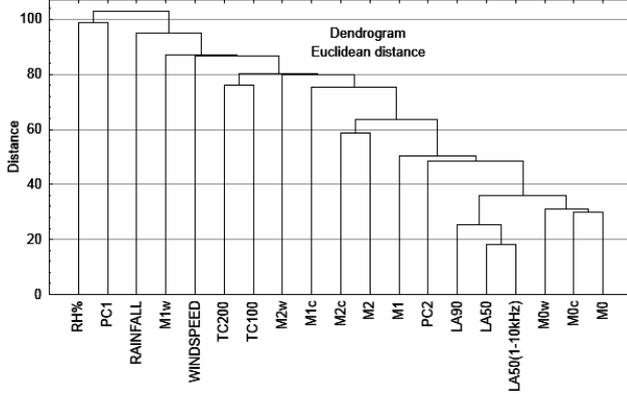


Fig. 6. As Fig. 5 for remaining cases including the acoustic background.

This indicates that features in the above-listed groups duplicate to a high degree the contributed information; therefore it seems justified to leave only one of the features in each group. Ultimately the authors abandoned determination of spectral moments of the same order for various frequency bands, the L_{A50} and $L_{A50(1-10)}$ Hz levels, as well as the M_0 moment, leaving only the L_{A90} level. Thus the 19 dimension space of the feature vector (4) has been reduced to 10 elements and can be finally represented symbolically by

$$X_n = \langle M_1, M_2, PC_1, PC_2, TC_{100}, TC_{200},$$

$$RH, RF, WS, L_{A90} \rangle. \quad (5)$$

4. Artificial neural network classification

An artificial neural network has been used for automated signal classification. Selection of the network architecture (the number of layers and number of neurons in each layer), as well as the activation function and the network training method was made experimentally, in two independent calculation environments, MATLAB and STATISTICA. In the experiment, 4460 records of the collected material were used, of which 70% for the network training and about 30% for verification of the method (of which 15% for validation and 15% for testing).

Two variants of the training vector were used — vector of four classes related to “0” (no corona effect), “0.25” (low corona effect), “0.5” (average corona effect) and “1” (high corona effect) and a vector of two classes related to “1” (high corona effect) and “0” (no corona effect). Two feature vectors composed of 19 or 10 parameters of the feature space were used.

Tables II–V present the top 5 results of the artificial neural network architecture selection for the considered problem in different variants of the feature vector and CAN classes.

TABLE II

Summary of the neural network selection process (19-parameter feature space, 4 classes of CAN).

ID	Network	Quality			Training algorithm	Error function	Activation	
		training	testing	validation			(hidden)	(output)
1	MLP 19-13-4	92.4	95.3	91.1	BFGS 95	SOS	Exponential	Tanh
2	MLP 19-16-4	90.0	93.5	91.1	BFGS 49	SOS	Exponential	Logistic
3	MLP 19-11-4	93.3	87.6	87.8	BFGS 55	SOS	Exponential	Tanh
4	MLP 19-15-4	93.3	86.2	86.7	BFGS 70	Entropy	Linear	Softmax
5	MLP 19-5-4	91.1	86.4	84.4	BFGS 41	SOS	Linear	Tanh

TABLE III

Summary of the neural network selection process (19-parameter feature space, 2 classes of CAN).

ID	Network	Quality			Training algorithm	Error function	Activation	
		training	testing	validation			(hidden)	(output)
1	MLP 19-14-2	97.6	97.1	98.9	BFGS 69	SOS	Tanh	Logistic
2	MLP 19-13-2	96.2	96.4	97.8	BFGS 46	SOS	Linear	Linear
3	MLP 19-11-2	93.1	94.4	94.4	BFGS 71	Entropy	Logistic	Softmax
4	MLP 19-5-2	92.4	92.3	92.2	BFGS 47	Entropy	Tanh	Softmax
5	MLP 19-9-2	90.2	91.1	91.1	BFGS 52	Entropy	Tanh	Softmax

TABLE IV

Summary of the neural network selection process (10-parameter feature space, 4 classes of CAN).

ID	Network	Quality			Training algorithm	Error function	Activation	
		training	testing	validation			(hidden)	(output)
1	MLP 10-10-4	96.1	95.5	94.4	BFGS 89	Entropy	Tanh	Softmax
2	MLP 10-7-4	95.0	95.0	96.6	BFGS 50	Entropy	Tanh	Softmax
3	MLP 10-9-4	94.0	93.9	92.2	BFGS 56	Entropy	Logistic	Softmax
4	MLP 10-4-4	93.1	93.7	93.3	BFGS 48	Entropy	Logistic	Softmax
5	MLP 10-4-4	92.9	93.5	94.2	BFGS 59	SOS	Tanh	Tanh

TABLE V

Summary of the neural network selection process (10-parameter feature space, 2 classes of CAN).

