

# 音の差分説明を用いた機械の異常原因推定

前橋 祐斗<sup>1</sup> 国定 恭史<sup>1</sup> 藤枝 大<sup>1</sup>

<sup>1</sup> 沖電気工業株式会社

{maehashi737,kunisada564,fujieda424}@oki.com

## 概要

工場における機械の異常検知では、動作音から正常か異常かを判定する手法がしばしば用いられる。しかし、従来の異常検知手法では、異常と判定した根拠や異常原因の説明を行うことはできない。そこで本研究では、Audio-Language Model による音の差分説明と、大規模言語モデルの知識を活用した異常原因の推定手法を提案する。提案手法の有効性を検証するため、新規に作成した異常原因ラベル付きデータセットによる評価を行った。また、異常原因推定の成功例と失敗例を分析し、課題を明らかにした。

## 1 はじめに

工場における機械の異常検知は、安全性や品質管理の観点から重要である。異常検知手法では、機械の動作音がしばしば用いられる。動作音は、非接触で取得可能であり、センサ設置のコストが低いという利点がある。また、動作音には異常の兆候が含まれていることが多く、異常検知に有用である。しかし、従来の動作音を用いた異常検知手法は、正常か異常かを判定するのみであり、異常と判定した根拠や異常原因の説明を行うことはできない。これは、機械学習による異常検知手法がブラックボックス化しやすいことや、異常原因に関するラベル付きデータが不足していることが主な要因であると考えられる。工場での異常検知において、異常の根拠や原因に関する説明が無い場合、現場作業者が適切な対処を行うことが困難になる。

そこで本研究では、正常音と異常音の差分説明から異常原因の推定を行う手法を提案する。提案手法では、Audio-Language Model を用いて、正常音と異常音の差分説明を生成することにより、非専門家でも理解しやすい表現で異常の根拠を提示する。さらに、音の差分説明を大規模言語モデル (LLM) に入力することで、LLM の知識を活用した異常原因推

定を行う。提案手法の有効性を検証するため、新規に作成した異常原因ラベル付きデータセットによる評価を行った。

## 2 関連研究

### 2.1 音響データの異常検知

音響データを用いた異常検知は広く研究されており、様々な手法が提案されている。autoencoder による再構成誤差を用いた手法 [1] では、正常音のみで学習を行い、学習していない異常音の再構成誤差が大きくなることを利用して、異常検知を行う。また、近年では事前学習済みモデルを用いた異常検知手法が盛んに提案されている [2]。

### 2.2 Audio-Language Model

Audio-Language Model (ALM) は、音とテキストを統合的に扱うモデルであり、代表的なものとして CLAP [3] がある。CLAP は、音とテキストのペアデータを用いた対照学習によって、2つのモダリティを共通の埋め込み空間で紐づけることができる。これにより、強力な zero-shot 能力と、下流タスクへの適応能力を獲得している。さらに近年では、LTU [4]、Qwen-audio [5]、Audio Flamingo [6] などの、音とテキストを入力可能なマルチモーダル LLM も提案されている。これらのモデルは、LLM の強力な言語理解能力を音声理解に活用することで、様々な音声タスクで高い性能を示している。

### 2.3 Audio Difference Captioning

Audio Difference Captioning (ADC) [7] は、類似しているがわずかに異なる音のペアに対して、その意味的な違いを説明するタスクである。従来の Audio Captioning タスクでは、類似した音に対して似たような説明が生成されてしまい、内容の違いが十分に説明されないという課題があり、ADC はこれを解決するために提案された。近年では、LLM を用い

表 1 異常原因推定に用いたプロンプト

## 指示
ギアボックスの正常音 (Audio 1) と異常音 (Audio 2) の違いを説明した文から、異常原因を推定してください。
なお、異常原因は以下のいずれかです。
1. ギア欠け
2. 潤滑油無し
3. ギア欠けかつ潤滑油無し
## 回答形式
推定した異常原因および推定理由を回答してください。
推定理由は簡潔に説明してください。
異常原因:
推定理由:

て、詳細度の異なる 3 段階の差分説明を生成する手法である ADIFF [8] が提案されている。また、国定ら [9] は異常検知におけるドメイン依存性の課題に着目し、音の差分説明を用いた異常検知手法を提案している。

### 3 提案手法

本研究では、正常音と異常音の差分説明を生成し、その内容に基づいて異常原因の推定を行う手法を提案する。提案手法の概要を図 1 に示す。提案手法は、以下の 2 つのステップで構成される。

**正常音と異常音の差分説明** ALM を用いて、正常音と異常音の差分説明を生成する。本研究では ALM として、Qwen2-Audio-7B-Instruct<sup>1)</sup> [10] を用いた。

**異常原因の推定** ALM が生成した差分説明を LLM に入力し、LLM の知識を活用した異常原因推定を行う。本研究では LLM として、GPT-4o [11] を用いた。LLM に入力したプロンプトを表 1 に示す。プロンプトでは、異常原因の候補を列挙し、差分説明に基づいた異常原因推定と、推定理由の説明を行うように指示している。

なお、本研究では ALM は正常音と異常音の差分説明の生成に特化し、異常原因の推定には、より強力な推論能力を持つ LLM を用いる構成とした。

## 4 実験

### 4.1 データセット

1) <https://huggingface.co/Qwen/Qwen2-Audio-7B-Instruct>

表 2 データ拡張の一覧

データ拡張	処理内容
AddBackgroundNoise	別の音を追加
AddGaussianNoise	ガウシアンノイズを追加
PitchShift	ピッチを変更
TimeStretch	再生速度を変更
Gain	ボリュームを変更

**学習データセット** ALM が生成する差分説明の品質向上を目的として、先行研究 [9] と同様の手順で差分説明データセットを作成した。データセットの作成には、DCASE 2022 Challenge Task 2 [12] で提供された機械音データセット (以下、DCASE データセット) を用いた。DCASE データセットには、7 種類の機械音 (Fan, Gearbox, Bearing, Slide rail, Toy car, Toy train, Valve) が含まれており、これら全ての機械音を使用した。

差分説明データセットは、2 つの音のペアとそれらの差分説明テキストから構成される。音のペアと差分説明テキストの作成方法を図 2 に示す。DCASE データセットに含まれる各機械音に対して、ランダムにデータ拡張を適用し、元の音と拡張後の音をペアとした。データ拡張の一覧を表 2 に示す。データ拡張には、Python ライブラリの audiomentations<sup>2)</sup> を用いた。AddBackgroundNoise では、ESC-50 データセット [13] の環境音を別の音として追加した。音のペアに対する差分説明テキストは、データ拡張の内容に基づいて、GPT-4o を用いて生成した。

以上の手順で作成した差分説明データセットを、学習データ 33600 件、検証データ 8400 件、評価データ 8400 件に分割し、LoRA [14] による ALM のファインチューニングを行った。ファインチューニングでは、2 つの音のペアと、それらの違いの説明を指示するプロンプトを ALM に入力し、差分説明テキストを正解文として出力するように学習を行った。

**評価データセット** 提案手法の評価には、新規に作成したギアボックスの動作音データセット (以下、ギアボックスデータセット) を用いた。ギアボックスデータセットは、38 件の正常音と異常音のペアから構成される。これらの音は、市販のギアボックスを雑音と反響の少ない部屋で動作させて、マイクで録音した。異常音は、正常なギアボックスに対して、以下のいずれかの異常を付与したときの動作音である。

- ギア欠け (12 件)

