Hindawi
Journal of Sensors
Volume 2018, Article ID 5828074, 13 pages
https://doi.org/10.1155/2018/5828074

Research Article

A Two-Level Sound Classification Platform for Environmental Monitoring

Stelios A. Mitilineos, Stelios M. Potirakis, Nicolas-Alexander Tatlas, and Maria Rangoussi

Department of Electrical and Electronics Engineering, University of West Attica, Campus 2, 250 Thivon and P. Ralli, Aigaleo, 122 44 Athens, Greece

Correspondence should be addressed to Stelios A. Mitilineos; smitil@gmail.com

Received 21 November 2017; Accepted 6 May 2018; Published 3 June 2018

Academic Editor: Eduard Llobet

STORM is an ongoing European research project that aims at developing an integrated platform for monitoring, protecting, and managing cultural heritage sites through technical and organizational innovation. Part of the scheduled preventive actions for the protection of cultural heritage is the development of wireless acoustic sensor networks (WASNs) that will be used for assessing the impact of human-generated activities as well as for monitoring potentially hazardous environmental phenomena. Collected sound samples will be forwarded to a central server where they will be automatically classified in a hierarchical manner; anthropogenic and environmental activity will be monitored, and stakeholders will be alarmed in the case of potential malevolent behavior or natural phenomena like excess rainfall, fire, gale, high tides, and waves. Herein, we present an integrated platform that includes sound sample denoising using wavelets, feature extraction from sound samples, Gaussian mixture modeling of these features, and a powerful two-layer neural network for automatic classification. We contribute to previous work by extending the proposed classification platform to perform low-level classification too, i.e., classify sounds to further subclasses that include airplane, car, and pistol sounds for the anthropogenic sound class; bird, dog, and snake sounds for the biophysical sound class; and fire, waterfall, and gale for the geophysical sound class. Classification results exhibit outstanding classification accuracy in both high-level and low-level classification thus demonstrating the feasibility of the proposed approach.

1. Introduction

European countries display one of the richest cultural legacies in the world. With millions of tourists drawn each year to landmark cultural heritage sites, the economic and financial impact of European cultural heritage is considered to be a priority for policymakers but also for the people of Europe [1–4]. Therefore, the conservation of European cultural heritage is critical in order to preserve the European identity but also because cultural heritage may boost economic impact. Alas, heritage sites are exposed to both anthropogenic activity (noise, vandalism, and pollution) and environmental phenomena or natural hazards that may compromise their value. Therefore, preventive measures need to be taken in order to mitigate the negative effects of anthropogenic activity and climate change and preserve cultural heritage artefacts and sites.

In this context, many European institutions have carried out substantial work on preventive strategies aimed at protecting the EU cultural buildings and sites. One of the first related projects, “Carta del Rischio” (“Risk Map”), was carried out in Italy in the early 1990s and completed a long and complex survey of territorial-based environmental and human-caused risks in order to develop the first ever risk map for cultural heritage across Italy [5]. Thereinafter, more countries followed Italy’s example and created similar works. An example is the HAR Programme (“Heritage at Risk”), produced by the Historic England organization [6], which resulted in two surveys at 1998 [former “Monuments At Risk Survey (MARS) 1998”] and 2008, helping to establish priorities for action and monument management. The “Carta de Risco do Património Arquitectónico” produced in Portugal by the Direcção-Geral dos Monumentos Nacionais is a similar project but one that is specifically targeted to architectural monuments. The EMERIC programme in Greece was a subproject of the CRINNO project for innovative actions in the island of Crete and included an activity for the tectonic and seismic risk assessment of the historical centers of the main cities of Crete [7]. Finally, the PROHITECH project investigated a series of available and novel technologies for the protection of historical buildings against earthquakes and other threats in the Mediterranean and Balkan areas [8].

Our work is part of a larger project ("STORM") that aims in developing a complete platform of technical and managerial resources for cultural heritage sites' safeguarding [4]. STORM will build upon previous work and combine upgraded legacy sensor systems with novel sensing technologies, like wireless acoustic sensor networks (WASNs), in order to provide a novel framework over which to determine how vulnerable structures are affected by risks associated to climatic conditions and anthropogenic activity. Part of the protective mechanisms is the implementation of a WASN platform that will monitor and continuously store sound samples originating from acoustic nodes' surroundings. The collected samples may correspond to environmental sounds regarding natural phenomena or human actions. The WASN-captured data will be transmitted to a central server to populate sound maps and create a database with history of the occurred events, thus alarming stakeholders in cases of potentially malevolent or hazardous events. Herein, we designate as high-level sound classification the act of classifying a sound sample to three classes, namely, *anthropogenic*, *biophysical* (other than human), and *geophysical* sound classes [9, 10]. We further contribute to previous work by extending the proposed classification platform to perform low-level classification too, that is, classify sounds to further subclasses that include airplane, car, and pistol sounds for the anthropogenic sound class; bird, dog, and snake sounds for the biophysical sound class; and fire, waterfall, and gale for the geophysical sound class. We examine two different approaches regarding second-level classification. The first approach is to directly classify each sound sample to one of the nine available classes. The other approach is to first classify each sound sample at high level and then perform classification in a smaller set of available classes depending on the high-level classification result. We report results that indicate the latter method to perform better. Furthermore, we present an integrated platform that includes sound sample denoising using wavelets, feature extraction from sound samples, and Gaussian mixture modeling of these features, as well as the proposed two-layer neural network classifier for the automated classification of incoming sound samples. Numerical results exhibit satisfactory classification accuracy in both high-level and low-level classification levels, thus demonstrating the feasibility of the proposed approach.

The rest of the paper is organized as follows: First, a short literature review of recent sound classification approaches is given in Section 2, providing a justification of the classification approach followed thereinafter. The proposed classification platform is described in Section 3 along with the necessary definition of the employed signal processing tools. The simulation results are presented in Section 4, and the conclusions are summarized in Section 5.

2. Sound Classification Approaches

In the literature, sound classification is performed using carefully selected sound features that feed a classifier tool like a neural network. The selection of sound features directly affects the performance of the classification procedure and is a demanding task since recorded sounds are typically nonstationary signals while there is also a superimposed background "noise" that originates from natural ambient sounds. Furthermore, sound events are overlapping in space and time and signals originating from neighboring sensors are typically highly correlated [11]. A variety of sound features have been proposed in the literature in order to perform environmental sound monitoring. These features may be either related to the time-domain representation of the signal, for example, zero-crossing rate (ZCR), linear prediction coefficients (LPC), audio signal energy function, and volume, or to the frequency-domain representation, for example, pitch, bandwidth, fundamental frequency, spectral peak track, brightness, mel-frequency cepstral coefficients (MFCC), and short Fourier transform coefficients. There are also many statistical features like the variance, skewness, kurtosis, median, and mean value, as well as various complexity measures (entropies, information) of the signal [12–17]. Other spectral features used in the literature include the 4 Hz modulation energy, percentage of low frames, spectral centroid, spectral roll-off point, spectral frequency, mean frequency, and high and low energy slopes [18–21]. Furthermore, automatic identification is necessary in order to monitor large areas of interest while keeping the operating costs low. Straightforwardly, researchers start up by deploying a network of wireless microphone sensors (WASN) over a large area that capture and transmit environmental sound data samples to a central server. These samples are partitioned into frames and are being processed in order to identify the sound source that created them [11, 17, 22]. Most often, *soundmaps* are created in order to visualize the audio content of large areas, as for example in [23].

