

MIMII DATASET: SOUND DATASET FOR MALFUNCTIONING INDUSTRIAL MACHINE INVESTIGATION AND INSPECTION

*Harsh Purohit, Ryo Tanabe, Kenji Ichige, Takashi Endo,
Yuki Nikaido, Kaori Suefusa, and Yohei Kawaguchi*

Research and Development Group, Hitachi, Ltd.
1-280, Higashi-koigakubo, Kokubunji, Tokyo 185-8601, Japan
{harsh_pramodbhai.purohit.yf, yohei.kawaguchi.xk}@hitachi.com

ABSTRACT

Factory machinery is prone to failure or breakdown, resulting in significant expenses for companies. Hence, there is a rising interest in machine monitoring using different sensors including microphones. In scientific community, the emergence of public datasets has been promoting the advancement in acoustic detection and classification of scenes and events, but there are no public datasets that focus on the sound of industrial machines under normal and anomalous operating conditions in real factory environments. In this paper, we present a new dataset of industrial machine sounds which we call a sound dataset for malfunctioning industrial machine investigation and inspection (MIMII dataset). Normal and anomalous sounds were recorded for different types of industrial machines, i.e. valves, pumps, fans and slide rails. To resemble the real-life scenario, various anomalous sounds have been recorded, for instance, contamination, leakage, rotating unbalance, rail damage, etc. The purpose of releasing the MIMII dataset is to help the machine-learning and signal-processing community to advance the development of automated facility maintenance.

Index Terms— Machine sound dataset, Acoustic scene classification, Anomaly detection, Unsupervised anomalous sound detection

1. INTRODUCTION

The increasing demand for automatic machine inspection is attributable to the need for a better quality of factory equipment maintenance. The discovery of malfunctioning machine parts mainly depends on the experience of field engineers. However, shortage of field experts due to the increased number of requests for inspection has become an important problem in the industry. Therefore, an efficient and affordable solution to this problem is highly desirable.

In the past decade, industrial Internet of Things (IoT) and data-driven techniques have been revolutionizing the manufacturing industry, and different approaches have been undertaken for monitoring the state of machinery; for example, vibration sensor-based approaches [1–4], temperature sensor-based approaches [5], pressure sensor-based approaches [6], etc. Another approach is to detect anomalies from sound by using technologies for acoustic scene classification and event detection [7–13]. A remarkable advancement has been made in classification of acoustic scenes and detection of acoustic events, and there are many promising state-of-the-art studies [14–16]. We know that the emergence of numerous open benchmark dataset [17–20] is essential for the advancement of the research field. However, to the best of our knowledge, there is no

public dataset which contains different types of machine sounds in real factory environments.

In this paper, we introduce a new dataset of machine sounds in normal and anomalous operating conditions in real factory environments. We include the sound of four machine types: (i) valves, (ii) pumps, (iii) fans, and (iv) slide rails. For each type of machine, we consider seven kinds of product models. We assume that the main task is to find an anomalous condition of the machine during a 10-second sound segment in an unsupervised learning situation. In other words, only normal machine sounds can be used in the training phase, and one has to correctly distinguish between a normal machine sound and an abnormal machine sound in the test phase. The main contributions of this paper can be summarized as follows: (1) We created an open dataset for malfunctioning industrial machine investigation and inspection (MIMII), first of its kind. (We will release this dataset by the workshop.) This dataset contains a total of 26,092 sound files for normal conditions of four different machine types. It also contains real-life anomalous sound files for each category of the machines. (2) Using our developed dataset, we have explored an autoencoder-based model for each type of machine with various noise conditions. These results can be taken as a benchmark to improve the accuracy of anomaly detection in the MIMII dataset.

The rest of the paper is organized as follows. In Section 2, we describe the recording environment and the setup. The details of the dataset content are given in Section 3. The autoencoder-based detection benchmark and results are discussed in Section 4. Section 5 concludes the paper.

