Acoustic Noise Detection and Classification Based on Support Vector Machines

Sanja Vujnović, Goran Kvaščev, Branko Kovačević and Lazar Cokić

Abstract—Acoustic signals are shown to be very informative for a wide variety of applications; however, they are highly susceptible to the surrounding noise. This paper proposes a new algorithm which uses cepstral coefficient extracted from acoustic recordings in order to detect whether the useful acoustic signal is contaminated with noise. Apart from noise detection, this algorithm also performs noise classification if the contamination is indeed present. The algorithm used for this purpose is a clustering scheme based on support vector machines and the performance of the algorithm is tested on nominal recordings taken from fan mills in thermal power plants in Serbia. The contamination has been performed using four different types of noise, and the algorithm has been tested for different values of signal-to-noise ratio.

Index Terms—Acoustic signals, support vector machines, noise detection and isolation, cepstral coefficients, pattern recognition.

I. Introduction

IN recent years there has been a rapid growth of intelligent systems, from predictive maintenance applications in the industry to smart buildings and internet-of-things technologies. This trend to collect a wide variety of data for the purpose of monitoring, estimation, control, or simply statistical analysis requires a careful choice of sensors and types of recorded signals.

Acoustic signals present a crucial element in modern intelligent systems due to their versatility and the tendency to easily absorb a wide variety of different information. One example of this is the problem of scene analysis which is usually tackled using acoustic event classification algorithms [1]. Another use for acoustic signals can be found in industrial predictive maintenance area of research. It has been shown that acoustic signals can be as informative as the dominantly used vibration signals when it comes to fault detection and state estimation of rotating machines [2]. In some cases acoustic signals can even detect a fault much sooner, but due to their high susceptibility to the surrounding noise they are rarely used in the industry for this purpose.

As informative as acoustic recordings may be, the fact that they are easily affected by disturbances can greatly influence the quality of useful information than can be obtained from them. For that reason it is very important to be able to detect whether the noise is dominant in the acoustic signal and, if so, what type of noise is in question. This would not only extend the applicability of acoustic signals to real industrial environment, but it would also improve the robustness of the existing algorithms based on sound recordings.

In this paper an algorithm is proposed for noise detection in acoustic signals, as well as classification of noise if it is detected. Since the authors see the main application of this research in the area of predictive maintenance in the industry, the nominal informative sound used for the experiments is the sound of fan mills in thermal power plant Kostolac A1 in Serbia. One of four different types of contamination are added to this signal so to artificially pollute it for the purpose of testing the algorithm. These contaminations are: the sound of metallic bell ringing, the sound of coins being mixed together, the sound of paper being torn and the sound of crumbling of the paper.

Due to the fact that support vector machine (SVM) has recently been successfully implemented in a number of audio classification tasks [3] it has been used in this paper for both noise detection and noise classification purposes. In its nature SVM classifies data not by estimating conditional densities, but by creating boundaries between classes, so it requires less data to perform accurate classification. Since SVM is a binary classifier, a strategy has been proposed to extend it to a multiclass problem by modifying a standard tree structure [4] so both noise detection and four class classification can be conducted.

In this research the standard acoustic detection features are chosen for pattern recognition: 14 mel-frequency cepstral coefficients (MFCCs) plus the energy of the frame [5]. Since the contaminations are artificially added to the nominal sound, the signal-to-noise ratio (SNR) can be controlled, and the algorithm has been tested for different intensity of contamination. The performance has been analyzed using confusion matrices for both noise detection and noise classification problems.

This paper is structured as follows. In Section 2 a detailed description of used signals and features extracted from them is given. Basic theory of SVM is presented in Section 3, while Section 4 presents the results of the experiments as well as the discussion. The conclusion is given in Section 5.

Sanja Vujnović is with the School of Electrical Engineering, University of Belgrade, 73 Bulevar kralja Aleksandra, 11020 Belgrade, Serbia (e-mail: svujnovic@etf.bg.ac.rs).

Goran Kvaščev is with the School of Electrical Engineering, University of Belgrade, 73 Bulevar kralja Aleksandra, 11020 Belgrade, Serbia (e-mail: kvascev@etf.bg.ac.rs).