Although spectral features are useful in audio classification, they do not provide any information about the temporal evolution of the signal. Therefore, spectral features alone are not enough to represent environmental audio signals that are highly nonstationary in nature. Time-frequency (TF) features have been introduced in order to capture the temporal variation of the spectra of such signals. TF features are effective for revealing nonstationary signal aspects such as trends, discontinuities, and repeating patterns. The usual approach is to extract spectral features for each frame, allowing a certain percentage of overlap between adjacent frames, to produce one of the well-known TF representations like spectrograms, scalograms, or different representations belonging to Cohen's class. However, this approach results in huge feature spaces. Different solutions for the reduction of the resulting data have been proposed. For example, spectrum flux, defined as the average variation value of spectrum between two adjacent frames, can be used.

An effective proposed solution is to use Gaussian mixture models (GMMs) to estimate the probability distribution function of spectral features over all frames. On the other hand, Ghoraani and Krishnan proposed to construct a so-called time-frequency matrix (TFM) of audio signals using a matching pursuit time-frequency distribution technique [18]. Chu et al. also used matching pursuit but with Gabor atom signal representation in order to obtain effective TF features [13]. We consider that the last two approaches impose a prohibitive processing cost (at the embedded level) for our distributed sensor nodes and propose to replace matching pursuit and time-frequency features with probability distribution fitting of temporally varied frequency features (in essence a 1D GMM as it will be explained in detail in the following subsections). After careful consideration and overview of the available literature, we chose to use the following features in our platform: zero-crossing rate, pitch, bandwidth, MFCCs, spectrogram coefficients, and a variety of statistical features, namely, different complexity measures (Shannon, Tsallis, wavelet, and permutation entropies). In order to capture the temporal variation of spectral features, we calculated the GMM of each one of them. Our goals for this selection of features were to keep a high level of performance and robustness together with ease of implementation and low complexity level. Numerical results that are presented in Section 4 of this work justify this approach.

The sound classification task is based on the assumption that every sound source exhibits a specific pattern of distributing its energy over frequency and time. A successful sound classifier should be able to categorize sounds that belong to nonconvex classes of the feature space. Sound classifiers broadly fall under two varieties: *discriminative* and *nondiscriminative*. Examples of the former include the k-means classifier, the polynomial classifier, the multilayer perceptron (neural network), and the support vector machines; such classifiers try to designate a boundary among training data input and match its test input to a specific data class. On the other hand, nondiscriminative classifiers like the hidden Markov model (HMM) attempt to model the underlying distribution of the training data [15, 24, 25]. For the proposed classification platform of the STORM project, we selected to use a generic discriminative classifier, that is, a neural network (NN), since NNs are well-known classifiers that have been extensively used for signal and audio classification purposes. Even though the ANN training needs a high processing power, we assume this to be made available by a central processing server while GMM modeling is adopted at a sensor node level to keep transmission data volume, from sensor node to server, to a minimum.

As long as the specific classes of sounds to be identified are concerned, the available literature is specifically oriented to sound classification for the purposes of environmental monitoring. For example, often we are not interested in identifying a specific bird species or subspecies but rather identify whether or not birds are present at a specific point of an area. The *first level of identification hierarchy* consists in identifying the general sound type, for example, whether it originates from human (anthropogenic) or animal (biogenerated other than human) activity or whether it is an ambient natural sound (e.g., waterfall and fire). This categorization is very popular in the respective literature [26–28]. The *second level of identification hierarchy* consists in further identifying for each sound type a more focused sound origin; for example, whether an animal activity is actually a bird, a snake, or, say, a dog. Deeper levels of identification hierarchy can also be defined where for each next hierarchy step an even narrower and more specific sound class is defined. In this context, a first approach could be to directly detect the second-level sound class. However, we propose to use a two-step approach, where the sound is firstly broadly classified as anthropogenic, animal, or natural ambient, and then it is further classified in a more detailed manner. Numerical results (see Section 4) indicate that this approach delivers much higher performance compared to direct second-level (one-step) sound classification.

Finally, it is worthwhile noting that wavelet analysis has also been used in environmental monitoring for audio signal analysis or signal denoising [25]; herein, we selected to employ wavelet for signal denoising where necessary.

In the following sections, we discuss the proposed sound classification platform approach and present numerical results that demonstrate its applicability in terms of high achieved performance and robust sound classification results.

3. Classification Platform Overview

An overview of the proposed classification platform, together with a discussion on its main components and their interconnectivity, is included in the subsections below.

3.1. General Presentation of the Proposed Platform. The functionality of the proposed platform is illustrated by the flowchart depicted in Figure 1. Every time a sound signal is fed to the platform, there is a decision as to whether the signal will be subject to denoising via wavelet analysis (decomposition and reconstruction) or not. Afterwards, the selected signal features are calculated either for the reconstructed or for the original signal. The features list includes the zero-crossing rate, the pitch, the bandwidth, the MFCCs, the spectrogram coefficients, and a variety of complexity measures including the Shannon, Tsallis, wavelet, and permutation entropies. The zero-crossing rate, the pitch, the bandwidth, and the entropies are scalar features and therefore very efficient in terms of computational cost during classification. On the contrary, the MFCCs and the spectrogram coefficients are TF and thus multidimensional features. More specifically, the signal sample is partitioned into frames and the MFCCs and spectrogram coefficients are calculated for each particular frame. Thereupon, if for example we employ 13 MFCCs and 16 spectrogram coefficients (a popular choice in the literature) for each frame, and a signal is split into 2048 frames (also, a not uncommon case), then we need 2048 vectors of 29 dimensions each (i.e., 59,392 elements); this is a huge number of elements to be used as classifier input. The approach adopted herein as a solution to this problem is to perform a statistical fit of spectral features to a sum of Gaussian probability distribution functions (PDFs) that are fully characterized by only their mean and standard deviation

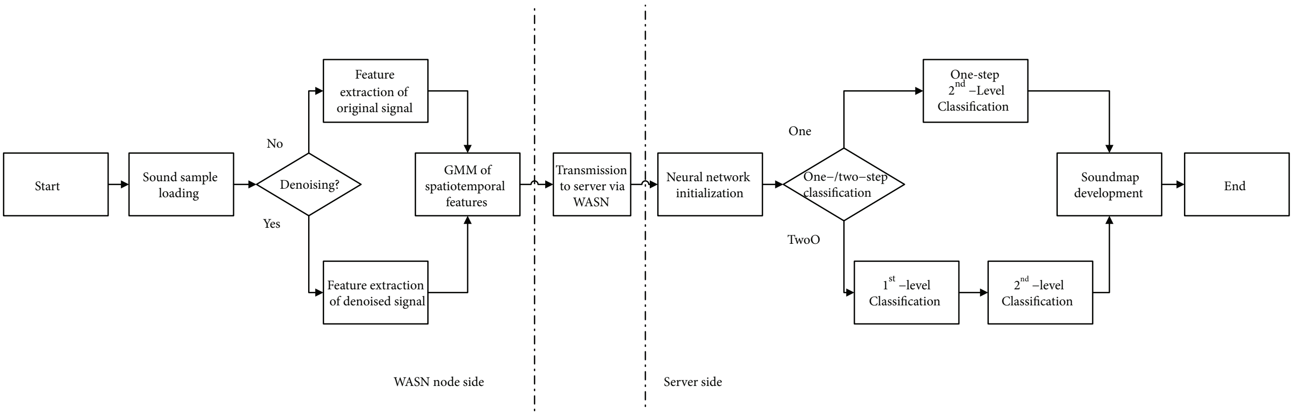

Start
Sound sample loading
Denoising?
No
Yes
Feature extraction of original signal
Feature extraction of denoised signal
GMM of spatiotemporal features
Transmission to server via WASN
Neural network initialization
One–/two–step classification
One
TwoO
One-step 2nd –Level Classification
1st –level Classification
2nd –level Classification
Soundmap development
End
WASN node side
Server side

Figure 1: Proposed sound classification platform functionality.

