

Title	音色属性に係る評価指標を用いた異常音検知の研究
Author(s)	倉, 誠吾
Citation	
Issue Date	2021-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/17135
Rights	
Description	Supervisor: 鷓木 祐史, 先端科学技術研究科, 修士 (情報科学)

修士論文

音色属性に係る評価指標を用いた異常音検知の研究

1910245 倉 誠吾

主指導教員 鵜木 祐史
審査委員主査 鵜木 祐史
審査委員 赤木 正人
党 建武
吉高 淳夫

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3年2月

Abstract

Maintenance and inspection work is essential for the safe operation of machine in factories. This work is usually carried out by machine maintenance technicians, but there are various problems such as ensuring safety and maintaining costs. In order to solve these problems, technology for anomaly detection is being developed. This technology generally consists of feature extraction and detectors. In particular, for sound-based anomaly detection, emphasis has been placed on the development of highly accurate detectors using machine learning. However, in order to establish an anomaly detection method with higher performance than existing techniques, it is necessary to examine what acoustic features and perceptual factors are important for anomaly detection.

A sound consists of three attributes: loudness, pitch, and timbre, which represent the three aspects of sound as an auditory impression. Unlike loudness and pitch, the psychological properties of timbre are expressed in a multidimensional manner and are complex. The problem of difficulty in anomaly sound detection may be due to the fact that only skilled technicians can identify the acoustic features and perceptual factors that are important for anomalous sound detection.

The purpose of this study is to clarify whether the evaluation index related to timbre is important for anomaly sound detection. In this study, Timbral Attribute used to investigate this issue. A timbre attribute is an adjective that describes the characteristics of a timbre, and is an index that is associated with the psychological quantity of a person's perception of sound. In this study, skilled engineers is considered that they rely on the complex differences in timbre between normal and abnormal sounds to judge abnormalities, and investigate which of the Timbral Attributes are important in anomaly sound detection.

The dataset of ToyADMOS and MIMII is used in this paper. ToyADMOS is a dataset created by NTT Media Intelligence Laboratory and designed to detect abnormal machine operation sounds. It contains normal and abnormal sounds for three types of toys. In this study, the ToyTrain is used, which is data for model trains. The MIMII dataset was used in DCASE together with ToyADMOS, and is an industrial equipment sound dataset for anomaly sound detection.

The proposed method consists of a feature extraction part and a detector. In the feature extraction part, eight timbral attribute are calculated using Timbral models. For each of the eight calculated values, all combinations (255 possible combinations) are extracted and input to the detector. In the detector, a support vector machine is used to detect between normal and abnormal.

In the study on ToyADMOS, the purpose is the timbre attribute can be used for anomaly sound detection using a simple detector such as threshold judgment by Receiver Operatorating Characteristic curve. In addition, the effectiveness of the

proposed method is clarified by comparing the proposed method with the aforementioned results. In addition, the results of the proposed method is compared with the results of Mizuno’s research to clarify whether the timbre attribute can be used for anomaly sound detection as well as the sound quality metrics.

In the study on the MIMII dataset, the purpose is the proposed method is effective for industrial equipment, instead of the toy failure sound used in the ToyADMOS study. In order to evaluate the effectiveness of the proposed method, we compared it with Baseline, which uses log-Mel-spectrograms as features. All the sound data in the MIMII dataset was used for the evaluation.

The results for the study on the ToyADMOS, Brightness and Sharpness showed the difference between normal and abnormal sounds except for one abnormal sound. The proposed method was able to detect all the normal and abnormal sounds.

Thre results for the study on the MIMII dataset, the proposed method was found to be more effective than the log-mel-spectrogram for all the individuals in the Valve and some of the individuals in the Pump and Slide rail.

By using Timbral attribute as a feature, high identification accuracy was achieved. Through the two studies, it is found that Brightness was the most frequently used evaluation metric in the proposed method. This indicates that Brightness is the most important of the timbral attributes. In addition, the number of dimensions of the features used in the proposed method is eight at most. From this, it is concluded that the timbral attribute represents the difference between normal and abnormal in some models better than other features that use a huge number of dimensions.

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	2
1.4 論文構成	3
第2章 関連研究	6
2.1 DCASE	6
2.1.1 概要	6
2.1.2 ToyADMOS	7
2.1.3 MIMII データセット	10
2.2 異常音検知の研究	14
第3章 特徴の抽出	15
3.1 音響特徴	15
3.2 音色属性に基づく特徴	15
第4章 提案法	18
4.1 特徴抽出部	20
4.2 識別器	20
4.3 提案法のコンセプト	20
4.4 提案法の評価方法	20
第5章 ToyADMOS に対する検討	22
5.1 目的	22
5.2 方法	22
5.3 結果	24
5.4 考察	31
第6章 MIMII データセットに対する検討	32
6.1 目的	32
6.2 方法	32
6.3 結果	32

6.4 考察	36
第7章 結論	37
7.1 明らかにしたこと	37
7.2 残された課題	37
参考文献	37
謝辞	42
研究業績	43

目次

1.1	本論文の構成	5
2.1	ToyTrain のスペクトログラムの例 (上:正常音, 下:異常音)	9
2.2	MIMII データセットのスペクトログラムの例 ((a-1,2) :Fan の正常・異常音, (b-1,2) :Pump の正常・異常音, (c-1,2) :Slide rail の正常・異常音, (d-1,2) :Valve の正常・異常音)	13
4.1	提案法のブロックダイアグラム	19
5.1	各評価指標の算出結果 (青:正常音, 赤:異常音)	25

表 目 次

2.1	ToyADMOS の構成 [20]	8
2.2	MIMII データセットの構成	11
2.3	MIMII データセットの操作の内容と異常の条件	12
5.1	分析に用いた音源と異常状態 [24]	23
5.2	音色に係わる各評価指標による識別精度	26
5.3	音色に係わる各評価指標によるテストデータに対する識別精度	27
5.4	水野の報告による識別精度 [24]	28
5.5	提案法と Baseline の比較評価結果	29
5.6	特徴選択の結果	30
6.1	特徴選択の結果	34
6.2	提案法と Baseline の比較評価結果	35

第1章 序論

1.1 はじめに

人の身の回りには常に、音声や音楽、車の音や動物の鳴き声などあらゆる音が存在している。他人との会話や音楽は、人の生活を豊かにする。また、車の警笛音や人の悲鳴を聞いて危険を予測したり、自然の音や人々が発する生活音を聞いて周囲の状況を把握したりできる。このように人は聴覚によって、目的とする音だけでなく、その音を取り巻く音環境などのさまざまな情報を取り入れて周囲の状況を把握している。

人の聴覚機能を計算機によって実現することを目指す研究分野として、音環境理解がある [1]。音環境理解とは、音響信号に基づいてその場の状況を把握する課題全般のことであり、音源の分離技術を中心に研究が進められている [2]。音環境理解では、音声に限らず、音楽や環境音などあらゆる音を検出・識別する技術についての研究もさかんに行われている。例えば、音の種類を分析するような技術は音響イベント分析、周囲の環境を分析するような技術は音響シーン分析と呼ばれる [1]。音響シーンとは、音が収録された状況や場所、周囲にいる人の行動などの情景のことである。また、音響イベントとは音響シーンよりも細かい、音の種類のことである。音響イベント・音響シーンを説明するための例として、会議の場で収録された音を挙げる。この場合、討論をしている状況や会議室などが音響シーンにあたり、キーボードの打鍵音や特定の人物の発言などが音響イベントにあたる。

音響イベント・音響シーン分析には、人間の生活を音を用いて記録するライフログ [3] や、身の回りの音を収集し機械の保全や防犯等を目的とした監視システム [4]、音を用いて高齢者や幼児に降りかかる危険の回避を目的とした見守りシステム [5]、映像や音楽への自動タグ付与 [6] など、様々な活用例がある。活用例の一つに、音からその場や対象そのものの異常を判断する、異常音検知がある。異常音検知の検知対象としては、銃声 [7] や転倒音 [8,9]、産業用機器の稼働音 [10-12] などがあげられる。特に産業用機器に対する異常音検知については、故障した状態での稼働による機械トラブルや製造機会損失などを避けるという研究意義があり、機械産業を支える上で重要視されている。

1.2 研究背景

工場の機械設備を安全に稼働し続けるには、保守・点検作業が必要不可欠である。この作業は、通常、技術士や機械保全技能士 [13] らのように資格を有する人など、熟練の技術者の手によって行われている。熟練技術者は、機械設備が発せられる音のかすかな違いを知覚することができ、異常を判断している [14]。人の五感の中でも聴覚による異常発見件数は多いことから、聴覚による異常検知は保守・点検作業において重要であることがわかる。

しかし、人の手でこのような作業を行う場合には問題点もある。例えば、工場稼働している機械設備から発せられる騒音による騒音障害 [15] がある。騒音環境下では、技術者同士の連絡や合図が妨害され、安全に作業を行えないことがある。また騒音に晒され続けると耳の機能が損なわれることがあり、最悪の場合、騒音性難聴を発症してしまうことがある。騒音による人体への影響は甚大であり、聴力に対する障害だけでなくその他の生体機能に対する影響もあると報告されている [16]。また、作業をするにあたって機械設備に近づくことが多く、安全性の確保ができないという問題もある。他にも、技術士や機械保全技能士のように技術を有する人材を雇用するコストの維持が難しいことや、熟練技術者の後継者が不足していることなど、さまざまな問題が挙げられる。

