Audio Anomaly Detection on Rotating Machinery Using Image Signal Processing

Thiago de M. Prego*†, Amaro A. de Lima*†, Sergio L. Netto† and Eduardo A. B. da Silva†
*PPEEL, Federal Center of Tech. Edu. Celso Suckow da Fonseca (CEFET-RJ),
Rio de Janeiro, Brazil
{thiago.prego, amaro.lima}@cefet-rj.br
†PEE/COPPE, Federal University of Rio de Janeiro,
Rio de Janeiro, Brazil
{sergioln, eduardo}@smt.ufrj.br

***Abstract*—This paper addresses the problem of anomaly detection on rotating machinery in industrial environments using single channel audio signals. The proposed algorithm is based on image processing feature analysis obtained from the image representation of the Short-time Fourier Transform of reference and degraded audio signals. In order to assess the potential of the algorithm, a 8 signals database is recorded. The proposed algorithm is able to separate signals of machinery normal behavior from signals of machinery anomalous behavior with 100% hit rate using the recorded database.**

***Keywords*—Image processing, audio anomaly detection, feature extraction.**

I. Introduction

Machinery condition based monitoring (CBM) [1], [2], fault and failure classification (FC) [3], [4], and remaining useful life(RUL) [6] are challenges that are gaining much interest by both academic and industrial communities. Table I [5] shows the expected improvements of using automatic or semi-automatic systems for condition monitoring, where one can observe a significant improvement on maintenance costs and total productivity, which are of great interest for the industry.

In order to develop techniques for CBM, FC and RUL, one must first determine whether an anomaly has occurred, then separate faults from failures and isolate the problems to be studied and/or modelled. One of the most sophisticated and reliable human sensors are the ears and it is known how machine operators have the capability of identifying a problematic machine just be hearing how it sounds.

The work described in this paper is part of the DORIS project, described in details in [7], [8].

This paper proposes a three stage algorithm for automatic audio anomaly detection using two simple features extracted from the image representation of Short-time Fourier Transform of audio signals. In order to show the potential of the proposed approach, an 8 audio signals database was recorded using a moving robot on a rail inside a closed environment.

The paper is organized as follows: In Section II the recorded database is described; Section III describes the three stages of the proposed algorithm; Section IV shows the experimental results of the proposed algorithm applied to the recorded database; and finally Section V is devoted to the conclusions.

TABLE I
Expected improvements of using systems for CBM. [5]

Evaluation metric	Improvement
Maintenance costs	Reduction of 50%–80%
Equipment damages	Reduction of 50%–60%
Extra hours expenses	Reduction of 20%–50%
Machine life expectancy	Increase of 50%–60%
Total productivity	Increase of 20%–30%

II. Database

In order to develop an algorithm that is able to discriminate audio signals with different time-frequency signatures, an audio database was recorded. This database was recorded using a robot moving on a fixed track in an enclosed room, as illustrated at Fig. 1. Attached to the robot is one microphone with 16 kHz sampling frequency and 24 bits/sample analog-to-digital converter. Every recorded audio is corrupted by ambient noise and moving robot sound.

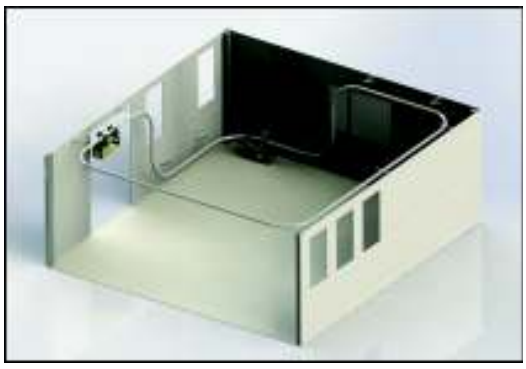


Fig. 1. Schematic design of the robot track inside an enclosed environment.

One loudspeaker was placed in a fixed position along the track to simulate 4 different rotating machinery configurations:

I) Loudspeaker plays no sound.
II) Loudspeaker plays the sound of machine 1.
III) Loudspeaker plays the sound of machine 2.

ISBN 978-1-4673-7835-2/16/$31.00©2016 IEEE

IEEE Catalog Number CFP16LAS-ART

IV) Loudspeaker plays the sound of both machines at the same time.