2. RECORDING ENVIRONMENT AND SETUP

The dataset was collected using TAMAGO-03 microphone, manufactured by *System In Frontier Inc* [21]. It is a circular microphone array which consists of eight distinct microphones; the details of the microphone array are shown in Figure 1. By using the microphone array, not only single-channel-based approaches but also multi-channel-based ones can be evaluated. The microphone array was kept at a distance of 50 cm from the machine (10 cm in case of valves); 10-second sound segments were recorded. The dataset contains eight separate channels for each segment. Figure 2 depicts the recording setup with the direction and distance for each kind of machine. It should be noted that each machine sound was recorded in separate session. In running condition, the sound of the machine was recorded as 16-bit audio signals sampled at 16 kHz in a reverberant environment. Apart from the target machine sound, background noise in multiple real factories was continuously recorded

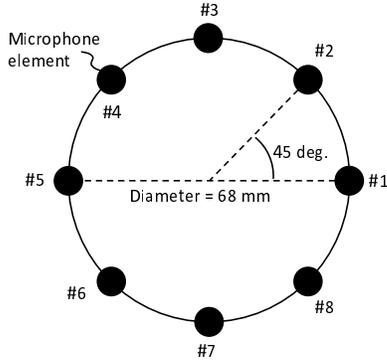


Figure 1: Circular microphone array

to mix it with the target machine sound for simulating real environments. For recording the background noise, we used the same microphone array as for the target machine sound.

3. DATASET CONTENT

The MIMII dataset contains the sound of four different types of machines: valves, pumps, fans, and slide rails. The valves are solenoid valves that are repeatedly opened and closed. The pumps are water pumps, which drained water from a pool and discharged water to the pool continuously. The fans represent industrial fans, which are used to provide a continuous flow or gas of air in factories. The slide rails in this paper represent linear slide systems, which consist of a moving platform and a stage base. The types of the sounds produced by the machines are stationary and non-stationary, have different features, and different degrees of difficulty. Figure 3 depicts a power spectrogram of the sound of all four types of machines, clearly showing that each machine has its unique sound characteristics.

The list of sound files for each machine type is reported in the Table 1. Each type of machines consists of seven individual machines. Individual machines may be of a different product model. We know that large datasets incorporating real-life complexity are needed to effectively train the models, so we recorded a total of 26,092 normal sound segments for all individual machines. In addition to this, different real-life anomalous scenarios have been considered for each kind of machine, for instance, contamination, leakage, rotating unbalance, rail damage, etc. Various running conditions are listed in Table 2. The number of sound segments for each anomalous sound for each different type of machine is small because we regard the main target of our dataset as an unsupervised learning scenario and regard the anomalous segments as a part of test data.

As explained in Section 2, the background noise recorded in multiple real factories was mixed with the target machine sound. Eight channels are considered separately during mixing the original sounds with the noise. For a certain signal-to-noise ratio (SNR) γ dB, the noise-mixed data of each machine model was made by the following steps:

1. The average power over all segments of the machine models, a , was calculated.
2. For each segment i from the machine model,

Table 1: MIMII dataset content details

Machine type/ model ID	Segments for normal condition	Segments for anomalous condition	
Valve	00	991	119
	01	869	120
	02	708	120
	03	963	120
	04	1000	120
	05	999	400
	06	992	120
Pump	00	1006	143
	01	1003	116
	02	1005	111
	03	706	113
	04	702	100
	05	1008	248
	06	1036	102
Fan	00	1011	407
	01	1034	407
	02	1016	359
	03	1012	358
	04	1033	348
	05	1109	349
	06	1015	361
Slide rail	00	1068	356
	01	1068	178
	02	1068	267
	03	1068	178
	04	534	178
	05	534	178
	06	534	89
Total	26092	6065	

Table 2: List of operations and anomalous conditions

Machine type	Operations	Examples of anomalous conditions
Valve	Open/close repeat with different timing	More than two kinds of contamination
Pump	Suction from/ discharge to a water pool	Leakage, contamination, clogging, etc.
Fan	Normal work	Unbalanced, voltage change, clogging, etc.
Slide rail	Slide repeat at different speeds	Rail damage, loose belt, no grease, etc.

