Features extraction from non-destructive testing data in cyber-physical monitoring system of construction facilities

V Kats and A Volkov
Moscow State University of Civil Engineering, Yaroslavskoe shosse, 26, Moscow, 129337, Russia
vladk_94@mail.ru

Abstract. Current paper presents a features extraction method when analyzing the acoustic emission (AE) control data of the technical condition of construction facilities as a part of decision support in cyber physical monitoring system. A set of statistical parameters that describe local properties of AE time series in time and frequency domains has been proposed. Furthermore, a features calculation method based on utilizing sliding windows with overlays in two time scales was introduced. Presented method has been pilot-tested during the technical diagnostics of the oil tank containing defects of different hazard classes.

1. Introduction
Technical condition diagnostics of construction facilities under operating conditions is a relevant problem. Its solution is connected to the necessity of conducting constant inspection of constructions’ technical condition, to the development of complex diagnostic monitoring system and furthermore to the effective accident management upon the inspecting objects. Failures forecasting and damage estimation based on cyber physical [1] decision support systems (DSS) become an important problem within the scope of practical implementation of diagnostic monitoring.

Machine learning techniques are widely employed in modern DSS. It is known from [2] that technical state estimation based on specified methods utilizing non-destructive testing data are prospective in particular results of acoustic emission (AE) diagnostics. The purpose of machine learning methods in DSS is structural defects (AE signal sources) detection and the evaluation of their hazard classes [3].

2. Feature extraction problem during non-destructive testing data analysis
A necessary requirement to apply machine-learning techniques to a danger class evaluation [4] is the extraction of features from the registered AE signal that describe evolution of the defects [2]. However, the desired AE signal has a set of peculiarities that complicate feature extraction process:

- Acoustic emission impulses from a single defect typically have small duration [2]. The duration of AE impulse tends to be no more than several dozens of milliseconds. Accordingly, data records from sensors under high sampling frequency no less than 1-2MHz that leads to an enormous volume of monitoring data.
- The process of defects (AE sources) evolution carries out slowly [5]. A typical time interval of the defect evolution to the pre-destructive state is from several weeks to several years in...
dependence of technical state of a certain building construction and its operating conditions. Meanwhile, it is necessary to record and analyze long-term trends of corresponding features for the evaluation of danger class of the defects and identification of their kinetics.

The feature extraction approach also appears in speech recognition problems but with less strict requirements. An effective method of extracting features from audio time series is based on consequent calculations of statistical parameters that describe an impulse form on short-term and mid-term time scaled basis [6].

This paper proposes an application of a step-by-step features extraction technique from AE signals time series via sliding windows of different width with overlays to the analysis of experimental AE time series of long duration gathered from construction facilities.

A standard scheme of AE inspection results analysis utilizes a threshold method of AE data collection and based on the calculation of several AE flux integral parameters [5]. However, these parameters do not consider local properties of AE flux and are not sustainable to noise since they are calculated on the basis of data obtained under the condition of complete noise suppression below the amplitude discrimination threshold specified by the operator at the hardware level.

3. Methods
A method of extraction of diagnostic features that signify a presence of the defect in structural element is being researched in the current paper. This method is implemented by using statistical values that are linked to parameters of the shape and local structure of AE and noise impulses sequence registered under monitoring of technical state of a construction facility. It is essentially that the process of noisy AE time series collection should be implemented in a way to provide true shape detection of the desired AE signal against noise. It is shown in [7] that current condition is met under the non-threshold AE data collection process. Detailed description of non-threshold AE data collection method is presented in the work [8]. The object of the proposed study is the method of further AE analysis connected with the extraction of diagnostic features that is, as mentioned above, conducted in two subsequent steps.

3.1. Feature extraction on short time scale.
At the first step informative features both in time and frequency domains are being calculated in sliding windows with overlays in order to identify on a small time scale signals’ local shape changes which are associated with defects (AE sources) occurrence.

Informative features that are calculated in analyzing windows should satisfy the additivity property in relation to noise and signal components of the AE time series. It should also demonstrate the sensitivity to the signal’s shape changes during the evolution process of the hazard class of defects and robustness with respect to shape changes within the corresponding hazard class. On the grounds of preliminary research of experimental AE time series it has been concluded that on short time scales the most informative feature set in the time and frequency domain is the following.

Time domain energy. It is defined as a normalized sum of amplitudes’ squares.

\[
E = \frac{\sum_{n=1}^{N} X^2 (k)}{N}
\]

where \(X(k)\) – amplitude value, \(N\) – window width.

