

Title	音色関連特徴量に着目した教師なし機械学習による異常音検知
Author(s)	小倉, 稜也
Citation	
Issue Date	2025-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/19788
Rights	
Description	Supervisor: 鶴木 祐史, 先端科学技術研究科, 修士 (情報科学)

修士論文

音色関連特徴量に着目した教師なし機械学習による異常音検知

小倉 稜也

主指導教員 鵜木 祐史

北陸先端科学技術大学院大学
先端科学技術研究科
(情報科学)

令和7年3月

Abstract

Experienced factory inspectors can identify abnormalities in equipment by listening to the operating sounds of the machines. However, due to the aging workforce, there is a lack of successors. Therefore, investigating computer-based anomalous sound detection (ASD) is necessary. ASD system consists of a feature extractor and a discriminator. The feature is mainly log Mel-spectrogram. The discriminators primarily use deep learning models such as autoencoder (AE) and convolutional neural network (CNN). Recent research has proposed methods using timbral features for ASD. Ota & Unoki reported high performance using acoustical features related to timbral attributes. However, this method employs supervised learning, where anomalous sounds are used together with normal sounds. Covering all anomalous sounds for learning is challenging, and collecting sufficient anomaly samples proves difficult. Therefore, unsupervised learning that learns only normal sounds is desirable. This paper proposes an ASD system based on unsupervised learning using acoustical features related to timbral attributes, addressing three limitations in Ota & Unoki’s previous method.

First, discriminators other than support vector machines (SVM) are explored. Better models than SVM for acoustical features related to timbral attributes could improve performance. However, these features are 6-dimensional or 7-dimensional data. Deep learning models such as CNNs typically handle high-dimensional data. These models are difficult to apply to low-dimensional features, such as acoustical features related to timbral attributes. Therefore, the investigation focused on supervised and unsupervised learning models that can handle relatively low-dimensional data. For supervised learning, models such as LightGBM and TabPFN were investigated. These models handle tabular data. For unsupervised learning, outlier detection methods such as the gaussian mixture model (GMM) and k-Nearest Neighbors (k-NN) were investigated. The models were evaluated using MIMII dataset employed in previous method. MIMII dataset contains four Machine Types: Fan, Pump, Slider, and Valve. Each Machine Type contains four Machine ID sounds. The evaluation results showed that TabPFN achieved the best performance, surpassing the previous method. GMM demonstrated the best performance in unsupervised learning.

The second limitation addresses performance degradation under low signal-to-noise ratio (SNR) conditions. In industrial equipment ASD, sounds from surrounding machines become environmental noise. Therefore, robust system performance is required even under low SNR conditions. The performance of the previous method significantly deteriorates as machine noise SNR decreases. The previous method employ the average of acoustical features related to timbral attributes calculated from multiple frames of machine sound. However, the frames are averaged

in non-stationary machine sounds, including those containing only environmental noise without machine sounds. This averaging significantly impacts performance for non-stationary sounds. Therefore, a feature extraction method called Timbral frames selection (TFS) was developed as a robust feature extraction for low SNR conditions. This method selects frames only containing non-stationary machine sound events. Evaluation of the MIMII dataset showed that TFS improved the performance of non-stationary sounds such as Slider and Valve. The effect of TFS was particularly significant under low SNR conditions.

The third limitation is implementing unsupervised learning. The previous method used supervised learning, requiring normal and anomalous sounds. However, collecting anomalous sounds for learning is challenging due to their rare occurrence in machines. This necessitates an unsupervised machine-learning system capable of learning from normal sounds alone. The solution estimates the distribution of anomalous sounds using various machine sounds prepared in advance. The anomalous sound distribution is defined as the distribution obtained by subtracting the normal sound distribution of the target machine from the distribution of various machine sounds. The estimated anomalous sound distributions were then sampled to generate pseudo-anomalous sound data. This approach enables model training with authentic normal sounds and pseudo-anomalous sounds. It satisfies unsupervised learning conditions while enabling supervised learning.

The proposed system incorporates all three solutions described above. The evaluation used the DCASE 2020 Task 2 development dataset, which includes six Machine Types: Fan, Pump, Slider, Valve, ToyCar, and ToyConveyor. The area under the curve (AUC) was used as the evaluation metric. These results showed that the proposed system using TabPFN achieved higher overall AUC than the AE baseline. The AUC scores for Fan and Valve improved by over 20%. However, the ToyConveyor system exhibited decreased AUC. Therefore, further investigation of timbral features is necessary.

In conclusion, the three solutions presented effectively address limitations in ASD systems based on unsupervised learning using acoustical features related to timbral attributes. The proposed system demonstrates superior discriminative performance compared to baseline methods. Future work should address domain generalization and first-shot anomalous sound detection.

目次

第1章 序論	1
1.1 はじめに	1
1.2 研究背景	2
1.3 研究目的	3
1.4 本論文の構成	3
第2章 関連研究	6
2.1 DCASE	6
2.2 深層学習からのアプローチ	7
2.3 聴知覚からのアプローチ	8
2.4 音色関連特徴量による異常音検知	10
2.4.1 Ota & Unoki の方法の概要	10
2.4.2 問題点	11
第3章 音色関連特徴量に適した識別器の選択	13
3.1 教師あり学習	13
3.2 教師なし学習	14
3.3 実験条件	15
3.4 実験結果	19
第4章 低 SNR 条件下においても機能する特徴量抽出法	22
4.1 異常音検知における環境騒音	22
4.2 音色関連特徴量の時系列変化	22
4.3 音色特徴の選択処理	26
4.4 実験条件	27
4.5 実験結果	27
第5章 教師なし学習による異常音検知	31
5.1 擬似異常音データを使用した異常音検知	31
5.2 擬似異常音データの生成	32
5.3 実験条件	34
5.4 実験結果	34

第 6 章	提案システムの総合評価と考察	36
6.1	提案システムの概要	36
6.2	実験条件	36
6.3	実験結果	38
6.4	考察	38
第 7 章	結論	42
7.1	本論文における成果	42
7.2	今後の展望	42
付 録 A	機械音の環境騒音の除去	43
A.1	実験条件	43
A.2	実験結果	44
謝辞		48
参考文献		49
研究業績		55

目 次

1.1	計算機による異常音検知のフロー	4
1.2	本論文の構成	5
4.1	MIMII dataset に含まれる機械音の対数メルスペクトログラム . . .	24
4.2	Valve における音色関連特徴量の時系列変化	25
4.3	音色特徴の選択処理のフロー	28
5.1	異常音の分布の推定	33
6.1	提案システムの学習フロー	37
6.2	音色関連特徴量の分布と各サンプルの異常スコア	41

表 目 次

2.1	音色関連特徴量を構成する指標	12
3.1	MIMII dataset の構成	17
3.2	各 Machine Type の異常音判別に使用する音色関連特徴量の指標 . .	18
3.3	MIMII dataset (SNR 6 dB) における各モデルの F 値評価	20
3.4	MIMII dataset の全 SNR 条件における各モデルの AUC [%] 評価 . .	21
4.1	各 Machine Type, Machine ID において TFS によるフレーム選択 が行われた指標	29
4.2	MIMII dataset における TFS を適用した場合の各モデルの AUC [%] 評価	30
5.1	擬似異常音データを学習した場合の AUC [%] 評価	35
6.1	提案システムの AUC [%] 評価	40
A.1	NRR [dB] による評価	45
A.2	NRR-TRF による評価	46
A.3	騒音除去前後のデータによる AUC [%] 評価	47

略語一覧

略語	正式名称
MFCC	Mel-frequency cepstrum coefficients
IM	inlier modeling
OE	outlier exposure
ROC	receiver operating characteristic
AUC	area under the curve
AS	振幅強調 shimmer amplified shimmer
APF	振幅強調優勢周波数 amplified predominant frequency
SNR	信号対雑音比 signal-to-noise ratio
SVM	サポートベクターマシン support vector machine
AE	オートエンコーダ autoencoder
CNN	畳み込みニューラルネットワーク convolutional neural network
k-NN	k 近傍法 k-nearest neighbor
GMM	混合ガウスモデル gaussian mixture model
OC-SVM	one-class support vector machine
Deep SVDD	Deep Support Vector Data Description
SFS	spectrogram frames selection
TFS	timbral frames selection
WF	ウィーナーフィルタ Wiener filtering
SS	スペクトルサブトラクション spectral subtraction
NRR	雑音抑圧量 noise reduction rate

第1章 序論

1.1 はじめに

人間の生活において、風の音や電話の着信音、ドアのノック音といった様々な環境音が発生する。環境音分析は、音声や音楽に限定せず、より多様な音を分析する技術のことである [1]。環境音分析の主なタスクとして音響シーン分類、音響イベント検出、音響キャプション生成があり、主に機械学習による方法が数多く研究されている。音響シーン分類では、対象となる音の音響シーンの推定を行う。音響シーンは、公園や電車といった音が収録された場所などを指す。音響イベント検出では、対象の音に存在する音響イベントの種類や、イベントがいつ発生したかを推定する。音響イベントとは環境音の種類のことであり、家庭内の食器の音などが該当する。音響キャプション生成では、環境音を説明する文章を生成する。これらの環境音分析技術は、室内の高齢者の見守りシステムや自動運転といった様々な応用が考えられる。

環境音分析のタスクの1つに、異常音検知がある。異常音検知は、対象の環境音が異常な状態かを判別するタスクである [2]。例えば、公共空間において、爆発音や銃声などといった通常の生活では発生しない異常音を検出する技術が検討されている。また、家庭内の転倒音や叫び声の検出についても研究されている。これら防犯やホームセキュリティに関する事例以外にも、医療分野における肺音や心音の異常検知など、異常音検知技術は幅広い分野へ応用されている。

近年では、産業機器を対象にした異常音検知技術の検討が行われている [1]。第4次産業革命において、人工知能を用いた異常音検知技術による機械故障の自動検出は不可欠である。2019年に機械用の異常音検知データセットが公開されて以降、数多くのシステムが提案されている。

1.2 研究背景

工場の安全な稼働を維持する上で、機械の設備点検業務は非常に重要である。通常、設備点検は熟練点検員が機械の稼働音を聴くことで行われる。熟練点検員は機械音のわずかな変化を聴き取ることができるため、機械音からその機械が異常な状態かどうか判断可能である [3]。しかし、熟練点検員の高齢化に伴う後継者不足といった問題があり、人の手に替わる設備点検システムが求められている。

以上の課題を踏まえ、現在、計算機による異常音検知システムの検討が盛んに行われている [1]。計算機の導入により、点検員に頼ることなく安定した検査が可能になると考えられる。異常音検知システムは、図 1.1 のように特徴抽出部と識別器という 2 つの機構から成る。特徴抽出部は入力信号から特徴量を抽出する。識別器は、特徴量を基に異常スコアを出力する。異常スコアは入力信号が異常であるほど大きくなる。そのため、異常スコアが十分に大きい場合、入力信号を異常とみなすことができる。特徴量には、主に MFCC (Mel-frequency cepstrum coefficient) や対数メルスペクトログラムが使用される。識別器には、autoencoder などの深層学習モデルが用いられる。

近年、機械音の音色の違いによる異常音検知システムについて検討されている [4, 5]。音の知覚的属性はラウドネス、ピッチ、音色の 3 つに分類され [6]、音色はラウドネスやピッチに比べ多次元的な性質を持つ [7]。熟練点検員は、機械音の複雑な音色の違いを聴き取っていると考えられる。そのため、異常音検知技術に音色を用いることで、熟練点検員の感覚に近いシステムの構築が期待できる。また、対数メルスペクトログラムといった従来の特徴量を用いるシステムに比べ、人間が判別ロジックの理解が容易になると考えられる。Ota & Unoki は、音色に係る複数の指標を音色関連特徴量として用いる方法を提案している [8]。この方法では識別器にサポートベクターマシン (support vector machine: SVM) が採用されており、産業機器データセットによる評価において高い異常音判別性能を示している。

一方、音色による異常音検知の課題として、正常音と異常音を両方学習に使う教師あり機械学習が行われていることが挙げられる。教師あり学習では、正常音と異常音の違いを学習することができる。しかし、機械の異常音の発生は稀であるため、学習に十分な数の異常音の収集は難しい。そのため、異常音を使用せずに学習する教師なし機械学習によるシステムが求められる。また、音色によるシステムにおいて、従来の方法では SVM 以外のモデルについて未検討である。そのため、音色特徴に有効な性質を持つモデルについて検討する必要がある。

1.3 研究目的

本研究の目的は、音色関連特徴量に着目した教師なし機械学習による異常音検知システムを提案することである。教師なし機械学習により、音色関連特徴量による異常音検知システムを正常音のみで学習することができる。提案システムの構築は3つの手順に分けて行う。まず、音色関連特徴量に適すると考えられる複数のモデルの判別性能を比較し、提案システムに使用するモデルを選択する。次に、環境騒音下においても有効な特徴抽出法として、機械音区間から音色関連特徴量を抽出する機構を導入する。最後に、様々な種類の機械音を基に生成した擬似異常音データを、実異常音の代わりに学習に使用する。

1.4 本論文の構成

本論文は全7章である。また、図 1.2 に本論文の構成を示す。

- 第1章:** 異常音検知の背景や、音色による異常音検知について述べる。また、本研究の目的を述べる。
- 第2章:** 関連研究について述べる。また、音色関連特徴量による異常音検知システムの概要と、従来法に残された3つの課題について述べる。
- 第3章:** 1つ目の課題の解決法として、音色関連特徴量に適する識別器の選択を行う。複数の識別器と、従来法の判別性能の比較を行う。
- 第4章:** 2つ目の課題の解決法として、低 SNR 条件下でも機能する特徴量抽出法を提案する。
- 第5章:** 3つ目の課題の解決法として、擬似異常音データの生成と、それを利用した教師なし機械学習の実装を行う。
- 第6章:** すべての課題解決法を導入した提案システムの評価と考察を行う。
- 第7章:** 結論と今後の展望を述べる。

