

Evaluating Similarity of Spectrogram-like Images of DC Motor Sounds by Pearson Correlation Coefficient

Dejan G. Ciric¹, Zoran H. Peric¹, Marko Milenkovic², Nikola J. Vucic^{1,*}

¹Faculty of Electronic Engineering, University of Nis,
Aleksandra Medvedeva 14, 18000 Nis, Serbia

²Faculty of Arts, University of Nis,
Kneginje Ljubice 10, 18000 Nis, Serbia
nikola.vucic@elfak.ni.ac.rs

Abstract—Three main approaches on how audio signals can be used as input to a deep learning model are: extracting hand-crafted features from audio signals, mapping audio signals into appropriate images such as spectrogram-like ones, and using directly raw audio signals. Among these approaches, the usage of spectrogram-like images represents a compromise regarding the bias enforced by the processing (seen in hand-crafted features) and computational demands (seen in raw audio signals). When any of the spectrogram-like images is used as a deep learning model input, then different techniques for image processing become available and can be implemented. They include techniques for assessing the image similarity, implementing image matching, and image recognition. The topic of this paper is similarity of spectrogram-like images obtained from DC motor sounds. In that respect, relevant measures of image similarity are first reviewed, and then one of them - the Pearson correlation coefficient - is applied for evaluating the similarity within the same class and between two classes of different spectrogram-like images.

Index Terms—DC motor sounds; Spectrogram-like images; Image similarity; Pearson correlation coefficient.

I. INTRODUCTION - ASSESSING SIMILARITY OF IMAGES IN AN OBJECTIVE WAY

Image quality, image comparison, and image matching have attracted a huge attention during previous several decades, and they have been used in various applications. In those related to registration and stereo pair matching, the images are aligned to obtain the highest similarity between them [1]. Visual tracking and robotic navigation have predefined target templates or features searched for in the current video frame, associating the best match location with the true target location. There are also applications where an image is retrieved from a large database by finding the best match to a given pencil sketch. In some recognition applications, predefined templates are compared with the extracted content of images.

A time-consuming and costly, but reliable way to assess

image quality is to perform visual experiments under controlled conditions, in which observers grade which image provides better quality [2]. An easier solution is to use an objective measure (in literature usually denoted as a metric) capable of quantifying the image quality or similarity between images. For this purpose, hundreds of algorithms and measures have been developed [3]. Unfortunately, there is no current standard on which objective measure(s) to use for such an application. Besides, there is no single similarity measure that works well for all tasks, but instead the measures to be used depend on the application.

Image similarity measures have played a very important role in image quality assessment (IQA). Here, one of the aims these measures should accomplish is to have as large a correlation as possible with subjective quality evaluations, i.e., to quantify the strength of the perceptual similarity (or difference) between the test and the reference images. IQA measures are divided into three groups: full-reference (FR), no-reference (NR), and reduced-reference (RR) measures [3], [4]. Among them, the NR measures are the most desirable, although their universality and correlation with subjective opinions given as mean opinion scores (MOS) or differential MOS (DMOS) values are typically significantly lower compared to FR measures [4].

Image similarity measures could also be divided into similarity-based measures (Pearson correlation, Spearman correlation, Kendall's Tau, Cosine similarity, and Jaccard similarity) and distance-based measures (Euclidean distance (ED) and Manhattan distance) [5]. The basic region-based image similarity measures are divided into two groups [1]: ED-based and correlation-based measures, where the latter ones provide superior performance in many cases, but they are more computationally demanding. The correlation coefficient and correlation could be used to solve the classification problems [6], [7], as well as for the recognition of radio transmission systems [8]. Correlation is also useful for compression of the speech signal by applying the delta modulation technique [9]. In the field of detection and classification, different signal processing techniques are deployed to obtain the features and other useful information

Manuscript received 27 December, 2021; accepted 31 March, 2022.

This work has been supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia and by the Science Fund of the Republic of Serbia under Grant No. 6527104, AI-Com-in-AI.

(MFCC, DCT, FFT) [6]. It is worth noting that some of the previously mentioned measures (such as zero-mean normalized cross-correlation (ZNCC) and median NCC) are not suitable for matching the scaled or rotated images. This is why rotation- and scaling-invariant correlation measures are proposed to deal with this limitation.

Many of the image similarity measures [10] do not take into account the spatial relationship between different pixels of an image. This limitation can cause unwanted results of image matching, e.g., images that look not similar to a human observer may have a measure of high similarity [1]. It is even more important, especially in the case of using spectrogram-like images, to emphasize one more limitation related to the fact that gradual deformation of the image may exhibit abrupt changes in the similarity measure.

The “classical” IQA measures are pixel-based methods, such as mean square error (MSE) or peak signal-to-noise ratio (PSNR). The introduction of the universal image quality index (UQI) can be considered a start for the development of modern visual quality measures [11]. This is followed by further improvement seen in measures such as the widely known structural similarity index (SSIM) implemented also as a multi-scale version (MS-SSIM) [11]. The SSIM index (designed as an improvement of traditional methods, such as MSE and PSNR) is one of the most popular FR measures in the image processing society [12]. The SSIM index has been used for various purposes such as image quality assessment, image enhancement, and image recognition, as well as a basis for some other similarity-based measures that led to a further increase in the correlations between objective quality scores and subjective MOS or DMOS values [11].