Meanwhile, the energy of a noisy AE signal in the window represents the additive sum of acoustic emission and noise components.

Spectral centroid. Defined as an average-weighted amplitude of the signal frequency spectrum.
The number of Fourier transform coefficients \( N \) can be calculated as:

\[
N = \sum_{k=1}^{W_L} X_i(k)
\]

where \( W_L \) is the number of coefficients, \( X_i(k) \) is the amplitude of the k-th coefficient, and \( i \) is the time window index.

Centroid is a measure of spectrum shape. The raise of centroid values indicates the domination of high-frequency components in the spectrum of the noisy signal, which are caused by noise sources appearance.

Spectral spread. It describes standard deviation of the spectrum distribution which is related to the width and the position of the spectrum signal bandwidth [9].

\[
S_i = \sqrt{\frac{\sum_{k=1}^{N}(k-C_i)^2X_i(k)}{\sum_{k=1}^{N}X_i(k)}}
\]

Larger values of spectral spread typically have noisy impulses while the AE impulses occurrence related to the structural defect leads to smaller values of spectral spread.

Spectral roll-off. It is defined as a frequency below which the q-th part of the area under the spectrum amplitude distribution function is concentrated (the corresponding frequency \( R_q \) is an analog of percentile in statistics [10]):

\[
\sum_{k=1}^{R_q} X_i(k) = q \sum_{k=1}^{N} X_i(k)
\]

The spectral roll-off also describes spectrum shape and may be employed for the identification of desired AE signal and noise and for the separation of signals related to defects of different classes of danger.

Spectral flux. It is defined as a squared difference between normalized spectrum magnitudes of two consequent windows.

\[
F_{(i,i-1)} = \sum_{k=1}^{N} \left( EN_i(k_L) - EN_{i-1}(k_L) \right)^2
\]

Given measure is sensitive to local changes of spectrum amplitude linked with shape evolution dynamics of the AE time series under the appearance of the defects that are related to different hazard class.

Entropy in frequency domain. This feature is defined as the entropy of normalized spectrum energy. Spectral entropy is a measure of spectrum variety.

\[
H(i) = -\sum_{f=0}^{L-1} n_f \cdot \log_2(n_f), \quad n_f = \frac{E_f}{\sum_{f=0}^{L-1} E_f}
\]

where \( E_f \) is the spectrum energy on frequency \( f \). It reveals the differences between the noise component and the AE signal. Larger values of the entropy are associated with random noise while smaller values are related to AE signal component.
3.2. Feature extraction on long-term scale
The second step is related to the extraction of AE impulses patterns that are caused by the development of the defects. The features which are calculated on the first step are then aggregated in time. The aggregation is carried out in the sliding windows that have window width up to several seconds by utilizing averaging on multiple statistical parameters such as mean, standard deviation, median, extreme values, variation coefficient, skewness and kurtosis. These windows with overlays allow to monitor dynamics of the spectral and time features changes.

Averaging on each parameter results in time series which sample is determined by the texture window of given width. The resulting matrix is employed as the input variable of machine learning algorithms during the classification of identified defects.

As shown below, presented technique allows to detect long-term alterations of the AE impulses statistical parameters that are induced to defects degradation processes.

4. Results and discussion
Figure 1 presents the experimental AE time series were obtained under the diagnostics of technical state of the oil tank, where defects of the different danger classes are indicated as a), b) and c). AE data acquisition was conducted by non-threshold collection method described in [8]. The hazard class of the defect was assessed by the results of independent inspection by means of NDIS criterion [11].

It can be noticed that the average amplitude of the desired AE signal grows with the raise of hazard class of AE source. Meanwhile, mean AE activity also increases.

4.1. The relationship of diagnostic features and hazard classes of defects.
Figure 2 includes density function plots of informative features that describe AE signal shape and its changes due to the transition from one hazard class to another. As an example the calculation results for three features in the frequency domain are presented namely by spectral centroid, spectral roll-off and spectral flux. Meanwhile, calculation on the short time scale is carried out with the following set of arguments: the width of the sliding time window is 18ms, overlay is 14ms, the width of spectral window – 32768 samples, type of the spectral window – Hamming window [9].

It follows from fig. 2 that probability density functions of enumerated in the previous section statistical features essentially depend on the hazard class of AE source and consequently on the waveform of emitted AE signal. Namely, mode of the spectral flux distribution shifts in the direction of larger values with the growth of the hazard class, however the value of the shift is slight (Fig. 2a). Meanwhile, the value of the density function significantly increases.
Figure 1. Fragments of the noisy AE time series. The inspected object: corner weld joint of the building construction (vertical steel oil tank #3. Volume – 1000 cubic meters, fuel handling facility, AO “NTEK”). Sampling frequency – 2.5MHz. a) The defect of the 1st hazard class. b) The defect of the 2nd hazard class. c) The defect of the 3rd hazard class.