これらの問題を解決するために人間にとって代わる、計算機を用いた異常状態検知技術の開発が行われている。異常状態検知とは、対象となるものが正常な状態にあるのか、あるいは異常な状態にあるのかを判断して検知する技術のことである。この技術は、一般に、特徴抽出部と識別器で構成される。

特に、音を対象とした異常状態検知（異常音検知）では、機械学習による高精度な識別器の開発に重点が置かれている。しかし、異常音検知をする際には、対象とする音の特徴も考慮する必要があると考えられる。音は大きさ、高さ、音色の3属性からなり、これらは聴覚的印象としての「音」が有する3つの側面を表す [17]。その中でも音色が持つ心理的な性質は、音の大きさや高さとは異なり多次元的であり、複雑な性質を持つ。異常音検知が難しい問題は、正常音と異常音の複雑な音色の違いを、熟練技術者にしか聞き分けられないということに起因すると考えられる。そこで本研究では、識別器よりも特徴抽出部に着目し、深く検討する必要があると考えた。

1.3 研究目的

本研究の目的は、音色に係わる評価指標が異常音検知において重要かを明らかにすることである。そこで本研究では、音色属性 (Timbral Attribute: TA [18]) を用いて検討を行う。音色属性とは、音色の特徴を形容詞等で表したものであり、音に対する人の心理量と結びつきがある指標である。本研究では、熟練技術者は正

常音と異常音の複雑な音色の違いを頼りに異常を判断していると考え、音色属性のうちどのような指標が異常音検知において重要かを検討する。

1.4 論文構成

本論文は7章で構成される。図 1.1 に本論文の構成図を示す。

第1章

音響イベント・音響シーン分析の一つである、異常音検知の研究背景と問題点を述べる。その上で、本研究の目的について言及する。

第2章

異常音検知の関連研究について述べる。ここでは、音響イベント・音響シーン分析分野のコンペティションである DCASE の概要や、本研究で用いるデータセットの説明を行う。また、異常音検知に関する従来の研究を説明する。

第3章

本研究では特徴の抽出に着目し、音色に係わる評価指標が異常音検知において重要かを明らかにする。そこで、従来の研究で用いられてきた音響特徴や、本研究で用いる音色属性に基づく特徴について述べる。

第4章

本研究の提案法について述べる。提案法は、従来法と同様に特徴抽出部と識別器で構成される。本研究で用いる特徴抽出部と識別器について述べたあと、提案法のコンセプトを述べる。

第5章

おもちゃの故障音が収録された、ToyADMOS のデータセットに対する検討を行う。音色属性に係る指標および提案法が、おもちゃの故障音に対して有効かどうかを検討する。

第6章

音色属性に係る指標および提案法が、実際の産業用機器のデータを用いた場合にも有効かどうかを検討する。そこで、産業用機器の異常音が収録された、MIMII

データセットに対する検討を行う。

第7章

第5章，第6章の結果を踏まえ，全体考察を行う。

第8章

本研究で明らかにしたこと，残された課題を述べる。

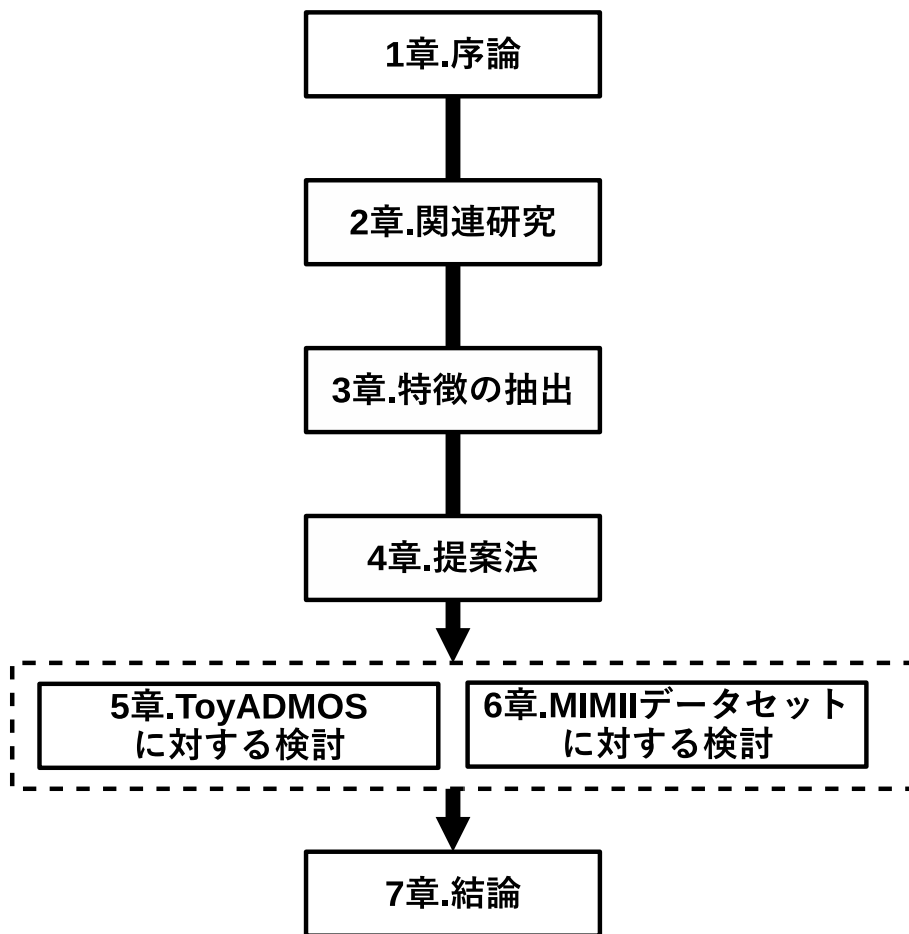


図 1.1: 本論文の構成

第2章 関連研究

2.1 DCASE

2.1.1 概要

DCASE (Detection and Classification of Acoustic Scenes and Events) とは, IEEE AASP TC (IEEE Audio and Acoustic Signal Processing Technical Committee) が公認する, 音響イベント・音響シーン分析の分野で最大の国際コンペティションである. DCASEでは多分野にわたったタスクが用意されており, それぞれで識別精度を競う. タスクのうちの1つである task2 では, 複数の機械を対象として, それらの機械音が正常か異常かを識別する, 異常音検知のタスクが用意されていた [19]. 既知データとして様々な機種 of 正常音のみを用いる教師なし学習を行うタスクであり, 未知データとなる正常音・異常音に対する識別精度が競われた.

タスクに使用されたデータは, ToyADMOS [20] に収録された ToyCar, ToyConveyor の2種類と, MIMII データセット [21] に収録された Fan, Pump, Slide rail, Valve の4種類の計6種類である. ToyADMOS はおもちゃを対象として正常音・異常音を収録したデータセットであり詳細な説明を本節の第2項で述べる, MIMII データセットは産業用機器を対象として正常音・異常音を収録したデータセットであり, 詳細な説明を本節の第3項で述べる. DCASE で用いられたデータは ToyADMOS と MIMII データセットの一部で構成されている.

DCASE2020 ではベースラインモデルが用意されており, 特徴量には対数メルスペクトログラムが, 識別器にはオートエンコーダが用いられた. DCASE2020task2 にて最も高い検出性能を示した Ritwik らの報告 [22] では, 特徴量として Baseline と同じく対数メルスペクトログラムを用いていた. 識別器として用いたオートエンコーダの出力を混合正規分布とみなし, 確率分布を用いた確率分布推定器 (Masked Autoencoder for Distribution Estimation) と MobileNetV2 を用いた自己教師あり学習を用いて異常音検知を行った. また, 多くの報告で特徴量として対数メルスペクトログラムが用いられていた.

2.1.2 ToyADMOS

DCASEで使用されたデータの一つである ToyADMOS は、NTT メディアインテリジェンス研究所が作成した、機械操作音 (ADMOS) の異常を検出するために設計された大規模データセットである。ToyADMOS には、ToyCar (自動車)、ToyConveyor (コンベア)、ToyTrain (列車) の3種類のおもちゃの正常音と故障させた上で動作させた際の異常音が収録されている。ToyCar は製品の検査、ToyConveyor は固定機械の故障診断、ToyTrain は移動機械の故障診断を想定して収録されている。

ToyCar にはミニ四駆と呼ばれるおもちゃが利用されており、2種のモータと2種のベアリングが用意されている。モータの動力を伝えるギアの故障やベントシャフトの歪みなど複数の異常の条件が設定されている。ToyConveyor は3種のおもちゃのコンベアが利用されており、コンベアベルトの緩みや異物の混入などの異常の条件が設定されている。ToyTrain は先頭車両と後部車両からなるおもちゃの列車が利用されており、車輪の破損や線路の非連結などの以上の条件が設定されている。

すべての機種に対して収録されているデータには2種類あり、機械の開始時間と終了時間が設定されている Individual (IND) と、機械を継続して動作させたデータを切り取った Continuous (CNT) が用意されている。ToyADMOS には、約540時間の正常音と、12,000以上の異常音のデータが含まれている。サンプリング周波数は48 kHzであり、4つのマイクを用いて収録されている。表2.1に ToyADMOS データセットの詳細な情報を示す。

3種の機器がどのような音であるかを表すスペクトログラムの例を図2.1に示す。窓幅が40ミリ秒のハミング窓で、スライド幅が20ミリ秒としてスペクトログラムを求めた。

表 2.1: ToyADMOS の構成 [20]

Sound	ToyCar	ToyConveyor	ToyTrain
IND normal sounds	1,350 samples	1,800 samples	1,350 samples
CNT normal sounds	≈ 150 samples	at least 124 samples	74 samples
IND anomalous sounds	≈ 250 samples	355 samples	270 samples
CNT anomalous sounds	72 samples	72 samples	72 samples

