JAIST Repository
https://dspace.jaist.ac.jp/

Title	音質評価指標による産業機器の異常音検知に関する研究
Author(s)	大田, 恭士
Citation	
Issue Date	2023-12
Type	Thesis or Dissertation
Text version	ETD
URL	http://hdl.handle.net/10119/18821
Rights	
Description	Supervisor: 鵜木 祐史, 先端科学技術研究科, 博士

JAIST
JAPAN
ADVANCED INSTITUTE OF
SCIENCE AND TECHNOLOGY

Japan Advanced Institute of Science and Technology

博 士 論 文

音質評価指標による産業機器の異常音検知に関する研究

大田 恭士

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]

2023 年 12 月

Abstract

This research aims to realize human-centric technology that can provide safety and security through intelligence that is easily understood and harmonized with users. Then this paper proposes an anomalous sound detection (ASD) method that uses a combination of timbral feature-based metrics and short-term features, which is tailored to industrial machine faults to identify whether the sound emitted from a target machine is anomalous.

Daily maintenance of industrial machines is essential to ensure safe operation for efficient production and business management. Inspectors use their knowledge to detect anomalous situations by using their senses, especially hearing. They have excellent skills in discriminating differences in sounds using their "noticeable difference in hearing". Due to a labor shortage, anomalous sound detection (ASD) is expected to support inspectors in identifying whether the sound emitted from a target machine is anomalous as automation and a remote solution.

Recent technical trends of ASD indicate advances in machine learning for discrimination, however acoustical features to capture characteristics of the anomaly of sound remain primitive measurements such as log-mel energies. This causes difficulties for users to understand the logic and to harmonize with ASD as a trustable technology in their fields.

We propose a timbre-feature-based ASD (TF-ASD) method that involves five sound quality metrics (SQMs), two short-term acoustic features, and a support vector machine (SVM) for classification. We develop two types of short-term features to estimate the change in the fluctuation of sound waves and pitch in terms of harmonics to improve the time resolution of the timbral analysis. This combination of SQMs and two dedicated short-term acoustical features is based on an investigation of timbral association with industrial machine malfunction from the viewpoint of "noticeable difference in hearing" that is the human ability to discriminate differences in sounds.

We evaluated the TF-ASD performance of our method in terms of SVM classification using the MIMII (Malfunctioning Industrial Machine Investigation and Inspection) dataset. The results indicate that the proposed method has excellent classification performance with an accuracy of 0.984 and an F-measure of 0.920 on average for emitted sounds of 16 machine types and models. This demonstrated that the combination of SQMs and our short-term features, which is derived from the "noticeable difference in hearing" is effective for ASD.

Furthermore, the timbre-based feature, which is a combination of SQMs and two short-term acoustical features, is verified on the classification capability for the abnormal sound of industrial machines using an anomalous score which estimates the rareness of probability of occurrence against the normal sound distribution. ROC-AUC (AUC) is used to measure the differentiation of the anomalous score of the abnormal sound from that of the normal sound, the further separated the anomalous scores each other, the closer the AUC value to 1.0. The result of an AUC of 0.856 shows the excellent discrimination ability of the timbre-based feature from the probability statistics.

In this paper, we derived a set of SQMs about an anomaly of the sounds emitted from industrial machines in terms of human noticeable differences in hearing and developed a timbre-based feature which is a combination of the SQMs and two original short-term acoustical features. With the timbre-based feature, human-centric ASD, which may be understandable and cooperative with users, is configured and demonstrates excellent classification performance.

Furthermore, by mutual analysis of the timbre-based feature value with human anomalous detection results, implicit knowledge of inspection for industrial machines can be digitized and be transferable assets in various industrial fields.

Keywords: anomalous sound detection, timbral feature, sound quality metrics, industrial machine faults, support vector machine.

概 要

日本のものづくりにおいては製造機器の安定的な運用が生産活動の基本であり，事業の成長を地道に支えている．ほとんどの製造機器は機械的な動作が伴うため経年劣化が避けられず，突然の故障等による稼働停止がひとたび起これば，生産停止によるビジネス上の損失のみならず，場合によっては事故や火災などの人的・物的被害を及ぼす可能性もある．この不測の事態を避けるため，製造の現場では日々の設備点検が欠かせない．現在でもほとんどの生産現場で主流となっている人（点検員）による巡回点検では，人が有する優れたセンシング機能（聴覚，視覚，触覚，嗅覚）の活用で迅速かつ統合的な判断が行われているが，実際には基準が属人的である上に日々の体調等でその正確性にはゆらぎが伴う．また，少子高齢化を背景とした長期的な人材不足への対応が課題となっており，熟練者ノウハウの顕在知化による共有化や若手への伝承の効率的方策の実現が急務となっている．

本研究は人に理解されやすい知能化で安心安全を提供できる ICT コア技術の実現を標榜し，人の聴覚にフォーカスして産業機器の動作音に対する人の聞こえを “音色” の観点で特徴量表現し，その数値を用いて異常を判別・分類することで人と協調しやすい異常音検知の実現を目指した．

異常音検知は音響イベント検出にルーツをもち，幅広い用途で研究が行われてきた．近年の機械学習手法の多様化やセンサやネットワークの拡充により IoT（Internet of Things）の実用化技術は格段の進歩を遂げている．これに産業界におけるインダストリアル 4.0 の気運の高まりを発端とした Digital transformation（DX）のトレンドが加わり，異常音検知の需要は急増している．異常音検知研究は 2020 年から国際的なコンペティションである Detection and Classification of Acoustic Scenes and Events（DCASE）で産業機器を対象とした異常音検知の task が新設されるなど，一層の活性化がはかられている．技術的には機械学習手法による判別・分類技術は飛躍的な進歩を遂げている反面，音響特徴量においては初歩的なレベルに留まっている状況にある．

本論文では，産業機器の動作音において異常時の音（異常音）に対する聴知覚上の差異から音色の指標（音質評価指標）と関連づけた独自の音色関連特徴量を考案した．この音質評価指標に基づく音色関連特徴量の正常音と異常音に対する数

値分布の間を機械学習技術を用いて分類することで，異常音検知手法を実現した．

具体的には，まず代表的な産業機器群として回転系（代表機種は送風機），摺動系（同，直動スライドレール），打付け系（同，電磁弁），流体移送系（同，給水ポンプ）の 4 つを対象に機器としての異常要因を調査し，異常発生時の音（異常音）の聴知覚上の差異を音色の観点から 5 種の音質評価指標に対応づけた．次に音質評価指標より高い時間分解能分析を行う補助的な特徴量として，波形振幅と音の高さの変動を捉える 2 種の短時間音響特徴量を考案することで音質評価指標と組合せた音色関連特徴量を定義した．そして，機器群ごとに選択した音質評価指標に短時間音響特徴量を組合せた音色関連特徴量に対してサポートベクターマシン（SVM）を用いた二値分類を適用する異常音検知システムを構築し，産業機器動作音データセット用いた評価で F 値が 0.920 と優れた異常音分類性能を示した．短時間音響特徴量の考案には複合音におけるバーチャルピッチの考え方，患者の声の主観評価を診察に活用する GRBAS 評価法や音でモノの材質を推測する psycho-mechanics 等，基礎から応用に渡る広範囲の知見を積極的に取り入れた．この短時間音響特徴量の産業機器動作音に対する有効性は，ターゲットとなる音質評価指標との間で同一の実機動作音から算出された数値同士が高い相関値を示すことで確認した．その上で音質評価指標と組合せた多次元の音色関連特徴量の異常音に対する分類貢献度を確率統計上の尤度尺度で検証し，1.0 に近づくほど貢献度が高くなる ROC-AUC 値で 0.856 と高い数値を示すことも確認でき，音色関連特徴量の産業機器の異常音検知に対する優れた有効性が複数の側面から実証された．

この音色関連特徴量による異常音検知は，分類貢献度が構成要素（音質評価指標と短時間音響特徴量）ごとに音色に関連づけながら数値分析できるため，現場点検員の聴感に照らした検出結果の検証や性能調整が行える利点を持つ．さらに分析数値を点検員の判定結果と関連づけることで異常音に対する点検員の聴感判断基準の数値化に展開できる可能性がある．本研究では機器の異常要因に関する事前知を想定し，人の聞こえに対応づけた独自の音色関連特徴量にて優れた分類性能の異常音検知手法を実現しただけでなく，利用者自身がロジック分析や性能カスタマイズにて協調できる基盤技術を具現化した．また本基盤上の数値蓄積を基に人の聴知覚上の差異を音色関連特徴量と関連づけることで，人（点検員）の知見の数値化で点検ノウハウの顕在知化がはかれる可能性も示した．

目 次

1 序論 2

1.1 背景 3

1.1.1 産業界における人による保守点検の重要性 3

1.1.2 保安業界の課題と IT 化への期待 3

1.1.3 IT 化に向けた研究動向と問題点 4

1.2 本研究の動機 8

1.3 課題設定 9

1.4 研究目的 9

1.5 論文構成 11

2 関連研究 14

2.1 聴覚心理学 15

2.1.1 音色 15

2.1.2 音質評価指標 16

2.1.3 ピッチ知覚 23

2.2 環境音分析 25

2.2.1 医療応用 28

2.2.2 産業応用 29

2.2.3 サウンドスケープ 32

2.3 異常音検知 36

2.3.1 DCASE 動向 38

2.3.2 産業機器を対象とした異常音検知 41

2.3.3 変化検知 42

3 産業機器異常音を対象とした音響特徴量表現の方策 45

3.1 聴知覚上の差異に基づく産業機器異常要因と音質評価指標との対応づけ 46

3.1.1 回転系について 47

3.1.2 摺動系 48

3.1.3 打付け系について 49

3.1.4 流体移送系について 49

3.2 音質評価指標の分析時間長の考察 53

3.3 音質評価指標を補助する短時間音響特徴量の検討 53

3.3.1 波形振幅変動特徴量 53

3.3.2 優勢周波数変動特徴量 56

3.4 音質評価指標と短時間音響特徴量の組合せによる音色関連特徴量の実現 57

3.5 異常音検知システム化検討 59

3.5.1 音色関連特徴量の分析時間長の検討 59

3.5.2 機械学習による異常分類仕様の検討 59

4 音質評価指標に基づく音響特徴量の有効性検証 62

4.1 産業機器動作音スペクトラムを用いた聴知覚上の差異の考察 63

4.2 短時間音響特徴量と音質評価指標の関連性検証 70

4.2.1 振幅強調 shimmer（AS）と roughness の相関について 70

4.2.2 振幅強調優勢周波数（APF）と boominess の相関について . 70

4.3 音色関連特徴量の確率統計分析による有効性検証：単次元正規分布仮定 74

4.3.1 異常度スコアの定義 74

4.3.2 異常度スコアの分布分離度評価 76

4.4 音色関連特徴量の確率統計分析による有効性検証：多次元正規分布仮定 80

4.4.1 異常度スコアの定義 80

4.4.2 異常度スコアの分布分離度評価 81

5 音色関連特徴量に基づく異常音検知システム:Timbre-feature-based ASD 86

5.1 基本構成 87

5.2 特徴量算出部 87

5.3 音色関連特徴量生成部 89

5.4 学習および分類部 89

5.5 分類性能の評価尺度 89

5.6 評価結果 91

6 統合考察 93

6.1 音色関連特徴量に基づく異常音検知システム性能評価 94

6.2 音色関連特徴量の SVM 分類性能への貢献と異常度スコア分布との関係 99

6.3 産業機器動作音データセットにおける正常音の正規分布性についての考察 103

6.3.1 音色関連特徴量分布の正規分布性の調査 103

6.3.2 正規分布性検定値を向上させたサブセットの抽出 105

6.3.3 正規分布性向上サブセットに基づく音色関連特徴量の有効性検証 108

7 結論 111

7.1 本論文における成果 112

7.2 今後の展望 115

謝辞 118

参考文献 119

本研究に関する発表論文 132

図 目 次

1.1 保安領域における業界課題と期待される効果 [1] 6
1.2 一般的な異常音検知の構成 7
1.3 本研究の動機：業界課題と解決アプローチ ※ [1] を基に加筆 10
1.4 本研究の課題 10
1.5 本論文の構成 13

