ANOMALOUS SOUND DETECTION METHOD WITH SIMPLE NOISE REDUCTION

Technical Report

Kosei Ozeki, Takeru Shiraga, Hideaki Terashima, Nobuaki Tanaka, and Takahiko Masuzaki

Mitsubishi Electric Corporation, Kamakura, Kanagawa, 2478501, Japan
Ozeki.Kosei@aj.MitsubishiElectric.co.jp
Shiraga.Takeru@ea.MitsubishiElectric.co.jp
Terashima.Hideaki@bp.MitsubishiElectric.co.jp
Tanaka.Nobuaki@ce.MitsubishiElectric.co.jp
Masuzaki.Takahiko@dc.MitsubishiElectric.co.jp

ABSTRACT

This paper presents anomalous sound detection methods for DCASE2026 Task 2. The goal of this contest is to identify whether the sounds emitted from target machines are normal or anomaly. First, we applied a simple noise reduction method. This method uses the far-channel as a noise reference signal and cancels the noise component mixed into the near-channel using a linear filter. Next, we applied the following four anomaly detection methods.
1. Pre-trained CED model + k-NN.
2. Pre-trained BEATs model + k-NN.
3. Baseline autoencoder model with Mahalanobis metric.
4. Pre-trained BEATs model + autoencoder + MMD
In the evaluation on the development dataset, the three systems using pre-trained feature extractors performed better on the target domain, with System 1 (CED + k-NN) performing particularly well.

Index Terms— noise reduction, data adaptation

1. INTRODUCTION

Anomalous sound detection (ASD) is an essential technique in machine condition monitoring. DCASE2026 Task 2 [1-4] is a data challenge focused on ASD. Participants, including the authors, aim to detect anomalous sounds using only normal data for training, reflecting real-world conditions.

The authors' group has been working on anomaly detection in time-series data and on improving ASD system performance by proposing various methods [5-10]. We participated in this challenge to evaluate our technical approach and gain further experience. This paper describes the algorithms and approaches we applied to DCASE2026 Task 2.

The structure of this paper is as follows. Section 2 explains the task and data of DCASE2026 Task 2. Section 3 presents the proposed algorithms. Section 4 shows the evaluation results. Section 5 concludes the paper.

2. PROBLEM DESCRIPTION

The task of DCASE2026 Task 2 is an advanced version of DCASE2025 Task 2. It includes the following five requirements, with the fifth requirement newly introduced in DCASE2026 Task 2:

1. Train the model using only normal sounds (unsupervised learning scenario).
2. Detect anomalies regardless of domain shifts (domain generalization task).
3. Train models for entirely new machine types.
4. Train the model with or without attribute information.
5. Training and inference with two-channel audio recorded at different distances from the target machine.

Next, we describe the provided data. There is a development dataset for seven types of machines and an evaluation dataset for five different types of machines. Both the development and evaluation datasets contain training data and test data. Each of the training and test datasets includes source data and target data, but the source/target information is concealed in the test data of the evaluation dataset. The training data contains only normal data. The test data includes both normal and anomalous data, but the normal/anomaly information is concealed in the test data of the evaluation dataset.

3. PROPOSED ALGORITHM

We first applied a simple noise reduction technique and then applied four anomaly detection methods (Figure 1).

First, the noise reduction technique is described. Assuming that the given far-channel contains only noise, we used the far-channel as a noise reference signal to remove noise from the near-channel. In the STFT domain, let $d_{\omega,\tau}$ denote the noise and $x_{\omega,\tau}$ denote the mixture of the signal and noise. The noise was canceled using a one-tap filter a_ω as follows:

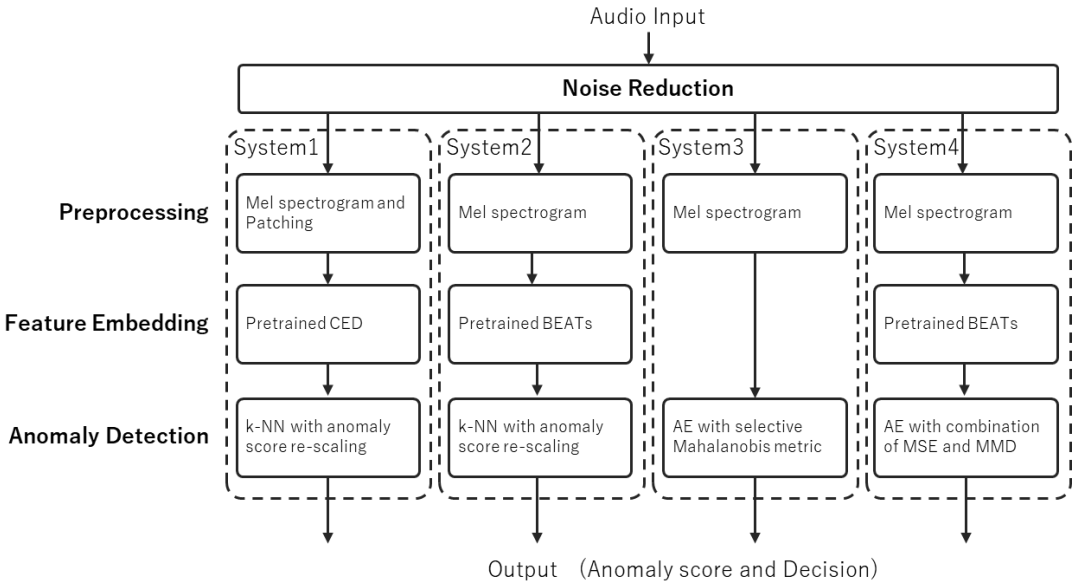

Figure 1: The flowchart of the proposed methods

$$y_{\omega,\tau} = x_{\omega,\tau} - a_{\omega} d_{\omega,\tau} \qquad (1)$$

Here, ω denotes the discrete frequency and τ denotes the frame index. The value of a_{ω} was estimated by complex linear regression to minimize the average energy of $y_{\omega,\tau}$.

Next, the four anomaly detection methods applied after noise reduction are described in the following four subsections.

3.1. Pre-trained CED model

In System 1, we used a pre-trained model based on CED [11] directly as a feature extractor. CED is a Vision Transformer (ViT)-based architecture proposed by Dinkel et al. for audio tagging tasks. It demonstrates strong feature extraction capabilities through pre-training on the large-scale AudioSet dataset. The pre-trained model is available for download from Hugging Face.

For preprocessing, we followed the CED method and applied mel-spectrogram extraction and patching.

For anomaly detection, we used the method by Wilkinghoff et al. [12] to handle a small number of target samples. In the following, we will refer to this method as Wilkinghoff's method. The number of samples used for k-NN rescaling, a hyperparameter, was set to 16, as suggested in the paper. We optimized the hyperparameters of Wilkinghoff's method using the training data. Then, we calculated the anomaly scores using the test data.

3.2. Pre-trained BEATs model

In System 2, we replaced the CED-based feature extractor in System 1 with BEATs [13] and used it in the same procedure. BEATs is a Transformer-based architecture that learns acoustic representations through self-supervised learning. The model is pre-trained on the large-scale AudioSet dataset and has strong ability to capture audio events and acoustic patterns. The pre-trained model is publicly available and can be downloaded for use. In this system, we employed the BEATs_iter3 model.

For preprocessing, following the BEATs method, we extracted log mel-spectrogram from the input audio waveform. We used the BEATs encoder as the feature extractor and employed its output embeddings as feature representations. For anomaly detection, we used Wilkinghoff's method with k-NN rescaling.

3.3. Baseline autoencoder model

In System 3, we used the baseline autoencoder [4] and Mahalanobis metric without modification, because, as described in Section 4 (Evaluation), it achieved the highest AUC_s among our systems. However, for its input audio, System 3 employed the noise reduction method described in Eq. (1). In addition, since a larger number of training epochs improved AUC scores in the target domain, we set the number of epochs to 200.

3.4. Pre-trained BEATs and autoencoder model

In System 4, we used the embeddings extracted with BEATs (System 2) as input features and employed the autoencoder introduced in System 3 for anomaly detection. The autoencoder was trained using a combination of mean squared error (MSE) loss and maximum mean discrepancy (MMD) loss to perform domain adaptation.

4. EVALUATION

The machine-averaged evaluation results on the development dataset are presented in Table 1. Here, hmean denotes the harmonic mean of the three metrics (AUC_s, AUC_t, and pAUC) and is reported as an overall score.

For the source domain, System 3 (AE) achieved the best performance. In contrast, for the target domain, System 1 (CED + k-NN), System 2 (BEATs + k-NN), and System 4 (BEATs +

Table 1: Evaluation of the development dataset

Method		AUC_s	AUC_t	pAUC	hmean
baseline	data_renameEmu_MAHALA	0.672	0.551	0.541	0.580
	data_renameEmu_MSE	0.675	0.527	0.545	0.570
proposed	system1 (CED+kNN)	0.610	**0.630**	0.554	0.587
	system2 (BEATs+kNN)	0.565	0.607	0.538	0.563
	system3 (AE)	**0.762**	0.596	**0.555**	**0.621**
	system4 (BEATs+AE+MMD)	0.606	0.619	0.527	0.576

AE + MMD), all of which use pre-trained feature extractors, yielded better results.

This suggests that the systems using pre-trained feature extractors are more effective in the target domain, likely because they were pre-trained on large-scale audio data and therefore have stronger generalization ability. In contrast, System 3 (AE) performed better in the source domain, which possibly because the pre-trained feature extractors are not specifically adapted to each machine. Although System 3 was trained on both the source and target domains, the source-domain data accounted for a much larger proportion of the training set, which may have limited its ability to generalize effectively to the target domain.