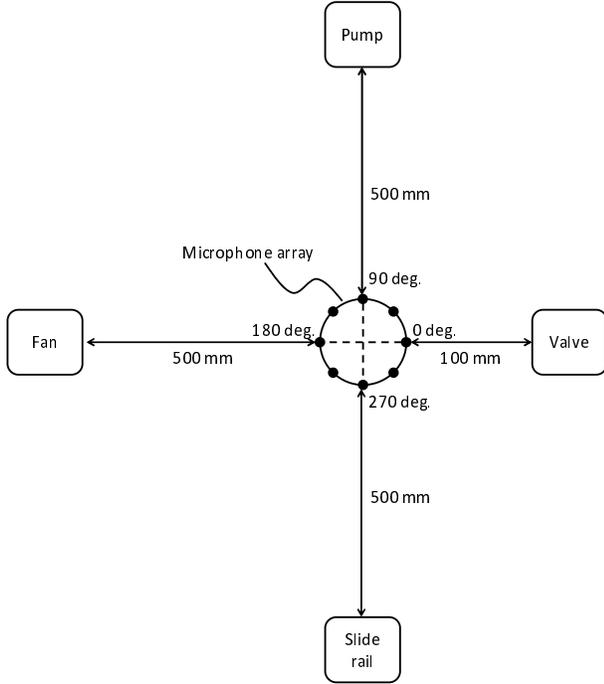


Figure 2: Schematic experimental setup for dataset recording

- (a) a background-noise segment j is randomly selected, and its power b_j is tuned so that $\gamma = 10 \log_{10}(a/b_j)$.
- (b) The noise-mixed data is calculated by adding the target-machine segment i and the power-tuned background-noise segment j .

4. EXPERIMENT

An example of benchmarking is shown in this section. Our main goal is to detect anomalous sounds in an unsupervised learning scenario as discussed in Section 1, and several studies have successfully used autoencoders for unsupervised anomaly detection [12, 22–24], so an autoencoder-based unsupervised anomaly detector was evaluated.

We consider log-Mel spectrogram as an input feature. To calculate the Mel spectrogram, we consider a frame size of 1024, a hop size of 512 and a mel filter banks of 64 in our experiment. Five frames have been combined to initiate our 320 dimensional input feature vector \mathbf{x} . The parameters of the encoder and decoder neural networks (i.e. $\theta = (\theta_e, \theta_d)$) are trained to minimize the loss function given as follows:

$$L_{AE}(\theta_e, \theta_d) = \|\mathbf{x} - D(E(\mathbf{x} | \theta_e) | \theta_d)\|_2^2 \quad (1)$$

Our basic assumption is that this trained model will give high reconstruction error for anomalous machine sounds. The autoencoder network structure for the experiment is summarized as follows: The encoder network ($E(\cdot)$) comprises $FC(Input, 64, ReLU)$; $FC(64, 64, ReLU)$; and $FC(64, 8, ReLU)$, and the decoder network ($D(\cdot)$) incorporates $FC(8, 64, ReLU)$; $FC(64, 64, ReLU)$

and $FC(64, Output, none)$ where $FC(a, b, f)$ means a fully-connected layer with a input neurons, b output neurons, and activation function f . The ReLUs are Rectified Linear Units [25]. The network is trained with Adam [26] optimization technique for 50 epochs.

For each machine type and model ID, all the segments were split into a training dataset and a test dataset. All the anomalous segments were regarded as the test dataset, the same number of normal segments were randomly selected and regarded as the test dataset, and all the rest normal segments were regarded as the training dataset. By using the training dataset consisting only of normal ones, different autoencoders were trained for each machine type and model ID. Anomaly detection was performed for each segment by thresholding the reconstruction error averaged over 10 seconds, and the area under the curve (AUC) values were calculated for the test dataset for each machine type and model ID. In addition to this, we also considered different levels of SNR (with factory noise) in the experiment, for example, 6 dB, 0 dB, and -6 dB.