Branko Kovačević is with the School of Electrical Engineering, University of Belgrade, 73 Bulevar kralja Aleksandra, 11020 Belgrade, Serbia (e-mail: kovacevic_b@etf.bg.ac.rs).

Lazar Cokić is with the School of Electrical Engineering, University of Belgrade, 73 Bulevar kralja Aleksandra, 11020 Belgrade, Serbia (e-mail: cokic@etf.bg.ac.rs).

II. ACOUSTIC SIGNALS AND FEATURE EXTRACTION

The main purpose of this research is to make acoustic signals more useful in an industrial environment for state estimation and noise detection. For that reason the nominal, informative signal which is recorded is the sound of the fan mill in thermal power plant Kostolac A1 in Serbia obtained while the coal grinding subsystem is in operation. The measurement is carefully taken in nominal surroundings when there is no nonstationary noise in the subsystem. This signal is artificially polluted with one of four different types of acoustic signals noise: the sound of bells ringing, the sound of coins, the sound of tearing and the sound of crumpling of paper. These sounds have been obtained from the RWCP database [6]. The goal, therefore, is to detect the whether the measurement from the mill is noisy or whether it has been polluted, and if the noise is detected to classify which of four types of noise is in question. The experiment has been repeated for different SNR values. An example of the frequency domain characteristic of nominal and noisy signal when SNR is 10dB is given in Fig 1. The PSD estimate indicates the different properties of the signal on different frequencies, but this distinction can be more or less pronounced depending on SNR (the larger the ratio, the less pronounced the noise is).

In order to reduce the computational complexity, all the excess information has been reduced in the preprocessing part of the algorithm. All the signals have been down sampled to 8kHz and divided into frames the length of which is 128 samples, with an overlap of 50%. Nominal mill recordings have been separated into 5 parts. One part has been used in its original form, and the other four parts have been polluted, each with a different type of noise. The features considered in this research are 14 mel-cepstral coefficients computed using 20 mel-scaled spectral bands, with frame energy added to the set. Since the number of characteristic signals in the database is scarce, there are 111 frames contaminated with each type of noise and there are 444 frames which are not contaminated.
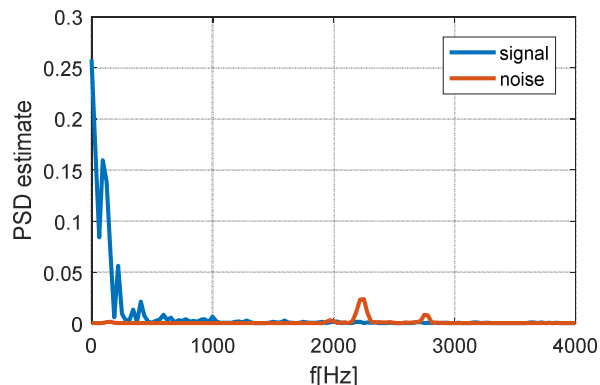

Fig. 1. Power spectral density of useful signal (recorded near the mill while it is in operation) and one type of noise (the sound of bell ringing) for SNR=10dB.

III. SUPPORT VECTOR MACHINE

The SVM is a pattern recognition algorithm which, given an n-dimensional training vector $x_i \in \mathcal{R}^n$ of two separable classes, constructs a hyperplane $\omega \cdot x + b = 0$ to separate the training data [7]. Here ω is the vector normal to the hyperplane H_0 (Fig. 2) and b determines the offset of H_0 from the origin. There are many possible solutions to this problem and the goal is to find the decision hyperplane H_0 which maximizes the distance between the adopted hyperplane and the nearest points of each class. The choice of H_0 therefore depends only on the vectors closest to the two hyperplanes H_1 and H_{-1} which are parallel to the decision hyperplane. These are called the support vectors.
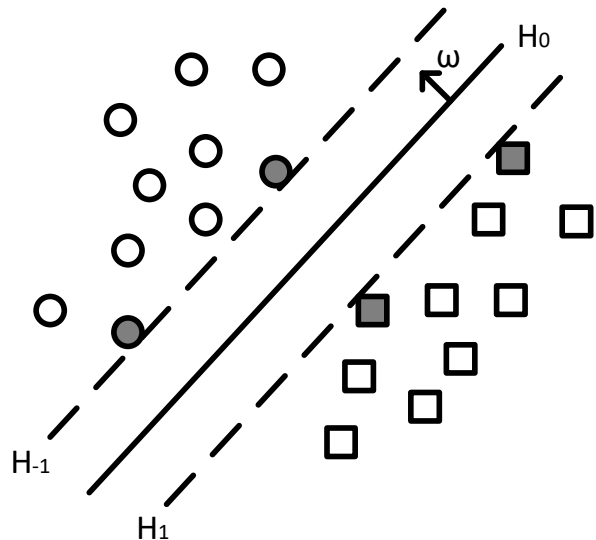