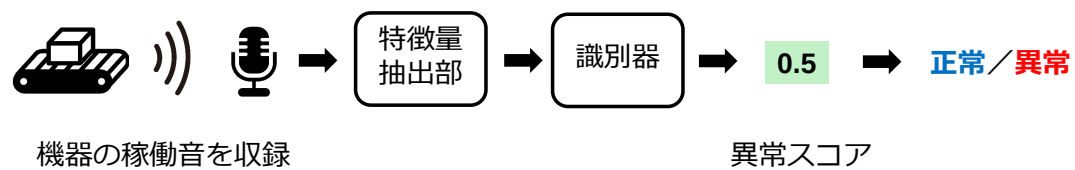


図 1.1: 計算機による異常音検知のフロー

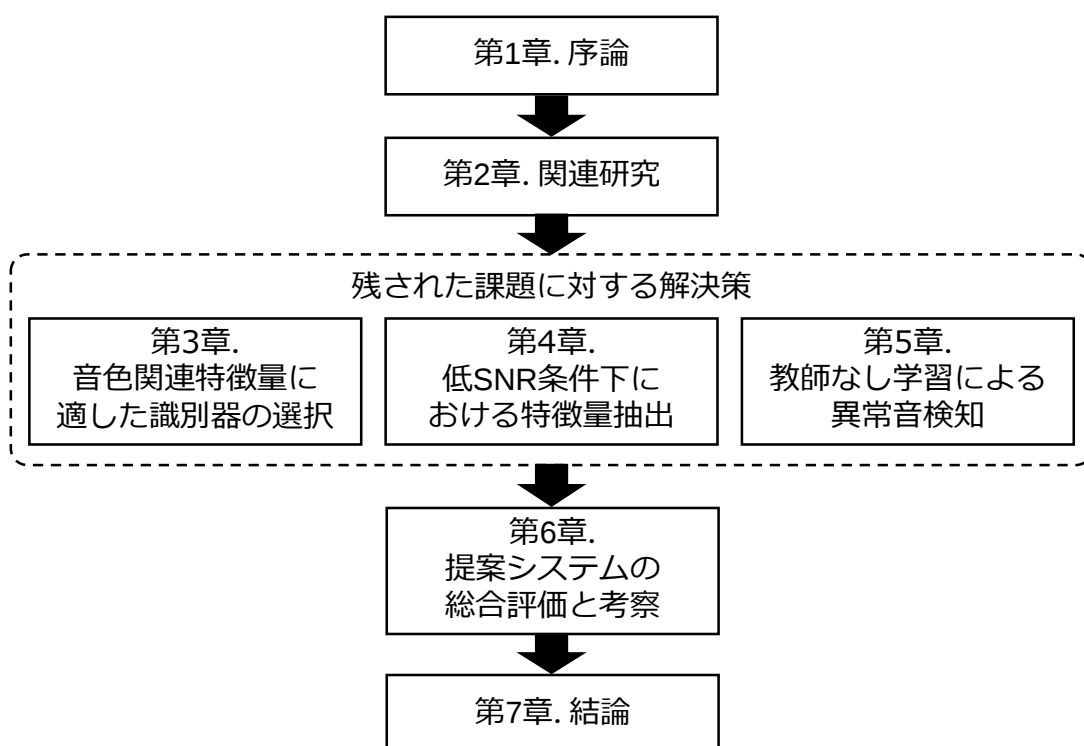


図 1.2: 本論文の構成

第2章 関連研究

2.1 DCASE

DCASE (Detection and Classification of Acoustic Scenes and Events) Challenge は環境音分析の国際コンペティションであり、音響シーン分類や音響イベント検出が扱われる [9]。参加者はそれぞれのタスクについてシステムを作り、性能を競う。

DCASE 2020 Challenge Task 2 は教師なし学習による異常音検知のコンペティションである [10]。DCASE において初めての異常音検知タスクであり、様々なシステムが提案された。DCASE 2021 / 2022 Challenge Task 2 [11, 12] ではドメインシフトへの対応が課題として追加された。ドメインシフトとは、稼働条件や運転条件の変化により、学習データと評価データ間の機械音の分布が変化することである。そのため、ドメインシフト条件下では、機械の正常／異常間の音の違いとドメインシフトによる音の違いを区別することが要求される。さらに、DCASE 2023 / 2024 Challenge Task 2 [13, 14] では First-shot 異常音検知が課題として追加された。DCASE 2022 以前は、構築したシステムを評価するために事前に少量の異常音を与えられていた。そのため、その異常音を使いシステムのパラメータを調整可能であった。しかし、現実的にはパラメータ調整に十分なデータを確保できるとは限らない。First-shot 異常音検知では、評価対象の機械とは異なる機種種の正常／異常音と、評価対象機種種の正常音のみで学習を行う必要があるため、パラメータを調整せずにシステムを構築する必要がある。

DCASE Challenge では、タスクに合わせたデータセットも毎年公開されている。例えば DCASE 2020 では、産業機器の稼働音が収録された MIMII dataset [15] の一部と、おもちゃの稼働音が収録された ToyADMOS [16] の一部で構成されるデータセットが公開された。機種は MIMII dataset に含まれる Fan, Pump, Slider, Valve と、ToyADMOS に含まれる ToyCar, ToyConveyor の 6 機種 (Machine Type) であり、機種ごとに 3 または 4 個体 (Machine ID) が含まれる。

DCASE Challenge では様々なシステムが提案されてきたが、研究分野に特に大きな影響を与えた方法は DCASE 2020 で提案された outlier exposure (OE) である。OE は、外部のデータを異常データとして使用する方法である。従来の異常音検知は外れ値検出によるアプローチが主流であったが、OE の登場により、クラス分類によるアプローチが主流となった。OE の詳細は 2.2 節で解説する。2021 年以降も、学習用データの生成や大規模事前学習モデルの利用など、様々なアプローチが提案されている。

2.2 深層学習からのアプローチ

本節では、教師なし学習による産業機器の異常音検知の動向について説明する。現在までに提案されている異常音検知システムの多くは、深層学習が利用されている。フローは図 1.1 に示した通りである。特徴抽出部は、正常音と異常音の違いをよく捉える特徴量を抽出する。例えば、対数メルスペクトログラムを特徴量として用いることで、機械音の時間周波数情報を利用することができる。識別器は特徴量を基に、入力サンプルの異常スコアを決定する。計算機による異常音検知において、特徴抽出器と識別器のどちらの性能も重要である。

異常音検知において、autoencoder は代表的な深層学習モデルの 1 つであり、DCASE Challenge においてもベースラインとして採用されている [17]。Autoencoder は入力データの次元削減を行う encoder と、encoder の出力から入力データを復元する decoder から構成される。この時、入出力の再構成誤差は、正常音の場合に小さくなり、異常音の場合に大きくなると考えられる。そのため、再構成誤差を異常スコアとして使用することができる。Autoencoder のように、判別対象の機械音のみをモデルの学習に使用した異常音検知法を inlier modeling (IM) と呼ぶ [1]。混合ガウスモデル (gaussian mixture model: GMM) や k-nearest neighbor (k-NN) による外れ値検出も IM の一種である。

IM に対し、DCASE 2020 では判別対象の機械音以外のデータを擬似的な異常クラスとして使用する OE による方法が提案されている [18]。例えば、ある機種特定の Machine ID について判別を行う場合、同じ機種の判別対象ではない Machine ID を異常クラスとしてクラス分類を行うことができる。OE により、識別器に外れ値検出モデルだけではなく、クラス分類モデルも用いることが可能になった。また、OE は IM に比べ高い判別性能を示しており、その後の研究動向に大きく影響した。

現在は、Machine ID 分類などの補助タスクを学習した分類モデルから得られる埋め込みを基に外れ値検出を行う、OE と IM の直列アプローチも盛んに研究されている [19,20]。このアプローチでは、クラス分類モデルが特徴抽出部、外れ値検出モデルが識別器の役割を担う。補助タスクを適切に設定することで、正常／異常が十分に分離された埋め込みを得ることができる。

一方、現実では OE を行うために十分な数の Machine ID や外れ値データがあるとは限らない。DCASE 2023 以降の First-shot 異常音検知では、各 Machine Type につき 1 個体のデータしか利用できない設定になっている。このような問題設定に対して、ID ではなく機械の稼働／非稼働を分類する補助タスクを解く方法 [21] が提案されている。また、機械の運転条件などの属性情報を分類する補助タスクを大規模事前学習モデルにより解く方法 [22] も提案されている。人工的に補助タスク用の擬似ラベルを作り出す、自己教師あり学習による方法もある [23]。なお、DCASE 2024 における 1 位のシステムは、擬似ラベルと大規模事前学習モデルの両方を使用している [24]。

近年では、DCASE の問題設定以外の研究も進められている．例として，クライアントが所持する機械音データを使用して学習する分散協調学習によるシステム [25] や，異常スコアが高いサンプルに人手でラベルを付けるアクティブラーニングによるシステム [26] が挙げられる．また，異常音と正常音の違いを説明する異常音キャプション生成についても検討されている [27]．

深層学習による異常音検知技術は，様々な企業の実務用の異常音検知システムに導入されている．例えば，DCASE Challenge Task 2 の主催である日本電信電話と日立製作所は，それぞれ実環境における産業機器の異常音検知システムの開発を行っている [28, 29]．その他の企業では，画像診断の補助として音による診断を行うシステムなどの開発も行っている [30]．

2.3 聴知覚からのアプローチ

熟練点検者は，機械音の知覚によりその機械が異常かどうか判断する能力に優れている．そのため，深層学習に限らず，人の聴知覚に基づく異常音検知技術も有効だと考えられる．

例えば，ガンマトーン聴覚フィルタバンクの時間変調特徴量を異常音検知に用いる方法が提案されている [31]．この方法では，対数メルスペクトログラムでは捉えること難しい時間情報の抽出に重点を置いている．ガンマトーンスペクトログラムを特徴量として用いる研究もされている [32]．

音質評価指標による異常音検知も研究されている．音質評価指標とは，人間が音を聴いた時の感覚を数値で表現した尺度であり，粗さや鋭さといった指標に対し様々なモデルが提案されている [33]．例えば，ベアリングの故障に対して，ラウドネスや変動強度などの音質評価指標を用いる方法が検討されている [34]．

音質評価指標の一種に，Pearce らの Timbral models がある [35]．Timbral models は，Pearce らが特定した効果音ダウンロードサイトのユーザーにとって有用であると考えられる音色の中から [36]，8 種類の音色を選択してモデル化したものである．8 種類の音色には音の固さ感 (hardness)，音のあたたかさ (warmth)，音の残響感 (reverb) などがあり，残り 5 種類の概要はそれぞれ以下の通りである．

音のブーミング感：boominess

ブーミングは，低音の振動として知覚されることが多い．そのため，低周波数を考慮して，以下の式で定義する．

$$\text{Booming index} = \text{Bandsum} \cdot \frac{S_l}{S_t} \quad (2.1)$$

ただし， Bandsum は $1/3$ オクターブ帯域信号のパワーの合計， S_t は全帯域のラウドネス， S_l は 280 Hz 以下の帯域のラウドネスを示している．Timbral models では，さらに線形回帰を用いて最終的な boominess を求めている．

音の明るさ感：brightness

人間の知覚する音の明るさは、スペクトル重心と、全周波数帯域に対する高周波数帯域の比率と相関がある．そのため、スペクトル重心 SC_{br} と、振幅の比を表す ER_{br} を以下の式で定義する．

$$SC_{br} = \frac{\sum_{k \in \{2,000 \text{ to } 8,000 \text{ Hz}\}} f(k)m(k)}{\sum_{k \in \{2,000 \text{ to } 8,000 \text{ Hz}\}} m(k)} \quad (2.2)$$

$$ER_{br} = \frac{\sum_{k \in \{2,000 \text{ to } 8,000 \text{ Hz}\}} m(k)}{\sum_{k \in \{20 \text{ to } 8,000 \text{ Hz}\}} m(k)} \quad (2.3)$$

ただし、 $f(k)$ は k 番目のビンの周波数、 $m(k)$ は k 番目のビンの振幅を示している． SC_{br} と ER_{br} を基に、最終的な brightness の値を決定する．

音の奥行き感：depth

音色としての奥行き感は、500 Hz 以下の振幅に影響を受ける．そのため、スペクトル重心 SC_{dp} と、振幅の比を表す ER_{dp} を以下の式で定義する．

$$SC_{dp} = \frac{\sum_{k \in \{20 \text{ to } 8,000 \text{ Hz}\}} f(k)m(k)}{\sum_{k \in \{20 \text{ to } 8,000 \text{ Hz}\}} m(k)} \quad (2.4)$$

$$ER_{dp} = \frac{\sum_{k \in \{20 \text{ to } 500 \text{ Hz}\}} m(k)}{\sum_{k \in \{20 \text{ to } 8,000 \text{ Hz}\}} m(k)} \quad (2.5)$$

ただし、 $f(k)$ は k 番目のビンの周波数、 $m(k)$ は k 番目のビンの振幅を示している．基本的な式の構成は brightness と同じだが、計算対象である周波数帯域が異なっている．

音の粗さ：roughness

音のざらざらした印象を表す roughness は、信号を構成する 2 つの正弦波を基に求める．50 ms ごとの roughness を以下の式で定義する．

$$R_{\text{frame}} = X^{0.1} \cdot 0.5(Y^{3.11}) \cdot Z \quad (2.6)$$

ここで、 X は正弦波信号の強さ、 Y は正弦波信号の振幅変動幅、 Z は正弦波信号の振幅変動率によって決定される数値である．

音の鋭さ：sharpness

音の中心周波数の高さに比例して、音の鋭さも大きくなる．中心周波数 1,000Hz, ラウドネスレベル 60 phon の狭帯域雑音の sharpness を 1 acum とした場合, sharpness は以下の式で定義される．

$$S = 0.11 \frac{\int_0^{24 \text{ Bark}} N'(z)g(z)z dz}{\int_0^{24 \text{ Bark}} N'(z) dz} \quad (2.7)$$