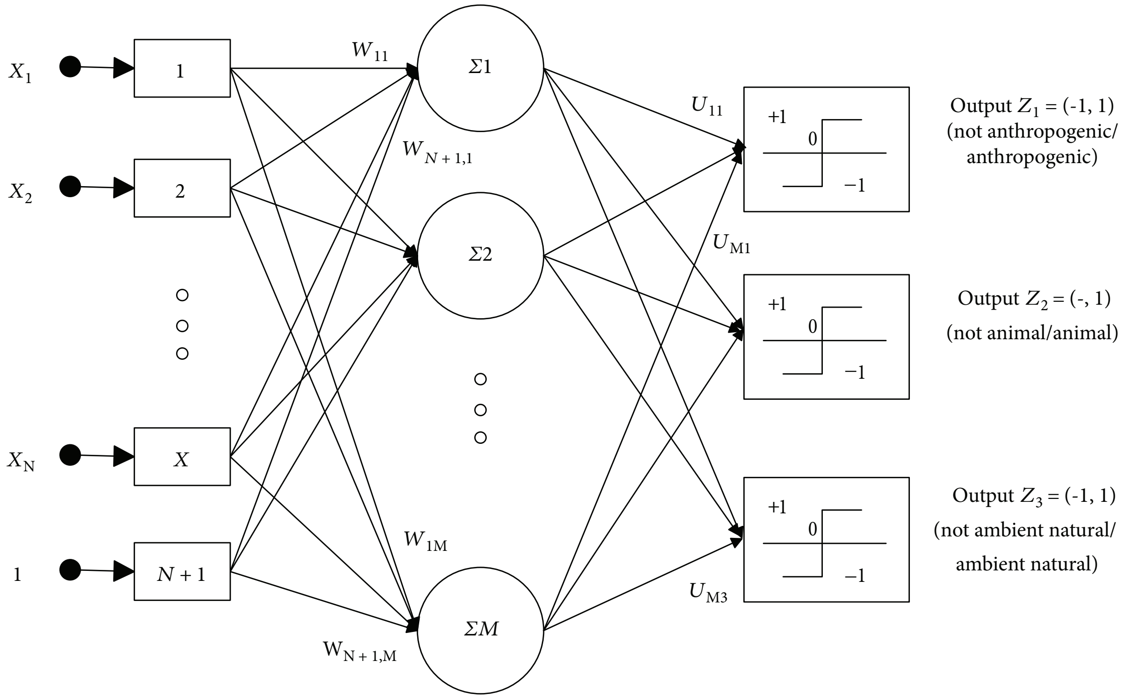


Figure 2: Neural network classifier for high-level sound classification.

values; this results to a dramatic reduction of the feature space dimensionality.

Then, a decision is made as to whether sound classification is going to be implemented as a two-step or one-step procedure. In the former case, the sound is first classified to a broad sound category or high-level hierarchy (anthropogenic, ambient natural sound, or animal-originated sound) and then further classified to a more specific class. In the latter case, the sound is directly classified to the more specific class in a one-step procedure. In particular, for the numerical results presented herein, we have used samples from the following sound subclasses: (i) anthropogenic: samples of airplanes', cars', and pistols' sounds; (ii) ambient natural sounds: samples of waterfall, gale, and fire sounds; and (iii) animal-originated sounds: samples of crows', dogs barking, and snakes rattling sounds.

As an example of neural network implementation, consider the high-level classification case. All calculated features are fed to a properly configured NN, as shown in Figure 2. We consider only feedforward artificial neural networks (FANNs) with the training function being an error backpropagation variant. The input layer of the network is used for data entry and weighting. The weights that multiply each data entry are subject to the network's training that is performed off-line and prior to classification. The weighted input features are then forwarded to an intermediate layer of neurons. The middle layer's number of nodes is tuned around the empirical rule-of-thumb value of one and a half times the number of input layer nodes. These neurons sum up all the weighted features and, essentially, configure all possible convex classes of data in the feature space. The output of the intermediate layer is then forwarded to three output neurons. These neurons at the output layer are essentially combining convex classes in order to configure nonconvex classes to classify the input data. Each output is taking a value of "+1" or "−1" that corresponds to "true" or "false" state, respectively.

It is worthwhile noting that 2nd-level classification is implemented in a straightforward manner similar to the procedure described in Figure 2. More specifically, in the second step there are three NNs, namely, the "anthropogenic sounds NN," the "animal sounds NN," and the "geophysical sounds NN"; each one of them is activated only when the respective output of the NN in Figure 2 is true and is fed by the same inputs as the NN of Figure 2. The anthropogenic sounds NN has three output classes, namely, airplanes, cars, and pistols; the animal sounds NN has the birds, dogs, and snakes output classes; finally, the geophysical sounds NN has the gale, rain, and waterfall output classes. This way, at the first step the general class is designated while at the second step the specific subclass of the input sound is identified. On the other hand, as long as one-step second-level classification is considered, there is only one NN with input and intermediate layer similar to that of Figure 2; however, in the specific case there are nine output nodes (i.e., nodes for each one of the subclasses of airplanes, cars, pistols, birds, dogs, snakes, gale, rain, and waterfall) and classification to subclasses is performed by one NN only. Both these cases are not presented using a figure for the sake of brevity.

Finally, postclassification performance results are derived in order to evaluate the applicability of our approach. For each feature's combination mentioned, above we implemented one hundred NNs in order to capture the statistical behavior of node weight assignment during training. For each sound sample and network implementation, we store the confusion matrix and the percentage of correct classifications. The performance of the network is thus evaluated by performance metrics of independent (total correct classifications) as well as dependent (percentage of correct classifications given that a particular sound type is loaded) random variable results.

3.2. Definition of Features

3.2.1. Zero-Crossing Rate. For discrete time signals, a zero-crossing is said to occur if successive samples have different signs [14]. The zero-crossing rate (ZCR) is defined as

$$\mathrm{ZCR} = \frac{1}{2(N-1)} \sum_{n=2}^{N} \left| \operatorname{sgn}(s(n)) - \operatorname{sgn}(s(n-1)) \right|, \quad (1)$$

where n is the discrete time index, N is the total number of time slots, $s(n)$ is the signal value at time index n, and $\operatorname{sgn}(x)$ is the sign function given by

$$\operatorname{sgn}(x) = \left\{ \begin{array}{c} 1, s(x) \geq 0 \\ -1, s(x) < 0 \end{array} \right\}. \quad (2)$$

3.2.2. Pitch. Pitch is a perceptual feature of the audio signal that depends on the fundamental frequency of the audio waveform. Pitch information can be extracted by using either temporal or frequency analysis. The temporal analysis method is based on the computation of the autocorrelation function or the average magnitude difference function, while with the frequency analysis method the pitch is determined from the periodic structure in the magnitude spectrum of the Fourier transform of an audio frame. The autocorrelation function of a signal is given by

$$R_n(i) = \sum_{n=1}^{N-i} s(n) \cdot s(n+i), \quad (3)$$

where i is the shift. A simple way to calculate the pitch is to estimate $i_{\max}$ that maximizes $R_n(i)$ (i.e., $|R_n(i_{\max})| \geq |R_n(i)|$); the pitch will then be equal to $1/i_{\max} \cdot dt$ where dt is the length of each time slot.

3.2.3. Spectrogram Coefficients. The spectrogram coefficients of a discrete time signal $s(n)$ are essentially the components of the discrete Fourier transform of the signal; the spectrum of such a signal is given by

$$F(f) = \mathrm{DFT}(s(n)), \quad (4)$$

where $\mathrm{DFT}(s(n))$ is the discrete Fourier transform of the signal $s(n)$. The spectrogram is the evolution of the spectrogram coefficients over time and is a TF feature of the signal.