Image similarity measures are used in various data science techniques. Thus, classification of new data objects can be done by applying the similarity approach to k -nearest neighbours. Another example is the use of ED in unsupervised learning, more precisely in k -means as a clustering method that computes the distance between the centroids of the cluster and its assigned data points [7].

In addition to IQA, similarity measures have also played an important role in image recognition. Here, it is necessary to find the similarity between two images, i.e., between a test image and its equivalent training database image. Several image recognition methods have been reported in the literature, including principal component analysis (PCA) and subspace linear discriminant analysis (LDA) [13]. In general, image recognition methods can be categorized as holistic, feature-based, and hybrid methods [13]. Image recognition can be applied in controlled (where the imaging conditions are fixed for both the trainee and probe images) and uncontrolled environments.

In the rest of this article, two image similarity measures are first described, and their main properties are presented. One of these measures, the Pearson correlation coefficient (PCC), is then applied to analyse the image similarity within the same class (intra-class) and image similarity between the classes (inter-class) of various spectrogram-like images. Based on the results, relevant conclusions on that topic are drawn.

II. IMAGE SIMILARITY MEASURES

Suppose that two non-negative images (either continuous or discrete) X and Y of the same size $P \times R$ whose similarity should be assessed are represented as $X = \{x_i | i = 1, 2, \dots, N\}$ and $Y = \{y_i | i = 1, 2, \dots, N\}$, where i is the sample index and $N = P \times R$ is the number of image samples (pixels). To compare two types of image similarity measures, a distance-based measure (ED) and a similarity-based measure (PCC) are presented below.

A. Euclidean Distance (ED)

The ED, which is typically described as the length of the straight line between the two given points [14], can be applied to quantify the distance between two points in Euclidean space and is used mainly when the data to be compared are continuous in nature. On the other hand, ED is not used very often in the context of natural language processing applications compared to some other measures such as cosine and Jaccard similarity [7].

The ED can get values between 0 and $\sqrt{N \times F}$, where F denotes the maximum value of x_i , that is, y_i (typically the maximum gray-level value, e.g., 255 for regular 8-bit images). A smaller distance indicates a better match between the compared images. The absolute distance depends on the size of the image. As in the case with all pixel-wise measures, the ED may be large for small deformations (e.g., shift of an image by one pixel), since the spatial connections between the pixels are not concerned [1]. In addition, the ED is sensitive to changes in noise and brightness [1]. Since the ED is not scale-invariant, scaling the data prior to computing this measure is recommended. Another property of ED is that it multiplies the effect of redundant information in the dataset [7].

A normalized version of the ED measure can also be applied and is called the “image Euclidean distance” (IMED) [14], [15]. Here, a smaller deformation leads to a smaller change in the distance. The IMED takes into account the spatial relationships between the pixels and is robust to small image deformations and perturbations. It is interesting that IMED can easily be embedded into other high- and medium-level matching procedures. The mentioned properties allow the IMED measure to be appropriate for image recognition and visual tracking applications [1].

B. Pearson Correlation Coefficient

The degree of probability that a linear relationship (both strength and direction) exists between two measured quantities (images in this case) is obtained by applying a relevant correlation method. More than a century ago (in 1895), Karl Pearson defined the Pearson product-moment correlation coefficient - the first formal correlation measure that provides the degree of correlation between two quantities (e.g., images) [13]. The calculation of this coefficient for monochrome digital images ρ , which is also often denoted by r , is given by

$$\rho_{X,Y} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}, \quad (1)$$

where σ_{XY} is the covariance between X and Y given as

$$\sigma_{XY} = E(XY) - E(X)E(Y), \quad (2)$$

here E is the expectation, σ_X and σ_Y are the standard deviations of X and Y :

$$\sigma_X = \sqrt{E(X^2) - (E(X))^2}, \quad (3)$$

$$\sigma_Y = \sqrt{E(Y^2) - (E(Y))^2}. \quad (4)$$

It should be noted that the covariance σ_{XY} is always smaller than the product of the standard deviations σ_X and σ_Y .

The PCC can also be represented as

$$r = \frac{\sum_{i=1}^N (x_i - x_m)(y_i - y_m)}{\sum_{i=1}^N (x_i - x_m)^2 \sum_{i=1}^N (y_i - y_m)^2}, \quad (5)$$

where x_i is the intensity of the i -th pixel in image X , y_i is the intensity of the i -th pixel in image Y , x_m is the mean intensity of image X , and y_m is the mean intensity of image Y . In the case where $x_m = y_m = 0$, the PCC becomes normalized cross correlation (NCC).

PCC has values in the range from -1 to $+1$, whereby the value $r = 1$ indicates that two compared images are absolutely identical, $r = 0$ indicates that they are completely uncorrelated, and the value $r = -1$ shows that the images are completely anti-correlated. On the one hand, the PCC is invariant to constant brightness changes, but, on the other hand, it is not defined for constant intensity images (it shows close to one correlation between approximately white and black images) [1]. In some applications, such as tracking, only positive correlation is of interest, and this is why $\max(0, r)$ is used as the similarity measure [1].