ここで, $N'(z)$ は臨界帯域率 z におけるラウドネス密度, $g(z)$ は重みづけ係数である．

Timbral models を異常音検知の特微量として用いた研究はいくつか存在する．例えば, Timbral models に加え, 人間の音声障害に係る特微量によって異常を検出する方法がある [37]．また, Timbral models の正常／異常間の数値の差を, 異常音のキャプション生成に利用する方法も研究されている [38]．

2.4 音色関連特微量による異常音検知

2.4.1 Ota & Unoki の方法の概要

聴知覚からのアプローチの 1 つに, Ota & Unoki の方法 (以後, 従来法と呼ぶ) がある [8]．この方法では, 音質評価指標と短時間音響特微量で構成される音色関連特微量を用いている．また, 識別器には SVM を用いている．

音色関連特微量の各指標を表 2.1 に示す．音質評価指標は, Timbral models に含まれる boominess, brightness, depth, roughness, sharpness の 5 種類である．短時間音響特微量は, 振幅強調 shimmer (amplified shimmer: AS), 振幅強調優勢周波数 (amplified predominant frequency: APF) の 2 種類であり, 音質評価指標を補助する役割を担う．Shimmer は音の振幅の変動を表し, 嗄声の評価に使われる指標である [39]．AS は音の知覚が振幅値に依存することを基に, shimmer を振幅値で強調している．APF は, 音の高さの変動を表す指標である．サブハーモニクス加算法を基に優勢周波数を求め, 振幅値で強調している．従来法では異常音をオノマトペで表現し, 音色関連特微量の中からオノマトペに適する指標を選択している．例えば, 送風機のガタガタという音には roughness, ブーンという音には boominess を割り当てている．

従来法は, MIMII dataset の信号対雑音比 (signal-to-noise ratio: SNR) 6 dB のデータによる F 値評価において 0.920 と高い判別性能を示している．また, 判別における各指標の貢献度を評価することで, 異常音の判別に重要な音色についての分析が可能となっている．

2.4.2 問題点

Ota & Unoki の方法は音色指標を異常音検知に用いることの有効性を示したが、主に 3 つの課題がある。

(1) 音色関連特徴量に適するモデルの検討

識別器として SVM 以外の音色関連特徴量により適するモデルを用いることで、判別性能の向上につながると考えられる。候補として、異常音検知において一般的に使われる autoencoder などの深層学習モデルが考えられる。しかし、これらの深層学習モデルは入力としてスペクトログラムなどの比較的次元数の大きいデータを必要とする。音色関連特徴量は次元の小さいデータであるため、深層学習モデルをそのまま適用することは難しい。

(2) 低 SNR 条件における判別性能の低下

工場では判別対象の機械の周囲にある機械が発する音が環境騒音となるため、産業機器の異常音検知システムは低 SNR 条件下でも頑健であることが求められる。しかし、従来法は SNR 6 dB のデータのみで評価されている。事前検討として、従来法を再現して MIMII dataset の SNR -6 dB のデータで評価したところ、SNR 6 dB の場合に比べ大きく性能が低下した。環境騒音への対策を行い、SNR 条件に頑健なシステムを構築する必要がある。

(3) 教師なし学習の適用

従来法は教師あり学習が行われており、正常／異常音をどちらも学習している。教師あり学習では、正常音と異常音両方の特徴を学習可能である。しかし、異常音の発生は稀であるため、学習用の異常音の収集は難しい。そのため、正常音のみで学習を行う教師なし学習が求められる。教師なし学習によるシステムは、学習していない未知の異常音を検出する必要がある。

音色関連特徴量に着目した教師なし学習による異常音検知システムの構築にあたり、これら 3 つの課題に対応する必要がある。

表 2.1: 音色関連特徴量を構成する指標

指標名	概要
boominess	音のブーミング感
brightness	音の明るさ感
depth	音の奥行き感
roughness	音の粗さ
sharpness	音の鋭さ
振幅強調 shimmer (AS)	音の振幅の変動
振幅強調優勢周波数 (APF)	音の高さの変動

第3章 音色関連特徴量に適した識別器の選択

3.1 教師あり学習

従来法は正常音と異常音を学習に使っており、これは教師あり学習と呼ばれる。教師あり学習による異常音検知は、正常／異常の二値分類タスクとみなすことができる。

例として、パラメータ θ を持つ深層学習モデルによる教師あり学習のフローを説明する。まず、学習用の正常音と異常音を使用して、損失関数 $E(\theta)$ を最小化するようにパラメータ θ を求める。ここで、損失関数 $E(\theta)$ を以下のように定義する。

$$E(\theta) = - \sum_{s=0}^1 \{g_{\mathbf{x}} \log(y_s)\} \quad (3.1)$$

s は 1 が異常、0 が正常を示す。 y_0 と y_1 はそれぞれのクラスに対するモデルの出力値であり、 $y_0 + y_1 = 1$ である。また、 $g_{\mathbf{x}}$ はモデルに入力されたデータ $\mathbf{X}$ に対する正解ラベル (1 または 0) である。次に、学習した θ を基に、異常スコア $A_{\theta}(\mathbf{X})$ を以下の式で求める。

$$A_{\theta}(\mathbf{X}) = \log \frac{y_1}{y_0} \quad (3.2)$$

最後に、サンプルごとに異常スコアと閾値を比較し、閾値を超える場合は異常、それ以外は正常と判別する。

従来法は識別器に SVM を用いており、モデルが異常スコアではなく正常／異常のラベルを直接出力する仕様になっている。一方、SVM 以外にも、クラス分類タスク用の機械学習、深層学習モデルは数多く存在する。そのため、SVM 以外のモデルを使うことで、判別性能が向上する可能性がある。例えば、音響シーン分類タスクでは CNN などの深層学習モデルが使用されている [1]。しかし、深層学習モデルは入力としてスペクトログラムなどの比較的次元数の大きいデータを必要とする。音色関連特徴量は 6 または 7 次元のベクトルであり、スペクトログラムに比べ次元の小さいデータである。そのため、環境音分析において一般的に使わ

れる深層学習モデルを、音色関連特徴量による異常音検知システムに適用することは難しい。

以上のことから、本研究では表データのような低次元データを扱うことのできるモデルに着目した。教師あり学習用のモデルとして、表データの処理に用いられる 3 種類の機械学習または深層学習モデルについて検討を行った。

XGBoost [40]

勾配ブースティング決定木によるモデルである。学習データから決定木を生成し、その出力を基に誤差をより小さくする新たな決定木を生成する。この手順を繰り返してできた複数の決定木の出力を基に、最終的な結果の予測を行う。

LightGBM [41]

XGBoost と同じく、勾配ブースティング決定木によるモデルである。XGBoost に比べ、決定木の葉の数が少なくなるように学習される。そのため、処理速度が速く、大規模なデータの処理に適している。

TabPFN [42]

Transformer [43] を用いた表データ向け深層学習モデルである。小規模な表データしか扱うことができないが、全体的に勾配ブースティング決定木によるモデルを上回る性能を示す。

これらのモデルを従来法の識別器として SVM の代わりに使用し、従来法と同様の手順で学習と評価を行い性能を比較した。

3.2 教師なし学習

教師あり学習の問題設定では、異常音データをモデルの学習に使用できる。しかし、実際の現場において機械の故障は稀であり、学習に十分な数の異常音の収集は難しい。そのため、異常音を使用せずに学習する教師なし学習によるシステムの検討が進められている。現在の異常音検知に関する研究のほとんどは教師なし学習によるものである。DCASE Challenge においても、教師なし学習が条件になる。

例えば、生成モデルを用いて学習用正常音の確率密度関数 $p(x)$ を推定した場合、入力データ $\mathbf{X}$ に対して次式で異常スコアを定義する。

$$A_{\theta}(\mathbf{X}) = -\log p(\mathbf{X}|\theta) \quad (3.3)$$

ここで、 θ はモデルのパラメータ、 $\log p(\mathbf{X}|\theta)$ はパラメータ θ が与えられた場合における $\mathbf{X}$ の対数尤度を示す。異常なサンプルほど、負の対数尤度による異常スコアは大きくなると考えられる。

教師あり学習と同様に、低次元データを扱うことのできるモデルに着目した。調査したモデルは、以下の 4 種類である。

K-nearest neighbor (k-NN)

評価データに対し、最も距離が近い k 個の学習データ（正常）を基に判別を行う。評価データが異常であるほど学習データとの距離は離れると考えられるため、 k 個の学習データとの距離の平均や最大値を異常スコアとして定義できる。

Gaussian mixture model (GMM)

学習データが m 個のコンポーネントで構成される混合正規分布に従うと仮定してモデル化し、モデルが評価データに対して出力する負の対数尤度に基づいて異常の判別を行う。

One-class support vector machine (OC-SVM)

SVM を外れ値検知に応用したモデルである。カーネルトリックにより別空間に射影した学習データの領域を学習し、領域外のデータを異常と判別する。

Deep Support Vector Data Description (Deep SVDD) [44]

OC-SVM を基にした深層学習モデルである。別空間の射影にニューラルネットワークを用いている。

教師あり学習の場合と同様に、これらのモデルを従来法の識別器として使用し、判別性能の比較を行った。ただし、従来法とは違い、正常音のみで学習した。

3.3 実験条件

各モデルの学習と評価には、従来法に則り MIMII dataet [15] を使用した。MIMII dataset の構成を表 3.1 に示す。Fan, Pump, Slider, Valve の 4 機種 (Machine Type) について、それぞれ 4 個体 (Machine ID) の機械音ファイルが含まれる。機械音ファイルの長さは 10 秒、サンプリング周波数は 16 kHz である。また、SNR 6 dB, 0 dB, -6 dB の 3 条件について、それぞれ表 3.1 と同様の構成で収録されている。モデルの学習と評価は SNR 条件, Machine Type, Machine ID ごとに行

い、1 種類のモデルにつき 48 種類 (3 つの SNR 条件 $\times$ 4 機種 $\times$ 4 個体) の学習済みモデルを生成した。正常音と異常音ファイルの各 7 割を学習データ、各 3 割を評価データとしてランダムに分割し、5 回の分割ごとの評価値の平均を最終的な評価値として採用した。ただし、教師なし学習を行う場合は、学習データに含まれる異常音は使用せずに学習を行った。

各 Machine Type における音色関連特徴量の指標の組み合わせを表 3.2 示す。指標の組み合わせは従来法と同様のものである。最終的な指標値は、機械音ファイルを複数のフレームに分割し、各フレームで算出した値の平均を使用した。音質評価指標のフレーム長は 1,024 ms、フレームシフトは 512 ms とした。AS と APF のフレーム長はそれぞれ 256 ms と 128 ms、フレームシフトはそれぞれ 128 ms と 64 ms とした。

GMM のパラメータ m は 10 に設定した。k-NN のパラメータ k は 1 に設定し、データ間の距離はマハラノビス距離で算出した。Deep SVDD のバッチサイズは 512、エポック数は 200 に設定した。また、SVM の異常スコアは従来法に則り、モデルが出力した 0 または 1 のラベルの値とした。

評価指標には、従来法で使用された F 値と、異常音検知で一般的に使用される area under the receiver operating characteristic (ROC) curve (AUC) を使用した。AUC は ROC 曲線の下部面積である。ROC 曲線は、異常スコアの閾値を変化させた場合の真陽性率と偽陽性率の変化を示す [45]。真陽性率は異常サンプルの中からモデルが異常と判別したサンプルの割合を示す。偽陽性率は正常サンプルの中からモデルが異常と判別したサンプルの割合を示す。AUC は以下の式で定義される [10]。

$$\text{AUC} = \frac{1}{N_- N_+} \sum_{i=1}^{N_-} \sum_{j=1}^{N_+} \mathcal{H}(A_{\theta}(\mathbf{X}_j^+) - A_{\theta}(\mathbf{X}_i^-)), \quad (3.4)$$

ただし、 N_- と N_+ はそれぞれ正常、異常サンプルの総数、 $\mathbf{X}_i^-$ と $\mathbf{X}_j^+$ はそれぞれ正常、異常サンプルを示す。関数 $\mathcal{H}(a)$ は $a > 0$ の場合に 1、それ以外の場合は 0 を出力する。AUC の範囲は 0% から 100% までであり、値が大きくなるほど判別性能が良く、50% に近いほど性能が悪い。F 値は適合率 (precision) と再現率 (recall) の調和平均であり、以下の式で定義される。

$$\text{F-measure} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3.5)$$

ただし、Precision はモデルが異常と判別したサンプルに含まれる、実際の異常サンプルの割合を示す。Recall は真陽性率と同義である。F 値は AUC と異なり、算出にあたり異常スコアの閾値を決める必要がある。そのため、ROC 曲線において真陽性率と偽陽性率の差が最も大きい箇所を閾値として採用した。F 値は 0 から 1 までの範囲であり、1 に近いほど判別性能が高いことを示す。

表 3.1: MIMII dataset の構成

Machine Type	Machine ID	正常音数	異常音数
Fan	00	1011	407
	02	1016	359
	04	1033	348
	06	1015	361
Pump	00	1006	143
	02	1005	111
	04	702	100
	06	1036	102
Slider	00	1068	356
	02	1068	267
	04	534	178
	06	534	89
Valve	00	991	119
	02	708	120
	04	1000	120
	06	992	120

表 3.2: 各 Machine Type の異常音判別に使用する音色関連特徴量の指標

	Fan	Pump	Slider	Valve
boominess	✓	✓	-	✓
brightness	✓	✓	✓	✓
depth	✓	-	✓	✓
roughness	✓	✓	✓	✓
sharpness	✓	✓	✓	✓
AS	✓	✓	✓	✓
APF	✓	✓	✓	✓

3.4 実験結果

MIMII dataset (SNR 6 dB) によって F 値評価を行った結果を表 3.3 に示す. 各 Machine Type の評価値は, Machine Type に含まれる全 Machine ID の評価値の平均を示している. SVM の評価値は文献 [8] から引用した. 教師あり学習の場合, 調査したモデルは全体的に SVM を超える判別性能を示しており, 従来法を上回っている. 特に TabPFN は全機種で SVM の評価値を超え, 全機種の平均の F 値は SVM を 0.042 上回る 0.962 となっている. 教師なし学習の場合, 調査したモデルは SVM やその他教師あり学習に使用したモデルの性能には及んでいない. これは, 異常音を学習に使用できない影響によるものだと考えられる. GMM の全機種の平均の F 値は 0.886 であり, 教師なし学習のモデルの中では最も高い判別性能を示している.