3.2.4. MFCCs. Mel-frequency cepstral coefficients (MFCCs) are very popular in speech/speaker feature extraction and aim at representing the hearing properties of the human ear by using a nonlinear scale of frequencies (i.e., the "mel-frequency" in mel units versus the conventional frequency i.e. measured in Hz) More specifically, the output of the human ear (i.e., output to the auditory processing cells of the human brain) is the convolution of the excitation signal (i.e., the sound under investigation) and the vocal tract filter. The mel transform essentially transforms the spectral coefficients of the sound signal to the mel-frequency domain; then, the cepstral coefficients (as opposed to the spectral coefficients) of the mel-frequency signal components are calculated.

An example of mel transformation is given by

$$M(f) = \left\{ \begin{array}{cc} f, & 0 \leq f \leq 1\,\mathrm{kHz} \\ 1127 \cdot \ln\left(1 + \dfrac{f}{700}\right), & 1\,\mathrm{kHz} < f \end{array} \right\}, \quad (5)$$

while an example of a cepstral function calculation formula is given by (the cepstral function is the real cepstrum as opposed to the spectral function i.e. the real spectrum of the signal)

$$c(\tau) = \mathrm{IDTFT}(S(M)), \quad (6)$$

where $\mathrm{IDTFT}(S(M))$ is the inverse discrete-time Fourier transform of the cepstrum magnitude; S(M) is the cepstrum magnitude of the discrete time signal $s(n)$ (i.e., in the mel-frequency domain).

MFCCs, like the spectrogram coefficients, are calculated for each particular frame, and their evolution over time is in essence a TF feature.

3.2.5. Complexity Measures. There is a wide variety of signal complexity metrics, including different kinds of entropic/information measures. We focus on different entropies that have recently been investigated as to their "insensitiveness" to specific signal compression schemes. As it has recently been shown [17], the precision of specific entropic/information metrics remains reasonably unchanged by certain compression schemes in the sense that the numerical values obtained for these metrics when applied to a compressed signal are very close to the corresponding ones that are obtained when applied to the specific signal in its unprocessed form. In the present paper, we used the Shannon, Tsallis, wavelet, and permutation entropies, the basic formulae of which are presented in the following.

Let $s_k = s(t_k)$ be a discrete measured variable, with $t_k = kT$, $k = 1, 2, \dots, K$, and T being the sampling period. One can then define a set of N disjoint but adjacent intervals (bins) spanning the observed range of values of the time-series $\{s_k\}$, denoted as $\{x_n\}$, $n = 1, 2, \dots, N$. Let also $P = \{p(x_1), p(x_2), \dots, p(x_N)\}$ be a finite discrete probability distribution, with $\sum_{n=1}^{N} p(x_n) = 1$, which describes the probabilities for the samples of the time-series to belong to each one of these N bins; the probability for a sample of the time-series to belong to the nth bin can be denoted as $p(x_n)$. The informational content of the normalized probability distribution P is given by Shannon's information measure as [29]

$$H_{sh} = -K_{sh} \sum_{i=1}^{N} p(x_i) \log[p(x_i)], \qquad (7)$$

where K_{sh} is a positive constant (it merely amounts to a choice of a unit of measure; however, it is usually set equal to 1). The choice of a logarithmic base corresponds to the choice of a unit for measuring information [29]. H_{sh} has been forwarded by Shannon as a measure of information, choice, and uncertainty. The decrease in Shannon entropy is attributed to an increase in the information content and order and, equivalently, to a decrease in complexity. Shannon entropy is recognized as a basic tool for the description of the information, behavior, and complexity of physical, sociological, economic, technological, and so on, systems and their observables, like time-series of measurable quantities that characterize them.

Another statistical representation of a time-series results from the probabilistic analysis of its spectrum. In this approach, instead of analyzing a time-series in terms of the probability of occurrence of its amplitude values, as in the case of Shannon entropy, a time-series is analyzed in terms of the distribution of its energy to frequencies or scales. The Shannon-like total wavelet entropy, or wavelet energy entropy, is defined in this context by [30]

$$H_{\mathrm{WT}} = -\sum_{j<1} p_j \ln p_j, \qquad (8)$$

where $p_j = E_j / E_{\mathrm{tot}}$ expresses the probability distribution of the energy at different scales of the wavelet spectrum of a signal as it results after the application of the continuous wavelet transform (CWT) on it; it holds that $\sum_j p_j = 1$ and the distribution $\{p_j\}$ can be considered as a time-scale density [30]. Note that the energy at resolution j is $E_j = \sum_k |C_{j,k}|^2$, while the total energy is $E_{\mathrm{tot}} = \sum_{j=-N}^{-1} \sum_k |C_{j,k}|^2$, and the signal is considered to be expanded as $y(t) = \sum_{j=-N}^{-1} \sum_k C_{j,k} \psi_{j,k}(t)$, where $j = -1, -2, \dots, -N$ is the number of resolution levels, corresponding to octave scales [30]. Like the other entropies, wavelet entropy decreases as a result of complexity decrease and order increase.

Symbolic dynamics refers to the mapping of the observables of a complex system (the real values of a time-series) to a sequence of symbols attempting to access useful information ([31] and references therein). The entropic analysis within the context of symbolic dynamics examines the probabilities of appearance of these symbols rather than the probabilities of appearance of the actual real values of the original time-series, thus providing a different kind of statistical representation of the system under analysis. Recently, a new form of symbolic mapping and a corresponding complexity metric has been proposed in the form of permutation entropy (PE) [32]. According to this approach, a time-series is first embedded to a m-dimensional space by building vectors $\mathbf{Y}_k$, each of which contains m values of the original time-series such that every two neighboring vector elements have a time distance equal to L in the original time-series. For every vector $\mathbf{Y}_k$, its m real values are then arranged in an increasing order. This way, each vector $\mathbf{Y}_k$ is uniquely mapped onto a new vector $\pi = [j_1, j_2, \dots, j_m]$, where π is one of $m!$ possible permutations of the vector of indices of $\mathbf{Y}_k$'s elements $[1, 2, \dots, m]$. If each of the $m!$ permutations is considered as a symbol, then the procedure allows the mapping of the original continuous time-series to a symbolic sequence [33]. The relative frequency of appearance of each possible permutation π in the time-series, as obtained during the sorting process of all vectors $\mathbf{Y}_k$, is denoted as [34]

$$p(\pi) = \frac{\text{the number of } \pi \text{ permutations found}}{K - (m-1)L}, \qquad (9)$$

while PE is defined according to the Shannon entropy way as

$$H_m = -\sum p(\pi) \ln p(\pi), \qquad (10)$$

where the sum runs over all $m!$ permutations of order m [33]. PE is a measure of regularity in the time-series. When the time-series is so irregular that all $m!$ possible permutations appear with the same probability $p(\pi) = 1/m!$ (completely random), then H_m reaches the maximum value $\ln(m!)$. On the other hand, with increasing regularity, that is, reduced complexity, H_m decreases. For convenience, we usually employ the normalized permutation entropy, by normalizing H_m by $\ln(m!)$ to handle entropy values in the interval $[0, 1]$.

Long-range spatial interactions or long-range memory effects may be observed in a vast variety of complex systems influencing their behavior. A very interesting class of such systems is formed by those characterized by nonextensive statistics. These systems share a very subtle property: they violate the main hypothesis of Boltzmann-Gibbs (B-G) statistics, that is, ergodicity. Inspired by multifractal concepts, Tsallis [35, 36] has proposed a generalization of the B-G statistical mechanics that covers systems that violate ergodicity, that is, systems of the microscopic configurations which cannot be considered as (nearly) independent. This generalization is based on nonadditive entropies, S_q, characterized by an index q which leads to a nonextensive statistics [36] as in

$$S_q = k \frac{1}{q-1} \left\{ 1 - \sum_{i=1}^{N} [p(x_i)]^q \right\}, \tag{11}$$

where $p(x_i)$ are the probabilities associated with the value bins x_i, as was previously defined for the Shannon entropy case, N is their total number, q is a real number, and k is Boltzmann's constant. The value of q is a measure of the nonextensivity of the system. Notice that in the limit where $q \to 1$, nonextensive statistics converges to the standard, extensive, B-G statistics [35]. Note that the parameter q itself is not a measure of the complexity of the system but measures the degree of nonextensivity of the system. The value of q represents the strength of the long-range correlations governing the dynamics of the system [37]. The cases $q > 1$ and $q < 1$ correspond to subextensivity or superextensivity, respectively. On the other hand, the time variations of the Tsallis entropy, S_q, for a given q quantify the dynamic changes of the complexity of the system. Lower S_q values characterize signals with lower complexity.