Two different runs along the track were recorded for each configuration, leading to 8 different audio signals. Every run used approximately the same start and end points and the same linear speed for the moving robot.

III. Proposed algorithm

The proposed algorithm is composed by three stages, as shown by Fig. 2.

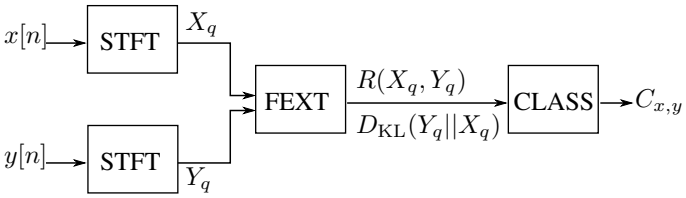


Fig. 2. Block diagram of the proposed algorithm.

A. STFT stage

At the first stage of the proposed algorithm, a Short-time Fourier Transform (STFT) is applied individually to the reference $x[n]$ and degraded $y[n]$ signals, leading to $X[k,m]$ and $Y[k,m]$. The STFT is calculated with time frame of size $L = 16000$ samples, with $L_{\mathrm{ov}} = 14400$ overlapping samples between consecutive frames. A Hanning window is used with DFT (discrete Fourier Transform) of size L.

As the anomalies to be detected are related to changes on the operating mode of the machinery, only frequency bins equivalent to up to 200 Hz are observed. Fig. 3 shows the image representation of $X[k,m]$ for configurations I and II and Fig. 4 shows the image representation of $X[k,m]$ for configurations III and IV. These 4 images are pixel matrices of dimensions $M \times N$ with pixel in the range $[0, 255]$, where 255 represents the maximum value of $X[k,m]$. The horizontal axis represents frequency bins k in ascending order from left to right and the vertical axis represents time frames m in ascending order from top to bottom.

It can be noticed that Fig. 3 (a) contains only the noise floor that is also present on Fig. 3 (b) and Fig. 4. In order to remove this noise floor, all values of $X[k,m]$ and $Y[k,m]$ that are below the threshold $T_q = 40$ dB are set to zero leading to the new image X_q. Fig. 5 shows the images X_q of configurations I and II and Fig. 5 shows the images X_q of configurations III and IV.

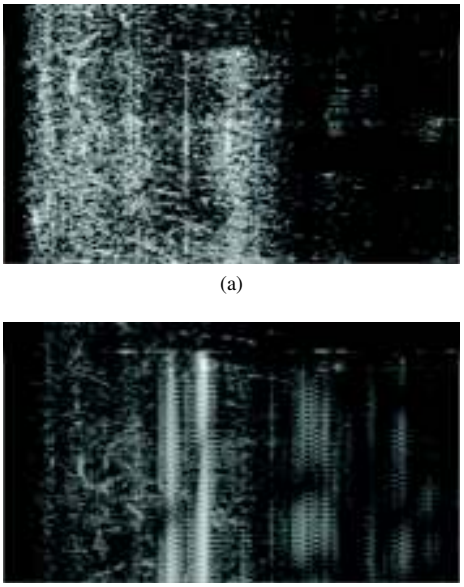
One can observe that the noise floor is removed from all images, emphasizing peaks on frequency bins around 90 Hz.

B. FEXT stage

At the second stage, two features are extracted from both reference and degraded images X_q and Y_q: maximum normalized 2D cross correlation with vertical lags $R(X_q, Y_q)$ and Kullback-Liebler Divergence $D_{\mathrm{KL}}(Y_q||X_q)$.

The normalized 2D cross correlation is a measure of similarity between two images and is largely used by image processing pattern recognition algorithms [9], [10], [11]. The maximum normalized 2D cross correlation with vertical lags is defined as:

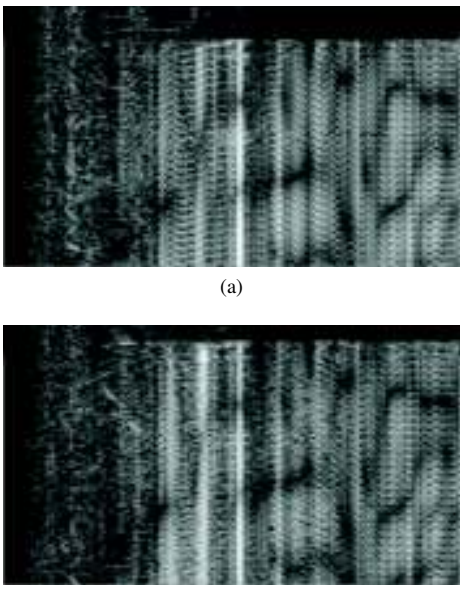
$$R(X_q, Y_q) = \frac{\max\left[\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} X_q(m,n)Y_q(m-j,n)\right]}{\sqrt{\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} X_q(m,n)^2}\sqrt{\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} Y_q(m,n)^2}}, \quad (1)$$



(a)

(b)

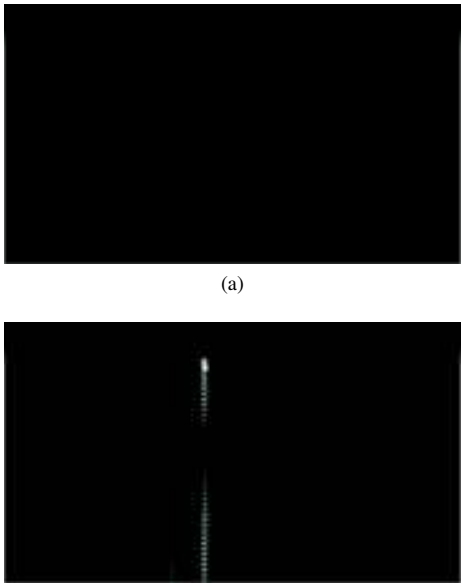
Fig. 3. Image representation of $X[k,m]$ for one signal of: (a) configuration I (b) configuration II.



(a)

(b)

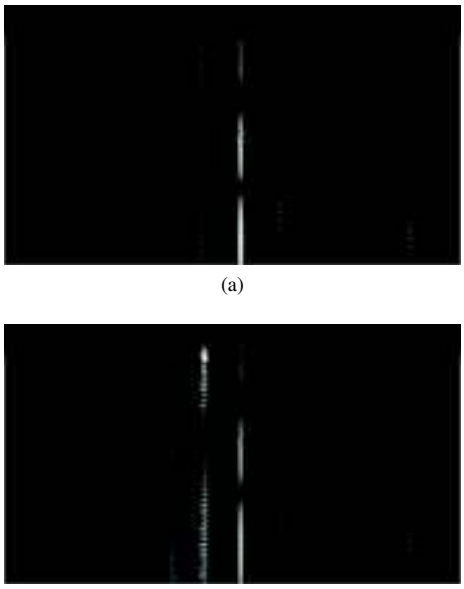
Fig. 4. Image representation of $X[k,m]$ for one signal of: (a) configuration III (b) configuration IV.



(a)

(b)

Fig. 5. Image representation of $X_q[k, m]$ for one signal of: (a) configuration I (b) configuration II.



(a)

(b)

Fig. 6. Image representation of $X_q[k, m]$ for one signal of: (a) configuration III (b) configuration IV.

where $0 \leq j < N$. As the maximal value of the cross correlation is obtained only for vertical lags (time dimension), this feature is robust to little time shifts between reference and to be tested runs.

The Kullback-Liebler Divergence is a measure of difference between statistical distributions and it is used by some image processing pattern recognition algorithms [12], [13]. It is defined as:

$$D_{\mathrm{KL}}(Y_q||X_q) = \sum_{i\in\mathcal{I}} p_y(i) \log\left(\frac{p_y(i)}{p_x(i)}\right), \qquad (2)$$

where $\mathcal{I}$ is defined such as $p_x(i) = 0$ leads to $p_y(i) = 0$. $p_x(i)$ and $p_y(i)$ are obtained by the histogram with 256 bins of each image. Little time shifts between reference and to be tested runs has little impact on this feature.

As both features are extracted from images representing the STFT of signals recorded from a microphone that is attached to a moving robot, it is important to ensure synchronism between reference and to be tested runs, which was guaranteed by the database recording setup.

C. CLASS stage

The third and final stage is responsible for the anomaly detection. Fig. 7 (a) and Fig. 7 (b) show the graphical representation of the features. The database described in Section II was divided in training and test sets.

