MACHINE-WISE DUAL-CHANNEL ANOMALOUS SOUND DETECTION WITH TARGET-AWARE THRESHOLDING

Technical Report

Yunhong Zhou

SUMERU ZOO, Beijing, China
zhouyunhong2025@gmail.com

ABSTRACT

We present a machine-wise anomalous sound detection system for DCASE 2026 Task 2. The proposed system addresses the challenges of first-shot unsupervised detection under severe source–target domain shift and dual-channel recording conditions. Instead of applying a single unified strategy to all machines, we separate the pipeline into two branches: conditional routing for attribute-aware machines and global healthy-reference modeling for no-attribute machines. The submitted system adopts a per-machine sub-band representation, followed by Mahalanobis-based anomaly scoring against a healthy reference. For decision-result generation, we further introduce a target-aware thresholding rule that relies only on legal train-normal data and helps reduce domain-induced false positives on machines with strong target shift. In addition, diagnostic side experiments on development machines were used to understand error modes and support system design. The resulting framework provides an interpretable and robust machine-wise solution for first-shot anomalous sound detection.

***Index Terms*—** anomalous sound detection, first-shot learning, dual-channel audio, machine-wise routing, target-aware thresholding

1. INTRODUCTION

Anomalous sound detection is an important component of machine condition monitoring, where the goal is to detect deviations from normal operation using acoustic observations. DCASE 2026 Task 2 formulates this problem in a challenging first-shot unsupervised setting: only normal training samples are provided, the system must generalize to unseen machine types, and strong source–target domain shifts are present [1]. In addition, the 2026 edition introduces dual-channel recordings, which create both opportunities and challenges for robust detection under realistic acoustic conditions.

The task builds on the ToyADMOS2 [2] and MIMII DG [3] datasets, which provide miniature-machine operating sounds under domain shift conditions. The first-shot formulation follows the domain generalization baseline established by Harada et al. [4], where only a small number of target-domain normal samples are available at inference time. A central difficulty of this task is that normal machine behavior is not a single fixed point in feature space. Instead, it is machine-specific, condition-dependent, and often shifted between source and target domains. In practice, some machines benefit from attribute-aware routing, while others are better handled by global healthy-reference modeling. This observation motivates us to adopt a machine-wise strategy rather than forcing all machines into a single detector design.

Table 1: Final submission pipeline by machine category.

Category	Routing and decision
Attribute-aware	Conditional routing to condition-aware anchors; Mahalanobis score; target-aware threshold
No-attribute	Global healthy-reference anchor; Mahalanobis score; target-aware threshold

Both categories use per-machine sub-band features.

Based on this view, our submission combines three key ideas: 1) per-machine sub-band anomaly scoring, 2) machine-wise branching for attribute-aware and no-attribute machines, and 3) a target-aware thresholding rule that uses only legal train-normal data. The resulting system is simple, interpretable, and compatible with the first-shot unsupervised constraints of the challenge.

2. SYSTEM OVERVIEW

Our system consists of four stages: representation extraction, machine-wise routing, anomaly score computation, and decision-result generation. Given an input waveform, we extract acoustic features, route the sample according to machine category, compute an anomaly score against a healthy reference, and finally apply a target-aware threshold to produce the decision result.

Instead of applying a single route to all machines, we divide the task into two branches. For machines with usable attribute structure in the training data, we use conditional routing and condition-aware reference anchors. For machines without explicit attribute information, we use a global healthy-reference model. Finally, a target-aware thresholding rule is applied to produce decision results from the anomaly scores.

Table 1 summarizes the final submission pipeline by machine category.

3. ACOUSTIC REPRESENTATION AND SCORING

3.1. Feature Extraction

The submitted system represents each clip using sub-band acoustic descriptors derived from log-mel energies together with spectral- and temporal-statistic summaries. We adopt a per-machine sub-band discriminative scheme: for each machine, the most informative sub-band representation is selected, and a Mahalanobis score is computed against its healthy reference. This per-machine refinement sharpens the normal/anomaly separation under machine-specific spectral characteristics, while keeping the specific per-machine choices fixed using only legal train-side information.

3.2. Anomaly Scoring

For each machine, the selected sub-band representation is scored by a Mahalanobis distance to a healthy reference distribution estimated from train-normal data. The Mahalanobis form measures how far a test sample lies from the healthy manifold while accounting for the covariance structure of normal operation.

For attribute-aware machines, the scoring pipeline is supported by conditional routing. In this branch, condition-dependent reference anchors are constructed from legal train-side information, and test samples are routed to the nearest condition reference. In contrast, no-attribute machines are handled using a global healthy reference model built from all available train-normal samples.