3.3. Wavelet Denoising. Denoising refers to processing a noisy signal aiming at the reduction of unwanted noise in such a way that this reduction is as high as possible while at the same time the useful signal is distorted as less as possible. One way to reduce the noise contaminating a signal is to decompose the noisy signal into a number of decomposition levels using the discrete wavelet transform (DWT) and an appropriate orthogonal wavelet basis [38] and then to reconstruct it using only the components that correlate to the useful signal. This is possible by (hard or soft) thresholding that reduces those components' coefficients that correspond to noise [39].

Both the decomposition and the reconstruction processes were performed using Mallat's fast algorithm [38]. According to this algorithm, a hierarchical multiresolution analysis of the signal is performed by using a set of consecutive low- and high-pass filters followed by a decimation; the outputs of these filters are usually referred to as the approximation coefficients and the detail coefficients, respectively. At each level of decomposition, the output of the low-pass filter of the previous level of decomposition (or the original signal for the first level) is fed to a new pair of low- and high-pass filters, the frequency band of which is the half of those of the previous level. As such, the output of each filter can be decimated (downsampled) by a factor of 2. Using this hierarchical approach results in a good time resolution at high frequencies (low scales) and good frequency resolution at low frequencies (high scales).

3.4. Definition of GMM. A mixture model is used in statistics in order to represent the presence of data subpopulations within an overall population without the need to identify such subpopulations explicitly. We are using Gaussian mixture models in order to statistically fit MFCC and spectrogram coefficient evolution over time to a PDF. A Gaussian mixture model (GMM) essentially dictates that the empirical PDFs of these coefficients are the weighted sum of Gaussian PDFs of different mean values and standard deviations. In the proposed platform, the user selects the number of Gaussian PDFs to configure the GMM, and the expectation maximization algorithm is used in order to calculate their parameters.

The PDF of a GMM is defined by

$$p(x) = \sum_{i=1}^{G} w_i g(m_i, \sigma_i), \tag{12}$$

where G is the total number of Gaussian PDFs participating in the GMM, w_ι is the weight of the ith Gaussian PDF, and $g(m_i, \sigma_i)$ is a Gaussian PDF of mean m_ι and standard deviation σ_ι.

It is worthwhile noting that, after multiple statistical fits of the empirical data to GMMs, we decided to abandon the GMM in favor of a simple Gaussian fit since the latter performs comparably with the former with respect to classification while being much less computationally intensive.

4. Numerical Results

In this section, we present numerical results on the performance of the proposed classification platform for several test configurations. We first examine the fundamental first-level classification and assess the performance of the proposed sound features in order to focus on the most high-performing and robust among them. Then, we present results on second-level classification and compare one-step versus two-step implementations.

For the numerical results presented herein, we have used feedforward artificial neural networks with one intermediate hidden layer. The neural network training and performance metric functions are a scaled conjugate backpropagation variant and a mean-square error function, respectively, while the output threshold function is a sigmoid function. The intermediate hidden neuron layer has a varying size according to the number of features used for classification. In the case where scalar features only are used, the number of features is 12 and the number of hidden nodes is 18, while in the case of MFCCs and spectrogram coefficients, the respective figures take a value of 26 features and 42 nodes in the one hand and 64 features and 85 nodes in the other hand, respectively. Furthermore, combinations of features were also used. For the scalar features and MFCC combination, there were 38 features and 40 nodes; for the scalar features and spectrogram coefficients, there were 76 features and 75 nodes; for the MFCCs and spectrogram coefficient combination, there were 90 features and 65 nodes; and, finally, for the scalar features plus MFCCs plus spectrogram coefficient combination, there were 102 features and 140 nodes. It is worthwhile

Table 1: Summary of classification performance of neural network with selected features: one-level classification.

	Scalar features	MFCCs	Spectrogram coefficients	Scalar feature + MFCCs	Scalar features + spectrogram coefficients	MFCCs + spectrogram coefficients	Scalar features + MFCCs + spectrogram coefficients
Average correct classifications	98.00%	91.06%	88.27%	97.42%	97.57%	91.82%	96.85%
Standard deviation of correct classifications	8.79%	14.30%	16.20%	9.90%	8.91%	16.61%	9.76%
Number of features/hidden layer nodes	12/18	26/42	64/85	38/40	76/75	90/65	102/140

Table 2: Confusion matrix of classification results using scalar features only.

	Geophysical samples to:	Animal samples to:	Anthropogenic samples to:	Correctly classified samples
Geophysical class	4446	123	92	4446
Animal class	0	5062	0	5062
Anthropogenic class	54	15	4408	4408
Total samples	4500	5200	4500	13,916/14,200 = 98.00%

noting that the number of hidden layer nodes was optimized after an exhaustive series of trial runs for varying numbers of hidden nodes.

4.1. Neural Network Performance with Selected Features Input: First-Level Classification. Numerical results obtained using three different classes, namely, anthropogenic, geophysical, and animal sounds, are presented in this subsection. Each of these classes was populated with sounds of three different subclasses. More specifically, we used airplane, car, and pistol sounds for the anthropogenic class; gale, waterfall, and fire sounds for the geophysical class; and dog, snake, and crow sounds for the animal sounds. For each subclass, we used a number of 15 different sample recordings of the specific sound type that were extracted from the "505 Digital Sound Effects" audio database [40].

From the list of available features, there exists a set of features (zero-crossing rate, pitch, and entropies) that are scalar and computationally light. Therefore, we group these features together under the label of "scalar features." On the other hand, as long as the MFCCs are concerned, we selected to use a set of 13 time-varying MFCCs for each sound sample and a simple GMM model for each one of them; this results to a set of 26 scalar feature inputs to be fed to the ANN (13 MFCCs; one mean and one standard deviation parameter for each one of them, thereupon a total of 26 scalar parameters). Similarly, we used 13 spectrogram coefficients with simple GMM modeling resulting to another 26 scalar feature inputs. Thereupon, it makes sense to partition the proposed features to three subsets (ZCR-pitch-entropies/MFCCs with GMM modeling/spectrogram coefficients with GMM modeling) and compare their performance. The results for the average and standard deviation of correct classifications are tabulated in Table 1. In the case where only the scalar features are used, the achieved accuracy is 98%; this figure is satisfactorily high and directly comparable (higher) to the respective figures of using scalar features in combination with either MFCCs, spectrograms, or both (see Table 1). Also, the respective standard deviation of accuracy is 8.69%, which is the smallest in Table 1.

Furthermore, confusion matrices are an information-rich and concise way to demonstrate the performance of a classification technique. Confusion matrices demonstrate the performance of a classification platform by illustrating the correct and incorrect classification results of all input samples while also illustrating the nature of the latter by depicting the type of incorrect classifications. Table 2 depicts a confusion matrix that demonstrates the performance of the proposed platform using only scalar features; it demonstrates that the overall classification accuracy is exceptional with a small number of classification errors that originate mostly from mistakenly classifying anthropogenic sound samples to the geophysical class.

4.2. Second-Level Classification of Sounds into Specific Subclasses: One-Step versus Two-Step Implementation. In this subsection, we demonstrate the performance of the proposed platform in the more demanding problem of second-level classification, that is, classifying sounds into more specific subclasses. Table 3 lists the performance of *one-step second-level* classification accuracy achieved by the proposed platform in various combinations of the aforementioned features. One-step second-level classification means that the classification of a sound sample into a specific subclass is performed directly by the neural network, that is, the network has 9 outputs and is directly fed with the signal features. The performance metrics include the average correct classifications and the respective standard deviation.