Mode of the spectral centroid (Fig. 2b) changes reverse: the value of density function remains practically constant but the mode value itself shifts noticeably in the direction of larger values of the centroid’s frequency. Moreover, the value of skewness also grows. The probability density function of the spectral roll-off appears multimodal behavior, which means that the plot of the function has two or more local maximums (Fig. 2c). Simultaneously the variance of the distribution reduces under the transition of the AE source to the higher hazard class.
Figure 2. Density functions of AE shape statistical features for the defects of the three hazard classes.

a) Spectral flux b) Spectral centroid c) Spectral roll-off

Calculation results of statistical parameters on minor intervals has been grouped in Table 1. Mode and median describe changes of distribution functions of the selected features such as skewness, kurtosis and corresponding time series according to the rank of the hazard class. It follows from results analysis presented in Table 1 that the probability density functions of selected features correlate with the hazard class rank. Consequently, we assume that time series of the specified features can be employed as input variables for the second step of the proposed method. Meanwhile, it is efficient to utilize parameters shown in Table 1 for averaging.

4.2. Mapping of defects hazard classes to the space of features set

The verification of informative content of the extracted diagnostic feature set within the framework of the problem of identification and classification of the defects assumes the implementation machine learning techniques in multidimensional feature space. Existing systems of AE diagnostic monitoring are based on adopting limited set of feature pairs that describe local shape of the signal: amplitude – AE signal duration in time domain [12], magnitude – frequency of the maximum in spectral domain [13]. Respectively, in this context machine learning is applied on two dimensional feature space. An alternative approach is also known based on dimensionality reduction of multidimensional feature space by using principal components analysis [14].
Both cases of results presentation are not the sufficient means for providing effectiveness of the classification and recognition procedures. First consideration is not suitable since local parameters of the unit impulse shape weakly correlate with integral parameters of impulse flux that are typically employed in classification of defects on hazard class ranks [5]. Second consideration has a problem related to the first two principal components that only partly contribute to the total variance of features that results in losses of informative content in comparison with the full set of components [14].

A Stochastic Neighbor Embedding method is adopted in the current paper for the problems of feature space reduction and multidimensional variables visualization [15].

SNE-method is a technique of nonlinear dimensionality reduction [16], which main advantage is preserving the local structure and probability properties of the original data in a low-dimensional space. This achieves by the fact that 1) dimensionality reduction in the SNE method is carried out by preserving distance proportions between points in original and final feature space, 2) by the transforming Euclidean metric to correspondent conditional probabilities [17]. Specified peculiarities allows visualizing the existence of grouping of the original data in the space of two or three dimensions [18].

Prior information about hazard classes of defects has been obtained by independent inspection method in order to assess the degree of the presented above features informativity.

Figure 4 presents features calculation results in the two-dimensional SNE space. Points that are corresponding to the texture window has been marked by one color: green – 1st class, yellow -2nd class, red – 3rd class. It follows from the Fig. 4 that features clustering in the space of SNE-components is generally consistent with colored marking of hazard classes. Furthermore, the classification error estimated by “Accuracy” metric does not exceed 2.5%. Thus, proposed set of features and its calculation method allow to separate AE signals on groups according to the hazard classes of the defects.

Table 1. Statistics of informative features on small time scale

| Feature / Distribution function statistics | The hazard class rank | Mode  | Median | Skewness | Kurtosis |
|------------------------------------------|----------------------|-------|--------|----------|----------|
| Spectral flux                            | I-class              | 0.0002| 0.0007 | 2.8611   | 18,8905  |
|                                          | II-class             | 0.0005| 0.0006 | 2.1951   | 12.4415  |
|                                          | III-class            | 0.0006| 0.0008 | 2.3149   | 14.1484  |
|                                          | I-class              | 0.0448| 0.0452 | 8.5160   | 95.4774  |
| Spectral centroid                        | II-class             | 0.0473| 0.0473 | 11.2270  | 176.0832 |
|                                          | III-class            | 0.0552| 0.0541 | 9.2873   | 133.9230 |
|                                          | I-class              | 0.0462| 0.0481 | 4.6831   | 98.8862  |
| Spectral spread                          | II-class             | 0.0534| 0.0532 | 1.0541   | 53.0021  |
|                                          | III-class            | 0.0581| 0.0584 | 1.0454   | 99.8923  |
Figure 3. Density functions of AE shape statistical features for the defects of the three hazard classes.  
a) Spectral flux  b) Spectral centroid c) Spectral roll-off

5. Conclusions
Current study presented an approach of diagnostic features extraction from AE time series in order to build a cyber physical DSS system on its bases which provides diagnostic monitoring of technical condition of building constructions.