4. DECISION RULE AND TARGET-AWARE THRESHOLDING

4.1. Threshold Formulation

Our diagnostics show that a source-only decision threshold can be too restrictive for machines with strong source–target domain shifts. On some machines, target-domain healthy samples are already far from source-domain healthy samples in the anomaly-score space. As a consequence, a threshold estimated only from source-normal samples can generate excessive false positives, even if the anomaly-score ranking itself remains informative.

To address this issue, we use a target-aware thresholding rule based only on legal train-normal data:

$$\tau = \max\left(p_{90}(s_{\mathrm{src}}),\ p_{50}(s_{\mathrm{tgt}})\right), \tag{1}$$

where s_{src} denotes source-train self-scores and s_{tgt} denotes target-train normal scores. This rule preserves source-based baseline behavior on machines with weak domain shift, while raising the threshold on machines where target healthy samples are systematically shifted away from source healthy samples. Importantly, the rule does not use evaluation labels or test ground truth.

4.2. Legality and Rationale

We emphasize three properties of the threshold rule:

- **Only legal train-normal data are used.** No evaluation labels, no test ground-truth, and no anomalous examples participate in threshold computation.
- **Conservative bias.** By taking the maximum of the two percentiles, the rule automatically elevates the threshold when the target domain exhibits a shifted score distribution, thereby suppressing domain-induced false positives.
- **Machine-wise adaptivity.** The threshold is computed independently for each machine, reflecting the machine-specific score landscape rather than a global constant.

5. DIAGNOSTIC SIDE EXPERIMENTS

During system development, we conducted diagnostic experiments on the development-set machines. These experiments were not designed as leaderboard-oriented optimizations, but rather to understand error modes and support design decisions in a controlled, no-leakage setting. In particular, we studied domain-induced false positives, the behavior of global versus conditional references, and the limits of coarse global spectral summaries.

The side experiments were useful in two ways. First, they helped us confirm that some machines are dominated by source–target mismatch rather than by purely anomalous structure. Second, they helped us reject unsupported hypotheses before moving them to the main submission pipeline. This diagnostic discipline was important for keeping the final system interpretable and avoiding unsupported heuristics.

6. DISCUSSION

Our final system reflects a practical trade-off between robustness, legality, and machine-wise flexibility. Attribute-aware machines benefit from routing and condition-aware references, whereas no-attribute machines are more sensitive to global domain shift and therefore benefit more from robust thresholding. This difference suggests that first-shot anomalous sound detection should not always be treated as a fully uniform problem across machine types.

Another important observation is that anomaly-score quality and decision-result quality are not always governed by the same mechanism. A machine may still have useful anomaly-score ranking even when a naive source-only threshold produces too many false positives. This is why our submission explicitly separates score computation from decision thresholding and introduces a target-aware correction only at the decision stage. Finally, our side experiments suggest that not every intuitively plausible refinement transfers well across machines. Some diagnostic directions were found to be useful, while others were rejected after controlled verification. This supports the use of a conservative machine-wise workflow when preparing challenge submissions under limited evaluation feedback.

7. CONCLUSION

We proposed a machine-wise dual-channel anomalous sound detection system for DCASE 2026 Task 2. The final submission combines per-machine sub-band features, Mahalanobis-based scoring, conditional routing for attribute-aware machines, a global healthy anchor for no-attribute machines, and a legal target-aware thresholding rule for decision-result generation. Our experiments and diagnostics indicate that a single unified detector is insufficient for all machines, and that machine-wise handling is important for robustness under source–target mismatch. Future work will focus on finer local structure modeling, more stable cross-domain healthy representations, and better short-duration handling for no-attribute machines with strong domain shift.

8. REFERENCES

[1] T. Nishida, N. Harada, D. Niizumi, D. Imoto, K. Dohi, K. 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," in *arXiv e-prints: 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 *Proc. Detection and Classification of Acoustic Scenes and Events Workshop (DCASE)*, pp. 1–5, 2021.

[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 *Proc. 7th Detection and Classification of Acoustic Scenes and Events 2022 Workshop (DCASE2022)*, 2022.

[4] N. Harada, D. Niizumi, D. Takeuchi, Y. Ohishi, and M. Yasuda, "First-shot anomaly detection for machine condition monitoring: A domain generalization baseline," in *Proc. 31st European Signal Processing Conference (EUSIPCO)*, pp. 191–195, 2023.