Table 3 shows the AUCs averaged over three training runs with independent initializations. In Table 3, It is clear that the AUCs for valves are lower than the other machines. Sound signals of valves are non-stationary, in particular, impulsive and sparse in time, and the reconstruction error averaged over time tends to be small. So, it is difficult to detect anomalies for valves. In contrast, it is easier to detect anomalies for fans than the other machines because sound signals of fans are stationary. Moreover, for some machine models, the AUC decreases rapidly as the noise level increases. These results indicate that it is important to solve the degradation caused by non-stationarity and noise for unsupervised anomalous sound detection.

5. CONCLUSION AND PERSPECTIVES

In this paper, we introduced the MIMII dataset, a real-world dataset for investigating the malfunctioning behavior of the industrial machines. We collected 26,092 sound segments of normal condition and 6,065 sound segments of anomalous condition and mixed the background noise recorded in multiple real factories with the machine-sound segments for simulating real environments. In addition, using the MIMII dataset, we showed an example of evaluation for autoencoder-based unsupervised anomalous sound detection. We observed that non-stationary machine sound signals and noise are the key issues for developing the unsupervised anomaly detector. These results can be taken as a benchmark to improve the accuracy of anomaly detection in the MIMII dataset.

We will release this dataset by the workshop. To the best of our knowledge, this dataset is the first of its kind to address the problem of detecting anomalous conditions in industrial machinery through machine sounds. As benchmarking is an important aspect in data driven methods, we strongly believe that our MIMII dataset will be very useful to the research community. We are releasing this data to accelerate research in the area of audio event detection, specifically for machine sounds. This dataset can be used for other use cases, for example, to restrict the training on specific number of machine models and then test on the remaining machine models. This study will be useful for measuring the domain adaptation capability of the different methods applied on machines from different manufactures. If the community finds interest in our dataset and validates its usage, we will improve the current version with the additional meta-data related to different anomalies.

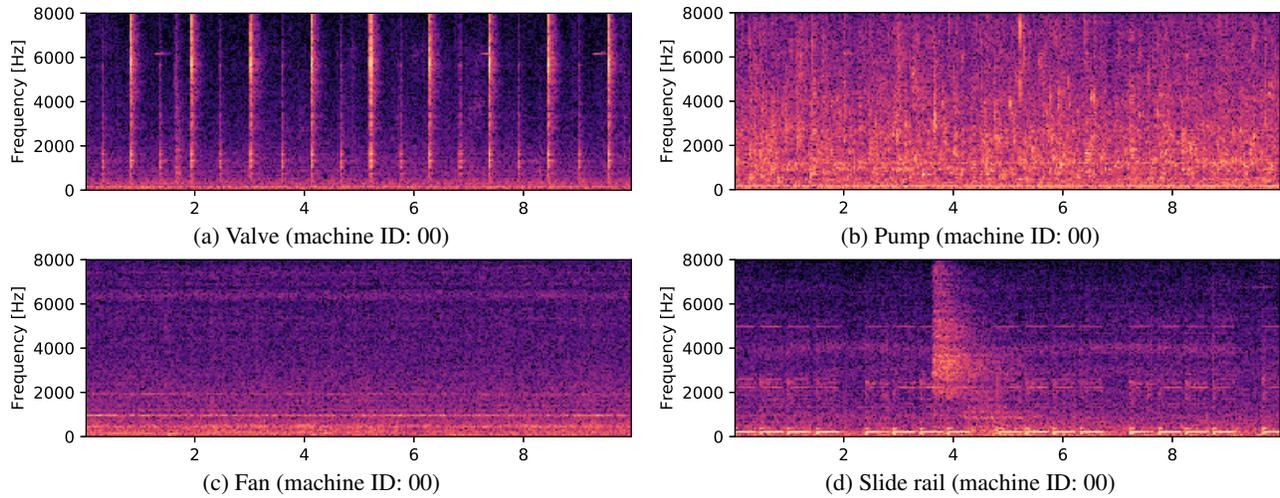


Figure 3: Examples of power spectrograms on a normal condition at 6 dB SNR.