Fig. 2. Two class linear classification using hyper planes with optimal separating hyperplane and support vectors. H_0 presents the separating hyperplane, H_1 and H_{-1} are parallel to it and gray points of each class are called the support vectors.

If there is a set of training data $X = \{x_1, \dots, x_m\}$, where $x_i \in \mathcal{R}^n$, and a set of corresponding classes $Y = \{y_1, \dots, y_m\}$, where $y_i \in \{1, -1\}$, finding the SVM separating hyperplane can be stated by the following optimization problem:

$$\text{minimize} \frac{1}{2}\|\omega\|^2,$$
$$\text{subject to } y_i(\omega x_i + b) \geq 1. \tag{1}$$

This optimization problem is solved by determining the saddle point of the Lagrange functional using numerical calculations:

$$L(\omega, b, \alpha) = \frac{1}{2}\|\omega\|^2 - \sum_{i=1}^{m} \alpha_i [y_i f(x_i) - 1]. \tag{2}$$

When the classes are not separable, slack variables can be introduced $\xi_i \geq 0$ which present a measure of

misclassification error for each training vector. This slightly modifies optimization problem from Eq. (1):

$$\text{minimize} \left(\frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^{m} \xi_i \right), \tag{3}$$
$$\text{subject to } y_i(\omega' x_i + b) \geq 1 - \xi_i,$$

where C is a positive constant which controls the extent in which slack variables are penalized.

Apart from linear classification, SVM can conduct nonlinear discrimination by mapping data onto higher-dimensional space, so that the classes become separable. This is done by kernel mapping. The two kernel functions which are most often used are radial basis function (RBF):

$$K(x_i, x_j) = \exp\left(-\frac{|x_i - x_j|^2}{2\sigma^2} \right) \tag{4}$$

and polynomial function:

$$K(x_i, x_j) = (x_i \cdot x_j)^d. \tag{5}$$

It has been mentioned previously that the SVM is a binary classifier. However, there are many cases in which this technique has been used for multiple-class problems. There are several methods used in the literature, the most common of which are *one against all* and *one against one* strategies. In this paper a mixture of the two is used since the problem which is tackled has to do with both detection and classification. The binary tree structure is shown in Fig. 3 and it classifies patterns into 5 categories. First SVM tests whether the noise has occurred or not. The second classifier is used if the contamination is detected and it tests whether the noise is metal (bell and coin) or paper (tear and crumple) in origin. After that, depending on classification results, the third or fourth SVM is used to determine which exact noise is in question. This way the number of binary classifiers is reduced and the algorithm examines a maximum of three support vector machines in each step.

IV. RESULTS

Since there are two different tasks the decision tree in Fig. 3 needs to conduct the testing of the algorithm has been done in two parts. First, the ability of the algorithm to detect the noise is examined and then its ability to classify it.