2) <https://github.com/iver56/audiomentations>

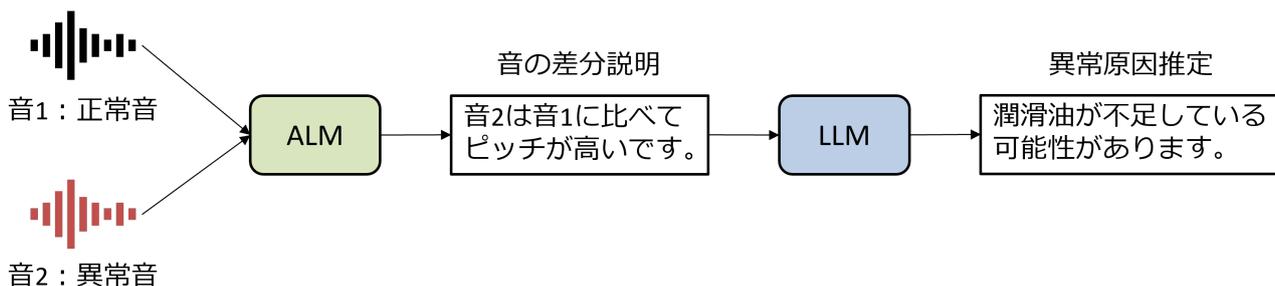


図1 提案手法の概要

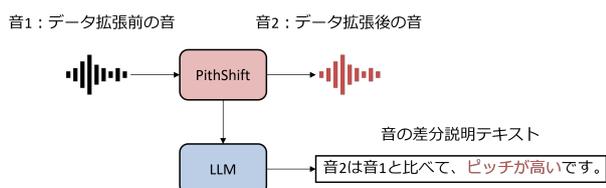


図2 差分説明データセットにおける音のペアと差分説明テキストの作成方法

- 潤滑油無し（14件）
- ギア欠けかつ潤滑油無し（12件）

ギア欠けは、ギアの一部の連続する2か所を削り取った状態である。また、潤滑油を塗ったものを正常、潤滑油を塗っていないものを異常とした。

## 4.2 評価方法

まず、差分説明データセットでファインチューニングした ALM の性能を確認するため、同データセットの評価用データを用いて評価を行う。評価指標として、BLEU [15] と LLM-as-a-Judge [16] を用いる。LLM-as-a-Judge では、GPT-4o を用いて、ALM が生成した差分説明と正解文の内容の一致度を以下の5段階で評価する。

1. 正解文の内容と全く一致しない
2. 正解文の内容とほとんど一致しない
3. 正解文の内容と部分的に一致する
4. 正解文の内容とほとんど一致する
5. 正解文の内容と完全に一致する

次に、ギアボックスデータセットを用いて、LLM による異常原因推定の評価を行う。評価指標として、異常原因の正解率を用いる。

## 4.3 実験結果

**ALM のファインチューニング** 差分説明データセットによってファインチューニングした ALM の評価結果を表 3 に示す。ALM が生成した差分説明

表3 差分説明データセットによってファインチューニングした ALM の評価結果

	BLEU-4	LLM-as-a-Judge
	0.331	3.98

表4 LLM による異常原因推定の評価結果

異常原因	正解件数/全件数	正解率
ギア欠け	5/12	0.417
潤滑油無し	14/14	1.00
ギア欠けかつ潤滑油無し	0/12	0.00
全体	19/38	0.500

が、正解文と単語レベルおよび意味的な観点で十分に一致することが確認できた。

**LLM による異常原因推定** LLM による異常原因推定の評価結果を表 4 に示す。潤滑油無しのデータは全て正しく推定できた一方で、ギア欠けは半数以上のデータで誤推定となった。また、ギア欠けかつ潤滑油無しのデータは全て誤推定となったが、これは ALM が生成した差分説明が1つの観点に着目したのようになっており、複数の異常原因を示唆する情報が含まれていなかったためと考えられる。

LLM による異常原因推定の成功例を表 5 に示す。正解の異常原因がギア欠けの場合に、ALM が生成した差分説明では、時計の針が刻むような音が特徴として挙げられており、LLM はこの情報に基づいて異常原因を正しく推定している。また、正解の異常原因が潤滑油無しの場合に、差分説明では音量の増加が特徴として挙げられており、LLM は同様に異常原因を正しく推定している。