2.1 Sharpness の重み $g'(z)$ [2] 20
2.2 振幅変調音の変調周波数と roughness の関係 [2] 20
2.3 主なサウンドスケープの構成要素 [3] 34
2.4 異常音検知のアプローチ分類 （文献 [4] を基に加筆） 37
2.5 変化検知の分類 （文献 [4] を基に編集） 44

3.1 （a）正常状態の直動スライドレールの動作音波形,（b）動作音波形から計算した shimmer 値,（c）動作音波形における 256ms 毎のフレーム振幅値 55
3.2 サポートベクターマシン （出典元：https://kit.socinno.com/2_3_ai/） 60

4.1 送風機（fan）の動作音スペクトラム：(a) 正常動作音，(b) 異常動作音 66
4.2 給水ポンプ（pump）の動作音スペクトラム：(a) 正常動作音，(b) 異常動作音 67
4.3 直動スライドレール（slider）の動作音スペクトラム：(a) 正常動作音，(b) 異常動作音 68
4.4 電磁弁（valve）の動作音スペクトラム：(a) 正常動作音，(b) 異常動作音 69
4.5 振幅強調 shimmer（AS）と roughness の相関評価 72
4.6 振幅強調優勢周波数（APF）と boominess の相関評価 73

4.7 正常音／異常音の特徴量数値分布の比較 [fan id04] 75
4.8 正常音／異常音の特徴量数値分布の比較 [slider id02] 75
4.9 異常度スコアにおける分布分離性の検証例１ 78
4.10 異常度スコアにおける分布分離性の検証例２ 82

5.1 TF-ASD のブロックダイヤグラム 88

6.1 特徴量追加に伴う異常度スコア分布と SVM 分類性能値の比較（fan id00）
100
6.2 多次元／単次元異常度スコア分布の遷移解析（fan id00） 102
6.3 サブセット抽出例（fan id00 の sharpness 分布） 106

7.1 研究アプローチと本論文における成果 114
7.2 今後の展望 117

表 目 次

2.1 音色の 3 因子 [5] 17
2.2 音色属性の客観評価指標（Timbral_models 実装例） 17
2.3 回転体の障害と振動周波数の関係 [6] 24
2.4 環境音分析における検出手法（代表例） 27
2.5 GRBAS 因子と相関する音響分析値と音質客観指標との関連づけ . . 30
2.6 GRBAS 因子と相関のある音響分析値 [7] 30
2.7 音響環境を説明する音響／心理音響インジケータ（ISO12913-1） . 35
2.8 DCASE challenge における Top10 方式採用の音響特徴量 40
2.9 産業機器動作音データセット収録機種とデータ数 40

3.1 音声生成系の調音種と発生音の対応 [8] 51
3.2 聴知覚上の差異に基づく機器の異常要因と音質評価指標の対応づけ 52
3.3 4 機器群に対する音色関連特徴量の割当 58
3.4 音色関連特徴量の構成要素と音色因子の関係について 58

4.1 異常度スコアの AUC 比較：単次元正規分布仮定 79
4.2 異常度スコアの AUC 比較（機種平均値）：単次元正規分布仮定 . . . 79
4.3 異常度スコアの AUC 比較：多次元正規分布仮定 84

5.1 SVM 分類性能評価結果 92

6.1 SVM 分類性能評価結果（F 値平均値） 97
6.2 SVM 分類性能評価結果（F 値機種個別） 98
6.3 正規分布性評価（Shapiro-wilk test）：MIMII データ全体 104
6.4 動作仕様の違いが表れると推測される音質評価指標 106
6.5 正規分布性評価（Shapiro-wilk test）：正規分布性向上サブセット . . 107

6.6　異常度スコアの AUC 比較（機種平均値）：正規分布性向上サブセット 109

第 1章

序論

1.1 背景

本章では本研究の背景として，主な適用先である保守点検業界が抱える課題や関連研究の動向から対処すべき問題点を指摘し，研究動機と課題設定を通じて研究の目的を述べる．

1.1.1 産業界における人による保守点検の重要性

ものづくりにおける日々の設備点検業務は生産の安定性に必須の重要な作業で，事業の成長を地道に支えている．日々稼働する製造機器は常に経年劣化のリスクにさらされており，故障による操業停止には多大な損害が伴うため，設備点検保守が欠かせない．特に近年では，操業停止リスクを最小化する目的で予防保全より予知保全が強く求められている．さらにコスト低減も目指して，製造機器を予め決められた周期で修理・交換を行う time based maintenance（TBM）から，機器の状態を監視してその状態に応じて修理・交換を行う condition based maintenance（CBM）への移行に対する需要も高まっている．この CBM 実現には点検員による日々の点検や監視が重要で，とくに広範な範囲で死角の影響が少なく即時性に優れる聴感監視が有効に活用されているが，その一方で人が行うが故に判断基準のブレや人材コストの維持などの問題を抱えている．

1.1.2 保安業界の課題と IT 化への期待

近年では，小型・高感度センサ等の高機能デバイス，高速 CPU 搭載の小型サーバ，柔軟な機器設置を可能にする高速無線アクセスネットワークおよび多機能 AI エンジン搭載の Web サービス等の複合的な技術発展の恩恵から，これらの有機的な結合で実現される IoT（Internet of Thing）に著しい高度化がはかられた [9]．日本においては特に無線通信技術の実用化が他国に比べて先行しており，例えば公衆網通信の 5G（第 5 世代移動通信システム）の「超高速」「超低遅延」「多点同時接続」の特長を活用するローカル 5G の適用が進んでいる．これらの技術進化を背景に世界的な DX（デジタルトランスフォーメーション）化の波を追い風として製

造工場やインフラ拠点における工場DX，いわゆるスマートファクトリ化の需要が高まっている [10].

この産業界の動向を踏まえて，経済産業省（経産省）はスマート保安を掲げ，産業保安力を強化する方策として，保安業務を対象にIoT やAI 等を活用する新技術の実証や，先進技術の活用を促す規制改革を進めている [1,11]．さらに経産省は電力やガスのインフラ保安を対象とした保安業界の課題として，設備の高経年化と併せて人材の高齢化と長期的な不足，技術・技能伝承力の低下を提起している．その上で，一層加速化されるデータを活用した人間中心社会（Society5.0）への移行を見据えて，設備の状態をデータによって常に把握・監視できる新技術への期待を示し，安全性の向上のみならず，人手不足等の課題に適切に対処していく考えも示している [1,12]．その中で保安領域における3つの業界課題とスマート保安とIT化で期待される4つの効果を挙げ，両者の対応関係を図1.1に整理している．ここでは，業務の自動化や遠隔化による作業の負荷軽減や省人化，安全性の向上だけでなく，IT化による熟練者ノウハウの定量的な評価から非熟練者の技術習得の円滑化も期待する効果として挙げられている．

1.1.3 IT化に向けた研究動向と問題点

産業機器の点検を担当する作業員（点検員）は，日々の点検において視覚，聴覚，嗅覚，触覚といった味覚以外の五感を駆使して機器異常を発見している．中でも音響信号からの聴覚に基づくモニタリングは，即時性，広い受容角度，広いダイナミックレンジと高い感度を持ち，比較的遮蔽物にも強い特性を持つ．点検員は，対象機器のある程度の知識と経験知を活用しながら，動作音の聴知覚上の差異を頼りに機器の異常性を識別する能力に優れていると考えられる．そのため，この能力の支援もしくは代替を目的とするIT化には機器から発せられた音に対して正常時の音と差異がないかを分析し，差異が一定の範囲を超えたものを異常として判定する異常音検知が核の技術となる．

一般的な異常音検知技術は，観測された信号（ここでは音）から目的に応じた特徴量を抽出するParameterization stage，大量の特徴量データを用いた学習によって異常検出に適切なモデル（学習モデル）を求めるTraining stage，そして学習モ

デルを基に入力された音（ここでは評価される観測音）に対するスコアを算出し，閾値以上となった場合に異常と判定する Detection stage の 3 つの要素から構成される [13].

この Training stage においては，一般的には機械学習技術が適用されるが，学習データとして異常音のサンプルが利用できるかどうかによって下記 2 つの問題設定に大別できる [14].

教師あり異常音検出

学習データとして正常音だけでなく，異常音のサンプルも利用できる場合

教師なし異常音検出

学習データとして正常音サンプルのみが利用できる場合

産業機器を対象とした異常音検知については，競争型ワークショップである Detection and Classification of Acoustic Scenes and Events（DCASE）でも 2020 年に専用の Task が設定されるなど，ここ 2〜3 年で研究が活発に進められてきた．詳細は 2.3 節で述べるが，DCASE における研究実績としては Training stage に相当する機械学習技術に飛躍的な進歩が見られるものの，Parameterization stage における音響特徴量としては log-mel energy や mel-band spectrogram 等の初歩的な特徴量がまだ主流となっている状況である．

そのため異常音検知性能が数値的に向上していても，どんな特長を捉えて異常を検知しているのか，性能をさらに向上させるための調整にどんな改善すれば適切なのかなど，性能根拠となる判定ロジックが利用者に理解が難しい “ブラックボックス” となってしまっている．

このように人間の統合知を反映し，安定した的確な判断により支援すべき IT 技術が，性能は向上するものの利用者である人間の理解や感覚から乖離している現状では，現場運用での協調性のみならず，信頼性に乏しい技術となってしまっていることが問題と考える．

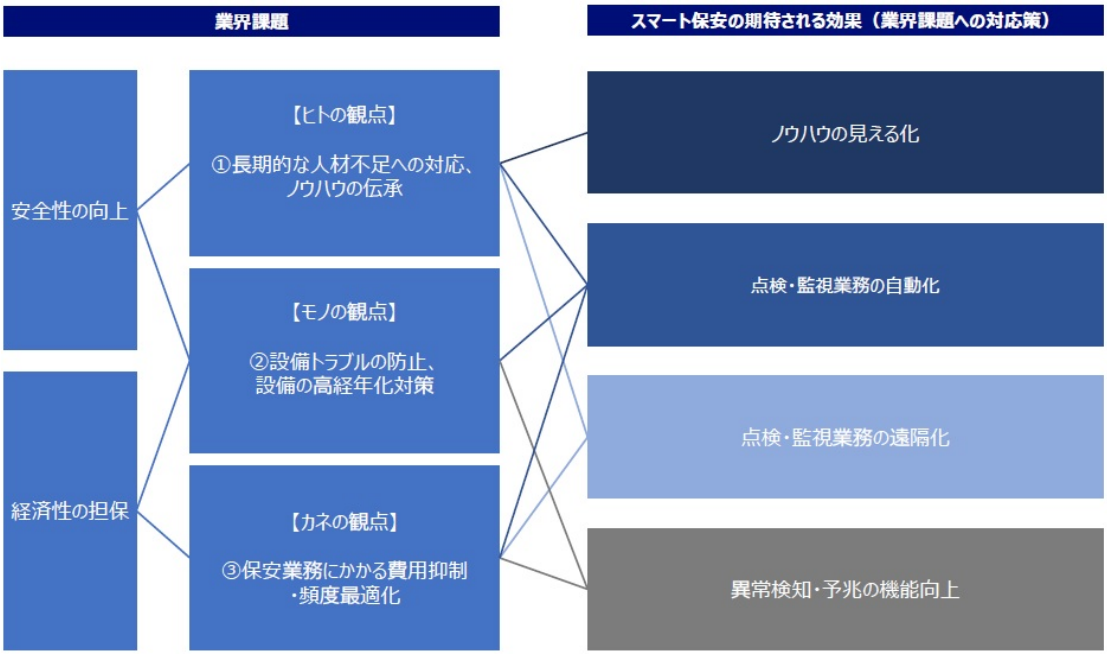


図 1.1: 保安領域における業界課題と期待される効果 [1]

図 1.2: 一般的な異常音検知の構成

1.2　本研究の動機

前節で述べた通り，現状の異常音検知の研究状況においては機械学習に基づく判別技術は飛躍的に進化している反面，音響特徴量としては初歩的なレベルにとどまっており，音響面で何を捉えていくことが必要なのかの聴知覚観点からの議論がなされていない．また，音響特徴量に音質評価指標（sound quality metric：SQM）を用いた異常音検知の研究も報告され始めているが，手法としては大規模観測データからルールを抽出するデータマイニング的アプローチが多く，人の聞こえとの関連づけまでは至っていない．