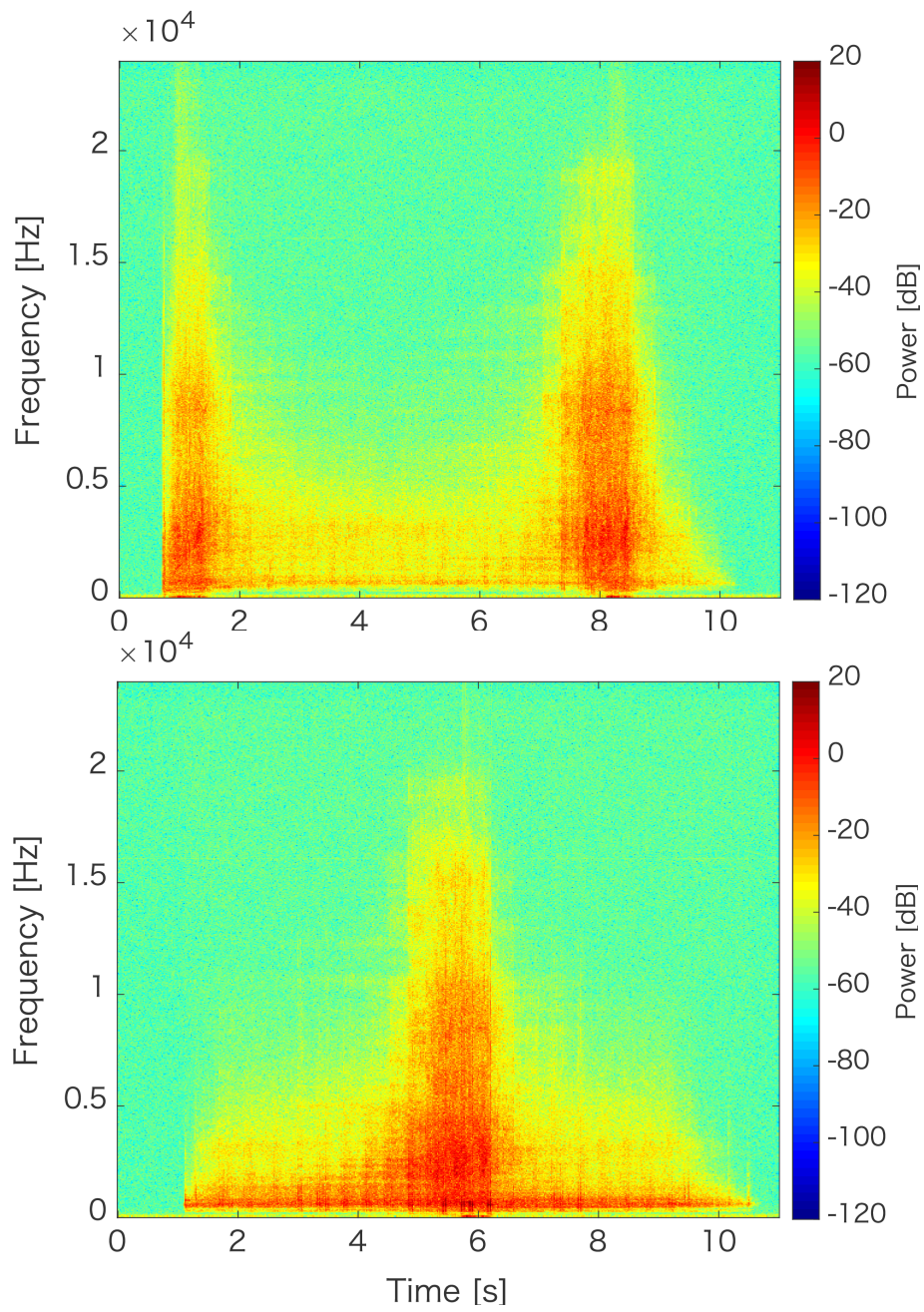


図 2.1: ToyTrain のスペクトログラムの例 (上:正常音, 下:異常音)

2.1.3 MIMII データセット

ToyADMOSと同様にDCASEで使用されたデータの一つであるMIMIIデータセットは、日立製作所が作成した、異常音検知用の産業機器音データセットである。MIMIIデータセットはオープンデータセットであり、機械学習と信号処理のコミュニティによる自動設備の開発支援を目的として公開されている。MIMIIデータセットには、各6種の個体の音データが収録されており、開発用データとしてFan, Pump, Slide rail, valveの4種類が公開されている.. 2.2にMIMIIデータセットの開発用データの詳細な情報を示す。Fanは産業用ファンを表しており、工場内のガスや空気の連続的な流れを作り出すために使用されている。Pumpはプールから水を吸入し、放水を連続的に行うポンプである。Slide railは、移動するプラットフォームと台座からなるスライドシステムを表しています。Valveはソレノイドの素早い開閉を繰り返すバルブが用いられている。機械によって停止状態を持つもの、持たないものが存在し、それぞれの機械が独自の特徴を持っている。

MIMIIデータセットには、4種類の機械の正常時と異常時の稼働音が収録されている。異常の条件は機種ごとに設定されており、異物の混入やレールの損傷など、さまざまな種類が含まれている。MIMIIデータセットにおける各機種の故障の条件を2.3に示す。これらの異常の条件は現実で起こりうる故障を想定して設定してある。

データセットは8個のマイクロホンからなる円形のマイクロホンアレイであるTAMAGO-03マイクを用いて収集されている。マイクロホンアレイは機械から50cm(バルブの場合は10cm)の距離に置いて収録されている。データ長はすべて10秒で、サンプリング周波数は16,000 Hzである。4種の機器がどのような音であるかを表すスペクトログラムの例を図2.2に示す。ToyADMOSの例と同様に、窓幅が40ミリ秒のハミング窓で、スライド幅が20ミリ秒としてスペクトログラムを求めた。

表 2.2: MIMII データセットの構成

Machine	ID	Normal	Anomalous	Samp.Freq. [Hz]	Duration [s]
Fan	00	1,011	407	16,000	10
	02	1,016	359	16,000	10
	04	1,033	348	16,000	10
	06	1,015	361	16,000	10
Pump	00	1,006	143	16,000	10
	02	1,005	111	16,000	10
	04	702	100	16,000	10
	06	1,036	102	16,000	10
Slide rail	00	991	119	16,000	10
	02	708	120	16,000	10
	04	1,000	120	16,000	10
	06	992	120	16,000	10
Valve	00	1,068	356	16,000	10
	02	1,068	267	16,000	10
	04	534	178	16,000	10
	06	534	89	16,000	10

表 2.3: MIMII データセットの操作の内容と異常の条件

機種	操作の内容	異常の条件
Fan	通常運転	不安定動作, 電圧変化, 目詰まり, etc.
Pump	プールからの水の吸入と放水	水漏れ, 異物の混入, 目詰まり, etc.
Slide rail	スライドの繰り返し	レール損傷, ベルトの緩み, 潤滑油不足, etc.
Valve	弁の開閉の繰り返し	2種類以上の異物混入