Some of the main benefits of the PCC as a dimensionless index can be summarized as the following [13]:

- It condenses the comparison of two 2-D images down to a single scalar value r ;
- It is invariant to linear transformations of variables x and y , which means that r is insensitive (within limits) to uniform variations in brightness or contrast in an image [16]. On the other hand, limitations of the PCC are [13], [16]: a) computational demands limiting its usefulness for image registration; b) extreme sensitivity to the image skewing, pincushioning, and vignetting that inevitably occur in imaging systems; c) undefined calculation due to the division by zero - if one of the test images has constant, uniform intensity.

PCC, also known as coefficient of correlation (CoC) [13], has been widely used in statistical analysis, pattern recognition, and image processing [6], whereby in image processing it is used to compare two images for image registration purposes, disparity measurement, etc.

It is of interest to see the performance of the PCC compared to some other image similarity measures. Thus, it is stated in [3] that PCC is superior to SSIM index for image

similarity analysis and is faster to calculate. This means that the PCC determines the similarity and dissimilarity of images more precisely than the SSIM index. Moreover, based on experimental demonstration, PCC is stated to be better than the SSIM index regarding coincidence with subjective MOS image estimates [3]. Furthermore, the PCC is connected to the Euclidean metric, but it is non-linearly related to this metric [3]. The better performance of PCC compared to SSIM index is also reported in [13] where it is stated that PCC has a high overall recognition rate and a low rejection rate compared to SSIM. In addition, it is concluded that, under normal lightning conditions, the recognition accuracy of PCC when used with and without a discrete wavelet transform on the rotated test images gives the best results compared to SSIM, PCA, and LDA.

In many image processing applications, images have different rotation, orientation, illumination, contrast, etc. As a consequence of these differences, the values of PCC, but also other image similarity measures, such as SSIM, can be changed depending on a particular change in an image. To overcome this problem and increase the recognition accuracy, the input test image can be pre-processed. However, this is not the case in the spectrogram-like images obtained from raw audio signals. If the duration and sampling frequency of raw audio signals are the same, which is typically the case or can easily be fulfilled by pre-processing, the spectrogram-like images will be perfectly aligned with each other - there will be no difference in the mentioned parameters (rotation, orientation, illumination, contrast, etc.).

III. METHOD OF ANALYSIS

The audio signals of DC motors are classified in two classes: OK motors (without failure or malfunction) of type A and direction of rotation 1 belong to class 1 and NOK motors (with certain failures/malfunctions) of type A and direction of rotation 1 belong to class 2. Class 1 contains 281 signals, while the class 2 contains 387 signals. The audio signals of DC motors are used for calculating various spectrogram-like images: lin-power spectrogram, mel-spectrogram, gammatonegram, constant Q transform (CQT) power spectrogram, short-term Fourier transform (STFT) chromagram, CQT chromagram and tempogram. First, the audio signals are pre-processing by extracting 5 s of useful signal [17], [18]. Each audio signal is then transformed into defined spectrogram-like images applying the mapping relevant for every particular image, with the aim of exploiting the maturity of the image technology. Thus, the STFT yielding complex values as a result are used for some of the images. The f -th component of the discrete Fourier transform (DFT) of the t -th frame of signal $x[n]$ is calculated as

$$X[t, f] = \sum_{k=0}^{M-1} w[k] x[tM + k] e^{-\frac{j2\pi kf}{M}}, \quad (6)$$

where $w[k]$ represents a window function, e.g., Hamming or Blackman, which is used to enforce continuity and periodicity at the edge of frames. In practice, the shift or hop size between consecutive frames is typically smaller than

the frame size, allowing smoother STFT and introducing statistical dependencies between frames [17]. A spectrogram is generated from the STFT results given as a matrix, where each column is the DFT result of a particular signal frame. More details on the mapping of audio signals to spectrogram-like images can be found in [17].

Further, the obtained images in the form of matrices are converted into arrays by concatenating the columns. These arrays representing numerical parameters of the concrete audio signal are used for the analysis of similarity among the spectrogram-like images within the same class (intra-class) and between the classes 1 and 2 (inter-class). This is done by calculating the PCC for each combination (pair) of a particular image (its array) and all other images of the same type belonging to the same class (intra-class similarity) and all other images belonging to another class (inter-class similarity). The numerical results obtained for the PCC are processed to extract the statistical quantities - mean, minimum, maximum, median, and standard deviation (STD) used for the analysis.

IV. PEARSON CORRELATION COEFFICIENT AS A SIMILARITY MEASURE OF DIFFERENT SPECTROGRAM-LIKE IMAGES

The lin-power spectrogram is considered here as a reference since it is a full-resolution representation in the time-frequency domain of a raw audio signal. Other spectrogram-like images are compared with the reference image with respect to the PCC and statistical quantities calculated from the obtained PCC. Only the mean value of the PCC within the class and between classes does not provide enough information to make relevant conclusions, instead other statistical quantities should be taken into account, too.

The statistical quantities of the PCC calculated within the same class (intra-class) for the mentioned spectrogram-like images are given in Table I for OK motors (class 1) and in Table II for NOK motors (class 2). The mean values among image types range from 0.124 for the STFT chromagram to 0.999 for the tempogram. The mean value depends on the fact that the PCC can have both positive and negative values. Thus, for some images, a minimum value of the PCC is positive, meaning that in this set of values there is not a single negative coefficient. On the other hand, the existence of the negative correlation coefficient reduces the mean value of this parameter, which can be overcome by taking the maximum of (0, PCC).