そこで本研究では産業機器の動作音を対象に，人が異常音を聞き分ける観点を音色の観点から考察した上で音響特徴量として表現していくアプローチを取ることで，利用者が理解しやすく協調しやすい人間中心，すなわち Human-centric な異常音検知の実現を目指すこととした．

現在の異常音検知アプローチは，発生頻度が低い異常な状態（＝音）を検出するには，正常状態（の音）の存在範囲をいかに的確に分布として抑えるかが外れ値としての異常を見つけ出す鍵と捉え，飛躍的な技術向上を遂げる機械学習がこの問題の唯一の解となり得ると考える傾向が見られる．そのため，上記の通り音響特徴量としては初歩的なものが好まれ，手法によっては音声信号そのものを入力とし，特徴量抽出から機械学習を適用するアプローチも見られる．そのため，音の性質をどんな観点（特徴量）から分析しているのか，何をもって異常と判断しているのか（判断ロジック）を外部から理解することが難しく，このような人の理解を寄せ付けない技術では利用者の立場からは可用性に乏しいものとなり，現場適用が難しくなる問題が生じる．

実際の点検現場で人は稼働している機器の種別や動作原理を把握しており，その事前知を無意識に使いながら，聞いた音から異常の有無を判断していることが容易に考えられる．この考え方を踏襲した手法であれば，利用者が自らの経験値を伝えながら育てていく見習い点検員と同じ立場で性能調整を受けながら実力を伸ばし，利用者の信頼度を高めていく現場協調型の技術となり得る．また本技術の性能調整を通して，熟練の点検員が自らの知見を特徴量に照らして可視化していくことで，潜在的な知見を顕在知化させることも可能となる．

さらにこの顕在知化を通じて，機器動作音に関する音色（音質評価指標）と聴知覚上の差異の関連性にフィードバックがかかり，音響聴覚心理学の深化への貢献も十分期待できる．そのため，本研究では機器動作音の異常を聞き分けることに適切な特徴量の探索と異常と判別するロジックの分かりやすさ（透明性）に重きを置いた異常音検知技術を探求する方針とした．

1.3 課題設定

上記動機の具体的検討にあたり，本研究では主に下記 2 点を本研究の課題（research questions）として取組んだ．

(1) 人の異常音聞き分け観点と機器の異常要因が音色関連特徴量を介して定量的に対応づけられるか?

(2) 機器の動作音の変化と関連づけた音色関連特徴量により優れた異常検知を実現できるか?

1.4 研究目的

本研究の目的は，機器動作音の変化と関連づけた音色関連特徴量による異常音検知の有効性を示し，人が異常音を聞き分ける観点と音色との関連性を明らかにすることである．そのため，本論文では機器動作音の変化と音色関連特徴量との関連づけの妥当性を客観的に示すことに努め，実機収録音データにて本異常音検知法が優れた分類性能を示すことで音色関連特徴量による異常音検知アプローチが適切かつ有望であることを示す．その上で，機器動作音に対する音色関連特徴量と人の異常音聞き分け観点の詳細解析にて稼働機器の異常原因や経年変化を推測するまでの将来展望を視野に入れた要件の抽出も目指す．

スマート保安による課題解決

期待される効果	効果によるメリット例
ノウハウの見える化	✓ 技術員の五感や経験則を頼りに実施してきた巡視点検をIoT/AI活用により自動化することで，技術の伝承が可能となる. ✓ 複雑かつ属人的で，熟練技術者にしか担当できなかったオペレーションがスマート保安技術により若手でも担当可能になる. ✓ システムによる異常検知時には熟練者の見解と併せてディスカッションすることで若手の育成が図られる.
点検・監視業務の自動化	✓ オペレータが24時間365日体制で，異常がないか監視するために張り付く必要がなくなり，工数としても精神的にも負荷が低減される. ✓ 大規模な点検業務であっても安全に工期を短縮できる. ✓ スマート化技術により，ボタン一つで再現性の高いオペレーションが可能となりエラーが抑制される.
点検・監視業務の遠隔化	✓ 遠隔から稼働データを取得することで，現地に赴かずとも状況を判断し，必要に応じて点検を実施できる. ✓ ドローン等を活用することで，高所・引火性ガス雰囲気下・海上サイトなど危険な場所で作業員が点検業務をする必要がなくなる. ✓ 点検業務を遠隔化することにより，同時に保安プロセスのデジタル化を推進することができる.

本研究の
解決アプローチ

人の聞こえの
デジタル化による
点検ノウハウの
顕在知化

機器の異常音と
人の聞こえを
対応づけた
異常音の検知

産業機器の
動作音分析による
自動点検の実現

図 1.3: 本研究の動機：業界課題と解決アプローチ ※ [1] を基に加筆

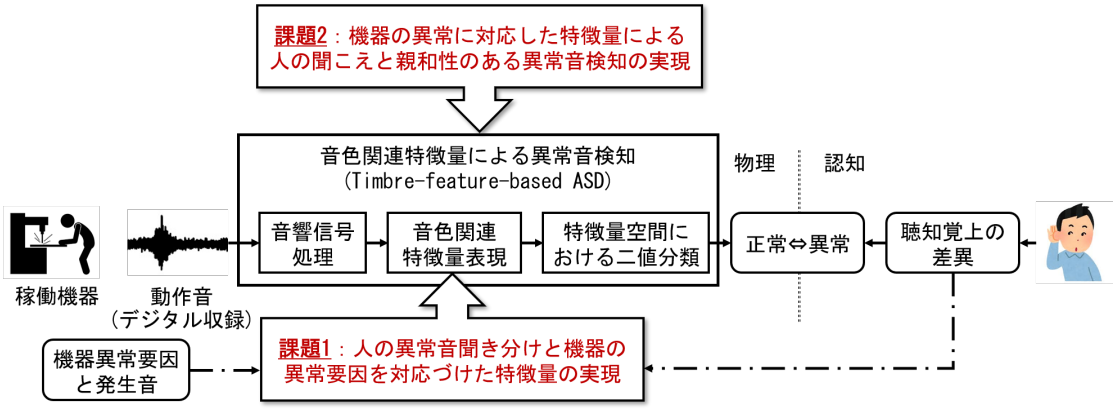


図 1.4: 本研究の課題

1.5　論文構成

本論文は 6 章で構成される．図 1.5 に本論文の構成を図示する．

第 1 章

本論文の研究対象である異常音検知に関する背景として，主な適用先である保守点検業界が抱える課題や関連研究の動向から対処すべき問題点を指摘し，研究動機と課題設定を通じて研究の目的を述べる．

第 2 章

人の聞き分けの重要な観点と考える音色，音質評価指標，およびピッチ知覚に関する聴覚心理上の研究知見を概説する．その上で，異常音検知の親分類となる環境音分析の研究状況とその医療と産業への応用例，およびサウンドスケープの最新の研究状況から音色や音質評価指標との接点を考察する．最後に異常音検知の最新動向をレビューしながら，音色観点の異常音検知実現に向けた方策検討のための課題を抽出する．

第 3 章

産業機器に対する異常音検知の土台となる音響特徴量として，異常音に対する聴覚上の差異を音色の観点から特徴量として表現するための方策を述べる．まず代表的な産業機器として 4 つの機器群を挙げ，各機器群ごとに調査した主要な異常要因を基に発生音から生じる聴覚的な差異を聞き分け観点としてオノマトペで表現し，この表現を介して音質評価指標を対応づけた．その上で音質評価指標の算出要件を考察し，それを補完する短時間音響特徴量の考案にて異常音知覚に近づける音色関連特徴量実現までの検討結果を述べる．

第 4 章

異常音の聴知覚上の差異に関連づけた音質評価指標 5 種と短時間音響特徴量 2 種の組合せによる音色関連特徴量について，産業機器動作音データセットの収録音を入力とした特徴量分布の評価および分析により異常音検知に対する有効性を検証する．

第 5 章

音色関連特徴量に基づく異常音検知手法の実現性を検証するため，サポートベクターマシンによる学習および分類部を組合せた異常音検知システムを構成し，産業機器動作音データセットを用いた分類性能の評価を行う．

第 6 章

共に産業機器動作音データセットを用いた第 4 章における音色関連特徴量の検証結果と第 5 章における音色関連特徴量に基づく異常音検知システム性能の評価結果を統合した検証を行う．

第 7 章:

本論文で得られた成果を要約し，明らかにしたことと今後の展望について述べる．

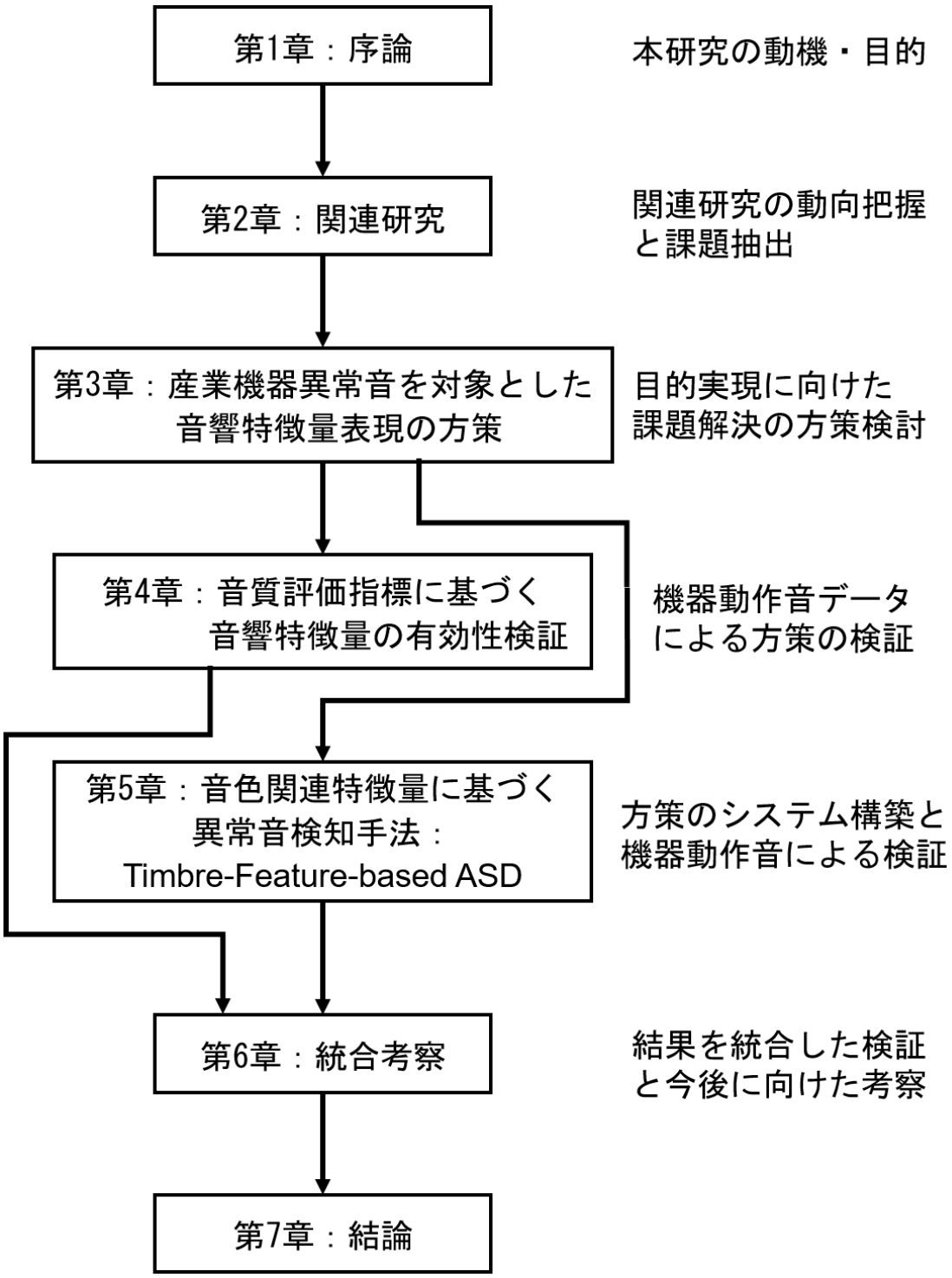

第1章：序論
本研究の動機・目的
第2章：関連研究
関連研究の動向把握
と課題抽出
第3章：産業機器異常音を対象とした
音響特徴量表現の方策
目的実現に向けた
課題解決の方策検討
第4章：音質評価指標に基づく
音響特徴量の有効性検証
機器動作音データ
による方策の検証
第5章：音色関連特徴量に基づく
異常音検知手法：
Timbre-Feature-based ASD
方策のシステム構築と
機器動作音による検証
第6章：統合考察
結果を統合した検証
と今後に向けた考察
第7章：結論