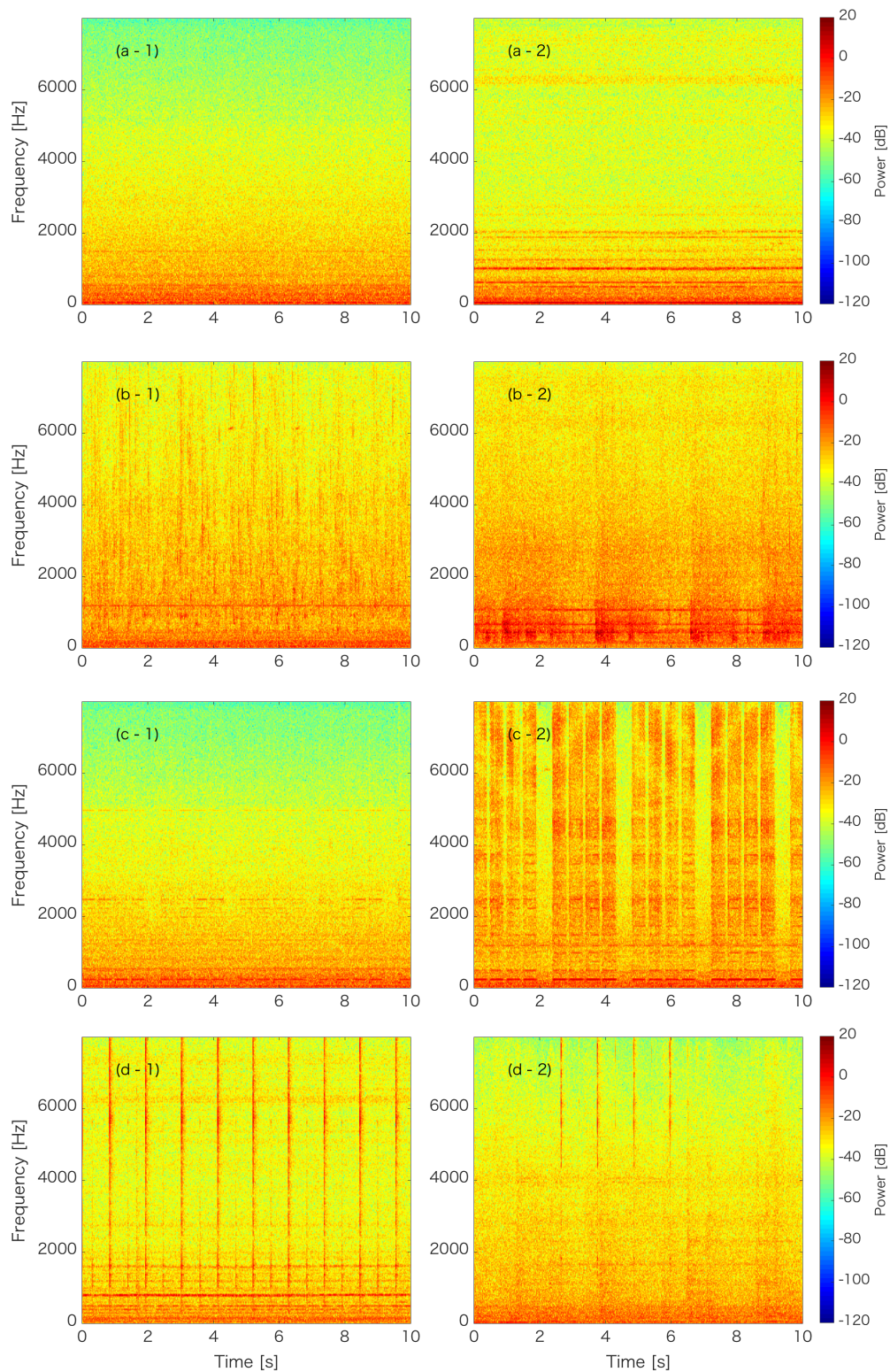


図 2.2: MIMII データセットのスペクトログラムの例 ((a-1,2) :Fan の正常・異常音, (b-1,2) :Pump の正常・異常音, (c-1,2) :Slide rail の正常・異常音, (d-1,2) :Valve の正常・異常音)

2.2 異常音検知の研究

現在までに、異常音検知に関する研究成果が、多数報告されている。異常音検知の諸問題として、異常音自体の大量な収集が困難なことが挙げられる。この問題を解決するために、正常音のみの大量なデータを用いて、機械学習により正常音の特徴を学習することで異常音検知を実現する研究が報告されている [10, 23]。植松らの深層学習を用いた異常音検知の研究 [10] では、スペクトログラムを特徴量として、放水ポンプ、3D プリンタ、給水ポンプに対して異常音検知を可能とした。従来の研究では、末房らの報告 [23] では、教師なしの異常音検知においてよく用いられるオートエンコーダが抱える問題についての検討が行われた。オートエンコーダは正常音を再合成するよう学習を行い、異常音を入力した際に異なる出力をさせることで、外れ値検出を行っている。しかし、対象となる機械音が非定常な場合には、再合成による予測が難しいという問題から、再構成誤差が異常とは無関係に大きくなる傾向にある。その問題を解決するために入力となるメルスペクトログラムをフレームごとに分割し、中心フレームを除いたものを入力とし、ディープニューラルネットワークを用いて設計されたオートエンコーダーに入力される。出力として再合成された特徴量を入力の際除いた中心フレームで補完する。結果、再合成誤差を正常音と異常音で一致させることができ、特に非定常機械音に対しての識別精度の改善がなされた。このように、音データからスペクトログラムやメルスペクトログラムを抽出し利用することが多く、どのようにこれらの代表的な特徴量を用いて異常音検知を行うかが報告として挙げられている。

異常音検知を行う際、特徴抽出部に着目した研究として、水野の研究がある [24]。この研究では、特徴量として音色に係る指標の一つである音質評価指標 [25–27] が用いられている。音質評価指標とは、客観的にヒトの知覚に係る主観的な音質を評価するために用いられる指標である。音質評価指標を代表する指標には、変動強度 (fluctuation strength)、ラフネス (roughness)、シャープネス (sharpness) などが挙げられる。シャープネスは音の鋭さや甲高さに対応し、ラフネスは音の粗さに対応し、変動強度は変動感の強さに対応する [28, 29]。この研究では、前述の ToyADMOS を用いて検討がなされた。その結果、高い精度で異常音検知が可能であった。また異常音は、ラフネスや変動強度といった音の変動に関する指標に関係することが明らかになり、異常音検知における音色の重要性が示された。しかしながら、一部の音色に係る指標のみの検討に限られているため、どのような音色に係わる指標が異常音検知に重要であるか、その解明には至っていないといえる。

第3章 特徴の抽出

3.1 音響特徴

異常音検知は、一般的に特徴抽出部と識別部に分かれており、特徴抽出部にて元データから特徴を抽出する。従来の研究では、パワーや周波数スペクトル、メル周波数ケプストラム係数 (MFCC) といった音響特徴が利用されてきた [30]。DCASE task2 のベースラインモデルでも用いられた対数メルスペクトログラムは MFCC を求める際に算出されるメルスペクトログラムを対数化したものである、MFCC は次の手順で求められる。

まず、音響信号に対して短時間フーリエ変換を用いることで、スペクトログラムを算出する。ここで、スペクトログラムを人の聴覚特性を考慮したメル尺度に変換する。メル尺度は式 (3.1) で表される。

$$f_{\text{mel}} = 2595 \log_{10} \left(1 + \frac{f}{700} \right) \quad (3.1)$$

MFCC を求める際によく用いられる方法として、周波数領域で定義された L 個の三角窓であるメルフィルタバンクを用いて、スペクトログラムからメルスペクトログラムに変換する。これに離散コサイン変換をかけることで、MFCC が得られる。式 (3.2) に MFCC の次数ごとの算出式を示す。

$$c_i = \sqrt{\frac{2}{N}} \sum_{l=1}^L A_l \cos \left[\left(l - \frac{1}{2} \right) \frac{i\pi}{L} \right] \quad (3.2)$$

N はフレーム長、 A_l は対数フィルタバンクの振幅、 i は MFCC の次数を表す。

3.2 音色属性に基づく特徴

音色に係るとされる音響特徴はいくつかある。代表的なものに、「明るさ」に係るスペクトル重心や、スペクトル傾斜などがあげられる。また、音響特徴ではなく、音の明るさや固さといった音に対する感覚を物理量である音響特徴から算出する指標がいくつか報告されている。音色属性 (Timbral attribute: TA [18]) とは、音色の特徴を形容詞等で表したものである。その中の一部が、サリー大学の AudioCommons というプロジェクトにて Timbral models として実装されている。特に、Hardness,

Depth, Brightness, Roughness, Warmth, Sharpness, Boominess, Reverb の8つの評価指標が実装されている。様々な音源を利用できる freesound.org で人が音源を検索する際に最も利用された検索語が属する音色属性であったため実装された。Reverbを除く各モデルは0から100までの値をとるように実装されており、例として Hardness なら、値が高いほどより硬い音とされる。

Hardness は音の固さを表す指標であり、Williams [31] や Freed [32] , Solomon [33] の報告をもとに実装された、検索語としては ‘hard’, ‘pillowy’, ‘soft’ が属している [34]。スペクトル傾斜やスペクトル重心などのスペクトルに関する情報や、アタックタイムなどから算出される。

Depth は音の奥行きを表す指標であり、Cartwright ら [35] の報告をもとに実装された。検索語としては ‘deep’ が属している [34]。depth は式 (3.3) で示す帯域制限をした spectral centroid や、式 (3.4) で示す特定の周波数帯域の比から算出される。

$$\text{Lower spectral centroid} = \frac{\sum_{n(30\text{Hz})}^{n(200\text{Hz})} f(n)x(n)}{\sum_{n(30\text{Hz})}^{n(200\text{Hz})} x(n)} \quad (3.3)$$

$$\text{Lower ratio} = \frac{\sum_{n(30\text{Hz})}^{n(200\text{Hz})} f(n)x(n)}{\sum_{n(30\text{Hz})}^{n(\text{Nyquist})} x(n)} \quad (3.4)$$

20 ~ 500 Hz の周波数成分が多く含まれ、基本周波数が低く、音の立ち上がり後にエネルギーの最大となる地点が早い音ほど奥行きを感じる。

Brightness は音の明るさ感を表す指標であり、Pearce [36] の報告をもとに実装された。検索語としては ‘bright’, ‘brilliant’, ‘dark’, ‘dull’, ‘polished’, ‘shining’ が属している [34]。式 (3.5) で示す帯域制限をした spectral centroid と式 (3.6) で示す特定の周波数帯域 RatioB から算出される。

$$\text{Frequency – limited spectralcentroid} = \frac{\sum_{n(3\text{kHz})}^{n(\text{Nyquist})} f(n)x(n)}{\sum_{n(3\text{kHz})}^{n(\text{Nyquist})} x(n)} \quad (3.5)$$

$$\text{RatioB} = \frac{\sum_{n(3\text{kHz})}^{n(\text{Nyquist})} x(n)}{\sum_{n(20\text{Hz})}^{n(\text{Nyquist})} x(n)} \quad (3.6)$$

3,000 Hz 以上の周波数成分が多く含まれる音ほど明るさを感じる。

Roughness は音の粗さを表す指標であり、Vassilakis [37] の報告をもとに実装された。検索語としては ‘rough’, ‘smooth’ が属している [34]。式 (3.7) が roughness を表し、式 (3.8), 式 (3.9), 式 (3.10) で示される X, Y, Z から算出される。Z の算出に用いる s は式 (3.11) で算出される。

$$r = 0.5X^{0.1}Y^{3.11}Z \quad (3.7)$$

$$X = A_{\min} * A_{\max} \quad (3.8)$$

$$Y = \frac{2A_{\min}}{A_{\min} + A_{\max}} \quad (3.9)$$

$$Z = e^{(-3.5s(f_{\max}-f_{\min}))} - e^{(-5.75s(f_{\max}-f_{\min}))} \quad (3.10)$$

$$s = \frac{0.24}{0.0207f_{\min} + 18.96} \quad (3.11)$$

50 ms ごとに時間分割し、窓ごとの周波数スペクトルのピークを求める。 A_{\max} と A_{\min} はピークどうしの距離が最大になる対と最小になる対の距離である。 f_{\max} と f_{\min} はピークとなる周波数が最大のもの、最小のものを指す。変調周波数が 70 Hz に近い音ほど粗さを感じる [29]。