ID	Network	Quality			Training algorithm	Error function	Activation	
		training	testing	validation			(hidden)	(output)
1	MLP 10-8-2	99.1	98.8	98.5	BFGS 105	Entropy	Logistic	Softmax
2	MLP 10-19-2	98.7	98.5	98.6	BFGS 61	Entropy	Tanh	Softmax
3	MLP 10-10-2	98.7	98.6	98.7	BFGS 62	Entropy	Logistic	Softmax
4	MLP 10-20-2	98.5	98.3	98.6	BFGS 54	Entropy	Exponential	Softmax
5	MLP 10-6-2	98.4	98.1	98.7	BFGS 65	Entropy	Logistic	Softmax

Symbols used in the tables denote: MLP — Multi-Layer Perceptron with the number of layers and the number of neurons in each layer; BFGS — Broyden-Fletcher-Goldfarb-Shanno training algorithm with the number of epochs; Logistic — logistic sigmoid activation function; Tanh — hyperbolic tangent activation function; Exponential — exponential activation function; Softmax — softmax activation function; SOS — sum of squares error function; and Entropy — cross entropy error function.

The results of CAN classification by using Multi-Layer Perceptron are very good. The highest quality testing value of neural network is 98.8% (MPL 10-8-2) for the 10-parameter vector and two classes of CAN. However, in the case of classification into 4 classes of CAN the quality testing value of neural network is 95.5% (MLP 10-10-4). For the 19-parameter feature vector classification quality testing values are smaller and equal to 95.3% (MLP 19-13-4) and 97.1% (MLP 19-14-2).

Better recognition of only the two classes does not mean more accurate final result, i.e. the estimation of long-term noise indicators. The quality of this result can be assessed by determining the measurement uncertainty for each of the classes.

5. Conclusions

The present paper presents a usefulness of artificial neural networks (Multi-Layer Perceptron) for automatic classification of measured data, collected in continuous acoustic monitoring stations designed for the corona audible noise generated by ultra-high voltage power lines. The dedicated feature vector consists of 10 features representing distinctive features of the CAN spectral density distribution, such as – spectral moments, spectral power coefficients, tonal components contribution coefficients and meteorological data (humidity, rainfall, wind

speed). It has been shown that it is possible to achieve more than 98% of correct qualifications (recognitions) for the registered samples. Better recognition (classification) level achieved for the samples registered in high corona intensity conditions (more distinct characteristic features of CAN) ensures higher accuracy of determination for the long-term indices of corona noise. A prerequisite (and also a disadvantage) for using these solutions is the necessity to collect a representative sample group of learners and successful validation of the method. The presented methodology of automated classification of the measured data can be also applied for monitoring measurements of other objects (noise sources).

The final choice of the method to the classification of the monitored data and the number of classes results should be preceded by an analysis of the uncertainty of determining the sought target parameters.

Acknowledgments

The paper has been written and the respective research work undertaken within the project 2011/01/D/ST6/07178 (National Science Centre).

References

- [1] Electric Power Research Institute, *Transmission Line Reference Book - 345 kV and Above*, second edition, F. Weidner&Son, Palo Alto 1982, p. 267.
- [2] Official Journal of the European Communities, *Directive 2002/49/WE of the European Parliament and of the Council of 25 June 2002 relating to the assessment and management of environmental noise* **L189**, 12 (2002).
- [3] T. Wszolek, M. Kłaczyński, *Archiw. Acoust.* **34**, 3 (2009).

- [4] M. Kłaczyński, T. Wszolek, *Acta Phys. Pol. A* **123**, 6 (2013).
- [5] V.L. Chartier, R.D. Stearns, *IEEE Trans. Power App. Syst.* **100**, 1 (1981).
- [6] T. Wszolek, *Electr. Rev.* **10**, (2008), (in Polish).
- [7] T. Wszolek, R. Tadeusiewicz, in: *50th Open Seminar of Acoustics*, Szczyrk–Gliwice 2003, p. 512.
- [8] T. Wszolek, R. Tadeusiewicz, in: *13th Int. Congress Sound Vibration*, Vienna 2006.
- [9] T. Kazuo, *IEEE Trans. Power Deliv.* **6**, 4 (1991).
- [10] R. Nisbet, J. Elder, G. Miner, *Handbook of Statistical Analysis and Data Mining Applications* Elsevier, 2009.
- [11] T. Wszolek, *Archiv. Acoust.* **34**, 1 (2009).
- [12] R. Tadeusiewicz, *Neural Networks*, AOW, Warszawa 1993, (in Polish).