表 3.4 に, MIMII dataset の全 SNR 条件のデータにおける AUC 評価を示す. 教師あり学習を行ったモデルは, 教師なし学習用のモデルに比べ高い AUC を示している. また, すべてのモデルにおいて, SNR の低下に伴い判別性能が低下している. 一方, TabPFN は他のモデルに比べ, すべての SNR 条件で全機種の評価値の平均が最も高い. また, SNR -6 dB において, TabPFN は唯一全機種で 90% を超える評価値を示している. 教師なし学習用のモデルについては, F 値評価の場合と同様に GMM が全体的に最も高い性能を示している.

以上の結果から, 表データのような次元の小さいデータを扱うことが可能なモデルが, 音色関連特徴量によるシステムの識別器として有効だとわかった. 特に, 教師あり学習モデルでは TabPFN, 教師なし学習モデルでは GMM が良い性能を示したため, この 2 種類のモデルを第 4, 5, 6 章の実験における識別器として使用する. 一方, 表 3.4 からわかるように, SNR の低下に伴いモデルの種類によらず判別性能が低下する傾向がみられた. これは環境騒音が音色関連特徴量の正常／異常間の分離度に影響したことが原因だと考えられる. 環境騒音に対する解決法は, 次章で検討を行う.

表 3.3: MIMII dataset (SNR 6 dB) における各モデルの F 値評価

		Fan	Pump	Slider	Valve	Avg.
教師あり	SVM [8]	0.985	0.959	0.937	0.798	0.920
	XGBoost	0.979	0.968	0.950	0.873	0.943
	LightGBM	0.984	0.975	0.949	0.879	0.947
	TabPFN	0.994	0.984	0.964	0.908	0.962
教師なし	k-NN	0.956	0.913	0.864	0.751	0.871
	GMM	0.970	0.916	0.882	0.777	0.886
	OC-SVM	0.875	0.738	0.827	0.542	0.746
	Deep SVDD	0.904	0.848	0.837	0.645	0.809

表 3.4: MIMII dataset の全 SNR 条件における各モデルの AUC [%] 評価

		SNR 6 dB				Avg.
		Fan	Pump	Slider	Valve	
教師あり	SVM	99.04	98.75	95.29	86.29	94.84
	XGBoost	99.74	99.92	99.05	98.02	99.18
	LightGBM	99.80	99.92	99.41	97.57	99.18
	TabPFN	99.99	99.99	99.64	98.77	99.60
教師なし	k-NN	99.43	97.28	95.15	94.01	96.47
	GMM	99.79	99.13	96.34	94.59	97.46
	OC-SVM	96.71	93.22	93.31	83.04	91.57
	Deep SVDD	95.85	93.25	92.75	88.20	92.51

		SNR 0 dB				Avg.
		Fan	Pump	Slider	Valve	
教師あり	SVM	91.98	86.70	89.35	81.25	87.32
	XGBoost	97.74	96.63	96.67	93.32	96.09
	LightGBM	97.87	96.76	96.45	93.88	96.24
	TabPFN	99.57	98.50	98.09	96.62	98.19
教師なし	k-NN	94.85	93.56	91.56	89.17	92.29
	GMM	97.27	94.22	91.80	90.45	93.43
	OC-SVM	85.57	86.67	88.24	77.67	84.54
	Deep SVDD	87.27	85.45	86.90	83.95	85.89

		SNR −6 dB				Avg.
		Fan	Pump	Slider	Valve	
教師あり	SVM	65.01	75.52	79.61	75.50	73.91
	XGBoost	86.75	87.89	90.72	89.13	88.62
	LightGBM	87.34	86.96	91.26	89.16	88.68
	TabPFN	92.93	92.45	93.78	93.57	93.18
教師なし	k-NN	81.65	83.65	83.35	81.83	82.62
	GMM	86.44	84.18	82.81	85.31	84.69
	OC-SVM	64.78	77.73	74.44	71.85	72.19
	Deep SVDD	74.33	78.60	81.58	76.18	77.67

第4章 低SNR 条件下においても機能する特徴量抽出法

4.1 異常音検知における環境騒音

工場内で異常音判別対象の機械音を録音する場合、周囲で稼働している機械音も同時に収録される。これが環境騒音となり、録音された機械音の SNR の低下につながる。そのため、異常音検知における環境騒音の除去についても研究が進められている。例えば、非負値行列因子分解により録音から機械音のみを抽出する方法 [46] や、深層学習による音源分離に基づく方法 [47] などが提案されている。しかし、いずれの方法も環境騒音のみの録音、または機械音のみの録音が必要であり、問題設定によっては適用できない。

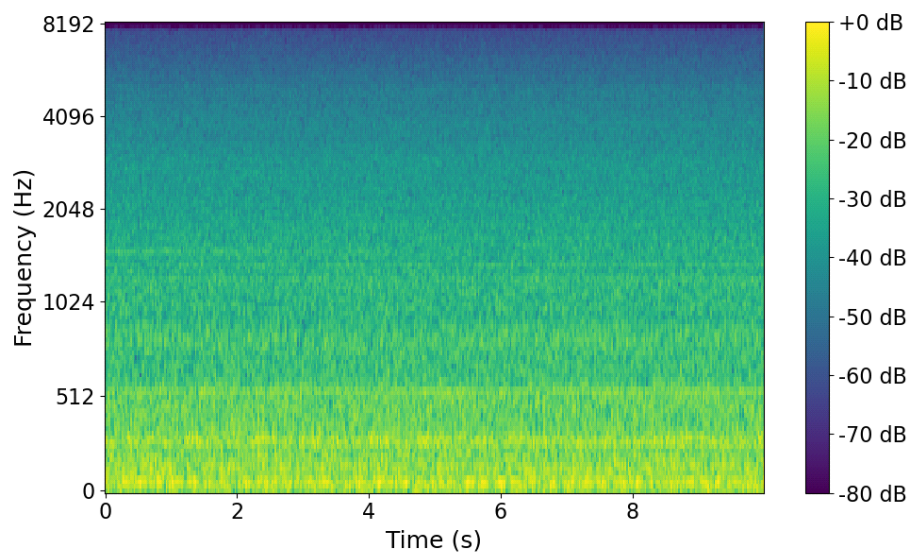
本研究では、雑音除去モデルにより処理した機械音による判別についての検討も行ったが、判別性能は除去前に比べて低下することがわかった。そのため、騒音除去以外のアプローチを検討する必要がある。環境騒音除去の検討については付録 A で述べる。

4.2 音色関連特徴量の時系列変化

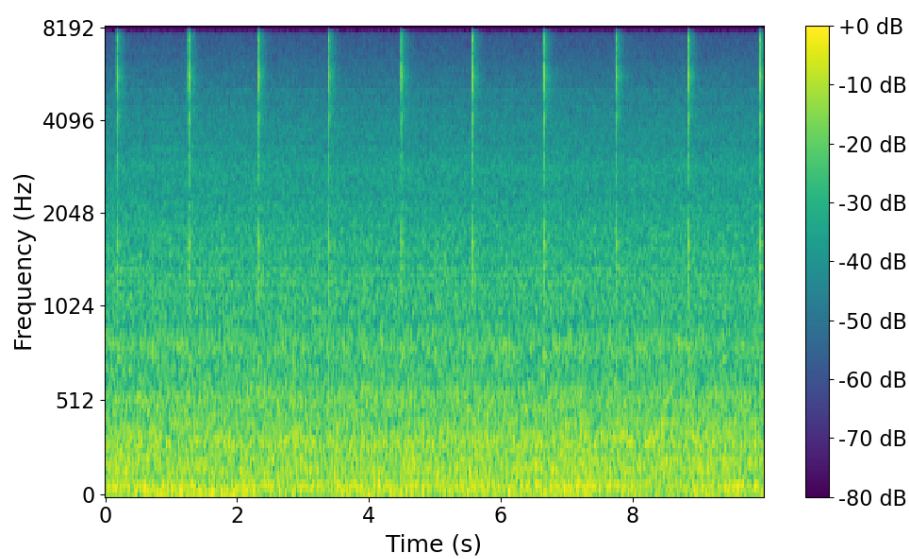
図 4.1 に、MIMII dataset の機械音の対数メルスペクトログラムを示す。Fan は定常な音であり、1 クリップ内で常に機械音が発生している。一方、Valve は断続的に機械音が発生しており、非定常である。そのため、環境騒音のみ発生している箇所も存在する。従来法では機械音クリップを複数フレームに分割し、フレームごとに音色関連特徴量の算出を行っている。しかし、非定常な機械音の場合、分析フレーム長によっては環境騒音のみのフレームが存在する可能性がある。

機械音の特定箇所のみを異常音検知に使用方法として、Wang らが提案した spectrogram frames selection (SFS) がある [48]。SFS は対数メルスペクトログラムを対象としており、共分散を基に非定常な機械音の発生箇所を含む時間フレームのみ選択する。Wang らは SFS により選択されたフレームのみ判別に使用することで、非定常音の判別性能が改善することを報告している。そのため、音色関連特徴量についても、機械音の発生箇所のみからの抽出が有効だと考えられる。

図 4.2 に、Valve の対数メルスペクトログラムと音色関連特徴量の時系列変化の例を示す。約 5 秒を過ぎた時点から機械音が発生しているが、AS の値は機械音の発生箇所で高くなっている。また、SQM の一部指標も AS と同じく機械音の発生に伴い変化しており、変化した箇所が機械音自体の音色関連特徴量を表していると推測される。したがって、非定常音では AS の値が高くなるフレームに機械音が含まれており、特に AS と相関がある変化をする指標が重要だと考えられる。



(a) Fan



(b) Valve

図 4.1: MIMII dataset (SNR 6 dB) に含まれる機械音の対数メルスペクトログラム：(a) Fan の正常音, (b) Valve の正常音

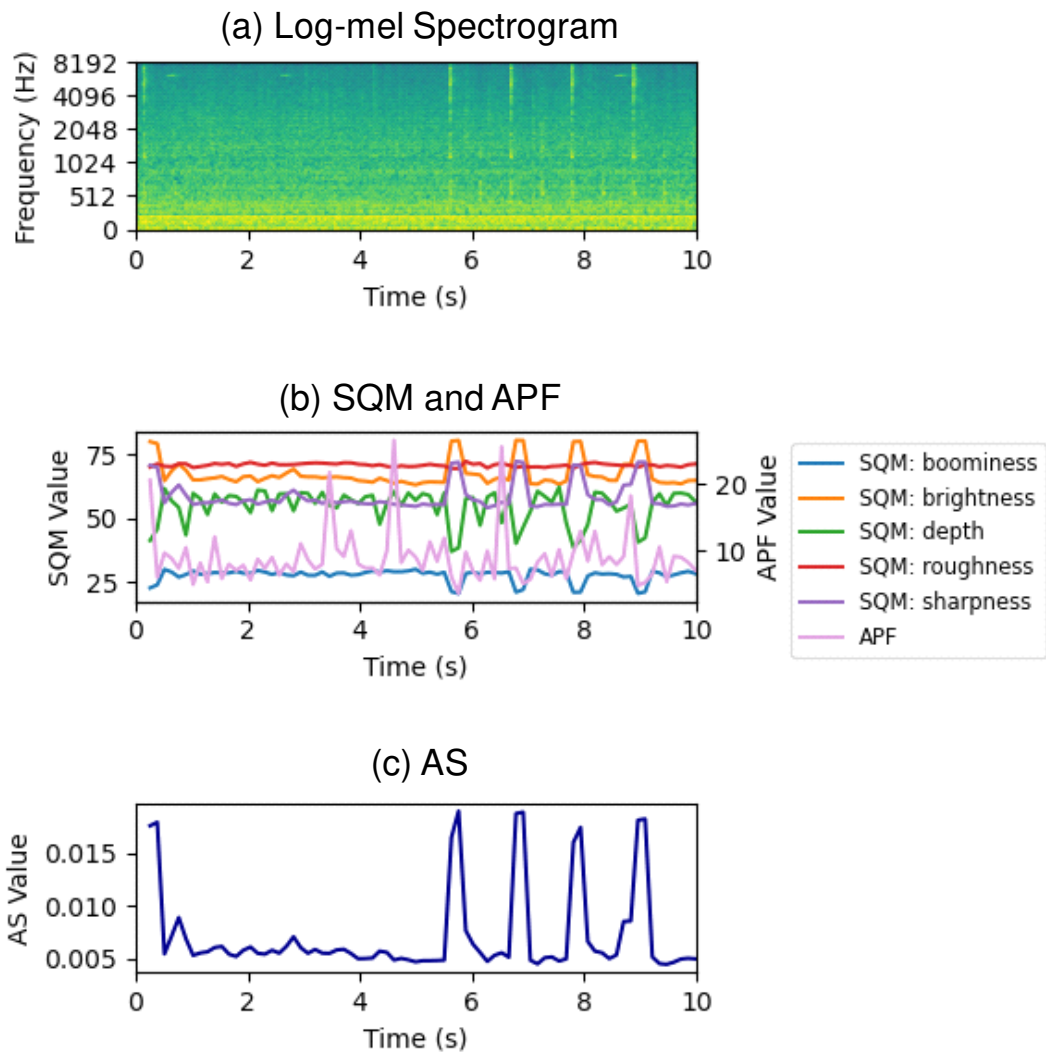


図 4.2: Valve における音色関連特徴量の時系列変化：(a) 対数メルスペクトログラム，(b) SQM と APF の時系列変化，(c) AS の時系列変化

4.3 音色特徴の選択処理

非定常な機械音における特徴量抽出法として、音色特徴の選択処理を行う Timbral frame selection (TFS) を提案する.