6. REFERENCES

[1] M. Yu, D. Wang, and M. Luo, “Model-based prognosis for hybrid systems with mode-dependent degradation behaviors,” *IEEE Transactions on Industrial Electronics*, vol. 61, no. 1, pp. 546–554, 2013.

[2] T. Ishibashi, A. Yoshida, and T. Kawai, “Modelling of asymmetric rotor and cracked shaft,” in *Proceedings of the 2nd Japanese Modelica Conference*, no. 148, 2019, pp. 180–186.

[3] E. P. Carden and P. Fanning, “Vibration based condition monitoring: A review,” *Structural health monitoring*, vol. 3, no. 4, pp. 355–377, 2004.

[4] G. S. Galloway, V. M. Catterson, T. Fay, A. Robb, and C. Love, “Diagnosis of tidal turbine vibration data through deep neural networks,” in *Proceedings of the 3rd European Conference of the Prognostics and Health Management Society*, 2016.

[5] G. Lodewijks, W. Li, Y. Pang, and X. Jiang, “An application of the IoT in belt conveyor systems,” in *Proceedings of the International Conference on Internet and Distributed Computing Systems (IDCS)*, 2016, pp. 340–351.

[6] R. F. Salikhov, Y. P. Makushev, G. N. Musagitova, L. U. Volkova, and R. S. Suleymanov, “Diagnosis of fuel equipment of diesel engines in oil-and-gas machinery and facilities,” *AIP Conference Proceedings*, vol. 2141, no. 1, p. 050009, 2019.

[7] Y. Koizumi, S. Murata, N. Harada, S. Saito, and H. Uematsu, “SNIPER: Few-shot learning for anomaly detection to minimize false-negative rate with ensured true-positive rate,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 915–919.

[8] Y. Kawachi, Y. Koizumi, S. Murata, and N. Harada, “A two-class hyper-spherical autoencoder for supervised anomaly detection,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 3047–3051.

Table 3: AUCs for all machines

Machine type/ model ID	Input SNR			
	6 dB	0 dB	-6 dB	
Valve	00	0.68	0.55	0.62
	01	0.77	0.71	0.61
	02	0.66	0.59	0.57
	03	0.70	0.65	0.44
	04	0.64	0.65	0.50
	05	0.52	0.48	0.44
	06	0.70	0.66	0.53
Avg.	0.67	0.61	0.53	
Pump	00	0.84	0.65	0.58
	01	0.98	0.90	0.73
	02	0.45	0.46	0.52
	03	0.79	0.81	0.75
	04	0.99	0.95	0.93
	05	0.66	0.66	0.64
	06	0.94	0.76	0.61
Avg.	0.81	0.74	0.68	
Fan	00	0.75	0.63	0.57
	01	0.97	0.90	0.70
	02	0.99	0.83	0.68
	03	1.00	0.89	0.70
	04	0.92	0.75	0.57
	05	0.95	0.90	0.83
	06	0.99	0.97	0.83
Avg.	0.94	0.84	0.70	
Slide rail	00	0.99	0.99	0.93
	01	0.94	0.90	0.83
	02	0.93	0.79	0.74
	03	0.99	0.85	0.71
	04	0.88	0.78	0.61
	05	0.84	0.70	0.60
	06	0.71	0.56	0.52
Avg.	0.90	0.80	0.70	