Among the statistical quantities of the PCC for the spectrogram-like images summarized in Table I, the closest results to the reference image are obtained for the mel and CQT power spectrogram. The mean and median correlation coefficient for the gammatonegram have slightly larger values, and the correlation can be negative. The mean and median values for two types of chromagrams are smaller, and there are negative correlation values, also. The results for tempogram are very specific, since the range of values of correlation coefficient is very small, from 0.949 to 1.

A situation similar to the one described above for the OK motors (class 1) exists in the NOK motors (class 2), as shown in Table II. The main difference in comparison to the

OK motors is that the mean and median values of the PCC are somewhat lower in the NOK motors in all spectrogram-like images except for the STFT chromagram. Such behavior can be expected, since NOK motors are less consistent in overall characteristics, as the failure or malfunction present in these motors can be of different nature.

Table III summarizes the statistical quantities of the PCC between OK and NOK motors (inter-class comparison) obtained for the observed spectrogram-like images. Here, the mean and median values of the PCC are between the values for OK motors (class 1) and NOK motors (class 2) for the lin-power spectrogram. The mean and median values of the PCC between OK and NOK motors obtained from the mel-spectrogram and CQT power spectrogram are smaller than the corresponding ones between the OK motors, and similar to those between the NOK motors. The distance between the intra-class and inter-class similarity in regard to mean and median values is the greatest in the lin-power spectrogram and mel-spectrogram; then the results for CQT power spectrogram follow.

TABLE I. PCCS BETWEEN OK MOTORS - CLASS 1 (INTRA-CLASS COMPARISON).

Spectrogram	Statistical quantity				
	Mean	Min.	Max.	Median	STD
Lin-power	0.620	0.170	1.0	0.622	0.136
Mel	0.697	0.046	1.0	0.705	0.071
Gammatonegram	0.881	-0.395	1.0	0.898	0.105
CQT power	0.658	0.314	1.0	0.660	0.063
STFT chromagram	0.124	-0.510	1.0	0.117	0.177
CQT chromagram	0.382	-0.328	1.0	0.401	0.279
Tempogram	0.999	0.949	1.0	0.999	0.004

TABLE II. PCCS BETWEEN NOK MOTORS - CLASS 2 (INTRA-CLASS COMPARISON).

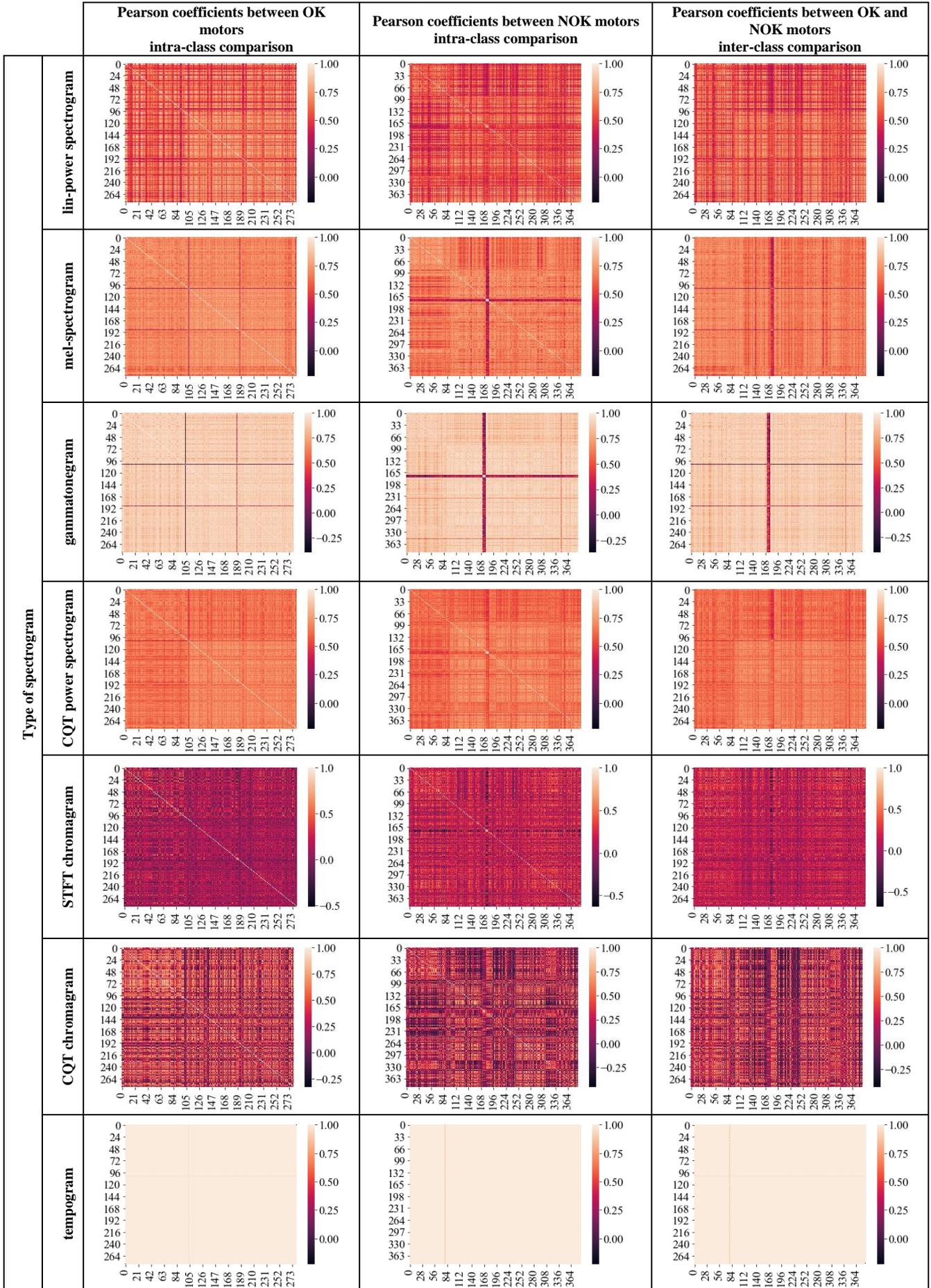
Spectrogram	Statistical quantity				
	Mean	Min.	Max.	Median	STD
Lin-power	0.549	0.142	1.0	0.534	0.125
Mel	0.656	-0.187	1.0	0.674	0.098
Gammatonegram	0.857	-0.368	1.0	0.890	0.141
CQT power	0.639	0.347	1.0	0.633	0.071
STFT chromagram	0.165	-0.630	1.0	0.165	0.191
CQT chromagram	0.213	-0.454	1.0	0.178	0.301
Tempogram	0.998	0.850	1.0	0.999	0.009