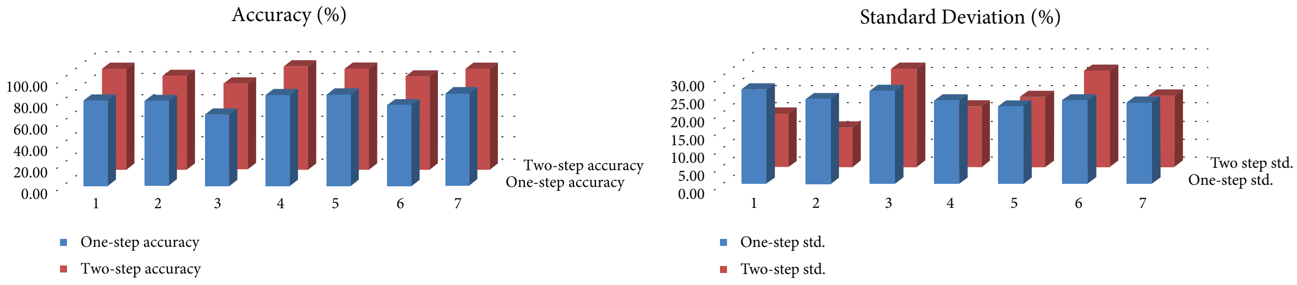
On the other hand, *two-step second-level classification* means that the classification into specific subclasses is performed in two steps. First, a sound is classified to a generic class using a 3-output neural network, as indicated in Figure 2. Then, according to the result of this first-level classification, the sound is fed to one of three subsequent neural networks each of which is optimized for classifying sounds of either anthropogenic, animal, or natural origin. The accuracy

Table 3: Summary of classification performance of neural network with selected features: second-level classification, one-step Implementation.

	Scalar features	MFCCs	Spectrogram coefficients	Scalar features + MFCCs	Scalar features + spectrogram coefficients	MFCCs + spectrogram coefficients	Scalar features + MFCCs + spectrogram coefficients
Average correct classifications	80.07%	79.55%	66.99%	85.34%	85.67%	75.77%	85.98%
Standard deviation of correct classifications	26.45%	23.77%	26.00%	23.38%	21.69%	23.42%	22.74%
Number of features/ hidden layer nodes	12/24	26/40	64/95	38/60	76/110	90/100	102/140

Table 4: Summary of classification performance of neural network with selected features: second-level classification, two-step implementation.

	Scalar features	MFCCs	Spectrogram coefficients	Scalar features + MFCCs	Scalar features + spectrogram coefficients	MFCCs + spectrogram coefficients	Scalar features + MFCCs + spectrogram coefficients
Average correct classifications	94.33%	87.80%	80.69%	96.73%	94.27%	86.99%	94.27%
Standard deviation of correct classifications	14.98%	11.12%	27.56%	16.96%	19.64%	26.92%	19.94%
Number of features/hidden layer nodes	12/18/18	26/42/40	64/85/100	38/40/70	76/75/140	90/65/135	102/140/150



(a)

(b)

Figure 3: Comparison of second-level classification hierarchy results using one-step and two-step implementations.

of each secondary network corresponds to the percent of correct second-level classification *given that first-level classification is correct*. These results are tabulated in Table 4. It is interesting to point out that both Tables 3 and 4 confirm that a balanced choice of features for reasonably good classification accuracy and reduced complexity is either to use the scalar features only or to use the scalar features in combination with MFCCs. Another interesting result is that second-level classification in two steps exhibits much higher accuracy of classification compared to one-step implementation.

Figure 3 demonstrates the comparative results of classification accuracy to the second level of hierarchy using either one-step or two-step implementation. It is interesting to point out that the accuracy achieved in the latter case is much higher compared to the former, for all types of features that were used (scalar, MFCCs, spectrogram coefficients, and combinations among them). Furthermore, the standard deviation of results in the two-step implementation case is much lower compared to the one-step implementation case; this implies that two-step implementation is more robust in terms of achieved accuracy compared to one-step implementation.

4.3. Confusion Matrices of First- and Second-Level Classification with Selected Features. Since scalar features are the most computationally effective yet yield satisfactorily accurate results, we consider the case of using only scalar features in order to combine both low computational cost and good enough accuracy. The computational effectiveness of using scalar features only is pointed out by the facts that (i) scalar features are easiest to calculate among all proposed features and (ii) the resulting neural network is fed with the minimum number of features and therefore exhibits the minimum number of hidden layer nodes. We also considered the case of scalar features combined with MFCCs (not presented herein) since the latter are the feature of choice most widely

Table 5: Second-level classification, one-step implementation, and scalar features only-confusion matrix.

	Airplane samples classified to:	Car samples classified to:	Pistol samples classified to:	Crow samples classified to:	Dog samples classified to:	Snake samples classified to:	Fire samples classified to:	Gale samples classified to:	Waterfall samples classified to:	Correctly classified samples
Airplane	1444	330	495	373	414	363	359	513	294	1444
Car	11	1125	0	0	0	0	1	2	15	1125
Pistol	0	0	990	0	0	0	0	13	0	990
Crow	0	0	0	1309	0	0	0	1	0	1309
Dog	5	15	0	3	1586	13	0	2	0	1586
Snake rattle	0	0	0	0	0	1123	0	0	0	1123
Fire	3	15	0	0	0	0	1125	0	0	1125
Gale	24	0	0	1	0	0	0	965	0	965
Waterfall	13	15	15	14	0	1	15	4	1191	1191
Total samples	1500	1500	1500	1700	2000	1500	1500	1500	1500	10,858/ 14,200 = 76.46%

Table 6: Second-level classification, two-step implementation, and scalar features only-confusion matrix results for second step classification of anthropogenic sounds to subclasses.

	Airplane samples classified to:	Car samples classified to:	Pistol samples classified to:	Correctly classified samples
Airplane	1485	30	30	1485
Car	4	1470	0	1470
Pistol	11	0	1470	1470
Total samples	1500	1500	1500	4425/4500 = 98.33%
Probability of correct classification of anthropogenic sounds during the first step (Table 2, row 3, column 3)				4408/4500 = 97.96%
Total accuracy of two-step classification for anthropogenic Sounds				96.32%

Table 7: Second-level classification, two-step implementation, and scalar features only-confusion matrix results for second-step classification of animal sounds to subclasses.

	Crow samples classified to:	Dog samples classified to:	Snake rattle samples classified to:	Correctly classified samples
Crow	1683	1	16	1683
Dog	17	1999	45	1999
Snake rattle	0	0	1439	1439
Total samples	1700	2000	1500	5121/5200 = 98.48%
Probability of correct classification of animal sounds during the first step (Table 2, row 2, column 2)				5062/5200 = 97.96%
Total accuracy of two-step classification for animal sounds				96.47%

used in the literature; however, numerical results demonstrated that using only scalar features is computationally much lighter and yields results that are comparable to those obtained by MFCCs or a combination of the two.

Confusion matrices are included in Tables 5–8, demonstrating the performance of the proposed platform using only scalar features. The network runs at these tables are different to the ones corresponding to the tables presented in Subsections 3.1 and 3.2. Table 5 illustrates the confusion matrix in the case of second-level classification in one step, that is, with one NN only with 9 outputs. The table clearly demonstrates the error sources in the various subclasses. It is evident that the main source of errors in one-step classification is due to sounds mistakenly classified as airplanes.