図 **1.5:** 本論文の構成

第 2章

関連研究

本章では，人の聞き分けの重要な観点と考える音色，音質評価指標，およびピッチ知覚に関する聴覚心理上の研究知見を概説する．その上で，異常音検知の親分類となる環境音分析の研究状況とその医療と産業への応用例，およびサウンドスケープの最新の研究状況から音色や音質評価指標との接点を考察する．最後に異常音検知の最新動向をレビューしながら，音色観点の異常音検知実現に向けた方策検討のための課題を抽出する．

2.1 聴覚心理学

2.1.1 音色

人は音の知覚から周囲の状況を把握することができる．特に音を監視の手段として用いる点検員などの専門職は，音の知覚から対象物の状態を推測する能力に優れる．彼らは聞いた知覚した音を主に音色として捉えていると考えられる．

この音色は音の大きさや音の高さとともに，音の 3 属性と呼ばれる [15]．音色は音の大きさ，音の高さと異なり多次元的な要素で，その印象の表現に形容詞句を用いる点が特長である [16]．音色には 2 つの側面があり，1 つ目は人が音から受ける印象の側面で，2 つ目が発音源の認知や識別の手がかりとなる側面である [17,18]．1 つ目の音から受ける印象に関しては，音色を表現するために通常，澄んだ，明るい，鋭いなどの形容詞句が音色表現語として使われるが，数種のグループに整理する研究により数十種にのぼる表現語が 3〜7 種の因子に集約されている [5,19]．現在では，金属因子，迫力因子，美的因子の 3 種に統合された因子が多く活用されている．(表 2.1)

音から受けた印象を他者に伝える手段としては擬音語，オノマトペがよく用いられる [20]．オノマトペで表現された擬音語からイメージされる音の印象が，美的因子，迫力因子，および金属因子の 3 つの主なる因子と関連づくことも実験にて明らかにされている [21]．さらに実際の音とオノマトペの印象を比較した実験により，オノマトペから受ける印象が実音から受ける印象と似通っていることを示す報告もある [22]．

そして，音色が持つ 2 つ目の側面は機械音の動作状況を知る手がかりに音色が

適していることを示している．模擬空調音などの機械音と多次元的な音色との関連をSD（semantic differential）法により分析した研究報告があり，美的因子，迫力因子，金属因子の3因子に対応づくことが示されている [23]．また，自動車製造の現場でも車室内エンジン音の客観音色評価法の検討に“ゴロゴロ音’’とオノマトペを用いてターゲット音色を特定し，音色表現語を用いた評価法にて主観品質を検証している [24]．

音色の持つ2つの側面，音から受ける印象と音からの発生源識別の手がかりの両面に関わる活用先として機械の快音化が挙げられる [25]．この報告では騒音レベル（音の大きさ）の低減だけでは喧噪感改善に繋がらないことが指摘されており，快音化には音色に踏み込んだ音設計が必要と主張している．

これらの研究報告から，人は産業機械の異常時の動作音である異常音に対する聞き分け，つまり聴知覚上の差異を音色の観点で捉えているとする前提は妥当であると考える．また，人が異常音から受ける印象の音色表現にはオノマトペが有効であることもわかる．

2.1.2 音質評価指標

この人が捉えた音色の種類や程度を数値的に表現する手段として，音質評価指標の活用が非常に有効である．音質評価指標とは，音色に関係した品質や価値判断までを示す音質に関して，人の感覚（主観量）を数値で表現した尺度である [15]．

音色の尺度化を目的としたモデル提案の研究は数多く報告されている [26–28]．中でも英国Surry大学が開発したTimbral modelsは，音色の種類（属性）における1,187種の記述子を調査し，相互関連性を精査することで145の音色属性を抽出 [29] しており，さらにその中から8種を選定してプログラム実装している [30,31]．選定した音色属性については提案文献の原理と性能を精査し，具体的なモデル記述がない場合は自らが創作し，追加の主観評価実験を通じた補正を加えることで算出される指標値の信頼性を高めている．本プログラム群はAudio Commonsとして公開されており，幅広い音色関連研究に活用されている．実装されている8種の指標を表2.2にまとめる．なお，本論文ではTimbral modelsに採用されている“音色属性の客観評価指標”も音質評価指標に含めて論じていく．

表 2.1: 音色の 3 因子 [5]

因子名	代表的な表現語対
第 1 因子 （金属性因子）	鮮やかな – ぼけた，鋭い – 鈍い，はっきりとした – ぼやけた，明るい – 暗い，派手な – 地味な，固い – 柔らかい
第 2 因子 （迫力因子）	迫力のある – もの足りない，豊かな – 貧弱な
第 3 因子 （美的因子）	きれいな – きたない，なめらかな – ざらざらした，きめの細かい – 粗い

表 2.2: 音色属性の客観評価指標（Timbral_models 実装例）

	指標名	説明	概要・算出法
1	boominess	音の**ブーミング感**を表す.	低い周波数（63Hz近辺）でブーミング感を感じる.
2	brightness	音の**明るさ感**を表す.	3,000Hz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.
3	depth	音の**奥行感**を表す.	20～500Hzの周波数成分が多く含まれ，基本周波数が低く，音の立上り後の減衰時間が長いほど奥行きを感じる.
4	roughness	音の**粗さ，ざらつき感**を表す.	変調周波数が70Hzに近い音ほど粗さを感じる.
5	sharpness	音の**かん高さ，鋭さ**を表す.	正弦波信号では，周波数が高いほどかん高さを強く感じる.
6	hardness	音の**固さ感**を表す.	立上り（onset）の強さなど6種の特徴量による算出.
7	warmth	音の**あたたかさ**を表す.	20～40Hzの周波数成分が多く含まれる音ほど，あたたかさを感じる.
8	reverb	音の**残響感**を表す.	ラプラス分布モデルによる残響時間推定に基づいた算出.

Timbral models には複数の過去文献から厳選された計算モデルが採用され，一部には主観評価実験に基づく回帰式による補正も行われている．また上述の通り，計算モデルが明示されていない指標については主観実験データに基づくロジック化も行われており，信頼性が高いと判断した [29–31]．さらに各算出指標の数値範囲が 100 までに正規化されており，指標間の相互関係の考察に有利である．この Timbral models の中から指標の代表として boominess, brightness, depth, roughness, sharpness の 5 つを取上げ，実装されている仕様を紹介する．

Sharpness

Sharpness は鋭さやかん高さの感覚に対応する評価指標である．その印象は周波数の中心が高周波に移動するにつれて増加する．この知見を基に Zwicker はラウドネスレベルが 60 phon の 1,000 Hz を中心とする狭帯域ノイズから感じる sharpness を 1 acum と定めた [32]．この acum を単位とした sharpness のモデルは式 (2.1) で表される．

$$S = 0.11\,\frac{\int_0^{24\mathrm{Bark}} N'(z)\,g\,(z)\,z\,dz}{\int_0^{24\mathrm{Bark}} N'\,dz} \tag{2.1}$$

ここで，S は sharpness の値で，$N'(z)$ が臨界帯域率 z におけるラウドネス密度，また $g\,(z)$ は z における S の重みづけ係数である．

音のラウドネスレベル（フォン）とは，周波数 1 kHz の純音と同等のラウドネスと判断される音圧レベル（dB）のことである．また，重みづけ係数は心理音響実験から 3,000 Hz までの周波数帯域では 1.0 の一定で，それ以上の周波数で 4.0 まで急激に増加している（図 2.1）．

Roughness

Roughness とは，粗さ，ざらざらやジージーといった印象の音質を表し，ラウドネスの変調周波数の変化に強く関係している [32, 33]．単位は asper である．この音の粗さの感覚は，音の振幅包絡（エンベロープ）の変動や周波数の変動によって生じる．このことから変調周波数（f_{mod}），搬送波周波数（f_c），変調度（m），音圧レベル（L）と主観的な粗さ（R）の関係は式 (2.2) で表される [2, 34]．

$$R = A\left(f_{\mathrm{mod}}, f_c\right) m^2\, 2^{\frac{L-40}{20}} \tag{2.2}$$

ここで，A は搬送周波数（f_c）ごとに変調周波数（f_{mod}）と roughness（R）の関係を示した係数である（図 2.2 参照）.

Roughness のモデル化には，様々な手法が提案されている [33,35,36]．Timbral models では，信号を構成する 2 つの主な正弦波要素の振幅変調度に着目した計算モデルを採用している [37]．この 2 つの正弦波要素の周波数（f_1, f_2）と振幅（A_1, A_2）から，周波数の最小値（$f_{\mathrm{min}} = \min\left(f_1, f_2\right)$）と最大値（$f_{\mathrm{max}} = \max\left(f_1, f_2\right)$）と振幅の最小値（$A_{\mathrm{min}} = \min\left(A_1, A_2\right)$）と最大値（$A_{\mathrm{max}} = \max\left(A_1, A_2\right)$）を求めて，これら 4 つの極値にて振幅変調度に着目した算出を行う.

Timbral models では roughness 値がまず 50 ms ごとの局所値（R_{frame}）として式 (2.3) で求められる.

$$R_{\mathrm{frame}} = X^{0.1} \times 0.5\left(Y^{3.11}\right) \times Z \tag{2.3}$$

ここで，第一項 $X^{0.1}$ は信号の intensity に対する依存要素として，式 (2.4) の X から算出される.

$$X = A_{\mathrm{min}} \times A_{\mathrm{max}} \tag{2.4}$$

第二項 $Y^{3.11}$ は，振幅変動幅に対する依存要素として，式 (2.5) の Y から算出される.

$$Y = 2 \times A_{\mathrm{min}} \,/\, \left(A_{\mathrm{min}} + A_{\mathrm{max}}\right) \tag{2.5}$$

第三項 Z は，振幅変動レートに対する依存要素として式 (2.6) により算出される.

$$Z = e^{-3.5\mathrm{c}\left(f_{\mathrm{max}} - f_{\mathrm{min}}\right)} - e^{-5.75\mathrm{c}\left(f_{\mathrm{max}} - f_{\mathrm{min}}\right)} \tag{2.6}$$

ここで，c は式 (2.7) で求められる.

$$c = \frac{0.24}{0.0207 f_{\mathrm{min}} + 18.96} \tag{2.7}$$

そして roughness 局所値の平均値から回帰式 (2.8) により，総合的な roughness 値が算出される [30].