TABLE III. PCCS BETWEEN OK AND NOK MOTORS (INTER-CLASS COMPARISON).

Spectrogram	Statistical quantity				
	Mean	Min.	Max.	Median	STD
Lin-power	0.574	0.114	0.950	0.563	0.132
Mel	0.659	-0.221	0.894	0.671	0.082
Gammatonegram	0.861	-0.402	0.965	0.885	0.121
CQT power	0.639	0.318	0.892	0.639	0.064
STFT chromagram	0.122	-0.679	0.836	0.119	0.175
CQT chromagram	0.259	-0.434	0.946	0.238	0.296
Tempogram	0.998	0.854	1.000	0.999	0.007

The PCCs for all analyzed spectrograms are presented as heat maps in Table IV.

TABLE IV. HEAT MAPS ILLUSTRATING THE PCCS FOR ALL ANALYZED SPECTROGRAMS.



First, we should note that all maps illustrating image similarity within the same class (intra-class) are symmetric in relation to the main diagonal, whereby the main diagonal represents the self-correlation, i.e., PCC equal to 1, and is therefore the brightest line in the heat map.

The presented PCC values for the lin-power spectrogram show a higher correlation within class 1 (OK motors) than between classes (1 and 2). It can also be concluded that the correlation between the classes is slightly higher than the correlation within class 2 (NOK motors), as explained above.

In spite of the smaller size of the mel-spectrogram images (it has 96 mel bands, i.e., 96 points on the y axis in comparison to about 40000 points in the lin-power spectrogram), all their heat maps are similar to the ones of the lin-power spectrogram. The correlation values are here slightly higher, and therefore the color in the heat maps is a bit brighter. Apart from a similar trend in these two types of images, each of them shows some unique correlations for some particular sound samples (audio signals from particular DC motors). In particular, some sound samples have darker horizontal and vertical lines in the heat maps for the mel-spectrogram. This is an indication that these samples (motors) should be analyzed in more detail.

Gammatonegrams provide even brighter heat maps than the mel-spectrogram, indicating higher values of correlation. The relative difference between the intra-class and inter-class correlation is slightly smaller than in the previous two spectrograms. Darker horizontal and vertical lines or strips in the heat maps exist here as well, and they are even more obvious than in the mel-spectrogram.

The pattern of the heat maps for the CQT power spectrogram is similar to the one for the mel-spectrogram. It is interesting to note that the correlation for samples of OK motors with numbers above 100 (lower right quadrant of the heat map image) is slightly higher, which is visible by brighter colors in the heat map. The pattern (distributions of correlation values among the sound samples located along both x and y axis) for the inter-class difference generated from the CQT spectrograms and mel-spectrograms is quite similar.

The heat maps obtained from both chromagrams (STFT and CQT) have a PCC smaller than in the previous spectrogram-like images, as a darker color covers the majority of the heat maps. In addition, the distinction between intra-class and inter-class similarity can hardly be noticed. It is interesting to mention that only in the STFT chromagrams the correlation is higher between NOK motors than between OK motors, while the correlation between OK and NOK motors is similar to the correlation between OK motors. These results show that the resolution and processing applied for obtaining the chromagrams are not adequate in this particular use-case to keep a significant difference between the classes while keeping a significant similarity within the same class.

The heat maps for the tempogram are opposite to those for the chromagrams, since the correlation has significantly higher values in tempograms, as denoted by bright colors in the heat maps. Unfortunately, the range of correlation coefficient values is too small and there is not a clear

distinction between intra-class and inter-class similarity.

V. ORDERING OF SAMPLES BASED ON IMAGE SIMILARITY

The previously presented results show that the difference of the PCC values among the sound samples (images) within the same class (even within the class 1 - OK motors) is rather large in most of the applied spectrogram-like images. Thus the PCC values for the lin-power spectrogram within the class 1 ranges from 0.170 to 1, see Table I and heat map given in the first row and column of Table IV. This situation is analyzed here in more detail using only the class 1 (OK motors) and lin-power spectrograms.