The training set is composed by signals from configuration I, III and IV as reference and to be tested signals. The test set is composed of signals from configuration II as reference signals and configuration I, III and IV as to be tested signals.

Fig. 7 (a) represents signals of configuration I as reference signals and signals of configuration III and IV as degraded signals. Fig. 7 (b) represents two signal of configuration IV as reference signals and signals of configuration I and III as degraded signals. It can be noted that when configuration I is either X_q or Y_q, $R(X_q, Y_q) = 0$ and $D_{KL} = 0$ only when X_q and Y_q are both from configuration I. When configuration I is neither X_q or Y_q, $R(X_q, Y_q) \neq 0$ and this feature alone can separate normal signals from anomalous signals.

These observations lead to the following classification algorithm:

1) If $R(X_q, Y_q) = 0$ and $D_{\mathrm{KL}}(Y_q||X_q) = 0$, the degraded signal is considered from the class NORMAL.
2) If $0 < R(X_q, Y_q) < 0.75$, the degraded signal is considered from the class ANOMALOUS.
3) If $R(X_q, Y_q) \geq 0.75$, the degraded signal is considered from the class NORMAL.

The thresholds were obtained in order to minimize intersection between NORMAL and ANOMALOUS classes.

IV. Experimental results

Two experiments were conducted, one for training and one for testing the algorithm. In the first experiment, the training set was used to obtain the algorithm parameters of STFT, FEXT and CLASS stages. In the second experiment, the test set was used to evaluate the algorithm performance.

Table II and Fig. 8 show the performance of the algorithm using the test set. It can be observed that the algorithm has a 100% hit rate.

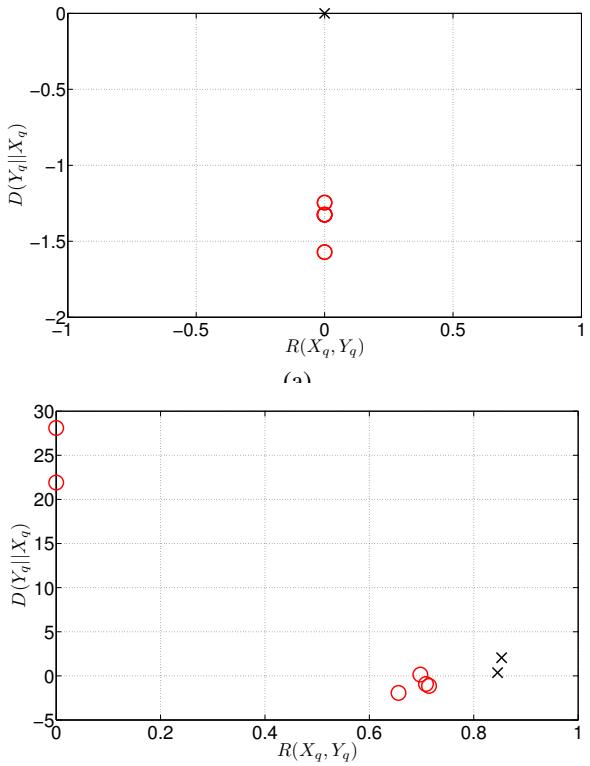


Fig. 7. Graphical representation of features $R(X_q, Y_q)$ and $D_{\mathrm{KL}}(Y_q||X_q)$: a) configuration I as reference b) configuration IV as reference. $(\times)$ represents signals from the NORMAL class and $(\circ)$ represents signals from the ANOMALOUS class.

TABLE II
ALGORITHM PERFORMANCE FOR CONFIGURATION II.

	Classification	
Original	Normal	Anomalous
Normal	2	0
Anomalous	0	12

V. CONCLUSION

This paper proposes a three stage algorithm for automatic audio anomaly detection of rotating machines using image processing features extracted from the image representation of Short-time Fourier Transform of audio signals.

It is shown that the proposed algorithm has the potential to automatically detect audio anomaly, as the performance with the 8 signals recorded database has a 100% hit rate, using simple image processing features.

ACKNOWLEDGMENT

This work is supported primarily by Petrobras S.A. and Statoil Brazil Oil & Gas Ltda under contract COPPETEC 0050.0079406.12.9 (ANP-Brazil R&D Program), and in part by CNPq and FAPERJ.