あらかじめ、判別対象機械の L 個の学習用正常音データ $\mathcal{D} = \{\mathbf{X}^{(l)}\}_{l=1}^L$ の各正常音サンプル $\mathbf{X}^{(l)}$ について、AS とそれ以外の M 個の音色指標 (TRF) の N フレームからなる時系列データ $\mathbf{AS}^{(l)} = (A_1^{(l)}, \dots, A_N^{(l)})$ と $\mathbf{TRF}_m^{(l)} = (T_{m,1}^{(l)}, \dots, T_{m,N}^{(l)})$ ($m = 1, \dots, M$) を抽出する. そして、 $\mathbf{X}^{(l)}$ ごとの AS とその他 M 個の指標間のピアソンの積率相関係数 $r_m^{(l)}$ ($m = 1, \dots, M$) と、全正常音サンプルの $r_m^{(l)}$ の平均 $\bar{r}_m$ を以下の式で計算する.

$$r_m^{(l)} = \frac{\frac{1}{N} \sum_{n=1}^N (A_n^{(l)} - \bar{A}^{(l)})(T_{m,n}^{(l)} - \bar{T}_m^{(l)})}{\sqrt{\frac{1}{N} \sum_{n=1}^N (A_n^{(l)} - \bar{A}^{(l)})^2} \cdot \sqrt{\frac{1}{N} \sum_{n=1}^N (T_{m,n}^{(l)} - \bar{T}_m^{(l)})^2}} \quad (4.1)$$

$$\bar{r}_m = \sum_{l=1}^L r_m^{(l)} \quad (4.2)$$

ただし、 $A_n^{(l)}$ と $T_{m,n}^{(l)}$ は、 $\mathbf{AS}^{(l)}$ と $\mathbf{TRF}_m^{(l)}$ の n フレーム目の値である. また、 $\bar{A}^{(l)}$ と $\bar{T}_m^{(l)}$ は、それぞれ $\mathbf{AS}^{(l)}$ と $\mathbf{TRF}_m^{(l)}$ の全フレームの値の平均値である. 式 4.1 と式 4.2 により、学習用正常音データ $\mathcal{D}$ に対して M 個の $\bar{r}_m$ が求まる. $\bar{r}_m$ の範囲は $-1 < \bar{r}_m < 1$ であり、 m 個目の音色指標と AS に相関がある場合、 -1 または 1 に近い値となる.

TFS の入力は、サンプル $\mathbf{Y}$ から抽出した AS と M 個の音色指標の N フレームの時系列データ $\mathbf{AS}' = (A'_1, \dots, A'_N)$, $\mathbf{TRF}'_m = (T'_{m,1}, \dots, T'_{m,N})$ ($m = 1, \dots, M$) と、相関係数 $\bar{r}_m$ ($m = 1, \dots, M$) である. ただし、 $\mathbf{Y}$ は $\mathcal{D}$ と同一の Machine Type, Machine ID であり、正常/異常は問わない. 入力に対し、以下の式で定義されるフレーム選択を行う.

$$I_k = \underset{n \in 1, 2, \dots, N}{\operatorname{argtop-k}} A'_n \quad (4.3)$$

$$\text{m-th timbral feature} = \begin{cases} \frac{1}{k} \sum_{i \in I_k} T'_{m,i} & \text{if } |\bar{r}_m| > p, \\ \bar{T}'_m & \text{otherwise} \end{cases} \quad (4.4)$$

ただし、 $\operatorname{argtop-k} X_j$ は j を 1 から J まで変化させた場合、 X_j が最大値となる上位 k 個のインデックス j の集合を返す関数である. また、 $\bar{T}'_m$ は $\mathbf{TRF}'_m$ の全フレームの平均値である. $|\bar{r}_m|$ が閾値 p を上回る場合、 $\mathbf{Y}$ は非定常音であり、 $\bar{r}_m$ に対応する指標 $\mathbf{TRF}'_m$ と $\mathbf{AS}' = (A'_1, \dots, A'_N)$ の間に強い相関があると考えられる. そのため、 A'_n の最大値上位 k 個のフレームと同じインデックスを持つ k 個のフレー

$\mu T'_{m,i}$ のみで平均値を計算する． $|\bar{r}_m|$ が閾値 p を下回る指標は全フレームを平均する．また，AS については， $|\bar{r}_m| > p$ となる指標が 1 つ以上ある場合， A'_n の最大値上位 k 個のフレームのみで平均し，それ以外の場合は全フレームを平均する．TFS により， M 個の音色指標と AS を加えた $M+1$ 個のスカラー値を取得し，音色関連特徴量として識別器の入力に使用する．

4.4 実験条件

従来法の特徴抽出部と識別器の間に TFS を適用し，適用前と適用後の判別性能の比較を行った．識別器には，TabPFN と GMM を使用した．TabPFN は教師あり学習，GMM は教師なし学習を行った．学習と評価には MIMII dataset を使用した．学習手順は 3.3 節と同様である．評価指標には AUC を使用した．

音色関連特徴量指標の抽出において，すべての指標でフレーム長を 256 ms，フレームシフトを 128 ms に設定した．そのため，TFS におけるフレーム数 N は 77 となった．また，閾値 p は 0.9 に設定した． $\operatorname{argtop-k} X_j$ における k は 5 に設定した．Machine Type ごとに使用する音色関連特徴量の指標の組み合わせは表 3.2 と同様のものを使用した．そのため，Fan と Valve において $M = 6$ ，Pump と Slider において $M = 5$ となった．

4.5 実験結果

表 4.1 は MIMII dataset の各 Machine Type，Machine ID において，TFS が適用された音色指標を示したものである．Slider と Valve の一部 ID で brightness と sharpness が選択されており，TFS が非定常音である Slider と Valve を識別できていると考えられる．

表 4.2 に MIMII dataset における AUC 評価を示す．GMM，TabPFN のどちらについても，TFS を適用することで Slider や Valve といった非定常音の判別性能が改善している．Slider と Valve では表 4.1 で示した通り TFS によるフレーム選択が行われているため，非定常音では AS と相関を持つ指標が異常の判別に大きく関わると考えられる．また，TFS の適用により，TFS 未適用の場合に比べ非定常音の SNR の低下に伴う判別性能の低下が抑えられていることがわかる．そのため，TFS は低 SNR 条件下で特に性能改善効果があると考えられる．一方，Fan と Pump ではすべての指標で全フレームの平均値が採用されたため，TFS の有無による評価値の変化は見られなかった．

以上のことから，提案した音色特徴の選択処理法である TFS が，SNR 条件の違いによらず非定常音の異常音検知に有効であることがわかった．

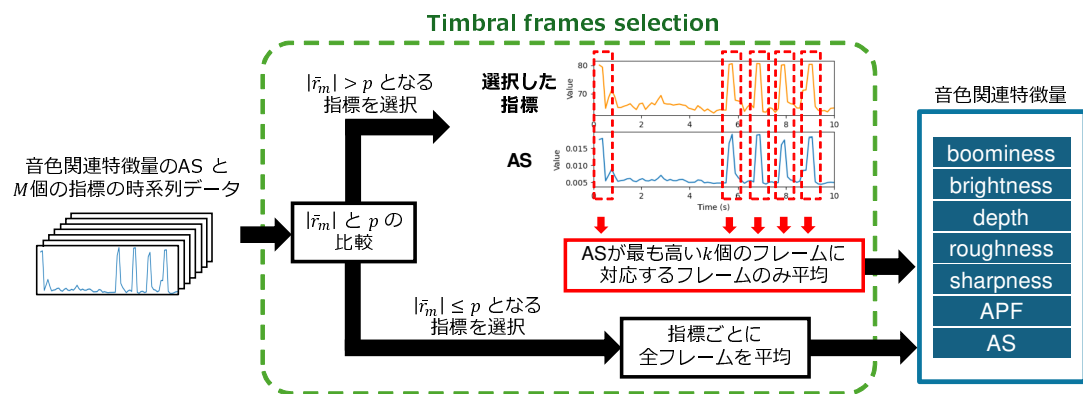


図 4.3: 音色特徴の選択処理のフロー

表 4.1: 各 Machine Type, Machine ID において TFS によるフレーム選択が行われた指標

	Fan				Pump			
	00	02	04	06	00	02	04	06
boominess	-	-	-	-	-	-	-	-
brightness	-	-	-	-	-	-	-	-
depth	-	-	-	-	-	-	-	-
roughness	-	-	-	-	-	-	-	-
sharpness	-	-	-	-	-	-	-	-
APF	-	-	-	-	-	-	-	-

	Slider				Valve			
	00	02	04	06	00	02	04	06
boominess	-	-	-	-	-	-	-	-
brightness	-	-	✓	✓	✓	✓	✓	✓
depth	-	-	-	-	-	-	-	-
roughness	-	-	-	-	-	-	-	-
sharpness	-	-	✓	✓	✓	✓	✓	✓
APF	-	-	-	-	-	-	-	-

表 4.2: MIMII dataset における TFS を適用した場合の各モデルの AUC [%] 評価

		SNR 6 dB				
		Fan	Pump	Slider	Valve	Avg.
教師あり	TabPFN	99.98	100.00	98.29	97.50	98.94
	TabPFN (TFS)	99.98	100.00	99.71	98.95	99.66
教師なし	GMM	99.59	98.84	93.68	91.47	95.89
	GMM (TFS)	99.59	98.84	94.87	94.85	97.04

		SNR 0 dB				
		Fan	Pump	Slider	Valve	Avg.
教師あり	TabPFN	99.63	98.09	94.97	93.37	96.52
	TabPFN (TFS)	99.63	98.09	99.26	96.95	98.48
教師なし	GMM	97.12	94.44	89.05	85.55	91.54
	GMM (TFS)	97.12	94.44	91.46	92.56	93.89

		SNR −6 dB				
		Fan	Pump	Slider	Valve	Avg.
教師あり	TabPFN	93.54	92.13	91.72	91.63	92.25
	TabPFN (TFS)	93.54	92.13	98.13	94.41	94.55
教師なし	GMM	87.08	85.71	83.14	81.24	84.29
	GMM (TFS)	87.08	85.71	89.15	87.44	87.35

第5章 教師なし学習による異常音検知

5.1 擬似異常音データを使用した異常音検知

従来法では正常／異常音を両方使用した教師あり学習を行っている．一方，異常音検知において異常音を学習することは現実的ではない．しかし，擬似的に学習用の異常音データを生成可能であれば，教師なし学習の条件を満たしたうえで正常／異常の二値分類による学習を行うことができる．本章では，OE の考え方を基にした擬似異常音データの生成について検討する．第2章でも説明したとおり，OE によるクラス分類アプローチは外れ値検出に比べ高い判別性能を示す．そのため，音色関連特徴量によるシステムにおいても OE は有効だと考えられる．

DCASE 2021 では，MobileNetV2 [49] を用いた OE による方法がベースラインとして使用されている [11]．特徴量には対数メルスペクトログラムが使用されている．ベースラインにおいて，例えば Fan ID00 の分類を行う場合，Fan ID02, 04, 06 の正常音を擬似的な異常クラスとして使用している．MobileNetV2 は，入力された Fan の正常音の ID (4 クラス) を分類するように学習する．この時，MobileNetV2 は異常音を学習していないため，Fan ID00 の異常音に対して正確に ID 分類することはできないと考えられる．そのため，モデルの出力を基に異常スコアを決定することができる．

OE における擬似異常音データには，判別対象機械の正常音と類似，または異なりすぎないデータが適している [11]．データが類似している場合は識別器の学習が難しくなり，データが異なりすぎる場合は識別器の学習が過度に簡単になる．そのため，どちらの場合も正常音の判別に有効な特徴を識別器が学習できない．以上を踏まえ，単純に Machine ID の分類を行うのではなく，モデルの学習に適する擬似異常音データを生成して学習に使用することを検討した．

本研究では，擬似異常音の音色関連特徴量のデータの分布を推定し，推定した分布からサンプリングすることで擬似異常音の音色関連特徴量データの生成する．異常音の分布は，判別対象機械の正常音以外の様々な機械音の分布だと考えることができる [50]．そのため，様々な Machine Type の機械音を使用し，判別対象機械の正常音と類似または異なりすぎない異常音の分布を推定する．

5.2 擬似異常音データの生成

擬似異常音データの分布推定に使用する機械音は、判別対象機械の正常音との分布の位置に近いものを使用する。事前に、 S クラスの機種音が収録された様々な機械音データセットを用意する。また、判別対象機械の正常音色関連特徴量の集合を $\mathcal{N}$ とする。まず、判別対象機種とその他機種の正常音色関連特徴量の平均値 $\mu_{\mathcal{N}}, \mu_i (i = 1, \dots, S)$ を計算する。次に、 $\mu_{\mathcal{N}}$ とその他 S 機種 μ_i との L2 距離を計算する。最後に、最も L2 距離に近い j 個の μ_i と対応する機種を、擬似異常生成用音色関連特徴量として選択する。この時選択された j 機種の正常音色関連特徴量と、判別対象の機種の正常音色関連特徴量をどちらも含む集合 $\mathcal{A}$ を用いて擬似異常音データの生成を行う。パラメータ j の設定により、判別対象機械の正常音と過度に異なる異常音データの生成を防ぐことができる。