Warmth は音のあたたかさを表す指標であり、80 の刺激にわたるあたたかさの主観的評価に基づいて実装された。検索語としては ‘cold’, ‘warm’ が属している [34]。20 ~ 40 Hz の周波数成分が多く含まれる音ほど暖かさを感じる。

Sharpness は音の甲高さを表す指標であり、Fastl のシャープネスモデル [38] をもとに実装された。検索語としては ‘blunt’, ‘sharp’ が属している [34]。Sharpness で利用する Fastl のシャープネスモデルは式 (3.12) で算出される。

$$S = \frac{\int_0^{24\text{Bark}} N'(z)g'(z)z dz}{\int_0^{24\text{Bark}} N'(z) dz} \quad (3.12)$$

$N'(z)$ は臨界帯域ごとのラウドネスを表し、 $g'(z)$ はシャープネスの重みを表し、 z は臨界帯域番号を示す。正弦波信号の場合、周波数が高いほどシャープネスを高く感じる [28]。

Boominess は音のブーミング感を表す指標であり、Hatano ら [39] の報告をもとに実装された。Hashimoto booming index を直接実装したものになり、動いている車内のブーミング感を測定するために設計された。検索語としては ‘booming’, ‘boomy’ が属している [34]。重みづけされた 1/3 オクターブバンドのパワーの合計を用いて算出される。シャープネスとは反対に低い周波数ほどブーミング感を感じる。

Reverb は音の残響感を表す指標であり、Jan ら [40] の報告をもとに実装された。検索語としては ‘dead’, ‘intimate’, ‘roomy’, ‘reverberant’, ‘dry’, ‘wet’ が属している [34]。

本稿ではこれらの音色属性を特徴量として用いた。

第4章 提案法

図4.1に提案法のブロックダイアグラムを示す。提案法は、従来法と同様に、特徴抽出部と識別器で構成される。Timbral attribute extractionが特徴抽出部を表し、Classifierは識別器を表す。

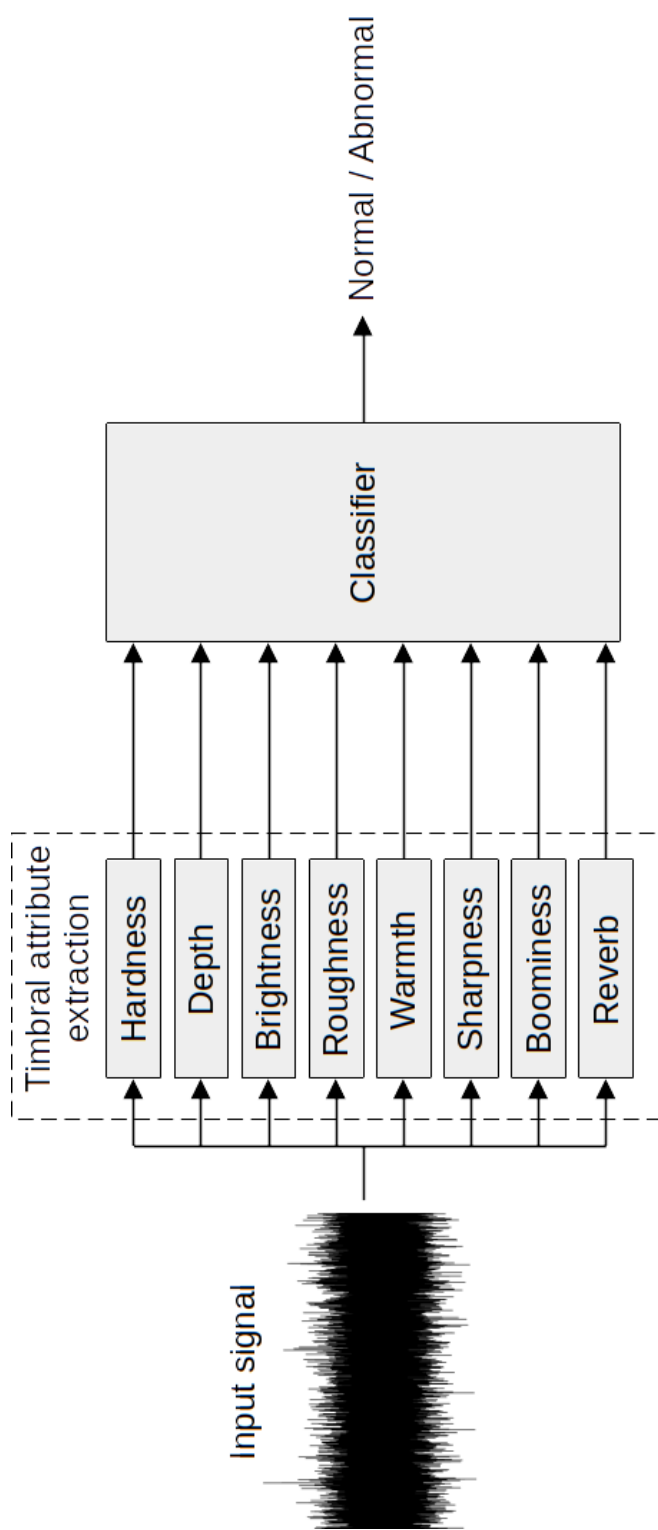


図 4.1: 提案法のブロックダイアグラム

4.1 特徴抽出部

特徴抽出部では2.2節で述べた Timbral models を用いて8つの指標を算出する。算出された8つの値に対してすべての組み合わせ（255通り）を抽出し識別器に入力される。

4.2 識別器

識別器にはサポートベクターマシン (SVM) [41] を用い、正常・異常を判別する。SVMは2クラスのパターン識別器の一つであり、カーネルトリックを用いて非線形の識別関数を構成できる。SVMの実装には、Pythonのオープンソース機械学習用ライブラリである scikit-learn を用いて実装した。識別器は学習するセクションと信号を入力して異常音判定するセクションに分かれている。学習するセクションでは、訓練データを入力し、用いる Timbral attribute のすべての組み合わせ（255通り）で学習を行い、最も精度が高くなった組み合わせで訓練済みモデルとして利用する。異常音判定するセクションでは、テストデータに対して Timbral models を用いて特徴抽出し、訓練済みの識別器を用いてテストデータに対して正常・異常の判定をする。

4.3 提案法のコンセプト

特徴抽出部において、音色属性のすべての組み合わせを抽出し、学習・テストを行うため、その機種において重要となる音色属性の組み合わせを明らかにすることができる。また、その機種に対して最も重要となる音色属性の組み合わせで評価を行うため、識別精度の向上が期待できる。

4.4 提案法の評価方法

提案法の識別精度を評価する際、正解率、偽陽性率、偽陰性率、F値を用いた。異常を正しく判定できた標本数を TP、正常を正しく判定できた標本数を TN、異常を誤って判定した標本数を FP、正常を誤って判定した標本数を FN とする。正解率 (Accuracy) は式 (4.1) で示される。

$$\text{正解率} = \frac{TP + TN}{TP + TN + FP + FN} \quad (4.1)$$

偽陽性率 (False positive rate:FPR) は異常の標本総数のうち誤って判定した割合を表し、式 (4.2) で示される。

$$\text{偽陽性率} = \frac{FP}{TP + FP} \quad (4.2)$$

偽陰性率 (False negative rate:FNR) は正常の標本総数のうち誤って判定した割合を表し、式 (4.3) で示される。

$$\text{偽陰性率} = \frac{FN}{TP + FN} \quad (4.3)$$

F 値 (F-measure) は式 (4.4) で示される適合率 (Precision) と、式 (4.5) で示される再現率 (Recall) の調和平均から求められる。適合率は異常を判定した標本のうち、実際に異常である標本の割合を意味し、再現率は実際に異常である標本のうち、異常であると判定された標本の割合である。また、適合率と再現率の間にはトレードオフの関係があり、それらをまとめて評価する際に F 値を利用する。F 値の計算式を式 (4.6) に示す。

$$\text{適合率} = \frac{TP}{TP + FP} \quad (4.4)$$

$$\text{再現率} = \frac{TP}{TP + FN} \quad (4.5)$$

$$\text{F 値} = \frac{2\text{Recall} \cdot \text{Precision}}{\text{Recall} + \text{Precision}} \quad (4.6)$$

本研究ではこれらの評価尺度を提案法を含めすべての評価対象に用いた。

第5章 ToyADMOSに対する検討

5.1 目的

本章ではROC曲線による閾値判定のような単純な識別器を用いて、音色属性を異常音検知に利用することが可能であるかどうかを明らかにする。

また、提案法と前述の結果の比較を行い、提案法の有効性を明らかにする。そのうえ、水野の研究で得られた結果と比較して音色属性が音質評価指標と同様に異常音検知に利用可能かどうかを明らかにする。最後に、対数メルスペクトログラムを用いたBaselineと提案法による識別精度を比較し、他の特徴量と比較した際の提案法の有効性を明らかにする。