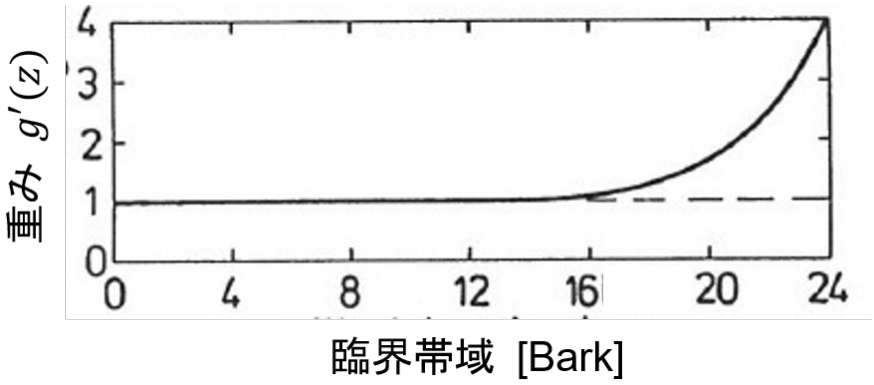


図 2.1: Sharpness の重み $g'(z)$ [2]

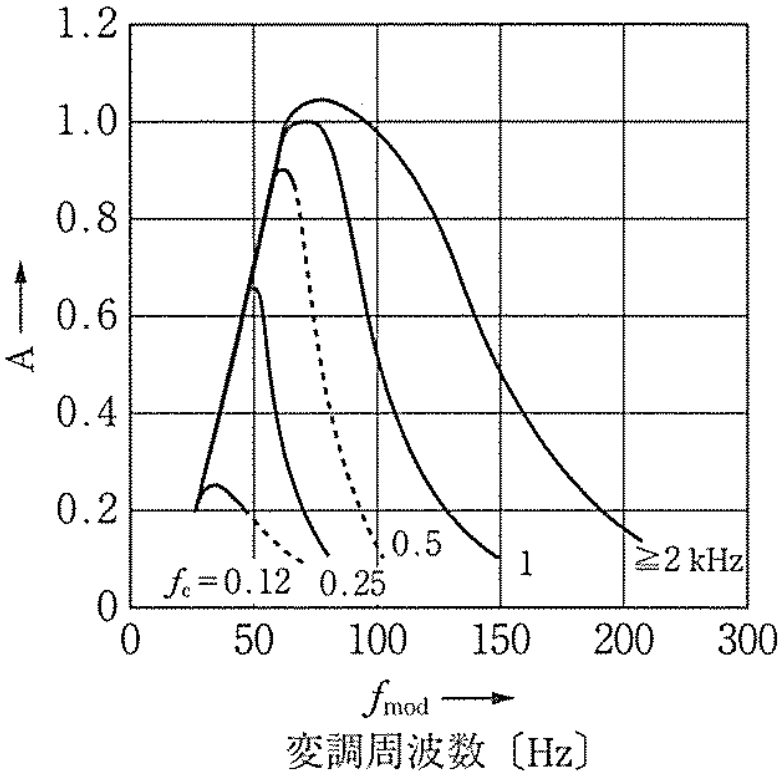


図 2.2: 振幅変調音の変調周波数と roughness の関係 [2]

$$R_{\mathrm{total}} = \log_{10}\left(average[R_{\mathrm{frame}}]\right) \times 13.988 + 48.976 \tag{2.8}$$

Boominess

Boominess はブーミング感を評価する指標で 1/3 オクターブ帯域の信号パワーに基づき式 (2.9) にて算出される [38, 39].

$$\mathrm{Booming\,index} = Bandsum \times \left(S_{\mathrm{low}} \,/\, S_{\mathrm{total}}\right) \tag{2.9}$$

ここで，$Bandsum$ は 1/3 オクターブ帯域信号の電力の合計値を表し，S_{total} が全帯域を対象としたラウドネス値，また S_{low} が 280 Hz 以下のラウドネス値を表す.

Timbral models の実装では，さらに 70 Hz 以下の帯域信号電力（S_{ll}）を求め，Booming index 値の実効値（root mean square：RMS）である $RMS_{\mathrm{Booming\,index}}$ とともに回帰式 (2.10) にて boominess の総合値が求められる [31].

$$\mathrm{Boominess}_{\mathrm{total}} = \log_{10}\left(RMS_{\mathrm{Booming\,index}}\right) \times 43.674 - \log_{10}\left(S_{\mathrm{ll}}\right) \times 10.901 + 26.837 \tag{2.10}$$

Brightness

高周波数帯域における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の 2 つが brightness の印象と相関を持つことが報告されている [40]. そのため，Timbral models では下記 2 つを組み込んだ計算モデルが導入されている.

2,000 Hz 以上の高周波数帯域で求められたスペクトル重心 SC_{br} と，20 Hz 以上の全帯域の振幅に対する 2,000 Hz 以上の高周波数帯域の振幅の比 ER_{br} が式 (2.11)，式 (2.12) でそれぞれ求められる.

$$SC_{\mathrm{br}} = \frac{\sum_{f(i)\in\{2{,}000\,\mathrm{to}\,8{,}000\,\mathrm{Hz}\}} f(i)\, m(i)}{\sum_{f(i)\in\{2{,}000\,\mathrm{to}\,8{,}000\,\mathrm{Hz}\}} m(i)} \tag{2.11}$$

$$ER_{\mathrm{br}} = \frac{\sum_{f(i)\in\{2{,}000\,\mathrm{to}\,8{,}000\,\mathrm{Hz}\}} m(i)}{\sum_{f(i)\in\{20\,\mathrm{to}\,8{,}000\,\mathrm{Hz}\}} m(i)} \tag{2.12}$$

ここで，$f(i)$ は第 i 番目の bin の周波数を示し，$m(i)$ は第 i 番目の bin の振幅を示している．

Timbral models の実装では，回帰式 (2.13) にて brightness の総合値が求められる [30].

$$\mathrm{Brightness}_{\mathrm{total}} = \log_{10}\left(ER_{\mathrm{br}}\right) \times 4.613 + \log_{10}\left(S_{\mathrm{br}}\right) \times 17.379 + 17.435 \tag{2.13}$$

Depth

空間的な属性ではなく，音色属性として定義される depth は具体的な算出モデルや音響特徴量との相関性の報告がないことから，Surry 大独自のアルゴリズムが実装されている [30]．人が奥行感を感じる音が 500 Hz 以下に高い振幅を有する傾向があることから，Timbral models では 20〜2,000 Hz の帯域におけるスペクトル重心 SC_{dp} と 20〜8,000 Hz の全帯域における振幅合計値に対する 20〜500 Hz の低周波帯域の振幅合計値の比率 LR_{dp} の 2 つを主要な特徴量として抽出している．

$$SC_{\mathrm{dp}} = \frac{\sum_{f(j)\in\{20\,\mathrm{to}\,2{,}000\,\mathrm{Hz}\}} f(j)\,m(j)}{\sum_{f(j)\in\{20\,\mathrm{to}\,2{,}000\,\mathrm{Hz}\}} m(j)} \tag{2.14}$$

$$LR_{\mathrm{dp}} = \frac{\sum_{f(j)\in\{20\,\mathrm{to}\,500\,\mathrm{Hz}\}} m(j)}{\sum_{f(j)\in\{20\,\mathrm{to}\,8{,}000\,\mathrm{Hz}\}} m(j)} \tag{2.15}$$

ここで，$f(j)$ は第 j 番目の bin の周波数を示し，$m(j)$ は第 j 番目の bin の振幅を示している．

SC_{dp} と LR_{dp} を信号振幅で正規化した後，基本周波数 f_0，音の立上り後の減衰時間 t_{decay}，LR_{dp} に sigmoid 関数をかけた $LR_{\mathrm{dp_sig}}$ を加えた 5 種を用いた線形回帰式で depth 値を求める計算モデル（式 (2.16)）が実装されている．

$$\begin{aligned}\mathrm{Depth}_{\mathrm{total}} = &-0.0044 \times SC_{\mathrm{dp}} + 32.837 \times LR_{\mathrm{dp}} + 4.751 \times t_{\mathrm{decay}} \\ &- 14.217 \times f_0 + 3.878 \times \left(LR_{\mathrm{dp}} \times t_{\mathrm{decay}}\right) \\ &- 0.854 \times \left(LR_{\mathrm{dp_sig}} \times t_{\mathrm{decay}}\right)\end{aligned} \tag{2.16}$$

この式から 20〜500 Hz の周波数成分が多く含まれて基本周波数が低く，音の立上り後の減衰時間が長いほど奥行感を高く評価されることがわかる．

上記の通り，Timbral models は主観評価実験に基づく回帰式を導入して入力音源のファイル単位での総合値を算出する点を特長としている．このことは Timbral models が算出する音質評価指標値は，ある一定時間長を想定した総括（マクロ）的な分析値であることを示している．

2.1.3 ピッチ知覚

音の高さも機器の異常音検知に大きく関係する．モータが使われている送風機や給水ポンプ，ベアリングが使われている直動スライドレールなど構成部品が回転する機器には，その回転周波数に基づく振動や音が発生する [6]．機器で発生する振動周波数は部品の回転周波数の関数として整理されている．（表 2.3）

この回転周波数の高低が音の高さ（ピッチハイト）として知覚される．ここでの音の高さは低から高への尺度上に配列される 1 次元的性質を持つもので，音程に相当する巡回的な性質を持つピッチクラス（トーンクロマ）とは異なる [41]．回転体の代表的な障害にアンバランスやアラインメント不良，内外輪の損傷があるが，いずれも回転数低下に伴う音の高さの変化や回転数に応じた音の高さにおける粗さの変化として知覚される．そのため，音の高さの変化が異常音検知の重要な鍵となる．

産業機器の動作音は複数の音源から成る複合音となるため，その音の高さも複合音のピッチ知覚と捉えて分析する必要がある．実際，音声分析などで重用されている波形相関に基づくピッチ抽出法 [42] をいくつか試みたが，声帯振動を音源とする音声と異なり複数の発生源が複雑に関係する機器動作音では機能しなかった．複合音から複数のピッチを抽出する方法としては，優勢なピッチ（predominant pitch）を抽出して反復的にストリームを抽出する手法 [43] や瞬時瞬時に複数のピッチ候補を推定して，ハーモニクスに基づくピッチ変動モデルにより複数ピッチ候補から単一ストリームを形成する手法 [44]，時変正弦波モデルを反復法にて推定してモデル化する手法 [45] などが提案されているが，いずれも楽器など明確なピッチを持つ音源の複合体を前提にしているため，直接の適用は難しい．

そこで，本論文では複合音のピッチ知覚がレジデューピッチとして知覚される [41] 点を活用した．人の耳は複合音を知覚可能ないくつかの低次倍音成分に分離する

表 2.3: 回転体の障害と振動周波数の関係 [6]

障害	高い振動レベルが予想される周波数 （f_r：回転周波数=RPM/60）
アンバランス	f_r （主に半径方向の振動）
アラインメント不良	f_r，$2f_r$ （多くの場合，$2f_r$ は f_r より高く，軸方向の振動は半径方向の振動より高い）
機械的な緩みと摩擦	$0.5f_r$，f_r といくつかの高調波成分.
転がり軸受の破損	回転体（ボール）の通過周波数として： 外輪の欠陥 $= \frac{\mathrm{N}}{2} f_r \left(1 - \frac{\mathrm{RD}}{\mathrm{PD}} cos\,\alpha\right)$ 内輪の欠陥 $= \frac{\mathrm{N}}{2} f_r \left(1 + \frac{\mathrm{RD}}{\mathrm{PD}} cos\,\alpha\right)$ ボールの欠陥 $= \frac{\mathrm{PD}}{\mathrm{RD}} f_r [1 - (\frac{\mathrm{RD}}{\mathrm{PD}} cos\,\alpha)^2]$ N: ボールの数 RD: ボールの直径 PD: ピッチサークルの直径 α：接触角度
ギアの破損	歯の噛み合わせ周波数 $= \mathrm{N}f_r$ 側帯域：$\mathrm{N}f_r \pm \mathrm{k}f_r$ N: 歯の数 k = 1,2,3 ...
送風機の羽根の破損	羽根の通過周波数 $= \mathrm{N}f_r$ N: 羽根の枚数
ポンプの障害	$\mathrm{N}f_r$ と調波成分 N: 羽根の枚数
ポンプのキャビテーション	f_r，$\mathrm{N}f_r$，$2,000\,\mathrm{Hz}$ 以上の周波数における広帯域ピーク N: 羽根の枚数

ことができ，これらの倍音成分は純音と同じ音色を持つことが知られている．高次の倍音は分離して知覚することはできないが，まとまった一つのピッチをもった成分として知覚され，この高次倍音群によって生じるピッチ（≒基本周波数として知覚されるピッチ）がレジデューピッチと呼ばれる．そして，レジデューピッチはバーチャルピッチの大きな概念で説明される [46]．バーチャルピッチの強さはスペクトラルピッチの決定に適した重要な区間におけるサブハーモニックのピッチの強さで指定される [47].