- [9] M. Yamaguchi, Y. Koizumi, and N. Harada, “AdaFlow: Domain-adaptive density estimator with application to anomaly detection and unpaired cross-domain translation,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 3647–3651.
- [10] Y. Kawaguchi, R. Tanabe, T. Endo, K. Ichige, and K. Hamada, “Anomaly detection based on an ensemble of dereverberation and anomalous sound extraction,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 865–869.
- [11] Y. Koizumi, S. Saito, H. Uematsu, Y. Kawachi, and N. Harada, “Unsupervised detection of anomalous sound based on deep learning and the Neyman–Pearson lemma,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 1, pp. 212–224, 2018.
- [12] Y. Kawaguchi and T. Endo, “How can we detect anomalies from subsampled audio signals?” in *Proceedings of the IEEE 27th International Workshop on Machine Learning for Signal Processing (MLSP)*, 2017, pp. 1–6.
- [13] Y. Kawaguchi, “Anomaly detection based on feature reconstruction from subsampled audio signals,” in *Proceedings of the European Signal Processing Conference (EUSIPCO)*, 2018, pp. 2524–2528.
- [14] A. Mesaros, T. Heittola, E. Benetos, P. Foster, M. Lagrange, T. Virtanen, and M. D. Plumbley, “Detection and classification of acoustic scenes and events: Outcome of the DCASE 2016 Challenge,” *IEEE/ACM Transactions on Audio, Speech and Language Processing*, vol. 26, no. 2, pp. 379–393, 2018.
- [15] S. S. R. Phayre, E. Benetos, and Y. Wang, “SubSpectralNet—using sub-spectrogram based convolutional neural networks for acoustic scene classification,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 825–829.
- [16] Z. Podwinska, I. Sobieraj, B. M. Fazenda, W. J. Davies, and M. D. Plumbley, “Acoustic event detection from weakly labeled data using auditory salience,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 41–45.
- [17] J. F. Gemmeke, D. P. W. Ellis, D. Freedman, A. Jansen, W. Lawrence, R. C. Moore, M. Plakal, and M. Ritter, “Audio Set: An ontology and human-labeled dataset for audio events,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017, pp. 776–780.
- [18] E. Fonseca, J. Pons, X. Favory, F. Font, D. Bogdanov, A. Ferraro, S. Oramas, A. Porter, and X. Serra, “Freesound datasets: A platform for the creation of open audio datasets,” in *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, 2017, pp. 486–493.
- [19] G. Dekkers, S. Lauwereins, B. Thoen, M. W. Adhana, H. Brouckxon, B. V. den Bergh, T. van Waterschoot, B. Vanrumste, M. Verhelst, and P. Karsmakers, “The SINS database for detection of daily activities in a home environment using an acoustic sensor network,” in *Proceedings of the Workshop on Detection and Classification of Acoustic Scenes and Events (DCASE)*, 2017, pp. 32–36.
- [20] Y. Koizumi, S. Saito, H. Uematsu, N. Harada, and K. Imoto, “ToyADMOS: A dataset of miniature-machine operating sounds for anomalous sound detection,” in *Proceedings of the Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2019, to appear.
- [21] System In Frontier Inc. (http://www.sifi.co.jp/system/modules/pico/index.php?content_id=39&ml_lang=en).
- [22] T. Tagawa, Y. Tadokoro, and T. Yairi, “Structured denoising autoencoder for fault detection and analysis,” in *Proceedings of the Asian Conference on Machine Learning (ACML)*, 2015, pp. 96–111.
- [23] E. Marchi, F. Vesperini, F. Eyben, S. Squartini, and B. Schuller, “A novel approach for automatic acoustic novelty detection using a denoising autoencoder with bidirectional LSTM neural networks,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2015, pp. 1996–2000.
- [24] D. Oh and I. Yun, “Residual error based anomaly detection using auto-encoder in SMD machine sound,” *Sensors*, vol. 18, no. 5, p. 1308, 2018.
- [25] K. Jarrett, K. Kavukcuoglu, M. Ranzato, and Y. LeCun, “What is the best multi-stage architecture for object recognition?” in *Proceedings of the 12th IEEE International Conference on Computer Vision (ICCV)*, 2009, pp. 2146–2153.
- [26] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.