5.2 方法

分析対象は水野の研究と同様にToyADMOS [20]に収録された、ToyTrainのデータセットの一部のデータを利用した。異常の条件として、先頭車両および後部車両の車輪および直線路、曲線路の破損、曲線路への異物の配置、線路の連結解除させた場合が含まれている。本稿では、これらの条件に合致する異常音14個と同数の正常音14個の計24個のデータを利用した。データ長は11秒で、サンプリング周波数は16,000 Hzとした。

Timbral models [18]を用いて特徴抽出を行い、得られた8個の評価指標の分析と識別精度の評価を行った。識別精度の評価方法としては、評価指標の値からROC曲線を書き、真陽性率が1、偽陽性率が0の地点から最短となる地点を異常音検知の閾値として設定した。これらの閾値を利用して24個の音響信号に対し、識別精度を調査した。また、用いたデータは訓練データとテストデータとに分割し、テストデータを用いた際の識別結果を用いて、水野の研究との比較を行った。

最後に、提案法の有効性を評価するために、特徴量として対数メルスペクトログラム(LME)を用いたBaselineと比較を行った。1,024サンプルの窓で、512サンプルのスライド幅でセグメント数が313のスペクトログラムを算出した。その後、128バンドのメルフィルタによって、メルスペクトログラムを算出した。得られたメルスペクトログラムを対数化し、LMEを算出した。得られたLMEの次元数は44,032 (=128 × 344) 次元となった。

表 5.1: 分析に用いた音源と異常状態 [24]

Name	Straight railway track	Curved railway track	First carriage	Last carriage
ab01	Normal	Normal	Normal	Chipped wheel axle
ab03	Normal	Normal	Chipped wheel axle	Chipped wheel axle
ab05	Normal	Disjointed	Normal	Chipped wheel axle
ab07	Normal	Disjointed	Chipped wheel axle	Chipped wheel axle
ab09	Normal	Obstructing stone	Normal	Chipped wheel axle
ab13	Normal	Broken	Normal	Chipped wheel axle
ab15	Normal	Broken	Chipped wheel axle	Chipped wheel axle
ab17	Disjointed	Disjointed	Normal	Normal
ab18	Disjointed	Disjointed	Normal	Chipped wheel axle
ab20	Disjointed	Disjointed	Chipped wheel axle	Chipped wheel axle
ab22	Disjointed	Obstructing stone	Normal	Chipped wheel axle
ab24	Disjointed	Obstructing stone	Chipped wheel axle	Chipped wheel axle
ab26	Disjointed	Broken	Normal	Chipped wheel axle
ab28	Disjointed	Broken	Chipped wheel axle	Chipped wheel axle

5.3 結果

図 5.1 に各評価指標の算出値を示す。横軸はデータ番号，縦軸は異常音検知の閾値をベースとした評価値を示す。青い棒グラフは正常音を，赤い棒グラフは異常音を示す。Brightness と Sharpness が異常音と正常音で 1 つの異常音を除き，違いをよく表していた。また，異常音 ab17 が Hrdness, Brightness, Roughness, Sharpness で最低の値をとり，Warmth, Boominess で最大の値をとる他の異常音と比べ特徴が顕著な値となった。

表 5.2 に各評価指標で求めた閾値での異常音検知精度を示す。Brightness と Sharpness がともに f 値が 0.963 となり，高い精度で異常音を検知できることが分かった。Brightness と Sharpness がともに高い識別精度で異常音を検知できることが分かった。ab17 (直線・曲線レールの非連結) の異常音のみ，正確に検出できなかった。

表 5.3 に訓練データを用いて各評価指標で求めた閾値で，テストデータを対象とした識別精度を示す。PM は提案法における識別精度を示す。24 の音響信号を対象とした識別精度では，Brightness と Sharpness が最も識別精度が高くなったが，テストデータを対象とした識別精度は Brightness のほうが高くなった。

表 5.4 に水野の研究で得られた結果を示す。同じ手法で Brightness を用いた結果と比較して，ラフネスの平均と変動強度の分散がより識別精度が高く，変動強度の分散は提案法と同じく F 値が 1 となり，異常と正常を完全に判定することができていた。

表 5.5 に提案法と対数メルスペクトログラムを特徴量として用いた Baseline の識別精度を示す。PM が提案法を示し，LME が Baseline を示す。 F 値が PM の場合 1 となり，LME と比較してより高い識別精度で識別が可能であった。

提案法において， F 値が最大 (1) となった組み合わせは 68 個あり，それらの中で利用された指標の割合を表 5.6 に示す。Brightness が 0.75 と最も利用頻度が高くなった。

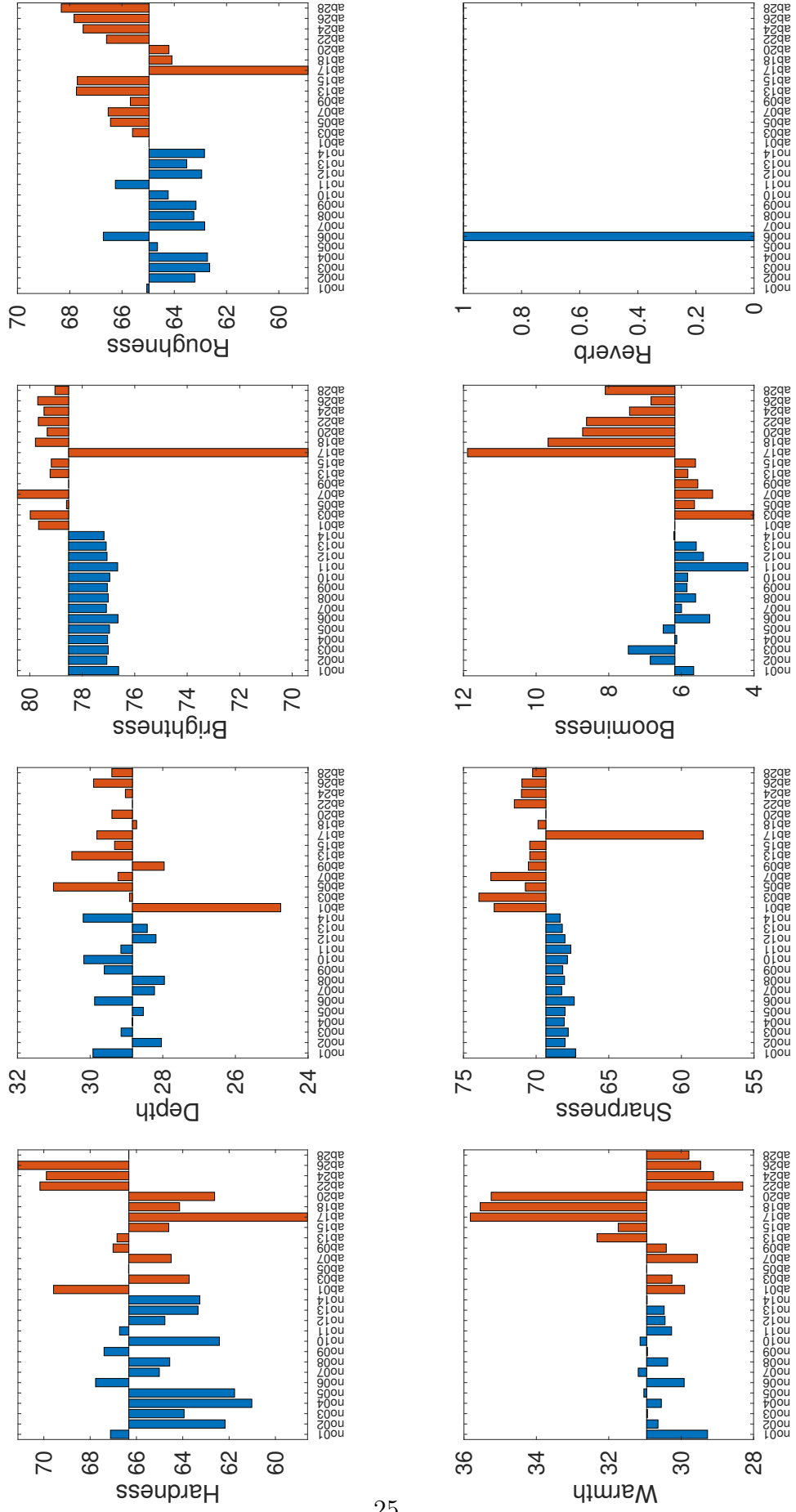


図 5.1: 各評価指標の算出結果 (青:正常音, 赤:異常音)

表 5.2: 音色に係わる各評価指標による識別精度

Timbral attribute	Accuracy [%]	FNR [%]	FPR [%]	F-measure
Hardness	64.3	42.9	28.6	0.615
Depth	64.3	21.4	50.0	0.688
Brightness	96.4	7.1	0.0	0.963
Roughness	78.6	21.4	21.4	0.786
Warmth	60.7	57.1	21.4	0.522
Sharpness	96.4	7.1	0.0	0.963
Boominess	64.3	42.9	28.6	0.615
Reverb	53.6	0.0	92.9	0.683
PM	100.0	0	0	1.0

表 5.3: 音色に係わる各評価指標によるテストデータに対する識別精度

Timbral attribute	Accuracy [%]	FNR [%]	FPR [%]	F-measure
Hardness	50.0	57.1	42.9	0.533
Depth	57.1	57.1	28.6	0.625
Brightness	92.9	0	14.3	0.923
Roughness	71.4	14.3	42.9	0.667
Warmth	64.3	14.3	57.1	0.545
Sharpness	71.4	0	57.1	0.600
Boominess	64.3	71.4	0	0.737
Reverb	50.0	100.0	0	0.667
PM	100.0	0	0	1.0