この原理を活用し，200 Hz 以下に振幅のピークを持つ基本周波数の候補の中から，その定数倍の周波数振幅を足し合わせた sub-harmonics summation（SHS）の値が最大となる周波数を最も有力な基本周波数（本論文では優勢な周波数，predominant frequency とする）として抽出した [48]．この predominant frequency が産業機器動作音の音の高さを最も適切に表しており，異常音検知に有効であると考えた.

2.2 環境音分析

人は声を介した他者とのコミュニケーションを通じて情報交換を行うとともに，周囲の環境音から様々な情報を感じ取りながら日々生活している．この基本的な人の能力を計算機で実現し，デジタル収録した音声信号の分析から周囲の状況を理解して，安心・安全な生活の支援に役立てる技術が求められている．特に音による状況理解は，画像と比べて広範な到来角度に対応でき，死角も少なく，プライバシーに対するハードルが低いとの利点がある.

本節では人の周りに存在する環境音から聴知覚を介して分析する技術を概説する．本技術の目的は大きく分けて，音からの状況推定，音からの分類，音環境の設計・修正に分けられると考える．この特長を活かした応用先として，高齢者の見守り [49,50]，ユーザ行動の識別 [51]，オーディオコンテンツへのタグ付け，音響シーン分析 [52,53] などが挙げられる.

この音環境の分析の中で重要なタスクの一つとして，事前に定義された音響的なイベントを対象に検出を行う音響イベント検出があり，様々な研究が行われている．この研究では統計的手法を用いた研究が盛んで，代表的な手法として確率的生成モデル（Gaussian mixture model：GMM，hidden Markov model：HMM），

非負値行列因子分解（non-negative matrix factorization：NMF），ニューラルネットワーク（convolutional neural network：CNN，recurrent neural network：RNN）などに基づく手法が提案されている [54].

音響イベント分析の中でも，いわゆる有事を知らせる音の検知への需要が高い．具体的には，叫び声，悲鳴，爆発，銃声，ガラス音，タイヤの滑り音の検出が挙げられる [55,56]．本応用においては，通常は正常状態の音が継続している中で稀に発生する音の変化を検出する必要性から，アプローチとしては外れ値検知が採られることが多い [57].

これらの応用先に向けては聞き分けのためのキーが予め特定できないこともあり，音響特徴量としても信号波形振幅の平均値や分散値，周波数スペクトルの中心周波数，重心や傾斜など基本的な統計素量 [56] や高次統計量を用いた研究 [58] が報告されている．また，周波数領域の特徴量として，mel スケールを利用した log-mel spectrum や MFCC（mel frequency cepstrum coefficient）が多く採用されている点が目立つ [55]．mel 自体は音の高さを示す尺度だが，音の高さの感度から聴覚の非線形特性（低い周波数の感度が高く，高い周波数の感度が低い）を表す mel スケールが得られる．すなわち log-mel spectrum は横軸が mel スケール変換された周波数，縦軸が対数振幅として入力信号を周波数スペクトル表現したものに過ぎず，特徴量としては非常に初歩的である．MFCC は log-mel spectrum を逆 FFT したもの（cepstrum）で，環境音というより音声の分析，特に音声の音源情報と包絡情報の分離に使われていた同じく初歩的な特徴量である [42]．この動向を示すために，サマリ文献 [56] を参考に環境音分析における代表的な検出手法を表 2.4 にまとめた.

このように現状の環境音分析においては，音響特徴量としては音信号の素性を表現した初歩的な特徴量にとどめて，主に確率統計分析やディープラーニングの技術により正常音の分布（傾向）をデータマイニング的に的確に捉えることで，優れた外れ値検出に繋げる方策と理解できる．しかし，これでは統計的に有意な正常音分布を得るために，質的，量的の両面で十分なデータを確保する必要がある上に，モデル化した正常音分布の完成度や妥当性の検証が難しい問題が残る.

この問題を解決するため，人が音色の観点で異常と捉える聴知覚上の差異と関連づけた音響特徴量とすることで，分布上の距離に聴感に関連づけた意味を持た

表 2.4: 環境音分析における検出手法（代表例）

対象音	特徴量	識別手法	備考	文献
足音, ノック, 叫び声	ZCR, LPC, LPCC, LFCC	GMM	音響のみ	[59]
転倒音, ガラス音	MFCC（segment）, パワー	多段階 GMM	音響のみ	[60]
転倒音, 警報, 叫び声	openSMILE	DAE+BLSTM	音響のみ	[61]
転倒音	Wavelet, ZCR	HMM	カメラも併用	[62]
転倒音	音響イベントの長さとパワー, MFCC, Shock response spectrum（SRS）	Gaussian	加速度センサも併用	[63]
転倒音	相対振幅，ピーク周波数	閾値	カメラ，加速度センサも併用	[64]

ここで，ZCR = zero crossing rate，LPC = linear predictive-coding coefficents,
LPCC = linear prediction cepstral coefficients,
LFCC = linear frequency cepstral coefficients,
DAE = denoising auto encoder，BLSTM = bidirectional long short term memory.

せることを考えることとした．その関連づけ探求のために，以下の項で人の発する音である音声から嗄声の診察を行う医療応用とモノが発する音に対して材質推定を行う産業応用の 2 方面での応用事例を挙げて，音色や音質評価指標との関連性を考察する．

2.2.1 医療応用

国内臨床の場で活用されている音質評価法に基づく検査には，音響分析法と聴覚心理的評価法の 2 種類が存在する [65]．その 1 つに GRBAS 尺度がある．これは日本音声言語医学会発声機能検査法委員会で作成された評価法である．医師が直接聞いた患者の声の聴覚心理的な評価結果を診療に用いる手法で，具体的には grade（嗄声の重症度），rough（粗ぞう性），breathy（気息性），asthenic（無力性），strained（努力性）の 5 つの側面で 0（正常）から 3（重度）までの 4 段階で評価が行われる（表 2.5）．本手法は評価用の機器が必要なく短時間での評価できる利点を持ち，運用面でも熟練者であれば再現性が高く検者間の差が小さい極めて有用な検査法である．このように本評価法は可能な限り客観的で再現性のある尺度として運用されているが，評価の信頼性に医師の経験値が影響するため，音響分析との対応づけの研究も行われている．具体的には，信号波形として観測した音声の周期性と振幅の変動を jitter および shimmer [66] と呼ばれる音響素性で表現して対応づける研究 [67] や音声の調波成分と雑音成分のエネルギー比である HNR（harmonic and noise ratio）を対象とした嗄声度の測定法の提案 [68] がある．

さらには人（医師）が実施した GRBAS 評価結果と音響素性量の関連性を定量的に調査した研究もある [7]．具体的には 90 種の音源に対して 10 名の医師が評価した GRBAS 結果と Dr.Speech や PRAAT など 4 種の音声分析ソフトウェアが算出した shimmer や HNR などの音響素性の数値との相関度を調査している（表 2.6）．この相関度調査結果によると，shimmer と GRBAS の rough（音質評価指標の roughness に相当）が特に高い相関を示すことがわかる．

GRBAS 評価を対象とした医療応用の探索より，音響素性量である shimmer が roughness と高い相関を有することが見出せ，音質評価指標を補助する特徴量候補となり得ると考えた．嗄声の粗ぞう性は，声帯周辺の病的障害などを理由に音声

の音源となる声帯の振動に問題が生じることで摩擦音が主体となる無声音化現象に相当する．この点が産業機器において部品間の摩擦が増加した摺動音と発生原理が対応しており納得性が高い．

これらの音質評価指標と関連づける研究は異常音検知開発の参考になるだけでなく，GRBAS 運用の更なる調査で医師の訓練法や診察結果の評価フィードバック法など，異常音検知の実地適用に役立つ知見が得られる可能性が高く有望である．

2.2.2 産業応用

音響を用いた産業機器診断は振動を用いた診断とともに，生産の効率化に大きく貢献している [69]．音響は振動と比べ，機器の振動そのものではなく音響放射が介在する間接的な測定法となってしまうが，原理的に非接触，広域，遠隔のセンシングを可能とし，単一機器の異なる部分や複数の機器の監視に有利となる．音響診断に適した機種としては，ベアリングを用いた回転体の診断，特に軸受けの診断は最も確立された技術の一つである．診断機器には各社のベアリングの諸元が登録されているものもあり，型名を入れると各種故障要因の発生周波数を自動計算し，その周波数でのスペクトルの大きさを診断に使用することも可能とのことである．さらに回転体の機器においては，前述のように機器が発する振動周波数が回転体の回転周波数の関数として関係が整理されている（表 2.3 参照）．

回転体に限らず機器故障の診断には振動信号の周波数分析が有効である．機械部品から発生する振動は特定の周波数から構成されるが，筐体中の伝達によりレベルは影響を受けるが周波数は変化しない．加えて異なる機械部品の振動周波数が混在しても，個々の部品の周波数に関する情報は失われないなど診断に有利な特性を有する [6]．このため振動スペクトルにおける特定の周波数の変化の分析が機械の部品と関連づけた機器の故障診断に有益な情報となり得る．人は機器の振動を空気を媒体に音の高さとして知覚している．そのため機器の振動に基づく周波数と機器の障害との関係を参考に，知覚される音の高さを機器の状態と関連づけて分析することが機器の診断に非常に重要な要素となり得る．特に先に述べた高調波（高い周波数領域のハーモニクス）から抽出した優勢周波数による診断が有望であると言える．

表 2.5: GRBAS 因子と相関する音響分析値と音質客観指標との関連づけ

	尺度	和文名	相関のある音響分析値	関連する音質評価指標（音色属性）
1	G(rade)	嗄声性	HNR，shimmer，SD F0	
2	R(ough)	粗そう性	shimmer，HNR，MeanF0，SD F0	roughness
3	B(reathy)	気息性	HNR，jitter，shimmer，MeanF0，SD F0	sharpness
4	A(sthenic)	無力性	HNR，SD F0，jitter	(anti)brightness
5	S(trained)	努力性	SD F0	depth，boominess

ここで，HNR = harmonic and noise ratio，F0 = 基本周波数，
SD F0 = 基本周波数の標準偏差，MeanF0 = 基本周波数の平均値

表 2.6: GRBAS 因子と相関のある音響分析値 [7]

TABLE 1.
Results of the Association of Acoustic Measurements in Different Software Programs to the GRBAS Parameters, Using a Multiple Linear Regression With Stepwise Variable Selection

Software Programs	G	R	B	A	S
Dr. Speech					
Mean F0 (Hz)		−0.344***	0.357***		
SD F0 (Hz)	0.329***	0.230***			0.492***
HNR (dB)	−0.687***		−0.784***	−0.817***	
Jitter local			−0.455**	−0.336*	
Shimmer local		0.650***	0.372***		
MDVP					
Mean F0 (Hz)			0.197***		
SD F0 (Hz)	−0.266*		−0.522***		0.420***
Jitter PPQ5			0.692***	0.546***	
Shimmer local	0.926***	0.742***	0.491***		
PRAAT					
Mean F0 (Hz)		−0.234***	0.372***		
SD F0 (Hz)		0.298***			0.673***
HNR dB	−0.838***	−0.580***	−0.836***	−0.875***	
Jitter local				−0.514***	
Voice Studio					
Mean F0 (Hz)		−0.269***	0.321***		
SD F0 (Hz)				−0.827***	0.526***
HNR (dB)				−0.427***	
Jitter PPQ5	0.344***			0.708***	
Shimmer local		0.646***			
ShimmerAPQ5	0.459***		0.696***		
Model	***	***	***	***	***
R^2_a *Dr. Speech*	0.61	0.64	0.66	0.35	0.23
R^2_a *MDVP*	0.53	0.55	0.68	0.29	0.17
R^2_a *PRAAT*	0.70	0.68	0.71	0.33	0.45
R^2_a *Voice Studio*	0.55	0.56	0.52	0.26	0.27

Abbreviation: R^2_a, adjusted coefficient of determination.
*$P<0.05$; **$P<0.01$; ***$P<0.001$.