LLM による異常原因推定の失敗例を表 6 に示す。これらの例では、同じギア欠けの異常音に対して、ピッチが高い、またはピッチが低いという正反対の差分説明が ALM によって生成されている。そして、LLM はどちらの場合においても、潤滑油無しと誤推定している。さらに、推定理由もそれぞれの例で

表5 LLMによる異常原因推定の成功例

正解の異常原因	ALMによる音の差分説明	LLMによる異常原因推定
ギア欠け	Audio 2 has an additional sound of a clock ticking in the background compared to Audio 1.	異常原因：ギア欠け 推定理由：ギア欠けがあると、ギアが回転する際に不規則な衝撃が発生し、それが時計の針が刻むような周期的な音として聞こえることがあります。このため、Audio 2で聞こえる時計の針の音はギア欠けによるものと推定されます。潤滑油の不足は通常、摩擦音や高音の金属音を引き起こしますが、時計の針のような音とは異なります。
潤滑油無し	Audio 2 sounds significantly louder than Audio 1.	異常原因：潤滑油無し 推定理由：潤滑油が不足していると、摩擦が増加し、ギアボックスの動作音が通常よりも大きくなる場合があります。音量の増加は潤滑油の不足を示唆しています。ギア欠けの場合は、音の質が変わることが多いですが、音量の増加が主な特徴ではありません。

表6 LLMによる異常原因推定の失敗例

正解の異常原因	ALMによる音の差分説明	LLMによる異常原因推定
ギア欠け	Audio 2 sounds slightly higher in pitch compared to Audio 1.	異常原因：潤滑油無し 推定理由：潤滑油が不足していると、摩擦が増加し、ギアの回転がスムーズでなくなるため、音のピッチが高くなる場合があります。ギア欠けの場合は、通常、異音や不規則な音が発生することが多いですが、ピッチの変化は潤滑油の不足に起因することが多いです。
ギア欠け	Audio 2 sounds slightly lower in pitch compared to Audio 1.	異常原因：潤滑油無し 推定理由：潤滑油が不足していると、摩擦が増加し、ギアの回転がスムーズでなくなるため、音のピッチが低くなる場合があります。ギア欠けの場合は、通常、異常音として高いピッチや不規則な音が発生することが多いです。したがって、音のピッチが低くなるという特徴は潤滑油の不足に起因する可能性が高いです。

一貫しておらず、矛盾した説明となっている。これは、LLMが異常原因を推定する際に、根拠となる知識が明示的に与えられていなかったため、もっともらしい推定理由を生成してしまったことが原因であると考えられる。以上の結果から、ALMが生成する差分説明と、LLMの知識に基づく推定の両方に課題があることを確認した。

## 5 おわりに

本研究では、音の差分説明を用いた機械の異常原因推定手法を提案した。実験の結果、ALMが生成した差分説明に基づき、LLMが自身の知識を活用して異常原因を推定できた例を確認した。一方で、もっともらしい推定理由に基づいて誤推定する例も見られた。今後の課題として、ALMが生成する差分説明の品質向上や、LLMへの異常原因に関する知識の追加が挙げられる。

## 参考文献

- [1] Yuma Koizumi, Yohei Kawaguchi, Keisuke Imoto, Toshiki Nakamura, Yuki Nikaido, Ryo Tanabe, Harsh Purohit, Kaori Suefusa, Takashi Endo, Masahiro Yasuda, et al. Description and discussion on dcase2020 challenge task2: Unsupervised anomalous sound detection for machine condition monitoring. **arXiv preprint arXiv:2006.05822**, 2020.
- [2] Lei Wang. Pre-trained model enhanced anomalous sound detection system for dcase2025 task2. Technical report, DCASE2025 Challenge, June 2025.
- [3] Benjamin Elizalde, Soham Deshmukh, Mahmoud Al Ismail, and Huaming Wang. Clap learning audio concepts from natural language supervision. In **ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)**, pp. 1–5. IEEE, 2023.
- [4] Yuan Gong, Hongyin Luo, Alexander H Liu, Leonid Karlinsky, and James Glass. Listen, think, and understand. **arXiv preprint arXiv:2305.10790**, 2023.
- [5] Yunfei Chu, Jin Xu, Xiaohuan Zhou, Qian Yang, Shiliang