分布の推定には、GMM を使用する。推定手順を図 5.1 に示す。GMM を用いて推定した $\mathcal{N}$ と $\mathcal{A}$ の確率密度関数をそれぞれ $f_{\mathcal{A}}(x)$ と $f_{\mathcal{N}}(x)$ とする。 $\mathcal{A}$ には $\mathcal{N}$ も含まれており、 $\mathcal{A}$ の分布は $\mathcal{N}$ を取り囲むような分布になる。そのため、 $\mathcal{A}$ から $\mathcal{N}$ を除いた分布が、異常音の分布だと考えられる。ただし、異常音の確率密度関数自体の推定は困難であるため、 $f_{\mathcal{A}}(x)$ からサンプル $\mathbf{Z}$ を生成し、 $f_{\mathcal{A}}(\mathbf{Z}) > k f_{\mathcal{N}}(\mathbf{Z})$ を満たす場合に $\mathbf{Z}$ を擬似異常音データとして採用する。パラメータ k の数値を上げることで、より正常音の分布から外れた $\mathbf{Z}$ のみ採用されるようになる。そのため、パラメータ k を適切に設定することで、判別対象機械の正常音と類似しすぎる擬似異常音データの生成を防ぐことができる。

以上の手順により、異常音の音色関連特徴量を擬似的に生成し、実正常音データと併せて学習に使用可能になる。

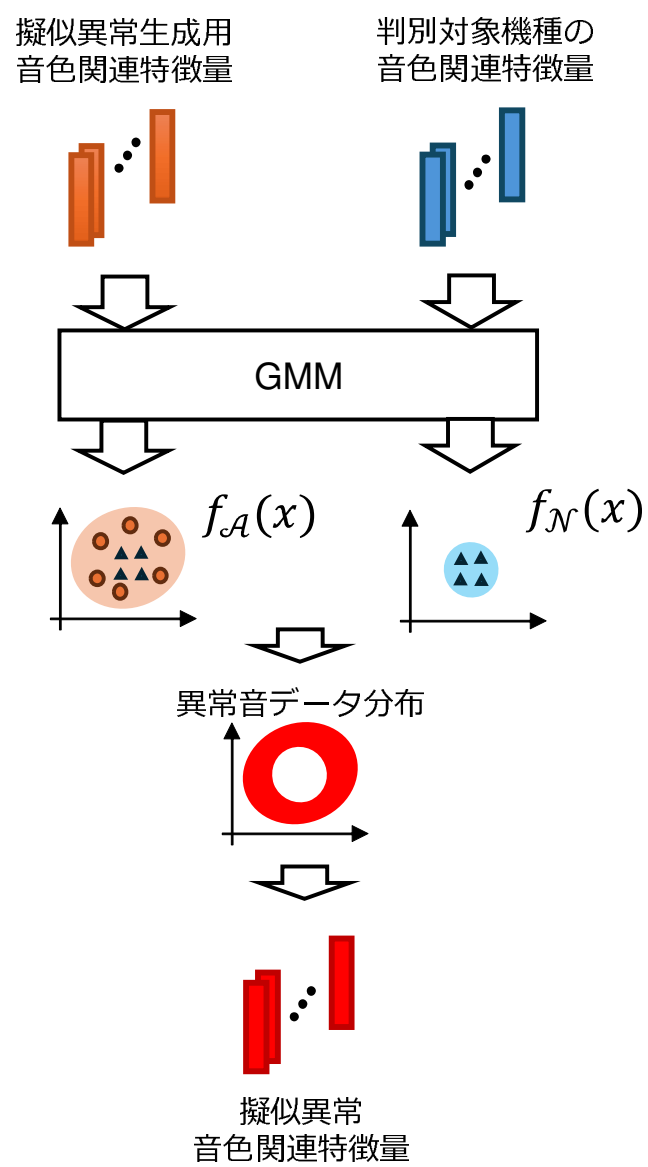


図 5.1: 異常音の分布の推定

5.3 実験条件

データセットに含まれる正常音と、生成した擬似異常音色関連特徴量を使用して TabPFN の学習を行い、正常音のみで学習された GMM との判別性能を比較した。学習または評価には DCASE 2020 Task 2 development dataset [10] を使用した。DCASE 2020 Task 2 development dataset は教師なし学習による異常音検知用のデータセットであり、MIMII dataset から Fan, Pump, Slider, Valve, ToyADMOS から ToyCar, ToyConveyor の合計 6 個の Machine Type が含まれる。また、各 Machine Type ごとに 3 または 4 個の Machine ID が収録されている。学習用データと評価用データがあらかじめ分割されている。学習用データには正常音のみ含まれる。モデルの学習は各 Machine ID ごとに行い、GMM の学習には各 ID のすべての正常音、TabPFN の学習には各 ID の正常音 512 個と、生成した異常音データ 512 個を使用した。評価指標には AUC を使用した。

音色関連特徴量指標の抽出手順において、すべての指標でフレーム長を 256 ms、フレームシフトを 128 ms に設定した。そのため、Machine Type ごとに使用する音色関連特徴量の指標の組み合わせは表 3.2 と同様のものを使用した。ToyCar と ToyConveyor に関しては、boominess, depth, roughness, sharpness, AS, APF の 6 指標を使用した。擬似異常音データの生成において、DCASE2020 Task 2 development dataset は全機種 of Machine ID の合計が 23 であるため、 S は 23 に設定した。また、 $j = 2$, $k = 1$ に設定した。

5.4 実験結果

表 5.1 に評価を示す。TabPFN (ID classification) は、DCASE 2021 で使用された MobileNetV2 によるベースラインの特徴量を音色関連特徴量、識別器を TabPFN に置き換えたものである。ID classification は、GMM に比べて Valve を除く機種で判別性能が低下している。そのため、深層学習による OE において一般的な Machine ID 分類は、音色関連特徴量に適さない可能性がある。

一方、GMM や ID classification と比較して、擬似異常音データを学習した TabPFN は高い判別性能を示している。擬似異常音データの生成に使用されたクラスを調査したところ、例えば Fan ID00 では Slider ID00 と Slider ID02, Pump ID00 では Pump ID02 と Valve ID02 が選択されていた。このように、同一の Machine Type のクラスに限定せず、様々な Machine Type, Machine ID のデータを基に擬似異常音データを生成したことで性能が向上したと考えられる。以上より、推定した異常音分布から生成した擬似異常音データが、音色関連特徴量による教師なし学習に有効だとわかった。

表 5.1: 擬似異常音データを学習した場合の AUC [%] 評価

	Fan	Pump	Slider	Valve	ToyCar	ToyConveyor	Avg.
GMM	83.58	79.63	75.88	76.55	69.90	55.41	73.50
TabPFN (ID classification)	79.88	72.37	75.25	77.86	65.35	54.36	68.58
TabPFN (pseudo-anomaly)	86.24	80.33	78.65	83.65	70.33	54.07	75.55

第6章 提案システムの総合評価と考察

6.1 提案システムの概要

従来法の残された課題に対し、第3章では識別器、第4章では特徴抽出、第5章では教師なし学習について解決法の検討を行った。それぞれの検討を踏まえ、すべての解決法を組み込んだ提案システムの学習フローを図6.1に示す。まず、学習データと様々な機械音データを用意し、それぞれ音色指標の抽出とTFSを行う。次に、TFSから出力された音色関連特徴量から、擬似異常音データ生成部により擬似異常音色関連特徴量を生成する。最後に、学習データと擬似異常音の音色関連特徴量を識別器で学習する。本章では、提案システムの評価と全体考察を行う。

6.2 実験条件

提案システムの識別器には、従来法で使用されたSVMに加え、第3章で高い判別性能を示したGMMとTabPFNを使用した。ただし、GMMには擬似異常音データは使わず、正常音のみで学習を行った。評価にはDCASE 2020 Task 2 development datasetを使用した。識別器の学習は各IDごとに行った。SVMの学習には各IDの正常音に加え、正常音と同数の擬似異常音データを生成して使用した。TabPFNの学習には各IDの正常音512個と、生成した擬似異常音データ512個を使用した。

音色関連特徴量指標の抽出手順において、すべての指標でフレーム長を256 ms、フレームシフトを128 msに設定した。Machine Typeごとに使用する音色関連特徴量の指標の組み合わせは第5章の実験と同じものを使用した。識別器、TFS、擬似異常音データ生成部の各パラメータは、すべて第3、4、5章の実験と同じものを使用した。

評価指標にはAUCを使用した。

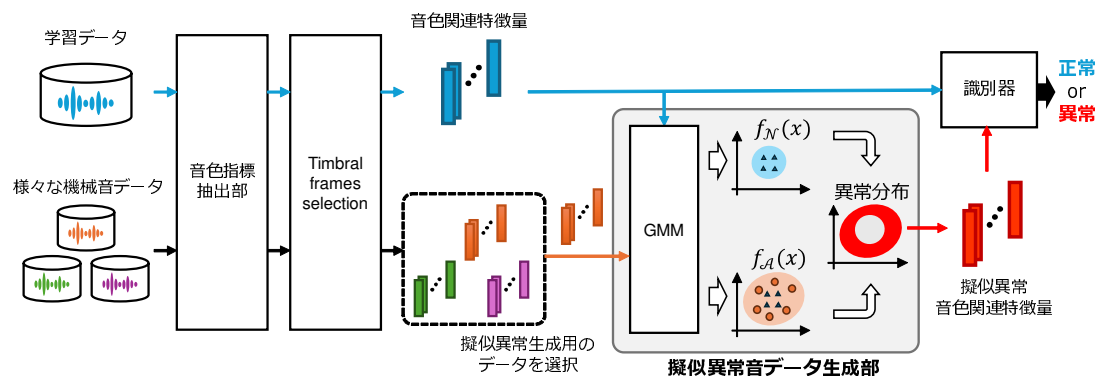


図 6.1: 提案システムの学習フロー

6.3 実験結果

提案システムの評価の結果を表 6.1 に示す。AE は DCASE 2020 で使用された autoencoder によるベースラインであり、特徴量は対数メルスペクトログラムである。AE の評価値は文献 [10] から引用した。各 Machine Type の評価値は、Machine Type に含まれる全 Machine ID の評価値の平均を示している。

GMM, SVM, TabPFN のすべてにおいて、TFS の適用により非定常音である Slider, Valve の判別性能が向上している。SVM と TabPFN では Pump の判別性能も上がっているが、これは擬似異常生成用データに TFS が適用されたことで、生成した擬似異常音データが変化したことによるものである。また、擬似異常データを学習することで性能が向上している。最も良い性能を示したシステムは、TabPFN に TFS を適用し、擬似異常音データを学習したものであり、特に Fan と Valve において AE によるベースラインを約 20% 上回る判別性能を示している。

図 6.2 は、Valve ID04 の音色関連特徴量の UMAP [51] による分布と、SVM と TabPFN が割り当てた異常スコアを示したものである。TabPFN は全体的に SVM を上回る性能を示している。また、TFS の適用により正常／異常の分布の分離度が上がっており、それに伴い判別性能も向上している。

6.4 考察

識別器については従来法で使用された SVM ではなく、GMM や TabPFN を使用することで性能が向上した。そのため、音色関連特徴量には、表データのような低次元データを処理可能なモデルが有効だと考えられる。TFS についてはモデルの種類にかかわらず Slider, Valve の性能の向上に寄与した。TFS を適用することで、図 6.2 (c), (f) のように音色関連特徴量自体の正常／異常の分離度が上がっていると考えられる。以上より、非定常音において、機械音の発生箇所から特徴量を抽出することが有効であった。ただし、TFS は定常音には影響しないため、今後は定常音用の特徴抽出法の開発も必要だと考えられる。また、擬似異常音データの学習により、ID 分類に比べ判別性能が向上した。SVM では、Pump において擬似異常音データの学習により性能が大きく低下したものの、全機種の評価値の平均は ID 分類に比べ向上した。そのため、学習に適した異常音データを生成できたと考えられる。以上より、本研究で検討を行った 3 つの解決法が、音色関連特徴量に着目した教師なし機械学習による異常音検知にそれぞれ有効であることが確認された。

また、提案システムが AE を用いた一般的な深層学習法を上回ることもわかった。特に TabPFN, TFS, PA を適用したシステムは、MIMII の 4 機種すべてで AE を上回っている。そのため、特定の機種においては対数メルスペクトログラム

のような典型的な特徴量に比べ、音色関連特徴量が有効である可能性が示唆された。一方、ToyCar と ToyConveyor において、提案システムは AE に比べ大きく性能が低下した。原因として、ToyADMOS の機種において音色関連特徴量として選択した指標が適していない可能性が考えられる。従来法において、Ota & Unoki は MIMII の 4 機種についてオノマトペを用いた指標の選択を行っており、本研究でも同じ指標の組み合わせを使用している。一方、Ota & Unoki は ToyADMOS については検討していない。そのため、本研究では従来法と同様の手順で ToyCar と ToyConveyor の指標の選択を行ったが、検討は簡易的なものである。機械ごとに適する音色関連特徴量指標の選択を行うことが、判別性能に大きな影響を与えらると思えられる。

表 6.1: 提案システムの AUC [%] 評価. ただし, TFS は Timbral frames selection, PA は擬似異常音データの学習の有無を示す. AE の特徴量は対数メルスペクトログラムである. PA が× の場合, DCASE 2021 [11] のベースラインを基に ID 分類を行っている.