産業機器の異常音は構成部品同士の摩擦や衝突により発生するため，モノから発せられた音の知覚から対象物の材質や構造を推定する psycho-mechanics と名付けられた研究分野が異常音検知は深く関連すると考える [70,71]．本研究分野は音源の力学的特性とその知覚特性との間に定量的な関係を確立することを目的に取り組まれている．力学と音響学の結びつきは決定論的であり，力学，音響学，聴覚の間には非常に密接な関係があると主張している．一例として，マレット（槌）の打音による材質分類を行った研究がある [72]．これによると，

- 固さの音色は叩くものの相対的な硬度に対応し，叩かれたもののサイズには無関係である．
- 固さの音色は周波数スペクトル全体の電力とスペクトル重心とともに増加し，スペクトルの傾斜とスペクトル重心の時間変化が緩やかになるにつれて減少する．
- 硬い構造物（マレット）は，より強く高いスペクトル重心と，より鋭く減少するスペクトル傾斜，より短い重心の時間変動を持つ．

とあり，対象物の大きさに関わらず，スペクトルの電力／重心／傾斜の時間変化で材質，特に硬度が分類できるとする知見が紹介されている．

また接触音から材質の感触を調査した文献によると，周波数の減衰が物性の粘弾性減衰特性と関連し，形状に依存せず材質のみに関連すると報告されている [73]

この音色と材質や形状との関係や材質分類に関連する音響特徴は，人が材質を推定する際に手がかりにしていることが容易に想定でき，主に金属素材から成る産業機器の異常音検知向けの特徴量検討に有力な情報となる．これらの研究成果より音の高さを含んだ音色が機器を構成する部品や機構上の特性に密接な関連を持つことが示されており，この点で機器の異常音検知に音質評価指標を用いることが有効であることが確認できた．さらに振動が伝搬した音響信号における波形変動やスペクトル特性（スペクトル重心やスペクトル傾斜）の変化の分析が異常状態など機構上の変化に対応づくことも重要な知見として得ることができた．

2.2.3　サウンドスケープ

サウンドスケープとは1960年代後半に生み出された用語で,「音の風景」と訳されることが多い．その定義は，個人，あるいは特定の社会がどのように知覚し，理解しているかに強調点の置かれた音の環境とある [74]．サウンドスケープにおいては音に関連する仕事に従事する人を対象に，特に耳で聞いた聴感を客観的に聞き分けるスキルを備えることを「イヤークリーニング」として重要視している．

実際に工場で働く点検者へのインタビューでも，熟練者からは様々な機器が動作している作業場（現場）に入った瞬間に産業機器の異常が分かるという声があった．おそらく点検者は日頃から現場で飛び交う音をサウンドスケープ的に捉えていて，普段と違うことを瞬時に感じ取っているものと推察される．

サウンドスケープの具体的な取組みとして，学際的な大規模サウンドスケープ研究である Positive Soundscape Project と国際規格である ISO12913 が挙げられる．

Positive Soundscape Project ではフィールドワークにより，サウンドスケープの知覚が認知的効果に影響されることが示され，特に人々がサウンドスケープを表現するときに使う言葉には，音源，音の記述子，サウンドスケープの記述子という 3 つの有意なクラスターが見いだされた [3]．音の記述子に音色のオノマトペ表現が見られる点が興味深い（図 2.3）．

一方，ISO12913 で定義されているサウンドスケープは音を発する全ての工業製品に当てはまるコンセプトであることが特長である．本標準では音響環境を説明する心理音響インジケータとして，sharpness や roughness など前節で述べた音質評価指標が採用されている（表 2.7）．

サウンドスケープ研究における音響心理指標の活用に関しては，loudness，sharpness，roughness に加えて，fluctuation strength と tonality が挙げられている [75]．Tonality は音の純音成分や狭帯域成分の印象の強さに対応する指標 [2] で主な算出法として，tone-to-noise ratio（TNR），prominence ratio（PR）および spectral flatness measure（SFM）がある [75]．TNR は周囲の雑音レベルに対するトーンの比を測定したもので，ISO 7729 として標準化されている [76]．PR は本質的に TNR と同じ意味を持つ指標だが，ターゲットとなる帯域のみならず隣接する帯域内の平均電力も考慮することで TNR の問題を解消している [77]．SFM は信号のパワースペクトルの geometric mean（Gm）と arithmetic mean（Am）の比とし

て開発された [78]．また tonality については新しい提案もある [79]．

これらのアクティビティは将来の異常音検知特徴量のさらなる発展研究における重要な知見となり得ると考える．

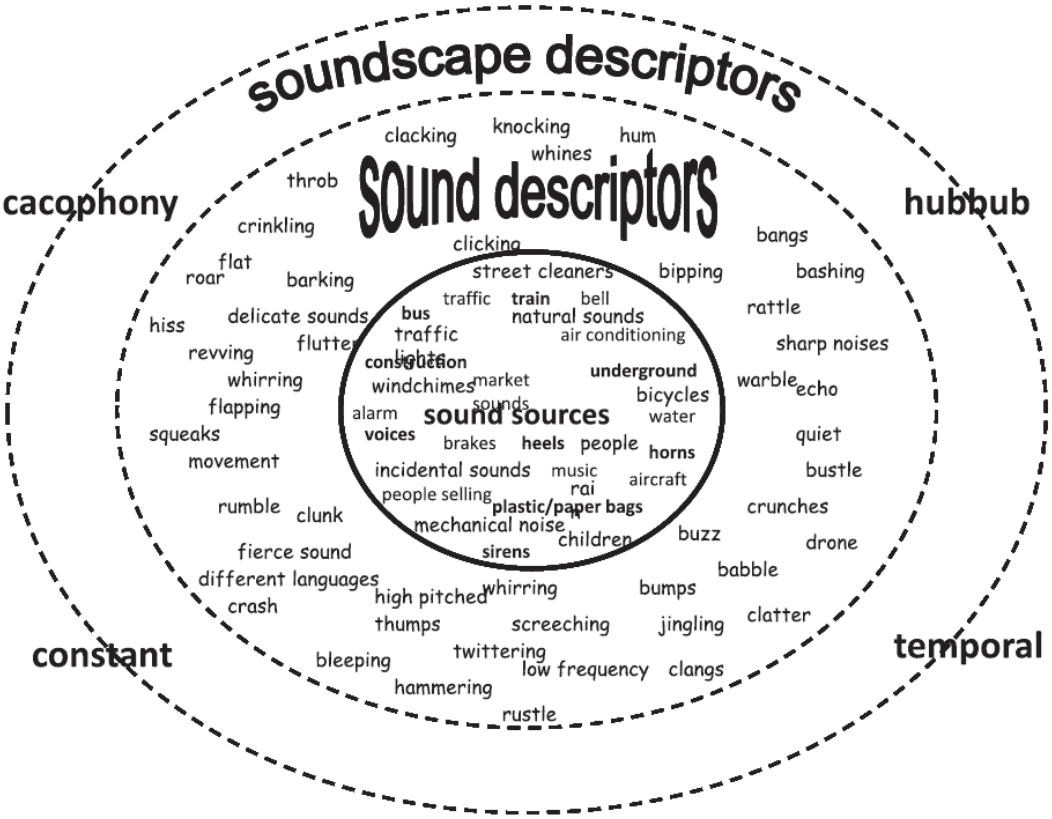


図 2.3: 主なサウンドスケープの構成要素 [3]

表 2.7: 音響環境を説明する音響／心理音響インジケータ（ISO12913-1）

Parameter	Metrics to be determined for each channel separately	Determination of representative single value	Reference
Sound pressure level	$L_{\mathrm{Aeq,T}}$，$L_{\mathrm{Ceq,T}}$，$L_{\mathrm{AF5,T}}$，$L_{\mathrm{AF95,T}}$	higher value of left and right metric value	ISO 1996-1
Loudness（time-variant loudness）	N_5，N_{average}，N_{rmc}，N_{95}，$\frac{N_5}{N_{95}}$	higher value of left and right metric value（or the average of left and right metric values）	ISO 532-1
Sharpness	S_5，S_{average}，N_{95}	higher value of left and right metric value（or the average of left and right metric values）	DIN 45692
Psychoacoustic tonality	T	higher value of left and right metric value（or the average of left and right metric values）	ECMA 418-2
Roughness	R_{10}，R_{50}	higher value of left and right metric value（or the average of left and right metric values）	ECMA 418-2
Fluctuation strength	F_{10}，F_{50}	higher value of left and right metric value（or the average of left and right metric values）	規格化未

※ https://ev-tech.jp/headline/20221227.html を基に作成.

2.3 異常音検知

異常音検知は，音により正常時の規則性から外れた不規則状態を異常として検知することである．この規則性から外れたことを検知する手法は下記に大別できる（図 2.4 参照）．

外れ値検知 正常音データの規則性を学習し，この規則性に対する観測値データの外れ値度合いを評価し，閾値以上であれば異常とする．

変化検知 データの正常時からの分布の変化を捉える．分布のパラメータ変化かモデルの変化として捉える．

本節では外れ値検知としての関連研究を中心にレビューし，生産現場（工場）に向けた適用を想定し日常点検業務の観点から既存手法の課題を抽出する．

異常音検知として，対象を特に特定しない代表的な手法提案としては音響特徴量抽出に variational auto encoder（VAE）を用いて，統計学的仮説検定（Neyman-Pearson）に基づく目的関数にて最適な特徴量を選定し，外れ値に基づく判別を行う報告がある [80]．統計学的な手法としては正常データの分布を推定し，その補完分布を異常とみなして VAE で学習する手法も報告されている [81]．

また，統計的仮説に因らない代表的な手法提案としては，音響データ取得のコスト削減を目的に非対称サンプリングの取得データに対して long short-term memory（LSTM）ベースの auto encoder（AE）を用いて判別する手法 [82]，異常音検出のために neural network（NN）を用いて複数の特徴量を抽出し，Gaussian mixture model（GMM）にて判別する手法 [83] や ArcFace 分類器と GMM を組合せた手法 [84] などもある．

いずれも対象を一般化した手法としては有用だが，対象を特定することでさらに有効な手法を実現できると考えられる．

産業機械の異常を対象とする異常音検知の研究分野としては，前述の通り Detection and Classification of Acoustic Scenes and Events（DCASE）と呼ばれる競争型ワークショップにおいて，専用の Task2 が 2020 年より新設され，オープンイノベーションスタイルでの研究が促進されている．

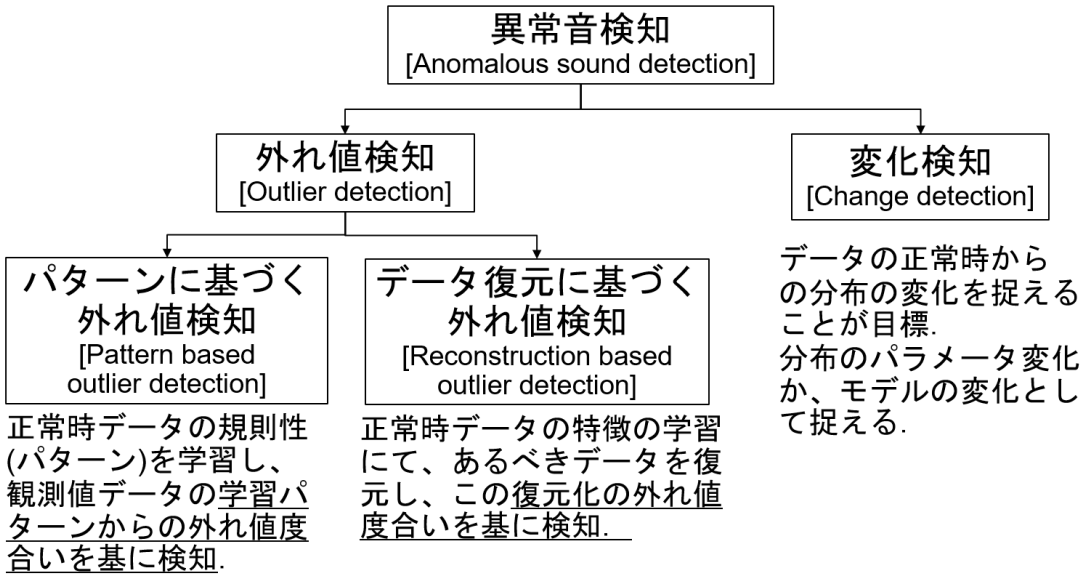


図 2.4: 異常音検知のアプローチ分類 （文献 [4] を基に加筆）

2.3.1 DCASE動向

概要

DCASEはその名の通り，音響シーンと音響イベントを対象とした検知や分類技術を競い合う，競争型ワークショップである [85]．2020年より，Task2として産業機器の状態監視向けの異常音検知のタスクが日立，NTT，同志社大のメンバで立ち上げられた．DCASE challengeと呼ばれるcompetitionでは，共通のデータセットが公開され，各参加者はこのデータセットを処理した出力を提出し，共通の評価尺度で順位づけされる．Task名が “Unsupervised detection of anomalous sound for machine condition monitoring” とある通り，教師なし学習を前提としている．