Tables 6–8 illustrate the results of second-level classification in two-step implementation. For example, Table 6 displays the confusion matrix in the case of anthropogenic sounds. Assuming that an anthropogenic sound has already been correctly classified as such in the first step, Table 6 demonstrates the results of the second step only. This means that in order to calculate the overall accuracy of second-level classification with two-step implementation, we need to

Table 8: Second-level classification, two-step implementation, and scalar features only-confusion matrix results for second-step classification of geophysical sounds to subclasses.

	Fire samples classified to:	Gale samples classified to:	Waterfall samples classified to:	Correctly classified samples
Fire	1485	15	15	1485
Gale	0	1470	15	1470
Waterfall	15	15	1470	1470
Total samples	1500	1500	1500	4425/4500 = 98.33%
Probability of correct classification of geophysical during the first step (Table 2, row 1, column 1)				4446/4500 = 98.80%
Total accuracy of two-step classification for geophysical sounds				97.15%

multiply any given probability with the probability that an anthropogenic sound is correctly classified during the first step. The latter probability can be found by dividing the number of correctly classified anthropogenic sound samples to the total number of anthropogenic sound samples. These numbers are equal to 4408 and 4500 respective (row 3, column 3 of Table 2). Similar calculations are also included in Tables 7 and 8.

Table 6 demonstrates that the main source of errors in the case of anthropogenic sounds and two-step implementation is the classification of different sounds to the airplane class. This result agrees with the results tabulated in Table 5. Furthermore, Table 7 demonstrates that a similar trend is present in the case of animal sounds but for dog sounds. However, Table 8 illustrates that both fire and waterfall sounds are more prone to classification errors compared to gale sounds. Finally, the comparison of Tables 6–8 to Table 5 verifies that two-step implementation is much more accurate compared to one-step second-level classification.

5. Conclusions

STORM is an ongoing project aiming at developing a platform for safeguarding cultural heritage sites across Europe. Part of the project objectives is to deploy wireless acoustic sensor networks over different historical and archaeological sites across Europe (the Diocletian Baths in Rome, Italy; the Mellor Heritage site in Manchester, UK; and the Roman Ruins of Tróia in Portugal) that will be used to monitor the sites and alarm stakeholders in the case of potential hazardous events. In this context, the proposed sound classification platform is a first step towards the accomplishment of this goal. The literature review revealed a number of popular approaches for sound feature selection together with denoising techniques and classification methods. In this paper, we presented the development of an integrated classification platform and evaluated its performance while the proposed classifier is extended to include the capability of classifying sounds within a hierarchy of two levels. First-level classification, or classifying sounds into generic classes like anthropogenic, animal, and geophysical, is sometimes critical; the proposed platform has been shown to deliver highly accurate results in this case. Also, it has been shown that the proposed scalar features are simple and computationally light, yet very accurate. As long as second-level classification is concerned, we showed that two-step classification may be more efficient compared to one-step implementation; in the presented numerical results, the achieved accuracy of the former was much higher compared to the latter. Furthermore, a confusion matrix analysis revealed that the main sources of errors are due to anthropogenic sounds mistakenly classified as geophysical sounds (first-level classification) or due to anthropogenic sounds mistakenly classified as airplanes (second-level classification). There is also a significant source of errors in second-level classification with other animal sounds mistakenly classified as dog sounds. In the future, we plan to apply the proposed classification approach in sound samples with varying signal-to-ratio values, as well as study the effect of noise on each sound feature separately and integrate our findings in the STORM platform at sensor and server level.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work is partially funded by EU and national funds under the research project "STORM (Safeguarding Cultural Heritage through Technical and Organizational Resources Management)" Grant agreement no. 700191, cofunded by the Horizon 2020 Programme of the European Union.

References

[1] EC, "Towards an integrated approach to cultural heritage for Europe," European Commission, 2015, http://www.europarl.europa.eu/sides/getDoc.do?type=REPORT&reference=A8-2015-0207&language=EN.

[2] InHERIT, "Promoting cultural heritage as a generator of sustainable development, European Research Project, co-funded by the ERASMUS+ action of the European Union," 2015, project website: http://www.inherit.tuc.gr/en/home/.

[3] G. Mergos and N. Patsavos, "Cultural heritage as economic value: economic benefits, social opportunities and challenges of cultural heritage for sustainable development," Report of the InHERIT European Research Project, 2017, http://www.inherit.tuc.gr/fileadmin/users_data/inherit/_uploads/%CE%9F2_Book_of_Best_Practices-f.pdf.

[4] STORM, *Safeguarding Cultural Heritage through Technical and Organizational Resources Management, co-funded by the*

Horizon 2020 Programme of the European Union, 2016, project website: http://www.storm-project.eu/, Grant Agreement No. 700191.

[5] Carte del Rischio, "Territorial information system for the protection of cultural heritage," ISCR, Italy, 1992, project websites: http://www.cartadelrischio.it/eng/index.html; http://www.icr.beniculturali.it/pagina.cfm?usz=1&uid=16.

[6] HAR, "HAR – Heritage at Risk Programme, Historic England," 1998, project website: https://historicengland.org.uk/advice/heritage-at-risk/types/.

[7] EMERIC, "EMERIC – Expert System for the Monitoring and Management of the Natural Environment of Crete," 2006, project website: http://emeric.ims.forth.gr/.

[8] F. M. Mazzolani, "The PROHITECH research project," appears in Structural Analysis of Historic Construction, Taylor and Francis Ed., 2008, http://www.hms.civil.uminho.pt/sahc/2008/CH123.pdf.

[9] S. A. Mitilineos, S. M. Potirakis, N.-A. Tatlas, and M. Rangoussi, "High-level sound classification in the ESOUNDMAPS project," in *Proceedings of the 4th International Conference on Materials and Applications for Sensors and Transducers (IC-MAST 2014)*, pp. 1–4, Bilbao, Spain, June 2014.

[10] S. A. Mitilineos, S. M. Potirakis, N.-A. Tatlas, and M. Rangoussi, "High-level sound classification in the ESOUNDMAPS project," *Key Engineering Materials*, vol. 644, pp. 83–86, 2015.

[11] M. Rangoussi, S. M. Potirakis, I. Paraskevas, and N.-A. Tatlas, "On the development and use of sound maps for environmental monitoring," in *128th Audio Engineering Society Convention*, London, UK, May 2010.

[12] R. Cai, Lie Lu, A. Hanjalic, Hong-Jiang Zhang, and Lian-Hong Cai, "A flexible framework for key audio effects detection and auditory context inference," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, no. 3, pp. 1026–1039, 2006.

[13] S. Chu, S. Narayanan, and C.-C. J. Kuo, "Environmental sound recognition with time-frequency audio features," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 17, no. 6, pp. 1142–1158, 2009.

[14] S. Despotopoulos, E. Kyriakis-Bitzaros, I. Liaperdos et al., "Pattern recognition for the development of Sound Maps for environmentally sensitive areas," in *International Scientific Conference eRA-7*, Piraeus, Greece, September 2012.

[15] A. J. Eronen, V. T. Peltonen, J. T. Tuomi et al., "Audio-based context recognition," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, no. 1, pp. 321–329, 2006.

[16] I. Paraskevas, S. M. Potirakis, and M. Rangoussi, "Natural soundscapes and identification of environmental sounds: a pattern recognition approach," in *2009 16th International Conference on Digital Signal Processing*, pp. 473–478, Santorini-Hellas, Greece, July 2009.

[17] N.-A. Tatlas, S. M. Potirakis, S. A. Mitilineos, and M. Rangoussi, "On the effect of compression on the complexity characteristics of wireless acoustic sensor network signals," *Signal Processing*, vol. 107, pp. 153–163, 2015.

[18] B. Ghoraani and S. Krishnan, "Time-frequency matrix feature extraction and classification of environmental audio signals," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 19, no. 7, pp. 2197–2209, 2011.