表 5.4: 水野の報告による識別精度 [24]

Threshold	Accuracy [%]	FNR [%]	FPR [%]	F-measure
Mean of Roughness	85.7	14.3	14.3	0.857
Variance of Roughness	100	0	0	1.000
Variance of Fluctuation strength	93.8	0	11.1	0.933

表 5.5: 提案法と Baseline の比較評価結果

Accuracy [%]		FPR [%]		FNR [%]		F-measure	
PM	LME	PM	LME	PM	LME	PM	LME
100	78.6	0	42.9	0	0	1.000	0.727

表 5.6: 特徴選択の結果

Hardness	Depth	Brightness	Roughness	Warmth	Sharpness	Boominess	Reverb
0.412	0.676	0.750	0.485	0.721	0.676	0.294	0.191

5.4 考察

高い精度で識別ができた Brightness と Sharpness で異常音 ab17 がともに極端に低い値をとった。また、ab17 は他の異常音と比べスペクトルパワーの最大値が小さい傾向にあった。これは、ab17 の異常の条件が直線・曲線路の非連結のため、走行時にのたびに線路の非連結部分で正常状態の走行時とは異なる音が鳴る。ab17 は低い周波数成分が大きく変化している。そのため、音の高さに係る Brightness, Sharpness が極端に低い値をとったと考えられる。音のブーミング感（低周波の音でブンブン響く感覚）に係る Boominess で Brightness や Sharpness とは逆に低い値をとり、ab17 の特徴が顕著に表れている。そのため、ab17 をうまく識別できない問題は閾値を複数設定し、識別器を高精度かすることや、複数の特徴量を用いて学習が可能な識別器を用いることで解決できると考えられる。その結果、提案法では ab17 を含めて完全に正常音・異常音を識別することができたと思われる。

水野が利用した音色に係る評価指標である音質評価指標は、時間情報を利用することができ、その統計量を用いて識別を行っている。このことから、ToyTrain に対する異常音検知においては音色の時間変化が大事なのではと考えられる。

対数メルスペクトログラムを特徴量として用いた Baseline よりも提案法のほうが識別精度が高くなった。これは、特徴量としての対数メルスペクトログラムの次元数が膨大で、少量のデータで用いるには適していないことに起因すると考えられる。

第6章 MIMIIデータセットに対する 検討

6.1 目的

第5章では音色属性に係る指標および提案法が、ToyADMOSに対して有効だということが分かった。本章では ToyADMOS の検討で用いたおもちゃの故障音ではなく、産業用機器でも提案法が有効かどうかを検討する。

6.2 方法

提案法の有効性を評価するために、特徴量として対数メルスペクトログラム (LME) を用いた Baseline と比較を行った。1,024 サンプルの窓で、512 サンプルのスライド幅でセグメント数が 313 のスペクトログラムを算出した。その後、128 バンドのメルフィルタによって、メルスペクトログラムを算出した。得られたメルスペクトログラムを対数化し、LME を算出した。得られた LME の次元数は 40,064 (=128 × 313) 次元となった。

評価には MIMII データセットに収録されているすべての音データを利用した。これらの各データを訓練データとテストデータとに分割し、その割合は 7 対 3 とした。分割する際の正常・異常データの割合は同じとした。訓練データを用いて学習を行い、テストデータを用いて識別精度の評価を行った。

6.3 結果

表 6.1 に各機種にて最も識別精度が高くなった TA の組み合わせを示す。表 6.1 中の 1 が特徴量として使用されているものを示し、0 は特徴量として使用されなかったものを示している。Fan に関して、Roughness と Boominess, Pump に関して Sharpness, Valve に関して Hardness と Brightness がすべての個体で特徴量として利用された。

提案法と Baseline の比較評価結果を、表 6.2 に示す。PM が提案法である最も F 値が高くなった TA 組み合わせでの識別精度を表している。LME が Baseline の結果を示す。表中のハイフン (-) は、F 値を求める際に真陽性率が 0 であるため

に F 値がマイナス無限大に発散するため計算できない場合を示す。評価結果から、提案法は Baseline と比べて 16 機種中 6 機種で F 値が高くなり、全体で 37.5 % 性能が向上することがわかった。

表 6.1: 特徴選択の結果

Machine	ID	Hardness	Depth	Brightness	Roughness	Warmth	Sharpness	Boominess	Reverb
Fan	00	0	1	1	1	1	1	1	0
	02	0	0	0	1	0	1	1	0
	04	0	1	0	1	1	1	1	0
	06	0	0	1	1	1	0	1	0
Pump	00	0	1	1	1	0	1	1	1
	02	0	1	0	1	0	1	1	1
	04	0	0	1	0	1	1	0	0
	06	0	0	1	0	1	1	1	0
Slide rail	00	0	1	1	1	1	1	0	1
	02	1	1	1	0	0	1	1	0
	04	0	1	1	0	1	0	0	0
	06	1	0	0	1	0	1	1	1
Valve	00	1	0	1	1	0	1	0	0
	02	1	1	1	0	1	0	0	1
	04	1	1	1	1	1	0	0	0
	06	1	1	1	1	0	0	1	1

表 6.2: 提案法と Baseline の比較評価結果

Machine	ID	Accuracy [%]		FPR [%]		FNR [%]		F-measure	
		PM	LME	PM	LME	PM	LME	PM	LME
Fan	00	83.3	90.8	50.8	31.1	3.00	0.30	0.63	0.81
	02	86.2	99.3	43.5	1.90	3.30	0.30	0.68	0.99
	04	87.5	98.3	41.0	6.70	2.90	0.00	0.70	0.97
	06	96.1	99.3	13.9	2.80	0.30	0.00	0.92	0.99
Pump	00	96.8	96.2	23.3	30.2	0.30	0.00	0.86	0.82
	02	94.6	97.0	51.5	30.3	0.30	0.00	0.64	0.82
	04	90.0	100.0	73.3	0.00	0.90	0.00	0.40	1.00
	06	97.4	98.8	25.8	12.9	0.30	0.00	0.84	0.93
Slide rail	00	97.7	99.5	7.50	1.90	0.60	0.00	0.95	0.99
	02	93.5	94.8	23.8	26.3	2.20	0.00	0.82	0.85
	04	88.8	95.8	31.5	17.0	4.40	0.00	0.76	0.91
	06	92.0	85.6	48.1	100.0	1.30	0.00	0.65	-
Valve	00	100.0	89.2	0.00	100.0	0.00	0.00	1.00	-
	02	98.4	85.5	11.1	100.0	0.00	0.00	0.94	-
	04	98.8	89.3	11.1	100.0	0.00	0.00	0.94	-
	06	91.0	89.2	83.3	100.0	0.00	0.00	0.29	-

6.4 考察

表6.1の結果から、Fanにてすべての個体でRoughnessとBoominessが利用されていた。これはFanは不安定動作という異常の条件によって起こる回転数の変化がRoughnessとBoominessに大きく係ると考えられる。表6.1の結果から、Pumpにてすべての個体でRoughnessとBoominessが利用されていた。これはPumpは水漏れによる流水量の変化がSharpnessに大きく係ると考えられる。表6.1の結果から、Valveにてすべての個体でRoughnessとBoominessが利用されていた。これは、Valveの異常によってカチカチという音色の変化がHardnessとBrightnessに係ると考えられる。

表6.1の結果から、ToyADMOSに対する検討と同様にBrightnessが全機種を通して最も利用頻度が高かった。この結果から、Timbral attributeの中ではBrightnessが最も異常音検知において重要だと考えられる。

Baselineは、Slide railのid06およびValveの全機種でFPRが1となり、すべてのデータを正常と判定していた。Slide railのid06はValveと似た傾向のある音であり、短い音が間隔をあけて鳴るような機種である。この結果から、LMEではValve等で鳴るカチカチという短い音を表現できないことに起因するからではないかと思われる。

Valveのid00, id02, id04がBaselineと比較して大きくF値の向上が見られた。この結果から、提案法は短い音が間隔をあけて鳴るような機種の異常音検知に対して有効ではないかと思われる。

提案法とBaselineの特徴量の次元数を比較すると、提案法の方が圧倒的に少ない。一部の機種でBaselineと比較して識別精度が上回ったことから、LMEと比較して一部の機種の正常・異常の違いをよく表すと考えられる。また提案法を用いて少ない次元数で異常音検知を行うことで、識別器で計算を行う際の計算コストの削減が期待できる。

第7章 結論

7.1 明らかにしたこと

本研究の目的は、音色に係わる評価指標が異常音検知において重要なかを明らかにすることである。そこで本研究では、おもちゃの故障音と産業用機器に対して異常音検知を音色属性を用いて検討を行った。その結果、おもちゃの故障音に対する検討では Brightness と Sharpness が異常と正常の違いをよく表していることがわかった。産業用機器に対する検討では、提案法が一部の機種に対して対数メルスペクトログラムを特徴量とした異常音検知法よりも識別性能が高くなることが分かった。ToyTrain の正常・異常音には周期性のある短い音が含まれており、異常音の場合にその音が大きくなる傾向にあり、MIMII の Valve は正常・異常音の両方でその傾向が見られた。このことから、短い音が間隔をあけて鳴る機器の異常に対して特に音色属性が有効であるということが言える。また、2つの検討を通じて Brightness が全機種を通して最も利用頻度が高かった。この結果から、Timbral attribute の中では Brightness が最も異常音検知において重要だと言える。