- Zhang, Zhijie Yan, Chang Zhou, and Jingren Zhou. Qwen-audio: Advancing universal audio understanding via unified large-scale audio-language models. **arXiv preprint arXiv:2311.07919**, 2023.
- [6] Zhifeng Kong, Arushi Goel, Rohan Badlani, Wei Ping, Rafael Valle, and Bryan Catanzaro. Audio flamingo: A novel audio language model with few-shot learning and dialogue abilities. **arXiv preprint arXiv:2402.01831**, 2024.
- [7] Daiki Takeuchi, Yasunori Ohishi, Daisuke Niizumi, Noboru Harada, and Kunio Kashino. Audio difference captioning utilizing similarity-discrepancy disentanglement. In **Proceedings of the 8th Detection and Classification of Acoustic Scenes and Events 2023 Workshop (DCASE2023)**, pp. 181–185, Tampere, Finland, September 2023.
- [8] Soham Deshmukh, Shuo Han, Rita Singh, and Bhiksha Raj. ADIFF: Explaining audio difference using natural language. In **The Thirteenth International Conference on Learning Representations**, 2025.
- [9] 国定恭史, 前橋祐斗, 藤枝大. 音響データの差分説明を利用した汎用型異常検知モデルの提案. 研究報告音声言語情報処理 (SLP), Vol. 2025-SLP-158, No. 20, pp. 1–6, 2025.
- [10] Yunfei Chu, Jin Xu, Qian Yang, Haojie Wei, Xipin Wei, Zhifang Guo, Yichong Leng, Yuanjun Lv, Jinzheng He, Junyang Lin, et al. Qwen2-audio technical report. **arXiv preprint arXiv:2407.10759**, 2024.
- [11] Aaron Hurst, Adam Lerer, Adam P Goucher, Adam Perelman, Aditya Ramesh, Aidan Clark, AJ Ostrow, Akila Welihinda, Alan Hayes, Alec Radford, et al. Gpt-4o system card. **arXiv preprint arXiv:2410.21276**, 2024.
- [12] Kota Dohi, Keisuke Imoto, Noboru Harada, Daisuke Niizumi, Yuma Koizumi, Tomoya Nishida, Harsh Purohit, Ryo Tanabe, Takashi Endo, Masaaki Yamamoto, and Yohei Kawaguchi. Description and discussion on DCASE 2022 challenge task 2: Unsupervised anomalous sound detection for machine condition monitoring applying domain generalization techniques. In **Proceedings of the 7th Detection and Classification of Acoustic Scenes and Events 2022 Workshop (DCASE2022)**, pp. 1–5, Nancy, France, November 2022.
- [13] Karol J. Piczak. ESC: Dataset for Environmental Sound Classification. In **Proceedings of the 23rd Annual ACM Conference on Multimedia**, pp. 1015–1018. ACM Press.
- [14] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, Weizhu Chen, et al. Lora: Low-rank adaptation of large language models. **ICLR**, Vol. 1, No. 2, p. 3, 2022.
- [15] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a method for automatic evaluation of machine translation. In **Proceedings of the 40th annual meeting of the Association for Computational Linguistics**, pp. 311–318, 2002.
- [16] Lianmin Zheng, Wei-Lin Chiang, Ying Sheng, Siyuan Zhuang, Zhanghao Wu, Yonghao Zhuang, Zi Lin, Zhuohan Li, Dacheng Li, Eric Xing, et al. Judging llm-as-a-judge with mt-bench and chatbot arena. **Advances in neural information processing systems**, Vol. 36, pp. 46595–46623, 2023.