First, the average PCC for each DC motor (lin-power spectrogram image) of class 1 is calculated across all motors of that class. The motors are then rearranged according to the average PCC value in a decreasing order (see Fig. 1). The largest average PCC is about 0.73, while the smallest average PCC is about 0.36. Although the PCC value decreases rather continuously, the PCC curve can be split in two parts, up to the 202nd motor (denoted by the vertical line in Fig. 1) and above that motor number. These two parts have different slopes, where the slope of the first part is smaller than that of the second part.

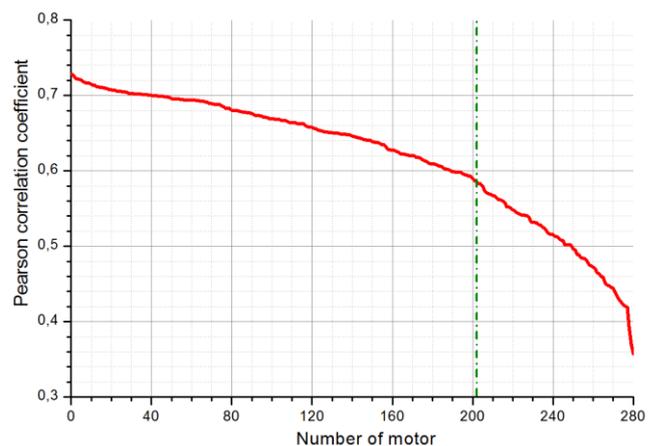


Fig. 1. Average (within the class) PCCs for DC motors of class 1 (OK motors), i.e., its lin-power spectrograms rearranged in decreasing order of PCC values.

The PCC values of two particular DC motors, having the largest (denoted as motor L) and smallest (denoted as motor S) average PCC within class 1, are presented in Fig. 2(a) before rearranging and in Fig. 2(b) after rearranging in a decreasing order of the average PCC. There are certain fluctuations in both PCC curves in Fig. 2(a), but there is also a prominent trend that the PCCs are considerably greater for the motor L than for the motor S. The PCC curve for motor L can also be split in two parts as in Fig. 1. As described above, the first part has a smaller slope than the second part. On the other hand, the PCC curve for motor S does not show this behavior. Instead, the PCC values along this curve become slightly larger, and also fluctuations of the PCC curve become somewhat larger from a certain point until the last motor. This point coincides with the limit between two parts of the PCC curve for motor L, also denoted by the vertical line in Fig. 2(b). The positions of the vertical lines in Fig. 1 and Fig. 2(b) representing a limit between two parts of the PCC curves are either exactly the

same or very close to each other. At the same time, this limit divides class 1 into two sub-classes (1.1 and 1.2), where the first one contains the motors with larger correlation coefficients.

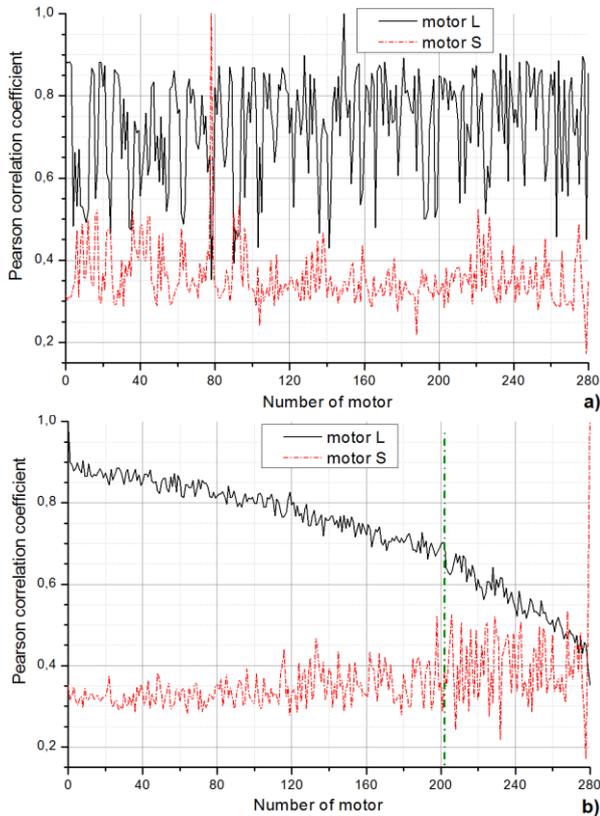


Fig. 2. PCCs for OK motor L (with the largest average PCC) and S (with the smallest average PCC) within class 1 (a) before rearranging and (b) after rearranging.

The previously described behavior can be seen in the heat map of the PCC values for the rearranged motors, too, see Fig. 3. The motor with the largest average correlation is located in the upper left corner, and correlation decreases going further away from this corner, which is denoted by a darker color. The introduced limit between two sub-classes is denoted here by both horizontal and vertical lines. The first sub-class (1.1) containing the motors up to the one with number 202 has larger PCC values.