TFS	PA	Model	Fan	Pump	Slider	Valve	ToyCar	ToyConveyor	Avg.
-	-	AE (baseline) [10]	65.83	72.89	84.76	66.28	78.77	72.53	73.51
×	-	GMM	83.58	79.63	75.88	76.55	69.90	55.41	73.50
✓	-		83.58	79.63	82.06	83.96	69.90	55.41	75.76
×	×	SVM	64.34	70.61	72.31	67.97	61.81	53.16	62.82
×	✓		64.46	74.34	66.64	68.75	62.77	53.90	65.14
✓	✓		64.46	78.41	77.77	84.27	62.77	53.90	70.26
×	×	TabPFN	79.88	72.38	75.25	77.86	65.35	54.36	68.58
×	✓		86.24	80.33	78.65	83.65	70.33	54.07	75.55
✓	✓		86.24	82.67	86.00	86.45	70.33	54.07	77.63

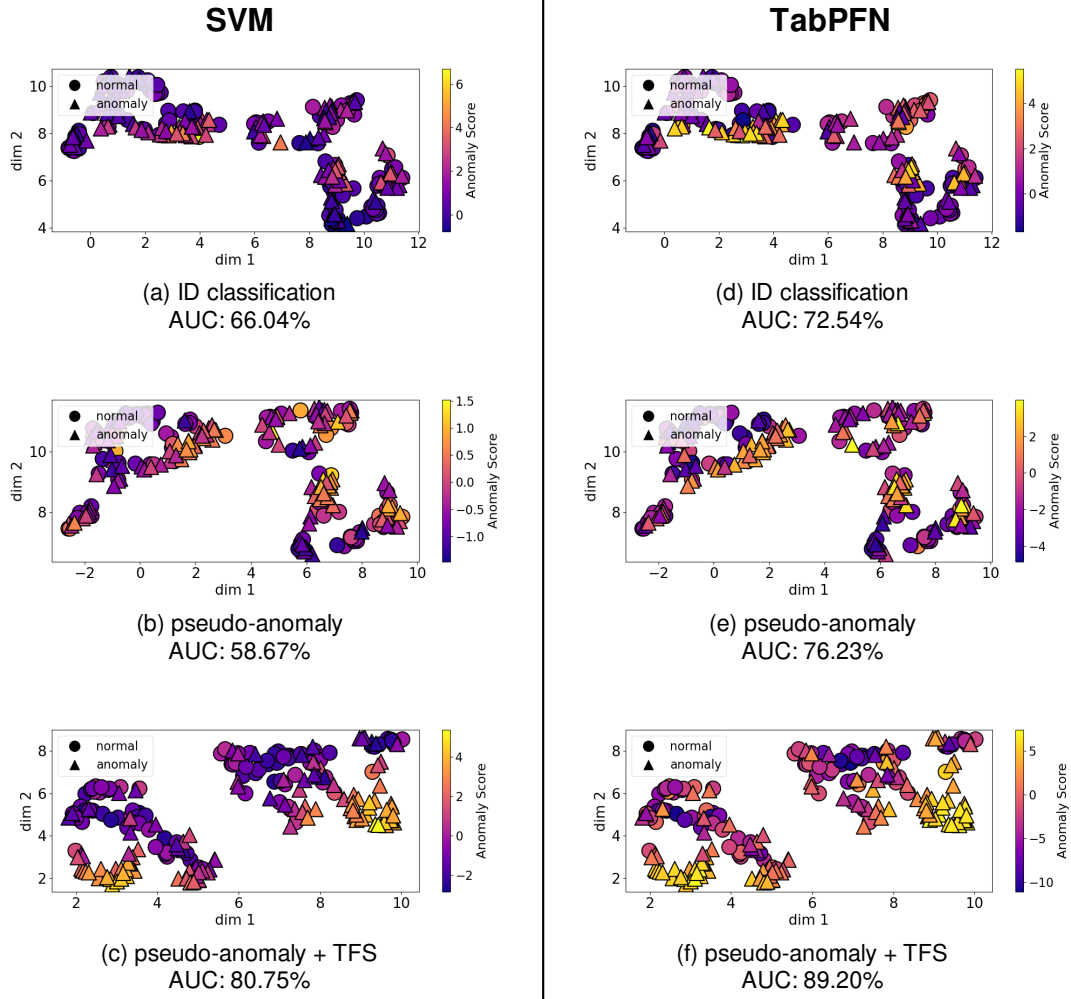


図 6.2: UMAP [51] による Valve ID04 の音色関連特徴量の分布の可視化と、識別器が各サンプルに割り当てた異常スコア。カラーマップが異常スコアを示す。ただし、(a), (b), (c) は SVM, (d), (e), (f) は TabPFN を識別器に用いている。また、(a) と (d) は ID 分類, (b) と (e) は擬似異常音データの学習, (c) と (f) は擬似異常音データの学習と TFS を適用した場合の結果である。

第7章 結論

7.1 本論文における成果

本研究では、音色関連特徴量に着目した教師なし機械学習による異常音検知システムの検討を行った。システムの構築にあたり、識別器、特徴量抽出、教師なし機械学習における問題に対する解決法を導入した。

識別器には、音色関連特徴量のような低次元データを処理できるモデルを用いることで性能が向上した。特徴量抽出については、非定常な機械音において、機械音の発生箇所を含むフレームのみ選択して特徴抽出を行うことが有効だとわかった。また、擬似異常音データによる教師なし機械学習により、正常音のみによる学習が可能になった。これら3つの解決法を組み込んだ提案システムは、autoencoderによるシステムを全体的に上回る性能を示した。

7.2 今後の展望

本研究の問題設定はDCASE 2020 までの範囲であり、現在のDCASEにおける課題であるドメイン汎化やFirst-shot 異常音検知については検討していない。今後、音色関連特徴量によるシステムを実環境で使用する場合は、これらの問題について対処する必要があると考えられる。特に、提案システムで導入した擬似異常音データ生成法はFirst-shot 異常音検知への適用が難しい。そのため、より多様な問題設定に対応できる教師なし学習の方法について検討する必要がある。

実験において使用した音色関連特徴量の指標の組み合わせについても課題がある。実験で使用した一部機種については、本研究で使用した7指標では音色を完全に表現できていない可能性がある。また、機械音の環境騒音除去についても課題が残っている。これらについて検討を進めることで、判別性能の向上につながると考えられる。

付 録 A 機械音の環境騒音の除去

A.1 実験条件

環境騒音の除去は、ウィーナーフィルタ (Wiener filtering: WF), スペクトルサブトラクション (Spectral subtraction: SS), NSNet2 [52] を用いて行った. WF と SS は古典的な雑音抑圧法である. NSNet2 は ICASSP 2021 Deep Noise Suppression challenge [53] のベースラインである. NSNet2 は事前に音声を使って学習されており, 未学習でも雑音除去が可能なモデルである.

データセットには MIMII dataset の Fan, Pump, Slider, Valve と, ToyADMOS の ToyCar, ToyConveyor の 6 個の Machine Type を使用した. SNR -6 dB の機械音から, WF, SS, NSNet2 によって環境騒音を除去した. 環境騒音除去の評価には雑音抑圧量 (Noise reduction rate: NRR) を使用した. NRR は以下の式で定義される [54].

$$\text{NRR} = 10 \log_{10} \frac{E[s_{\text{out}}^2]/E[n_{\text{out}}^2]}{E[s_{\text{in}}^2]/E[n_{\text{in}}^2]} \quad (\text{A.1})$$

ここで, s_{in} と s_{out} は入力/出力信号, n_{in} と n_{out} は入力/出力雑音信号である. NRR の単位は dB であり, 入力前後の SNR の向上量を表す. ここで, 雑音除去が目的信号に影響しない場合, NRR は入出力前後の雑音成分の強さの違いを表す. ただし, MIMII dataset には騒音がない機械音が含まれていない. そのため, 騒音除去前後の信号から, SNR 6 dB の信号を減算することで雑音成分 n_{in} と n_{out} を求め, NRR を $10 \log_{10}(E[n_{\text{in}}^2]/E[n_{\text{out}}^2])$ とした. 本実験の NRR は厳密な NRR とは異なることに注意されたい. また, NRR を基に, 音色関連特徴量が騒音除去前後でどれだけ SNR 6 dB の音色関連特徴量の値に近づくかどうかの評価も行った. 音色関連特徴量の評価指標は以下の式で行った.

$$\text{NRR-TRF} = \log_{10} \frac{\sqrt{\frac{1}{N} \sum_{n=1}^N (I_n - R_n)^2}}{\sqrt{\frac{1}{N} \sum_{n=1}^N (O_n - R_n)^2}} \quad (\text{A.2})$$

ここで, N は音色関連特徴量の各指標のフレーム数, I と O はそれぞれ入力/出力信号の音色関連特徴量の指標, R は SNR 6 dB の信号とした. NRR-TRF は, SNR 6 dB の信号の音色関連特徴量との二乗平均平方根誤差が, 騒音除去前後でどれだけ改善したかを示す. ただし, 改善していない場合, NRR-TRF は負の値を示す.

A.2 実験結果

各 Machine Type の ID00 からそれぞれ 50 サンプルを対象として、NRR と NRR-TRF による評価を行った。NRR による環境騒音除去後のデータの評価を表 A.1 に示す。WF や SS では SNR が悪化しているのに対し、NSNet2 は比較的高い除去性能を示している。また、NRR-TRF による騒音除去による音色関連特徴量の変化を表 A.2 に示す。WF や SS では音色関連特徴量が改善されなかった。WF や SS は除去によりハーモニックノイズが発生し、音色指標が変化したと考えられる。NSNet2 においても全体的に悪化しているが、Valve では改善する傾向が見られた。また、AS や APF は他の指標より比較的改善している。

WF, SS, NSNet2 の中で最も NRR と NRR-TFS の評価値が高い NSNet2 を使用して、全 ID を対象にして騒音除去前と騒音除去後のデータ間の異常音判別性能を比較した。モデルは TabPFN を使用し、教師あり学習を行った。AUC による判別性能の評価を表 A.3 に示す。騒音除去により性能が下がっており、特に Fan の評価値が大きく低下した。表 A.2 に示した通り、NSNet2 も音色指標を劣化させる傾向があるため、判別性能も低下したと考えられる。また、音色関連特徴量が改善した Valve においても評価値が下がった。

実験結果より、NSNet2 は機械音中の騒音を除去できるものの、騒音除去と同時に機械音成分も除去する可能性が高いと考えられる。

表 A.1: NRR [dB] による評価

	Fan	Pump	Slider	Valve	ToyCar	ToyConveyor
WF	-13.3	-13.7	-13.4	-5.26	-10.5	-3.12
SS	-14.9	-11.6	-15.7	-4.60	-8.69	-0.105
NSNet2	5.90	5.69	6.67	5.19	2.41	2.69

表 A.2: NRR-TRF による評価

WF						
	Fan	Pump	Slider	Valve	ToyCar	ToyConveyor
boominess	−0.447	−0.293	−0.436	−0.205	−0.277	−0.183
brightness	−0.309	−0.138	−0.260	−0.006	−0.375	−0.075
depth	−0.268	−0.238	−0.310	−0.232	−0.246	−0.144
roughness	−0.423	−0.728	−0.522	−0.318	−0.302	−0.293
sharpness	−0.350	−0.305	−0.359	−0.079	−0.464	−0.289
AS	−1.141	−1.168	−1.194	−0.687	−1.257	−0.691
APF	−0.465	−0.628	−0.221	−0.386	−0.613	−0.645

SS						
	Fan	Pump	Slider	Valve	ToyCar	ToyConveyor
boominess	−0.346	−0.214	−0.297	−0.164	−0.127	−0.065
brightness	−0.250	−0.087	−0.202	−0.008	−0.318	−0.045
depth	−0.165	−0.141	−0.250	−0.190	−0.105	−0.030
roughness	−0.381	−0.706	−0.479	−0.356	−0.183	−0.201
sharpness	−0.274	−0.260	−0.293	−0.048	−0.381	−0.262
AS	−1.047	−1.105	−1.157	−0.635	−1.132	−0.523
APF	−0.422	−0.510	−0.295	−0.404	−0.509	−0.568

NSNet2						
	Fan	Pump	Slider	Valve	ToyCar	ToyConveyor
boominess	−0.279	−0.006	−0.296	0.053	−0.152	−0.062
brightness	−0.474	−0.053	−0.421	0.063	−0.595	−0.484
depth	−0.073	0.039	−0.103	0.022	−0.089	−0.039
roughness	−0.241	−0.048	−0.237	−0.078	−0.312	−0.409
sharpness	−0.454	−0.232	−0.404	0.090	−0.653	−0.700
AS	0.243	0.081	0.252	0.011	−0.300	−0.345
APF	−0.085	0.036	0.078	0.012	−0.221	−0.133

表 A.3: 騒音除去前後のデータによる AUC [%] 評価

	Fan	Pump	Slider	Valve	ToyCar	ToyConveyor	Avg.
騒音除去前	92.93	92.45	93.78	93.57	85.15	88.23	91.02
騒音除去後	68.57	78.23	84.83	73.30	67.09	69.98	73.67
差分	-24.36	-14.22	-8.950	-20.27	-18.06	-18.25	-17.35

謝辞

研究生活や学生生活について，数多くの助言，指導をしていただいた鵜木祐史教授に深く感謝致します．研究計画の議論において様々な助言をしていただいた大田恭士博士にも深く感謝致します．本研究は大田博士の研究に大きく影響を受けたものになります．また，研究室のミーティングにおいて様々な議論していただいた木谷俊介講師，上江洲安史特任助教，磯山拓都博士や鵜木研究室の皆様にも心から感謝致します．

中間発表会，修了審査を担当して下さった岡田将吾教授，吉高淳夫准教授にも感謝致します．また，学会発表において議論していただいた皆様にもお礼申し上げます．

最後に，生活をはじめとした様々な面で支援していただいた家族に感謝します．

参考文献

- [1] 井本桂右, 川口洋平, “環境音分析・異常音検知の研究動向,” 電子情報通信学会 基礎・境界ソサイエティ *Fundamentals Review*, Vol. 15, No. 4, pp. 268-280, 2022
- [2] 伊藤彰則, “環境音から異常を検知する統計的手法,” 日本音響学会誌, Vol. 75, No. 9, pp. 538-543, 2019.
- [3] 寺島真介, “音響信号による設備監視,” 紙パ技協誌, Vol. 57, No. 6, pp. 870-877, 2003.
- [4] 水野滉介, “音質評価指標を用いた異常音検知の研究,” 北陸先端科学技術大学院大学 修士論文, 2020.
- [5] 倉誠吾, “音色属性に係る評価指標を用いた異常音検知の研究,” 北陸先端科学技術大学院大学 修士論文, 2021.
- [6] 日本音響学会編, 新版 音響用語辞典 (コロナ社, 東京, 2003) , p. 40.
- [7] 日本音響学会編, 新版 音響用語辞典 (コロナ社, 東京, 2003) , p. 297.
- [8] Ota, Y. and Unoki, M., “Anomalous sound detection for industrial machines using acoustical features rRelated to timbral metrics,” *IEEE Access*, Vol. 11, pp. 70884-70897, 2023
- [9] 井本桂右, “DCASE Challenge: 環境音分析・理解のための統合的コンペティション,” 日本音響学会誌, Vol. 79, No. 9, pp. 470-476, 2023.
- [10] Koizumi, Y. Kawaguchi, Y. Imoto, K. Nakamura, T. Nikaido, Y. Tanabe, R. Purohit, H. Suefusa, K. Endo, T. Yasuda, M. and Harada, N., “Description and discussion on DCASE2020 Challenge Task2: Unsupervised anomalous sound detection for machine condition monitoring,” *Proc. Detection and Classification of Acoustic Scenes and Events Workshop (DCASE Workshop)*, November 2020, pp. 81-85.