7.2 残された課題

今回は長時間の信号に対する音色属性を用いて検討を行った。しかし今回の特徴抽出の方法では、各音色そのものの変化を表現することはできず、音色の変化に以上の特徴が現れるような異常は検知することができない。そこで、入力信号を時間分割し、各分割後の信号に対して音色属性の値を算出することで音色の変化を表現することができると考えられる。

参考文献

- [1] 井本 桂右, “音響イベントと音響シーンの分析,” 日本音響学会誌, vol. 74, no. 4, pp. 198–207, 2018.
- [2] 柏野 邦夫, “小特集「音環境理解の近年の動向」にあたって,” 日本音響学会誌, vol. 75, no. 9, pp. 510–511, 2019.
- [3] 綾部 櫻子, 田野 俊一, 市野 順子, 岩田 満, 橋山 智訓, “イベントの内容, 感情, 話者情報をロギングするリッチなサウンドライフログの実装と評価,” 研究報告ユビキタスコンピューティングシステム, vol. 2013-UBI-37, no. 31, pp. 1–8, 2013.
- [4] A. Harma, M. F. McKinney, J. Skowronek, “Automatic surveillance of the acoustic activity in our living environment,” 2005 IEEE International Conference on Multimedia and Expo, 2005.
- [5] 児島 宏明, “音による見守りシステム”, 日本音響学会誌, vol. 75, no. 9, pp. 544–550, 2019.
- [6] Zhang, T., Kuo, C. C. J., “Audio content analysis for online audiovisual data segmentation and classification,” IEEE Transactions on speech and audio processing, vol. 9, no. 4, pp. 441–457, 2001.
- [7] G. Valenzise, L. Gerosa, M. Tagliasacchi, F. Antonacci, and A. Sarti, “Scream and Gunshot Detection and Localization for Audio-Surveillance Systems,” 2007 IEEE Conference on Advanced Video and Signal Based Surveillance, 2007.
- [8] Y. Zigel, D. Litvak and I. Gannot, “A method for automatic fall detection of elderly people using floor vibrations and sound? Proof of concept on human mimicking doll falls,” IEEE Trans. Biomed. Eng., vol. 56, no. 12, pp. 2858–2867, 2009.
- [9] C. N. Doukas, I. Maglogiannis, “Emergency fall incidents detection in assisted living environments utilizing motion, sound, and visual perceptual components,” IEEE Trans. Inf. Technol. Biomed., vol. 15, no. 2, pp. 277–289, 2011.

- [10] 植松 尚, 小泉 悠馬, 齊藤 翔一郎, 中川 朗, 原田 登, “動作音から機器の異常を検知する異常音検知技術,” NTT 技能ジャーナル, vol. 29, no. 6, pp. 24–27, 2017.
- [11] 酒井 隼樹, 小林 和矢, 東 孝幸, “機械学習を用いた加工異常検知技術の開発—ミリング加工におけるびびり振動の検知,” 2018 年度精密工学会秋季大会学術講演会講演論文集, pp. 160–161, 2018.
- [12] 渡辺 嘉二郎, 豊田 成人, “異常音による機械の診断-軸受呼び番号の同定と故障の診断-,” 計測自動制御学会論文集, vol. 29, no. 1, pp. 94–101, 1993.
- [13] 岩壺 卓三, 松田 博行, “機械状態監視に関する振動技術者の資格認証について,” ターボ機械, vol. 32, no. 6, pp. 321–327, 2004.
- [14] 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.
- [15] 労働省, 「騒音障害防止のためのガイドライン」, 基発 546 号, 1992.
- [16] 寺本 和幸, 藤田 正, 工藤 市兵衛, “プレス工場の騒音が生体機能に及ぼす影響,” 日本経営工学会誌, vol. 32, no. 3, pp. 217–218, 1981.
- [17] 日本音響学会編, 新版 音響用語辞典, コロナ社, pp. 40, 2003.
- [18] A. Pearce¹, T. Brookes, and R. Mason, “Timbral attributes for sound effect library searching,” AES Conference on Semantic Audio, Erlangen, pp. 22–24, 2017.
- [19] Y. Koizumi, Y. Kawaguchi, K. Imoto, T. Nakamura, Y. Nikaido, R. Tanabe, H. Purohit, K. Suefusa, T. Endo, M. Yasuda and N. Harada, “Description and Discussion on DCASE2020 Challenge Task2: Unsupervised Anomalous Sound Detection for Machine Condition Monitoring,” in arXiv e-prints: 2006.05822, June 2020, pp. 1–4. [Online]. Available: <https://arxiv.org/abs/2006.05822>
- [20] Y. Koizumi, S. Saito, H. Uematsu, N. Harada and K. Imoto, “ToyADMOS: A Dataset of miniature-machine operating sounds for anomalous sound detection,” in Proc. of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2019.
- [21] H. Purohit, R. Tanabe, K. Ichige, T. Endo, Y. Nikaido, K. Suefusa, and Y. Kawaguchi, “MIMII Dataset: Sound Dataset for Malfunctioning Industrial Machine Investigation and Inspection,” in Proc. of DCASE 2019 Workshop, 2019.

- [22] R. Giri, S. V. Tenneti, F. Cheng, K. Helwani, U. Isik, and A. Krishnaswamy, “Unsupervised Anomalous SoundDetection using Self-Supervised Classification and GroupMasked Autoencoder for Density Estimation,” Tech. report in DCASE2020 Challenge Task 2, 2020.
- [23] K. Suefusa, T. Nishida, H. Purohit, R. Tanabe, T. Endo, and Y. Kawaguchi., “Anomalous Sound Detection Based on Interpolation Deep Neural Network,” in Proc. of ICASSP, 2020.
- [24] 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 修士論文, 2020.
- [25] H. Fastl, E. Zwicker, “Psychoacoustics : Facts and models, third edition,” Springer, pp. 257-264, 2006.
- [26] 岩宮 眞一郎, 小澤 賢司, 小坂 直敏, 山内 勝也, 高田 正幸, 藤沢 望, “音色の感性学,” コロナ社, 2010.
- [27] 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.
- [28] G. von Bismarck, “Sharpness as an attribute of the timbre of steady sounds,” *Acustica*, vol. 30, pp. 159–172, 1974.
- [29] von E. Terhardt, “U ¨ber akustische Rauigkeit und Schwankungsst ¨arke (Acoustic roughness and fluctuation strength),” *Acustica*, vol. 20, pp. 215–224, 1968.
- [30] 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. 1–6, 2019.
- [31] Williams, D., “Towards a Timbre Morpher,” PhD thesis, University of Surrey, Department of Music Sound Recording, 2010.
- [32] Freed, D., “Auditory correlates of perceived mallet hardness for a set of recorded percussive sound events,” *J. Acoust. Soc. Am.*, vol. 87,no. 1,pp. 311–322, 1990.
- [33] Solomon, L., “Search for physical correlates of psychological dimensions of sounds,” *J. Acoust. Soc. Am.*, vol. 31, no. 4, pp. 492–497, 1959.
- [34] Pearce, A., Brookes, T., and Mason, R., “Hierarchical ontology of timbral semantic descriptor,” Audio Commons project deliverable D, 5, 2016.

- [35] Cartwright, M. and Pardo, B, “Social-EQ: Crowdsourcing an Equalization Descriptor Map”, in International Society for Music Information Retrieval.
- [36] Pearce, A., “Perceived differences between microphones,”” Ph.D. thesis, University of Surrey, 2017.
- [37] Vassilakis, P. N. and Fitz, K., “SRA: A web-based research tool for spectral and roughness analysis of sound signals,” Proceedings of the 4th Sound and Music Computing (SMC) Conference, pp. 319–325, 2007.
- [38] Fastl, H., and Zwicker, E., “Psychoacoustics: facts and models,” Springer Science Business Media, vol. 22, 2006.
- [39] Hatano, S., and Hashimoto, T. “Booming index as a measure for evaluating booming sensation,” The 29th International congress and Exhibition on Noise Control Engineering, 2000.
- [40] Jan, T., and Wang, W., “Blind reverberation time estimation based on Laplace distribution,” EUSIPCO. pp. 2050–2054, 2012.
- [41] 栗田 多喜夫, “サポートベクターマシン入門”, <https://home.hiroshima-u.ac.jp/tkurita/lecture/svm.pdf>, (2021,2,1 閲覧) .

謝辞

本研究に取り組むにあたって、研究のことだけにとどまらず、生活態度や取り組む姿勢まで多くのご指導をしていただいた主指導教員の鶴木祐史教授に深く感謝いたします。

研究室会議やミーティングなどの場において、数多くのご指摘、ご助言をいただいた赤木正人教授、木谷俊介助教授、小林まおり博士に深く感謝いたします。

研究へのご助言だけでなく、私生活でも多くのことを学ばせていただいた磯山拓都さん、鳥谷輝樹さん、水野滉介さん、森田翔太さんをはじめとする先輩の皆さまに深く感謝いたします。

研究などで行き詰った時に親身に相談に乗っていただき、心の大きな支えになった坂本貴望さん、藤田周平さん、木所晃利さん、王鋭さん、坂本湧暉さんをはじめとする同窓生の皆さまに深く感謝いたします。

研究発表練習などの場で数多くのご指定、ご助言をいただいた後輩の皆様に深く感謝いたします。

研究業績

1. 倉 誠吾, 水野滉介, 鵜木 祐史, “音色に係る評価指標を用いた異常音検知の検討,” 2020 年度電気・情報関係学会北陸支部連合大会, G-2, 2020.
2. 倉 誠吾, 鵜木 祐史, “音色属性に係る評価指標を用いた異常音検知の検討,” 音講論(春), 3-1P-14, 2021.