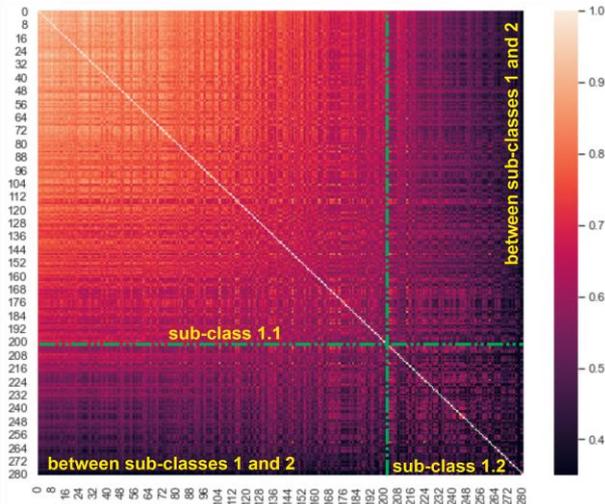


Fig. 3. PCCs within class 1 (OK motors) rearranged in decreasing order of average PCC values.

To show the difference between OK motor L (with the largest average PCC) and S (with the smallest average PCC) within class 1, their lin-power spectrograms are shown in Fig. 4.

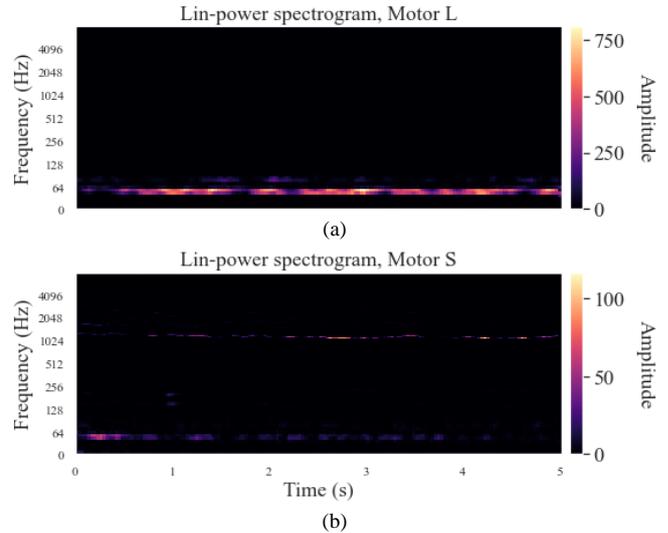


Fig. 4. Lin-power spectrograms for (a) OK motor L (with the largest average PCC) and (b) S (with the smallest average PCC) within class 1.

Statistical quantities of the PCCs calculated within the newly created sub-classes and between the sub-classes are given in Table V. Comparing these results with those summarized in Table I, it can be seen that the mean, minimal, and median PCCs within sub-class 1.1 are considerably greater than the corresponding ones within class 1. On the other hand, quite the opposite can be said for these values within the sub-class 1.2, where they are significantly smaller.

TABLE V. PCCs BETWEEN OK MOTORS IN SUB-CLASSES 1.1 AND 1.2, AS WELL AS BETWEEN THESE SUB-CLASSES FOR LIN-POWER SPECTROGRAM.

Sub-class	Statistical quantity				
	Mean	Min.	Max.	Median	STD
Sub-class 1	0.721	0.459	1.0	0.715	0.085
Sub-class 2	0.474	0.170	1.0	0.462	0.108
Between sub-classes 1 and 2	0.516	0.269	0.857	0.516	0.078

VI. CONCLUSIONS

Based on the results obtained related to the evaluation of the image similarity, the applied spectrogram-like images can be divided into two groups, where the first includes the lin-power spectrogram, mel-spectrogram, gammatonegram, and CQT power spectrogram, while the second group includes STFT and CQT chromagrams and tempogram. The images from the first group give larger PCC values between the OK motors (intra-class similarity for class 1) and smaller PCC between the NOK motors (intra-class similarity for class 2) as well as between OK and NOK motors (inter-class similarity). This is not the case for the images from the second group, where it is hard to see a significant distance between the intra-class and inter-class similarity. This is why the images from the first group are considered here as a better option for using them as input to a deep learning model.

Among the images from the first group, the lin-power

spectrogram represents an image with full resolution on the y axis, meaning that the size of an image is significantly larger than in the other three images from this group. Consequently, the preference for usage as a deep learning model input is given to the other three images of the first group. When they are compared, the greatest relative distance between the intra-class and inter-class similarity is obtained for the mel-spectrogram. This image is used in a number of other studies, including the IEEE AASP challenge on the detection and classification of acoustic scenes and events - DCASE 2021 [19].

One more important conclusion is that the distance between the intra-class and inter-class similarity is rather small, showing that the sound samples and images obtained from them are rather similar even when they belong to different classes. In addition, in each particular class used in this article (class of OK motors or class of NOK motors), there are sound samples having a stronger correlation (larger correlation coefficient) and those having a lower correlation coefficient within the same class. This property is further explored here by using the PCCs of the lin-power spectrogram to rearrange (order) the motors belonging to the class 1 - OK motors. Thus, the image having the largest similarity with other images within the same class (the largest average PCC) is placed at the topmost position, the next image according to the similarity at one position below, and so on. Apart from ordering the motors according to decreasing image similarity (correlation coefficient), it becomes possible to form a sub-class of class 1 containing motors with more similar lin-power spectrograms than it is the case within the whole class 1.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