A. Noise detection

In Fig. 3 only SVM1 is in charge of noise detection. There are 888 frames in total and one third of them has been used to train this SVM, while the others are used to test it. Results for SNR=10dB are shown in Table I in a form of confusion matrix. The number of false alarms (samples which are not polluted with noise but were classified as contaminated) is quite high: 23%. However, since the main reason of implementing this sort of noise detection procedures is to enhance the robustness of the algorithms which use acoustic signals for information gathering, this high false alarm rate can be tolerated as long as the polluted frames get classified as contaminated. This is indeed the case since the missed detection rate (the number of contaminated signals classified as uncontaminated) is only 2%. These results show that the proposed algorithm is quite rigorous towards contaminated frames.
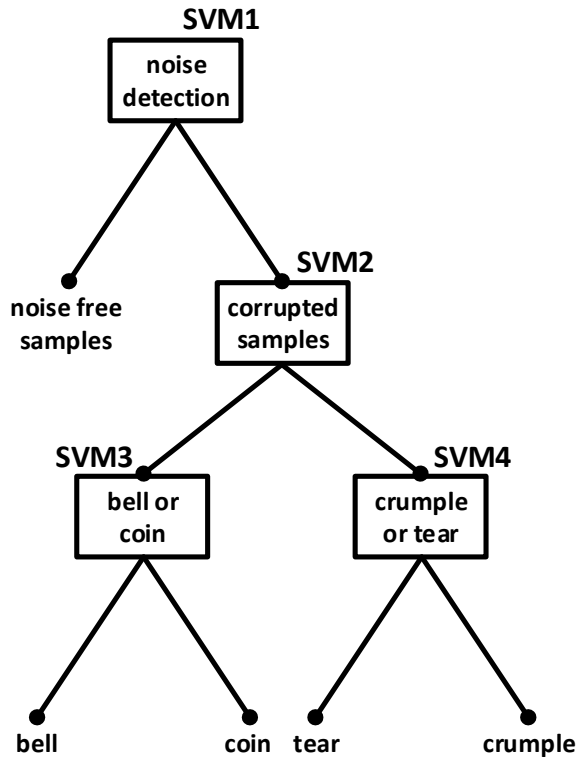

Fig. 3. Binary classification tree for fault noise detection and noise classification. There are 4 binary classifiers and 5 possible classes of objects.

TABLE I
CONFUSION MATRIX FOR CONTAMINATION DETECTION FOR SNR=10DB

	Classified as uncontaminated	Classified as contaminated
Uncontaminated samples	77%	23%
Contaminated samples	2%	98%

Figure 4 shows the change in false alarm and missed detection rate as the signal-to-noise ratio changes. For low values of SNR the noise is much more pronounced and it is easier to detect it. As long as SNR is below 5dB detection rate of the algorithm is 100%, albeit on account of high false alarm rate. When SNR rises above 20dB both false alarm and missed detection rates are quite high showing that the noise is not so pronounced and the algorithm will struggle to detect it.

This, as well, is not very problematic because the noise of low intensity will probably not significantly pollute the informative part of the acoustic signal and later signal processing applications will still be able to make the correct decisions.
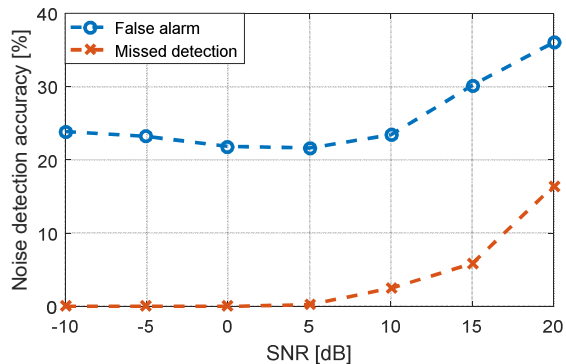

Fig. 4. False alarm rate and missed detection rate as a function of SNR.

B. Noise classification

The second part of the algorithm deals with noise classification. There are 111 frames contaminated with each type of noise. In Fig 3. SVM2 is in charge of separating metal based noise (bell and coin) from the paper based noise (tear and crumple). One half of all the contaminated frames is used for training this classifier. SVM2 and SVM3 are used to differentiate between the noise caused by bell or coin and between noise caused by tear and crumble of the paper, respectively. There are 222 frames used for training and testing of each of the classifiers (50% for training and 50% for testing). Table II shows the confusion matrix for noise classification with SNR=10dB. Bell and coin noises are classified with higher accuracy than paper-sourced noises (tear and crumble) and the reason for that might be that the paper-sourced noises have similar energy distribution, so cepstral coefficient might not be the best choice of feature vectors.

