NOISE-AWARE UNSUPERVISED ANOMALOUS SOUND DETECTION WITH UBM AND GLOBAL GMM

Technical Report

Xiao Mei[*]

Fund Intelligence Detection (Shanghai) Co., Ltd
Shanghai, China
xiao.mei@fund-a.com

Baiji Ren

Fund Intelligence Detection (Shanghai) Co., Ltd
Shanghai, China
penny.ren@fund-a.com

ABSTRACT

This report describes our submission to DCASE 2026 Task 2: Noise-aware Unsupervised Anomalous Sound Detection for Machine Condition Monitoring. We propose three handcrafted systems built on a 221-dimensional clip-level feature vector and Gaussian Mixture Models. A two-channel scaled-difference signal suppresses environmental noise. Two systems use a Universal Background Model likelihood-ratio framework; the third uses a per-machine GMM. No evaluation labels or public code are used.

Index Terms—Anomalous sound detection, domain generalization, Gaussian mixture model, universal background model, handcrafted features.

1. INTRODUCTION

DCASE 2026 Task 2 targets unsupervised anomalous sound detection under domain shift, using datasets such as ToyADMOS2 and MIMII DG. Each machine type has a large source-domain training set but few target-domain normal samples, motivating a handcrafted-feature approach. We exploit the two-channel audio (near-mic ch0 and far-mic ch1) to separate machine-specific information from shared background noise. Three systems are submitted: two UBM likelihood-ratio variants and one per-machine GMM.

2. PROPOSED METHOD

2.1. Feature Extraction

All clips are resampled to 16 kHz and fixed to 10 seconds. A 221-dimensional feature vector is extracted per clip: 117 base handcrafted statistics, 49 nonlinear features, and 55 phase-1 auxiliary features. Base features include log-mel band energy, spectral and temporal statistics, crest factor, and high-frequency variability. Nonlinear features cover harmonic ratios, sidebands, band ratios, modulation, and temporal shape. Phase-1 features are machine-specific statistics for each development machine type.

Table 1. Feature groups and dimensions.

Group	Sub-components	Dim
Base	band energy/shape, spectral/temporal stats, HF variability	117
Nonlinear	harmonic ratios, sidebands, band ratios, modulation, spectral/temporal shape	49
Phase-1	per-machine statistics for 7 dev machines	55
Total		221

2.2. Channel Compensation

Each recording contains a near microphone (ch0) and a far microphone (ch1). A machine-dependent coefficient α is estimated on the normal training set via ordinary least squares, forming the scaled-difference signal:

$$\alpha = \frac{ch0^{T}\, ch1}{ch1^{T}\, ch1 + \varepsilon} \tag{1}$$

$$x_{sd} = ch0 - \alpha\, ch1 \tag{2}$$

This suppresses environmental components common to both microphones while preserving machine-specific anomalies. The far microphone ch1 is kept separately to train the UBM background model.

2.3. Systems Overview

Table 2 summarizes the three systems. Systems 1 and 2 share the same UBM likelihood-ratio framework but use different background data; System 3 is a per-machine GMM without background.

Table 2. Overview of the three submitted systems.

System	Background	Score	Dev(%)
1 UBM eval only	eval ch1	LR	64.78
2 UBM eval+dev	eval+dev ch1	LR	65.08
3 Per-machine GMM	none	NLL	62.65

2.4. UBM Likelihood-Ratio System

Two systems model both the target machine and a general background with tied-covariance GMMs. A 130-component GMM is trained on scaled_diff features per machine, and a 512-component background GMM is trained on ch1 features. System 1 uses only evaluation machines for the background; System 2 additionally includes development-machine ch1.

$$s(x) = -\log p_m(x) + \log p_{bg}(x) \tag{3}$$

where p_m(x) and p_bg(x) are the likelihoods under the machine-specific model and the background model, respectively.

2.5. Per-Machine GMM

System 3 trains a separate 130-component tied-covariance GMM on each machine's own scaled_diff training features, standardized per machine. The anomaly score is the negative log-likelihood under the machine-specific normal model.

3. EXPERIMENTS

3.1. Dataset and Setup

Experiments use the DCASE 2026 Task 2 development set (7 machines) and evaluation set (5 machines). Hyperparameters were selected on the development set only; evaluation labels were never used.

3.2. Hyperparameters

All GMMs use tied covariance, reg_covar=5e-4, max_iter=300, tol=1e-3, random_state=42. Machine GMMs use 130 components; the UBM background uses 512. Each machine has its own StandardScaler. Decision thresholds are set to the median score of normal training clips per machine. The source code is not publicly released.

3.3. Development Results

Tables 3–5 report per-machine development-set results. The official overall score is the harmonic mean of the 21 per-machine values (AUC_src, AUC_tgt, pAUC for each of the 7 development machines). System 1 and System 3 achieved overall scores of 64.78% and 62.65%, respectively; System 2 reached 65.08%.

Table 3. Per-machine development results for System 1.

Machine	Src	Tgt	pAUC
ToyCar	75.0	75.5	55.1
ToyCarEmu	63.4	86.6	58.2
bearingEmu	63.0	61.8	60.0
fan	94.0	69.5	61.4
gearboxEmu	61.1	59.1	54.0
sliderEmu	62.6	56.6	50.7
valveEmu	94.9	76.9	61.7

Table 4. Per-machine development results for System 2.

Machine	Src	Tgt	pAUC
ToyCar	74.3	75.4	54.9
ToyCarEmu	63.1	86.7	58.3
bearingEmu	64.4	63.6	59.8
fan	93.7	67.7	59.6
gearboxEmu	62.6	59.1	54.2
sliderEmu	63.0	60.5	51.0
valveEmu	95.2	76.6	62.2

Table 5. Per-machine development results for System 3.

Machine	Src	Tgt	pAUC
ToyCar	74.6	74.7	54.8
ToyCarEmu	63.2	86.0	58.2
bearingEmu	64.2	61.6	59.9
fan	89.4	51.6	55.5
gearboxEmu	60.2	61.8	54.3
sliderEmu	56.7	52.9	49.1
valveEmu	93.8	75.5	61.1

4. CONCLUSION

We presented three handcrafted systems for DCASE 2026 Task 2. Scaled-difference channel compensation and the UBM likelihood-ratio framework address domain shift effectively. The best development result of 65.08% was obtained by enriching the UBM background with development data.

5. REFERENCES

[1] T. Nishida, N. Harada, D. Takeuchi, D. Niizumi, K. Imoto, K. Dohi, H. Purohit, R. Tanabe, T. Endo, M. Yamamoto, Y. Nikaido, and Y. Kawaguchi, "Description and Discussion on DCASE 2026 Challenge Task 2: Noise-aware Unsupervised Anomalous Sound Detection for Machine Condition Monitoring," 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)*, 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 *Proc. 7th Detection and Classification of Acoustic Scenes and Events 2022 Workshop (DCASE2022)*, Nancy, France, Nov. 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.
[5] NTT CSLab, "dcase2023_task2_baseline_ae," GitHub repository, 2026. https://github.com/nttcslab/dcase2023_task2_baseline_ae
[6] DCASE Community, "DCASE 2026 Challenge Submission," 2026. http://dcase.community/challenge2026/submission