[19] J. M. Kates, "Classification of background noises for hearing-aid applications," *Journal of the Acoustical Society of America*, vol. 97, no. 1, pp. 461–470, 1995.

[20] N. Mesgarani, M. Slaney, and S. A. Shamma, "Discrimination of speech from nonspeech based on multiscale spectro-temporal modulations," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, no. 3, pp. 920–930, 2006.

[21] E. Scheirer and M. Slaney, "Construction and evaluation of a robust multifeature speech/music discriminator," in *1997 IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 1331–1334, Munich, Germany, April 1997.

[22] S. M. Potirakis, B. Nefzi, N.-A. Tatlas, G. Tuna, and M. Rangoussi, "A wireless network of acoustic sensors for environmental monitoring," *Key Engineering Materials*, vol. 605, pp. 43–46, 2014.

[23] I. Paraskevas, S. M. Potirakis, I. Liaperdos, and M. Rangoussi, "Development of automatically updated soundmaps for the preservation of natural environment," *Journal of Environmental Protection*, vol. 02, no. 10, pp. 1388–1391, 2011.

[24] J. J. Aucouturier, B. Defreville, and F. Pachet, "The bag-of-frames approach to audio pattern recognition: a sufficient model for urban soundscapes but not for polyphonic music," *The Journal of the Acoustical Society of America*, vol. 122, no. 2, pp. 881–891, 2007.

[25] S. Ntalampiras, I. Potamitis, and N. Fakotakis, "Acoustic detection of human activities in natural environments," *Journal of Audio Engineering Society*, vol. 60, no. 9, pp. 686–695, 2012.

[26] A. D. Mazaris, A. S. Kallimanis, G. Chatzigianidis, K. Papadimitriou, and J. D. Pantis, "Spatiotemporal analysis of an acoustic environment: interactions between landscape features and sounds," *Landscape Ecology*, vol. 24, no. 6, pp. 817–831, 2009.

[27] Y. G. Matsinos, A. D. Mazaris, K. D. Papadimitriou et al., "Spatio-temporal variability in human and natural sounds in a rural landscape," *Landscape Ecology*, vol. 23, pp. 945–959, 2008.

[28] K. D. Papadimitriou, A. D. Mazaris, A. S. Kallimanis, and J. D. Pantis, "Cartographic representation of the sonic environment," *The Cartographic Journal*, vol. 46, no. 2, pp. 126–135, 2009.

[29] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423, 1948, 623–656.

[30] O. A. Rosso, S. Blanco, J. Yordanova et al., "Wavelet entropy: a new tool for analysis of short duration brain electrical signals," *Journal of Neuroscience Methods*, vol. 105, no. 1, pp. 65–75, 2001.

[31] S. M. Potirakis, G. Minadakis, and K. Eftaxias, "Analysis of electromagnetic pre-seismic emissions using Fisher information and Tsallis entropy," *Physica A: Statistical Mechanics and its Applications*, vol. 391, no. 1-2, pp. 300–306, 2012.

[32] C. Bandt and B. Pompe, "Permutation entropy: a natural complexity measure for time series," *Physical Review Letters*, vol. 88, no. 17, 2002.

[33] A. A. Bruzzo, B. Gesierich, M. Santi, C. A. Tassinari, N. Birbaumer, and G. Rubboli, "Permutation entropy to detect vigilance changes and preictal states from scalp EEG in epileptic patients - a preliminary study," *Neurological Sciences*, vol. 29, no. 1, pp. 3–9, 2008.

[34] X. Li, G. Ouyang, and D. A. Richards, "Predictability analysis of absence seizures with permutation entropy," *Epilepsy Research*, vol. 77, no. 1, pp. 70–74, 2007.

[35] S. Abe and Y. Okamoto, Eds., *Non-extensive Statistical Mechanics and its Applications*, Springer-Verlag, Heidelberg, Germany, 2001.

[36] C. Tsallis, “Non-additive entropy *Sq* and non-extensive statistical mechanics: applications in geophysics and elsewhere,” *Acta Geophysica*, vol. 60, no. 3, pp. 502–525, 2012.

[37] A. Rényi, “On measures of entropy and information,” in *Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability*, pp. 547–561, University of California Press, Berkeley, LA, USA, 1961.

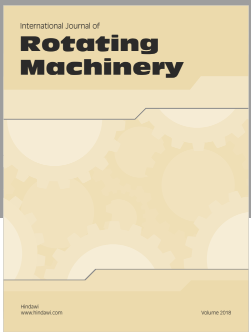
[38] S. G. Mallat, “A theory for multiresolution signal decomposition: the wavelet representation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, 1989.

[39] D. L. Donoho, “De-noising by soft-thresholding,” *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 613–627, 1995.

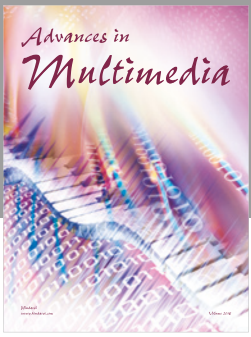
[40] *505 Digital Sound Effects*, CD Box Set, Laserlight Digital, 2006.

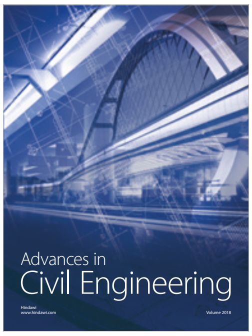
Journal of
Engineering

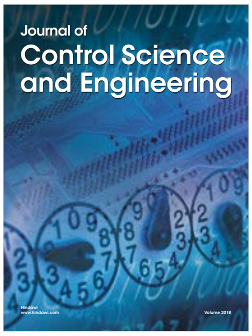
The Scientific
World Journal


International Journal of
Rotating
Machinery

Journal of
Sensors

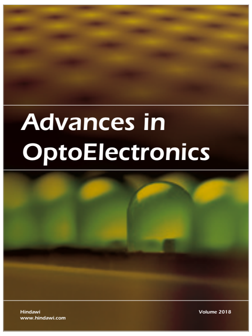

Advances in
Multimedia


Advances in
Civil Engineering


Journal of
Control Science
and Engineering

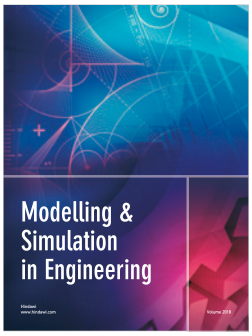
Journal of
Robotics

Journal of
Electrical and Computer
Engineering

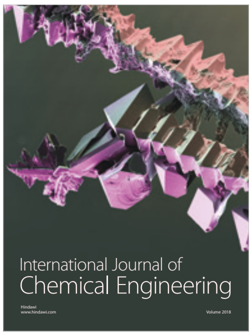

Advances in
OptoElectronics

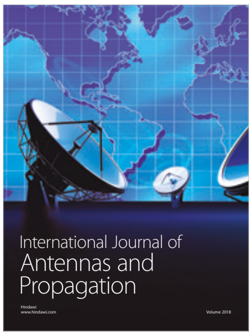
VLSI Design

International Journal of
Navigation and
Observation

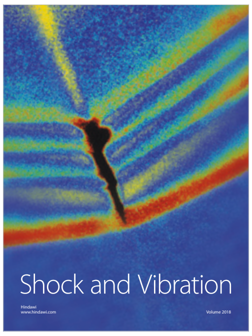

Modelling &
Simulation
in Engineering

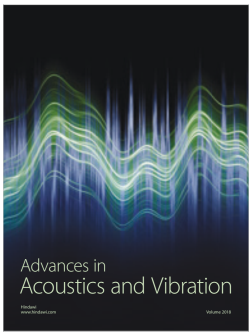
International Journal of
Aerospace
Engineering


International Journal of
Chemical Engineering


International Journal of
Antennas and
Propagation

Active and Passive
Electronic Components


Shock and Vibration


Advances in
Acoustics and Vibration