The authors of this paper would like to thank Guilherme Carvalho for the schematic design of the robot trail depicted in Fig. 1.

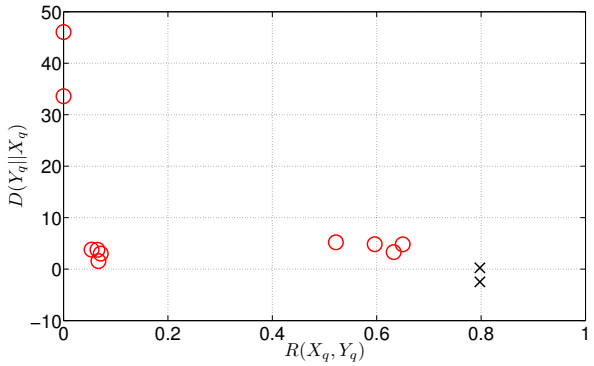


Fig. 8. Graphical representation of features $R(X_q, Y_q)$ and $D_{\mathrm{KL}}(Y_q||X_q)$ for configuration II as reference. $(\times)$ represents signals from the NORMAL class and $(\circ)$ represents signals from the ANOMALOUS class.

REFERENCES

[1] S. K. Yadav, K. Tyagi, B. Shah, and P. K. Kalra, "Audio Signature-Based Condition Monitoring of Internal Combustion Engine Using FFT and Correlation Approach," *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 4, pp. 1217–1226, April 2011.

[2] P. Henriquez, J. B. Alonso, M. A. Ferrer, and 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. 5, pp. 642–652, May 2014.

[3] A. A. de Lima, T. de M. Prego, S. L. Netto, et al., "On Fault Classification in Rotating Machines using Fourier Domain Features and Neural Networks," in: *Proc. IEEE Fourth Latin American Symposium on Circuits and Systems (LASCAS)*, pp. 1–4, Peru, February 2013.

[4] G. C. Pousa, D. F. Comesaa, J. Wild, "Acoustic particle velocity for fault detection of rotating machinery using tachless order analysis," in: *Proc. Inter-Noise*, pp. 1–8, August 2015.

[5] R. M. Vilela, J. C. Metrolho, and J. C. Cardoso, "Machine and industrial o monitorization system by analysis of acoustic signatures," in: *Proc. 12th Conf. IEEE Mediterranean Electrotech.*, pp. 277–279, Croatia, May 2004.

[6] P. Scanlon, A. M. Lyons, and A. O'Loughlin, "Acoustic Signal Processing for Degradation Analysis of Rotating Machinery to Determine the Remaining Useful Life," in: *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 90–93, October 2007.

[7] G. P. S. de Carvalho, G. M. Freitas, R. R. Costa, et al., "DORIS - Monitoring Robot for Offshore Facilities," in: *Proc. Offshore Technology Conference 2013 (OTC)*, pp. 1–13, October 2013.

[8] M. Galassi, A. Rϕyrϕy, G.P.S. de Carvalho, et al., "DORIS - A Mobile Robot for Inspection and Monitoring of Offshore Facilities," in: *Anais do XX Congresso Brasileiro de Automtica (CBA)*, pp. 3174–3181, September 2014.

[9] D.-M. Tsai, C.-T. Lin, and J.-F. Chen, "The evaluation of normalized cross correlations for defect detection," *Pattern Recognition Letters*, vol. 24, issue 15, pp. 2525–2535, November 2003.

[10] F. Zhao, Q. Huang, and W. Gao, "Image Matching by Normalized Cross-Correlation," *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 2, pp. 1–4, May 2006.

[11] M. Debella-Gilo, and A. Kb, "Sub-pixel precision image matching for measuring surface displacements on mass movements using normalized cross-correlation," *Remote Sensing of Environment*, vol 115, issue 1, pp. 130–142, January 2011.

[12] J. Goldberger, S. Gordon, and H. Greenspan, "An efficient image similarity measure based on approximations of KL-divergence between two gaussian mixtures," in: *Proc. IEEE International Conference on Computer Vision*, pp. 487–493, October 2003.

[13] R. Kwitt, and A. Uhl, "Image similarity measurement by Kullback-Leibler divergences between complex wavelet subband statistics for texture retrieval," in: *Proc. IEEE International Conference on Image Processing (ICIP)*, pp 933–936, October 2008.