TABLE II
NOISE CLASSIFICATION CONFUSION MATRIX FOR SNR=10DB

%	Bell	Coin	Tear	Crumple
Bell	90	8	1	1
Coin	0	95	0	5
Tear	1	10	83	6
Crumple	0	21	5	74

Figure 5 shows noise classification accuracy as a function of signal-to-noise ratio. Again, bell and coin have a higher classification rate for almost all the values of SNR. Also, there is no distinct trend which indicates how the classification ratio changes as the SNR changes. What is evident is the fact that for very high values of SNR the accuracy of detection

decreases for all types of noise. This is an expected result because high value of SNR indicates that noise is very weak, and all the characteristic elements of different types of noise are not very distinct. It might be possible to improve upon these results by extracting different features from the signal, but only up to a point.
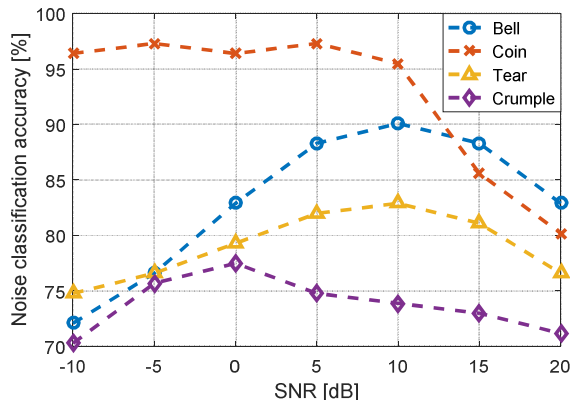

Fig. 5. Noise classification accuracy as a function of SNR.

V. CONCLUSION

In this paper a new algorithm for noise detection and classification has been presented. This algorithm uses mel-frequency cepstral coefficients as features and a binary tree of SVM classifiers for pattern recognition. The algorithm has been tested on real signals recorded in thermal power plant Kostolac A1 in Serbia which have been artificially polluted by four different types of noise using different signal-to-noise ratios.

The goal of this algorithm is to expand the use of the acoustic signals for information gathering in industrial surroundings and to enhance the robustness of existing algorithms by ensuring the use of noise free measurements in decision making process. While noise detection part of the algorithm can be useful for predictive maintenance applications, the part of the algorithm which provides classification of noise is important for statistical analysis, information gathering or event classification problems.

The results have shown that noise detection algorithm is quite rigorous in detecting contamination (which is characterized by a very low missed detection rate and higher false alarm rate). This is indeed very important for successful implementation in an industrial environment. Noise classification results vary, and for lower values of SNR metal based noises (ringing of bell and mixing of coins) are detected with very high accuracy. Paper based noises (tearing and crumpling of paper) are classified with less success for all SNR values. This can be due to the fact that the energy distribution of paper based noises is more similar than that of metal based ones.

Because of the limited number of specific signals in RWCP database, the training and testing sets are not very large and,

for further research, more experiments should be conducted to record the noise signals and expand the database. Also, classification results indicate that the choice of features may not be optimal for noise classification even though it has been shown that it is quite informative for noise detection. Further research on this subject will include more experiments with different feature sets, and different combinations of features for each classifier.

REFERENCES

[1] A. Temko, C. Nadeu, "Classification of acoustic events using SVM-based clustering schemes," *Pattern Recognition*, vol. *39,* pp. *682-694,* 2006.

[2] P. Henriquez, J.B. Alonso, M.A. Ferrer, C.M. Travieso, "Review of automatic fault diagnosis systems using audio and vibration signals," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. *44,* no. *5642-652*, 2014.

[3] G. Guo, Z. Li, "Content-based audio classification and retrieval using support vector machines," *IEEE Trans. Neural Networks*, vol. 14, pp. 209-215, 2003.

[4] C.W. Hsu, C.J. Lin, "A comparison of methods for multi-class support vector machines," *IEEE Trans. Neural Networks*, vol. 13*,* pp. 415–425, 2002.

[5] J.T. Foote, "Content-based retrieval of music and audio," *Proc. SPIE*, vol. 3229*,* pp. 138-147, 1997.

[6] S. Nakamura, K. Hiyane, F. Asano, T. Nishiura, T. Yamada, "Acoustical Sound Database in Real Environments for Sound Scene Understanding and Hands-Free Speech Recognition," in *LREC*, 2000.

[7] C. Cortes, V. Vapnik, "Support vector networks," *Mach. Learn.*, vol. 20*,* pp. 273-297, 1995.