- [1] A. Nakhmani and A. Tannenbaum, "A new distance measure based on generalized image normalized cross-correlation for robust video tracking and image recognition", *Pattern Recognit. Letters*, vol. 34, no. 3, pp. 315–321, Feb. 2013. DOI: 10.1016/j.patrec.2012.10.025.
- [2] E. Dumic, I. Bacic, and S. Grgic, "Simplified structural similarity measure for image quality evaluation", in *Proc. of 2012 19th International Conference on Systems, Signals and Image Processing (IWSSIP)*, 2012, pp. 442–447.
- [3] V. V. Starovoitov, E. E. Eldarova, and K. T. Iskakov, "Comparative analysis of the SSIM index and the Pearson coefficient as a criterion for image similarity", *Eurasian Journal of Mathematical and Computer Applications*, vol. 8, no. 1, pp. 76–90, 2020. DOI: 10.32523/2306-6172-2020-8-1-76-90.
- [4] A. M. Neto, A. C. Victorino, I. Fantoni, D. E. Zampieri, J. V. Ferreira,

- and D. A. Lima, "Image processing using Pearson's correlation coefficient: Applications on autonomous robotics", in *Proc. of 2013 13th International Conference on Autonomous Robot Systems*, 2013, pp. 1–6. DOI: 10.1109/Robotica.2013.6623521.
- [5] M. Lüthe, "Calculate Similarity — The most relevant Metrics in a Nutshell". [Online]. Available: <https://towardsdatascience.com/calculate-similarity-the-most-relevant-metrics-in-a-nutshell-9a43564f533e>
- [6] M. Sigmund and M. Hrabina, "Efficient feature set developed for acoustic gunshot detection in open space", *Elektronika ir Elektrotechnika*, vol. 27, no. 4, pp. 62–68, Aug. 2021. DOI: 10.5755/j02.eie.28877.
- [7] Y. Huang, Q. Xu, X. Jiang, Y. Yang, and G. Lin, "An analytic approach to probabilistic load flow incorporating correlation between non-Gaussian random variables", *Elektronika ir Elektrotechnika*, vol. 24, no. 3, pp. 61–67, Jun. 2018. DOI: 10.5755/j01.eie.24.3.20980.
- [8] L. Nowosielski and B. Uljasz, "The automatic radio transmission systems recognition", *Elektronika ir Elektrotechnika*, vol. 23, no. 6, pp. 77–84, Dec. 2017. DOI: 10.5755/j01.eie.23.6.19698.
- [9] Z. Peric, B. Denic, and V. Despotovic, "Multilevel delta modulation with switched first-order prediction for wideband speech coding", *Elektronika ir Elektrotechnika*, vol. 24, no. 1, pp. 46–51, Feb. 2018. DOI: 10.5755/j01.eie.24.1.20156.
- [10] A. Laucka, V. Adaskeviciute, D. Andriukaitis, "Research of the Equipment Self-Calibration Methods for Different Shape Fertilizers Particles Distribution by Size Using Image Processing Measurement Method," *Symmetry*, vol. 11, no. 7, p. 838, Jun. 2019. DOI: 10.3390/sym11070838.
- [11] K. Okarma, P. Lech, and V. V. Lukin, "Combined full-reference image quality metrics for objective assessment of multiply distorted images", *Electronics*, vol. 10, no. 18, p. 2256, 2021. DOI: 10.3390/electronics10182256.
- [12] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity", *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, 2004. DOI: 10.1109/TIP.2003.819861.
- [13] A. Kaur, L. Kaur, and S. Gupta, "Image recognition using coefficient of correlation and structural similarity index in uncontrolled environment", *Int. J. Comp. App.*, vol. 59, no. 5, pp. 32–39, 2012. DOI: 10.5120/9546-3999.
- [14] V. Singh, V. Devgan, and I. Anand, "Determining image similarity with quasi Euclidean metric", 2020. arXiv: 2006.14644.
- [15] L. Wang, Y. Zhang, and J. Feng, "On the Euclidean distance of images", *IEEE Tran. Pattern Analysis Machine Intelligence*, vol. 27, no. 8, pp. 1334–1339, 2005. DOI: 10.1109/TPAMI.2005.165.
- [16] E. K Yen and R. G. Johnston, "The ineffectiveness of correlation coefficient for image comparisons", Vulnerability Assessment Team, Los Alamos National Laboratory, New Mexico. [Online]. Available: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.47.399>
- [17] D. Ćirić, Z. Perić, J. Nikolić, and N. Vučić, "Audio signal mapping into spectrogram-based images for deep learning applications", in *Proc. of 2021 20th International Symposium INFOTEH-JAHORINA (INFOTEH)*, 2021, pp. 1–6. DOI: 10.1109/INFOTEH51037.2021.9400698.
- [18] D. Ćirić, Z. Perić, J. Nikolić, and N. Vučić, "Intra-class and inter-class differences in mel-spectrogram images of DC motor sounds", in *Proc. of 2021 15th International Conference on Advanced Technologies, Systems and Services in Telecommunications (TELSIKS)*, 2021, pp. 189–192. DOI: 10.1109/TELSIKS52058.2021.9606288.
- [19] DCASE2021 Challenge, IEEE AASP Challenge on Detection and Classification of Acoustic Scenes and Events, 1 March - 1 July 2021. [Online]. Available: <http://dcase.community/challenge2021/index>



This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution 4.0 (CC BY 4.0) license (<http://creativecommons.org/licenses/by/4.0/>).