A set of diagnostic features that are determined both in time and frequency domains has been proposed. They are spectral centroid, spectral flux, spectral spread, entropy and energy. Enumerated features describe the shape and local structure of AE signal sequence.

A method of calculation of diagnostic features containing two consequent steps has been described.

Based on numerical comparison of classification results with a priori known grouping of AE sources by hazard classes, it has been established that proposed features set and its calculation method allow to match numerical values of features vector and the hazard class.

Acknowledgments
This study was performed with the financial support of the RF Ministry of Education and Science, President Grant #NSh-3492.2018.8.

References
[1] Volkov A and Ovchinnikov A 2018 Sistemotehnika stroitelstva, kiberfizicheskiye stroitel’niiye sistemi Moscow 1 pp. 184-188
[2] Calabrese L, Campanella G, Proverbio E 2012 Noise removal by cluster analysis after long time AE corrosion monitoring of steel reinforcement in concrete. Construction and Building Materials. 34 pp. 362-371.
[3] Volkov A and Ovchinnikov A 2018 Sistemotehnika stroitelstva, kiberfizicheskiye stroitel’niiye sistemi Moscow 1 pp. 194-197
[4] Volkov A, Sedov A, Chelyshkov P, Titarenko B, Malyha G, Krylov E 2016 The theory of probabilities methods in the scenario simulation of buildings and construction operation. Research journal of pharmaceutical, biological and chemical sciences. 7 pp. 2416-2420
[5] Volkov A Vagapov R 2007. Informatsionnoye modelirovanie v zadachah upravleniya bezopasnost’yu zdaniy i soorujeniy. Vestnik MGSU 4 pp. 76-78
[6] Aggelis D 2016 Acoustic Emission Analysis for NDE in Concrete. Innovative AE and NDT Techniques for On-Site Measurement of Concrete and Masonry Structures 20 pp. 69-88.
[7] Tzanetakis G., Cook P. 2002 Musical genre classification of audio signals. IEEE Transactions on Speech and Audio Processing 10 pp. 293–302.
[8] Kuzmin A, Kats V, Inozemtsev V, Prokhorovsky A, Axelrod E 2018 Tehnologiya besporogovoy registratsii dannykh akusticheskoy emissii pri kontrole promishlennih objectov. Himicheskaya tehnika, 3, pp. 10-16.

[9] Inozemtsev V., Prokhorovsky A, Axelrod E, Kuzmin A. Sposob registratsii i analiza signalov akusticheskoy emissii. Patent RUS 2570592. More references

[10] Proakis J, Manolakis D 2009 Digital Signal Processing. London: Pearson Education, 4 1104 p.

[11] Giannakopoulos T, Pikrakis A 2014 Introduction to Audio Analysis - A MATLAB Approach. Elsevier Ltd 262 pp.

[12] Kuzmin A, Kats V., Axelrod E. 2017 Acustiko-emissionniy kontrol’ pri otsenke tehnichestkogo sostoyaniya oborudovaniya neftegazovogo kompleksa. J. V mire NK. 3, pp. 71-80.

[13] Moradian Z., Li B 2017 Hit-Based Acoustic Emission Monitoring of Rock Fractures: Challenges and Solutions/ Advances in Acoustic Emission Technology 179 pp. 357-370.

[14] Ogura N, Yatsumoto H, Chang K, Shiotani T 2017. Development of Damage Evaluation Method for Concrete in Steel Plate-Bonded RC Slabs. Advances in Acoustic Emission Technology 179 pp. 321-334.

[15] Soltangharaei V, Rafał A, Nolan W, Lateef A, Yann Le P 2018. Damage Mechanism Evaluation of Large-Scale Concrete Structures Affected by Alkali-Silica Reaction Using Acoustic Emission. Applied Sciences 8 pp. 2148-2167.

[16] Van der Maaten L, Hinton G 2008 Visualizing High-Dimensional Data Using t-SNE // Journal of Machine Learning Research 9 pp. 2579-2605.

[17] Barnes J, Hut P 1986 Hierarchical O(N log N) force-calculation algorithm. Nature 324 pp. 446–449.

[18] Bishop C 2007 Pattern Recognition and Machine Learning (Information Science and Statistics). New York: Springer 738 pp.