5. CONCLUSION

In this paper, we introduced an anomaly sound detection method for DCASE 2026 Task 2. We evaluated the following approaches. First, we applied a simple noise reduction method, in which the far-channel signal was used as a noise reference signal and a linear filter was employed to cancel the noise components mixed into the near-channel signal. We then applied four anomaly detection methods: Pre-trained CED model with k-NN, Pre-trained BEATs model with k-NN, Baseline autoencoder model, and Pre-trained BEATs model with autoencoder and MMD.

In the evaluation on the development dataset, System 3 (AE) achieved higher performance for the source domain, whereas the three systems using pre-trained feature extractors achieved higher performance for the target domain. In particular, System 1 (CED + k-NN) performed the best.

6. REFERENCES

[1] T. Nishida, N. Harada, D. Takeuchi, D. Niizumi, K. Imoto, K. Dohi, H. Purohit, T. Endo, and Y. Kawaguchi, “Description and discussion on DCASE 2026 challenge task 2: noise-aware unsupervised anomalous sound detection for machine condition monitoring,” arXiv: 2606.01578, 2026.

[2] N. Harada, D. Niizumi, D. Takeuchi, Y. Ohishi, M. Yasuda, and S. Saito, “ToyADMOS2: Another dataset of miniature-machine operating sounds for anomalous sound detection under domain shift conditions,” in Proceedings of the Detection and Classification of Acoustic Scenes and Events Workshop (DCASE), Barcelona, Spain, November 2021, pp. 1-5.

[3] K. Dohi, T. Nishida, H. Purohit, R. Tanabe, T. Endo, M. Yamamoto, Y. Nikaido, and Y. Kawaguchi, “MIMII DG: Sound dataset for malfunctioning industrial machine investigation and inspection for domain generalization task,” in Proceedings of the 7th Detection and Classification of Acoustic Scenes and Events 2022 Workshop (DCASE2022), Nancy, France, November 2022.

[4] N. Harada, D. Niizumi, D. Takeuchi, Y. Ohishi, and M. Yasuda, “First-shot anomaly sound detection for machine condition monitoring: A domain generalization baseline,” Proceedings of the 31st European Signal Processing Conference (EUSIPCO), pp. 191-195, 2023.

[5] T. Nakamura, M. Imamura, R. Mercer, and E. Keogh, “MERLIN: Parameter-Free Discovery of Arbitrary Length Anomalies in Massive Time Series Archives,” 2020 IEEE international conference on data mining (ICDM). doi:10.1109/ICDM50108.2020.00147

[6] T. Nakamura, R. Mercer, M. Imamura, and E. Keogh, “MERLIN++: parameter-free discovery of time series anomalies,” Data Mining and Knowledge Discovery, vol. 37, pp. 670-709. doi:10.1007/s10618-022-00876-7

[7] N. Tanaka, T. Shiraga, and Y. Itani, “Improving Anomalous Sound Detection by Distance Matrix-Based Visualization of Measurement Flaws,” Vol. 4 No. 1 (2023): Proceedings of the Asia Pacific Conference of the PHM Society 2023. doi:10.36001/phmap.2023.v4i1.3754

[8] T. Shiraga, H. Makimoto, et al. “Improving valvular pathologies and ventricular dysfunction diagnostic efficiency using combined auscultation and electrocardiography data: A multimodal AI approach.” Sensors 23.24 (2023): 9834.

[9] H. Makimoto, T. Shiraga, et al. “Efficient screening for severe aortic valve stenosis using understandable artificial intelligence: a prospective diagnostic accuracy study.” European Heart Journal-Digital Health 3.2 (2022): 141-152.

[10] T. Shiraga, K. Ozeki, T. Masuzaki, N. Tanaka, and T. Kuriyama, “Anomalous sound detection method using contrastive learning,” DCASE2025 Challenge, Tech. Rep., June 2025.

[11] H. Dinkel, Y. Wang, Z. Yan, J. Zhang, Y. Wang, “CED: Consistent ensemble distillation for audio tagging,” arXiv:2308.11957, 2023

[12] K. Wilkinghoff, H. Yang, J. Ebbers, F. G. Germain, G. Wichern and J. L. Roux, "Keeping the Balance: Anomaly Score Calculation for Domain Generalization," ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Hyderabad, India, 2025, pp. 1-5, doi: 10.1109/ICASSP49660.2025.10888402.

[13] S. Chen, Y. Wu, C. Wang, S. Liu, D. Tompkins, Z. Chen and F. Wei, "BEATs: Audio Pre-Training with Acoustic Tokenizers," in Proc. 40th Int. Conf. Mach. Learn. (ICML), Honolulu, HI, USA, 2023, pp. 5178–5193.