- [11] Kawaguchi, Y. Imoto, K. Koizumi, Y. Harada, N. Niizumi, D. Dohi, K. Tanabe, R. Purohit, H. and Endo, T., “Description and discussion on DCASE 2021 Challenge Task 2: Unsupervised anomalous detection for machine condition monitoring under domain shifted conditions,” *Proc. Detection and Classification of Acoustic Scenes and Events Workshop (DCASE Workshop)*, November 2021, pp. 186–190.
- [12] Dohi, K. Imoto, K. Harada, N. Niizumi, D. Koizumi, Y. Nishida, T. Purohit, H. Tanabe, R. Endo, T. Yamamoto, M. and Kawaguchi, Y., “Description and discussion on DCASE 2022 Challenge Task 2: Unsupervised anomalous sound detection for machine condition monitoring applying domain generalization techniques,” *Proc. Detection and Classification of Acoustic Scenes and Events Workshop (DCASE Workshop)*, November 2022, pp. 1–5.
- [13] Dohi, K. Imoto, K. Harada, N. Niizumi, D. Koizumi, Y. Nishida, T. Purohit, H. Tanabe, R. Endo, T. and Kawaguchi, Y., “Description and discussion on DCASE 2023 Challenge Task2: First-shot unsupervised anomalous sound detection for machine condition monitoring,” *Proc. Detection and Classification of Acoustic Scenes and Events Workshop (DCASE Workshop)*, September 2023, pp. 31–35.
- [14] Nishida, T. Harada, N. Niizumi, D. Albertini, D. Sannino, R. Pradolini, S. Augusti, F. Imoto, K. Dohi, K. Purohit, H. Endo, T. and Kawaguchi, Y., “Description and discussion on DCASE 2024 Challenge Task2: First-shot unsupervised anomalous sound detection for machine condition monitoring,” *Proc. Detection and Classification of Acoustic Scenes and Events Workshop (DCASE Workshop)*, October 2024, pp. 111–115.
- [15] Purohit, H. Tanabe, R. Ichige, T. Endo, T. Nikaido, Y. Suefusa, K. and Kawaguchi, Y., “MIMII Dataset: Sound dataset for malfunctioning industrial machine investigation and inspection,” *Proc. Detection and Classification of Acoustic Scenes and Events Workshop (DCASE Workshop)*, November 2019, pp. 209–213.
- [16] Koizumi, Y. Saito, S. Uematsu, H. Harada, N. and Imoto, K., “ToyADMOS: A dataset of miniature-machine operating sounds for anomalous sound detection,” *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2019, pp. 313–317.
- [17] Harada, N. Niizumi, D. Ohishi, Y. Takeuchi, D. and Yasuda, M., “First-shot anomaly sound detection for machine condition monitoring: A domain

- generalization baseline,” *Proc. European Signal Processing Conference (EUSIPCO)*, 2023, pp. 191–195.
- [18] Primus, P., “Reframing unsupervised machine condition monitoring as a supervised classification task with outlier-exposed classifiers,” *Detection and Classification of Acoustic Scenes and Events Challenge (DCASE Challenge)*, Technical Report, June 2020.
 - [19] Kuroyanagi, I. Hayashi, T. Takeda, K. and Toda, T., “Improvement of serial approach to anomalous sound detection by incorporating two binary cross-entropies for outlier exposure,” *Proc. European Signal Processing Conference (EUSIPCO)*, 2022, pp. 294–298.
 - [20] Wilkinghoff, K., “Design choices for learning embeddings from auxiliary tasks for domain generalization in anomalous sound detection,” *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023, pp. 1–5.
 - [21] Nishida, T. Dohi, K. Endo, T. Yamamoto, M. and Kawaguchi, Y., “Anomalous sound detection based on machine activity detection,” *Proc. European Signal Processing Conference (EUSIPCO)*, 2022, pp. 269–273.
 - [22] Han, B. Lv, Z. Jiang, A. Huang, W. Chen, Z. Deng, Y. Ding, J. Lu, C. Zhang, W.-Q. Fan, P. Liu, J. and Qian, Y., “Exploring large scale pre-trained models for robust machine anomalous sound detection,” *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2024, pp. 1326–1330.
 - [23] Wilkinghoff, K., “Self-supervised learning for anomalous sound detection,” *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2024, pp. 276–280.
 - [24] Lv, Z. Jiang, A. Han, B. Liang, Y. Qian, Y. Chen, X. Liu, J. and Fan, P., “AITHU System for First-Shot Unsupervised Anomalous Sound Detection,” *Detection and Classification of Acoustic Scenes and Events Challenge (DCASE Challenge)*, Technical Report, June 2024.
 - [25] Dohi, K. and Kawaguchi, Y., “Distributed collaborative anomalous sound detection by embedding sharing,” *Proc. European Signal Processing Conference (EUSIPCO)*, 2024, pp. 91–95.
 - [26] Ho, T.V. Dohi, K. and Kawaguchi, Y., “Stream-based active learning for anomalous sound detection in machine condition monitoring,” *Proc. Interspeech*, 2024, pp. 102–106.

- [27] Tsubaki, S. Kawaguchi, Y. Nishida, T. Imoto, K. Okamoto, Y. Dohi, K. and Endo, T., “Audio-change captioning to explain machine-sound anomalies,” *Proc. Detection and Classification of Acoustic Scenes and Events Workshop (DCASE Workshop)*, September 2023, pp. 201–205.
- [28] NTT テクノクロス 異常音検知ソリューション Mono-Note. ONLINE Accessed on: Jan. 29, 2025, <https://www.ntt-tx.co.jp/products/Mono-Note/>
- [29] 日立製作所 異常音検知ソリューション 設備保全や製品検査の属人化を AI で解決 自動化・省力化を支援する異常音検知ソリューション. ONLINE Accessed on: Jan. 29, 2025, https://deh.hitachi.co.jp/_ct/17465529
- [30] バーナードソフト エスカレイド導入事例: 會澤高圧コンクリート株式会社様 生コンクリートを練り混ぜる「音」の違いで流動性を判断. ONLINE Accessed on: Jan. 30, 2025, <https://www.barnardsoft.co.jp/product/skaleid/interview01/>
- [31] Li, K. Nguyen, Q.H. Ota, Y. and Unoki, M., “Unsupervised anomalous sound detection for machine condition monitoring using temporal modulation features on gammatone auditory filterbank,” *Proc. Detection and Classification of Acoustic Scenes and Events Workshop (DCASE Workshop)*, November 2022.
- [32] Hafiz, P.A. Mawalim, C.O. Lestari, D.P. Sakti, S. and Unoki, M., “Anomalous machine sound detection based on time domain gammatone spectrogram feature and IDNN model,” *Proc. Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 2024.
- [33] 高田正幸, “音質評価指標の計算法と適用事例,” *日本音響学会誌*, Vol. 75, No. 10, pp. 582-589, 2019.
- [34] Mian, T. Choudhary, A. and Fatima, S., “An efficient diagnosis approach for bearing faults using sound quality metrics,” *Applied Acoustics*, Vol. 195, 108839, 2022.
- [35] Pearce, A. Safavi, S. Brookes, T. Mason, R. Wang, W. and Plumbley, M., “Release of timbral characterisation tools for semantically annotating non-musical content,” *AudioCommons Deliverable D5.8*, 2019.
- [36] Pearce, A. Brookes, T. and Mason, R., “Timbral attributes for sound effect library searching,” *Proc. AES International Conference on Semantic Audio*, June 2017.

- [37] Rafsanjani, M.A.H. Mawalim, C.O. Lestari, D.P. Sakti, S. and Unoki, M., “Unsupervised anomalous sound detection using timbral and human voice disorder-related acoustic features,” *Proc. Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 2024.
- [38] Nishida, T. Purohit, H. Dohi, K. Endo, T. and Kawaguchi, Y., “Timbre Difference Capturing in Anomalous Sound Detection,” arXiv preprint arXiv:2410.22033, 2024.
- [39] 井手美稀, 川越仁, 湯本英二, 兒玉成博, “Multi-Dimensional Voice Program (MDVP) および Praat を用いた音響分析結果の比較,” *音声言語医学*, Vol. 60, No. 3, pp. 214-219, 2019.
- [40] Chen, T. and Guestrin, C., “XGBoost: A Scalable Tree Boosting System,” *Proc. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Vol. 11, pp. 785-794, Aug., 2016.
- [41] Ke, G. Meng, Q. Finley, T. Wang, T. Chen, W. Ma, W. Ye, Q. and Liu, T., “LightGBM: A Highly Efficient Gradient Boosting Decision Tree,” *Proc. Advances in Neural Information Processing Systems (NIPS)*, Vol. 30, 2017.
- [42] Hollmann, N. Müller, S. Eggenberger, K. and Hutter, F., “TabPFN: A Transformer That Solves Small Tabular Classification Problems in a Second,” *Proc. the International Conference on Learning Representations (ICLR)*, 2023.
- [43] Vaswani, A. Shazeer, N. Parmar, N. Uszkoreit, J. Jones, L. Gomez, A.N. Kaiser, L. and Polosukhin, I., “Attention is All you Need,” *Proc. Advances in Neural Information Processing Systems (NIPS)*, Vol. 30, 2017.
- [44] Ruff, L. Vandermeulen, R. Goernitz, N. Deecke, L. Siddiqui, S.A. Binder, A. Müller, E. and Kloft, M., “Deep One-Class Classification,” *Proc. International Conference on Machine Learning (ICML)*, Vol. 80, pp. 4393-4402, 2018.
- [45] 井手剛, 杉山将, 異常検知と変化検知 (講談社, 東京, 2015) , pp. 8-13.
- [46] Aiba, A. Yoshida, M. Kitamura, D. Takamichi, S. and Saruwatari, H., “Noise robust acoustic anomaly detection system with nonnegative matrix factorization based on generalized gaussian distribution,” *IEICE Transactions on Information and Systems*, Vol. E104.D, No. 3, pp. 441-449, 2021.
- [47] Shimonishi, K. Dohi, K. and Kawaguchi, Y., “Anomalous sound detection based on sound separation,” in *Proc. INTERSPEECH*, pp. 2733-2737, 2023.

- [48] Wang, Y. Zhang, Q. Zhang, W. and Zhang, Y., “A lightweight framework for unsupervised anomalous sound detection based on selective learning of time-frequency domain features,” *Applied Acoustics*, Vol. 228, Article 110308, 2025.
- [49] Sandler, M. Howard, A. Zhu, M. Zhmoginov, A. and Chen, L.-C., “MobileNetV2: Inverted Residuals and Linear Bottlenecks,” *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 4510–4520.
- [50] Koizumi, Y. Saito, S. Uematsu, H. Kawachi, Y. and Harada, N., “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, 2019.
- [51] McInnes, L. Healy, J. Saul, N. and Großberger, L., “UMAP: Uniform Manifold Approximation and Projection,” *Journal of Open Source Software*, Vol. 3, No. 29, 2018, pp. 861.
- [52] Braun, S. and Tashev, I., “Data augmentation and loss normalization for deep noise suppression,” arXiv preprint arXiv:2008.06412, 2020.
- [53] Reddy, C.K.A. Dubey, H. Gopal, V. Cutler, R. Braun, S. Gamper, H. Aichner, R. and Srinivasan, S., “ICASSP 2021 deep noise suppression challenge,” *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 6623–6627.
- [54] 高橋 祐, 宮崎 亮一, 猿渡 洋, “高次統計量に基づく非線形雑音抑圧処理の数理解析とその応用,” *日本音響学会誌*, Vol. 68, No. 11, pp. 578-583, 2012.

研究業績

国内発表

1. 小倉稜也, 井本桂右, 貴家仁志, 塩田さやか, “距離に基づく音源分離を用いたシングルチャンネル環境音分類,” 日本音響学会秋季研究発表会, 2023.
2. 大田恭士, 小倉稜也, 鵜木祐史, “音色関連特徴量を利用した産業機器の異常音検知,” 応用音響研究会, Vol. 124, No. 42, pp.23-28, 2024.
3. 小倉稜也, 鵜木祐史, “音色関連特徴量を用いた異常音検知における識別器の検討,” 電気・情報関係学会北陸支部連合大会, H2-3, 2024.
4. 小倉稜也, 鵜木祐史, “音色関連特徴量を用いた異常音検知システムにおける特徴抽出法と識別器の検討,” 信号処理シンポジウム, 2024.
5. 小倉稜也, 鵜木祐史, “音色関連特徴量に着目した教師なし機械学習による異常音検知技術の検討,” 日本音響学会春季研究発表会, 2025. (予定)

Preprint

1. Ogura, R. Tomoya, N. and Kawaguchi, Y., “Retrieval-Augmented Approach for Unsupervised Anomalous Sound Detection and Captioning without Model Training,” arXiv preprint, arXiv: 2410.22056, 2024.

受賞

1. 日本音響学会 第27回学生優秀発表賞, 2024.
2. 電気・情報関係学会北陸支部連合大会 音響部門優秀発表賞, 2024.