本タスクは2023年度も継続されているが，2022年度までの結果を見ても，前述の通り，検知および分類に適用されている機械学習手法に進化が見られるものの，音響特徴量は直近の2022年度でもlog-mel energiesなどの初歩的なものが主流となっている [86–88]

DCASEの公開情報から，2020年から2022年までのDCASE challengeにて成績が上位10位までの提案方式が採用した音響特徴量を表2.8にまとめた．

この初歩的な特徴量に基づく判定の手法には，対象の機種ごとに性能差が生じても原因考察が限定的となり，十分なフィードバックが行えていない傾向が見受けられる．このことは現場適用においても例えばマイクの設置の問題なのか，機器の異常音の問題なのかの切り分けなど判定性能の問題点解析や改善に向けた性能調整が難しくなる点を示唆しているものと考える．

直近のDCASEでは，ガンマトーン聴覚フィルタバンクの時間変調特徴を用いた教師なし学習の手法 [89] など特徴量に聴覚特性を組み入れる報告事例も見受けられるがまだ主流とは言えない．

2020年度のDCASE challengeでは後述する産業機器動作音データセットが共通の入力音の一部として指定された．性能上位の提案手法における判別器の動向としては，auto encoder（AE）やconvolutional neural network（CNN）などの機械学習手法の導入が目立つ．評価尺度にはDCASEが指定するanomaly scoreにおけるROC-AUC（Area under the receiver operating characteristic curve）が用いられていた．このDCASE challenge独自の尺度で対等な評価を行うためには彼らの

評価用方式（baseline）に基づく機械学習手法や評価尺度のロジックを取り込む必要があり，この制約が本来着目すべき音響特徴量の検討および評価に大きく影響するとの考えから，本論文ではDCASE2020の評価結果との直接の数値比較は行わないこととした．

産業機器動作音データセット

ここでDCASE向けに作成，公開された産業機器動作音データセット（sound dataset for malfunctioning industrial machine investigation and inspection. 以下，MIMII）を紹介する [90]. 本データセットには送風機（以下，fan），給水ポンプ（以下，pump），直動スライドレール（以下，slider），電磁弁（以下，valve）の4種の産業機器における正常時と異常時それぞれの実機動作音が収録されている．4種の機種の各々で4個体（id00,02,04,06）が用意され，全部で16種の音源で構成されている．収録データ数は正常動作音が14,719サンプル，機器異常時の動作音が3,300サンプルの合計18,019サンプルである（表2.9参照）．各音源は時間長が10秒で，サンプリング周波数は16kHzで収録されている．

対象機器4種の概要と異常時の障害種（異常条件）を以下に記す [90].

1. fan
   送風機．正常時は送風動作を継続している．4つの個体間では，主に回転数の違いにより，風切り音に音の高さの違いが知覚される．異常条件には回転軸のバランス不良，目詰まりと正常運転時の電圧変化も挙げられている．

2. pump
   給水ポンプ．正常時は水をプールに継続的に排出している．そのため，モータの動作音と水が水面を叩く音（流水音，ピチャピチャ音）が収録されている．4つの個体間では排出水量の違いから生じる流水音の混合具合の違いが知覚される．異常条件には漏れ，異物混入や詰まりが挙げられている．

3. slider
   直動スライドレール．正常時は継続的に金属性のレールの上を金属性の台座が往復動作している．そのため，レールと台座の間の摺動音が収録されてい

表 2.8: DCASE challenge における Top10 方式採用の音響特徴量

順位	DCASE2020	DCASE2021	DCASE2022
1	log-mel energies	log-mel energies	spectrum
2	log-mel energies	spectrogram	raw waveform
3	log-mel energies	spectrogram	spectrum
4	log-mel energies, spectrogram, mel energies, sqrt-mel energies	log-mel energies	log-mel energies
5	log-mel energies	spectrogram	magnitude spectrogram, magnitude spectrum
6	log-mel energies	log-mel energies	log-mel energies
7	log-mel energies	log-mel energies	log-mel energies
8	log-mel energies	log-mel energies	raw waveform
9	log-mel energies	log-mel energies	magnitude spectrogram, magnitude spectrum
10	log-mel energies	log-mel energies	log-mel energies

表 2.9: 産業機器動作音データセット収録機種とデータ数

	送風機 (fan)	給水ポンプ (pump)	直動スライドレール (slider)	電磁弁 (valve)	計
正常音データ数	4,075	3,749	3,204	3,691	14,719
異常音データ数	1,475	456	890	479	3,300
計	5,550	4,205	4,094	4,170	18,019

る．4 つの個体間では台座動作のスピードやレール間の摩擦の違いから，摺動音の時間長や音の粗さや高さの違いが知覚される．異常条件としてはレール損傷，潤滑油不足やベルトの緩みが挙げられている．

4. valve
電磁弁．正常時は弁の切り換えが継続している．そのため，金属同士が当たる音（打付け音，カチカチ音）が収録されている．4 つの個体間では打付け音の間隔や音の高さや粗さの違いが知覚される．異常状態として複数種の異物混入が挙げられている．

2.3.2 産業機器を対象とした異常音検知

産業機器を対象とした異常音検知においては，対象の機種やパーツ（部品）を特定した研究が数多く報告されている．

まず回転体の機器診断として，摩擦の影響による障害を wavelet 信号から変分的モード分解（variational mode decomposition: VMD）を用いて検知する手法の報告がある [91]．また，ベアリングの内輪，外輪，ボール異常を対象に振動信号を収録し，標準的な統計量 13 種（平均，分散，標準偏差等）からマハラノビス距離で最適な特徴量を選定してサポートベクターマシン（SVM）にて分類する手法 [92] や，同じく標準的な統計量 10 種の時間領域，周波数領域，時間-周波数領域の 3 つの観点からの特徴量を定義して SVM にて分類する手法 [93] も報告されている．さらに対象を電動ドリルやグラインダと具体化した手法として，FFT スペクトル差分を音響特徴量とし，最近傍（nearest neighbor）法や単純ベイズ（naïve bayes）法による分類を行う手法も提案されている [94]．

いずれも対象の特長を考慮しながらも信号素性や統計素量に基づく特徴量により分類を実現しているため，人の知覚とは関連づけが難しい．

収録音の “timbre”（音色）に着目した手法も複数報告されている．回転体機器の異常音検知を目的に離散フーリエ変換された周波数情報から得られた隣り合うピッチ周期の差分，つまりピッチ周期の変化を “timbre”（音色）として抽出し，差分が零でないことを異常として検知する手法がある [95]．心理音響特徴量に踏み込んだ手法としては，ギアボックスの障害診断として音響と振動信号を対象とし

た loudness, sharpness, roughness, tonality の特徴量を用いた手法 [96] やエンジンの不着火を loudness, roughness, fluctuation strength（変動強度）で判別する手法 [97] が提案されている．

最新の競合（いわゆる state of the art：SOTA）手法としては，ベアリングの 4 種の障害を対象に loudness，sharpness，roughness，変動強度など 6 種の “timbre”（音色）に関連する特徴量を sound quality feature（SQF）として抽出し，SVM で正常か異常かを判別する手法の報告がある [98]．本報告は正常と異常状態のベアリングを実際に治具を用いて回転させ，head and torso simulators（HATS）で収録した動作音に対して異常判別を行う．SVM 判別性能を正解率（accuracy）で評価して，左右の疑似耳からの入力音に対してそれぞれ 97％と 99.7％（両耳平均値で 98.4％）と高い数値を示している．このように高い分類性能を示す手法ではあるが，対象がベアリングの既知障害に限定されている点に加えて SQF と人の聞こえとの関連性への言及がなく，音色が適用された必然性が十分に示されていない．

これまでレビューした通り，異常音検知の研究は 2020 年あたりから急激に盛んとなっているが，その対象は機械学習を中心とした判別（classification）技術の進化が主流となっているように見受けられる．音響特徴量においても音質（timbre）に踏み込む報告も増えているが，1.1.3 項で指摘した通り，高い性能を示していても利用者にはそのロジックに対する理解が難しく，現場適用においては性能調整までの使いこなしに至らない可能性が高い．そのため，1.2 節で提起した通り，方式が判断した結果が人の聞こえや聴知覚上の差異に関連づけられて理解・分析できれば，現場での安定性能の実現に有効であると同時に，人は何を差異として知覚するのかを掘り下げる手立てとなり有用であると考える．

2.3.3 変化検知

異常音検知においては異常状態を示す音を発生と同時に検知することが第一義的な機能ではあるが，工場などの現場からは予防的観点より異常状態に繋がる兆候（予兆）の検知にも強い要望が寄せられている．この需要は異常音検知のアプローチ分類における変化検知に該当する（図 2.4 参照）．この変化検知はデータの背後にある分布の変化を捉える技術で，その変化がパラメータとしての分布の変

化なのか，モデルの変化なのかに大別できる [4].

本論文で採用する音色関連特徴量による異常音検知実現のアプローチを基点とすると，パラメータとしての分布の変化を捉える手法が関連する．この観点からの既存技術としては，任意の長さの局所的なダイナミクスを記述するための 1 次元シーケンスとしてメタ特徴を導出する手法 [99] や，空間埋め込み戦略によって多変量時系列を低次元空間に投影する手法 [100] が提案されている．これら 2 つのアプローチは，比較的低い計算コストで急激な変化を捉えるように設計されている．

このように変化検知は日頃の機器状態を継続監視して予兆を捉えるために必須の技術と考える．そのため本研究の将来の発展として，異なる時間分解能で局所的なダイナミクスを表現できる共通特徴を定義し，初期故障を事前に推測できるような時間的変化を捉える手法の開発が課題であると考えている．

本章では関連研究として聴覚心理学の研究知見を概説し，環境音分析の医療と産業の応用事例やサウンドスケープの研究事例から音色や音質評価指標との接点を探った．また異常音検知の最新動向をレビューにより，識別的側面を持つ音色観点からの特徴量の具現化，およびその特徴量を用いた異常音検知実現の重要性を改めて確認した．

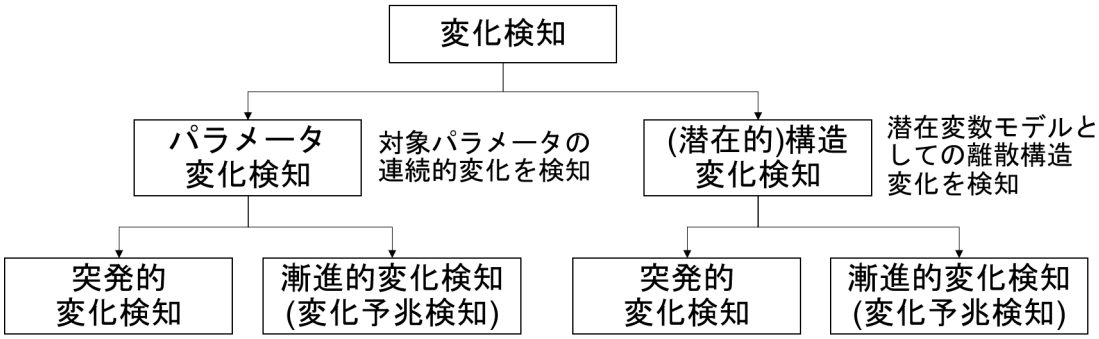


図 2.5: 変化検知の分類 （文献 [4] を基に編集）

第 3章

産業機器異常音を対象とした音響特徴量表現の方策

本章では産業機器に対する異常音検知の土台となる音響特徴量として，異常音に対する聴覚上の差異を音色の観点から特徴量として表現するための方策を述べる．まず代表的な産業機器として 4 つの機器群を挙げ，各機器群ごとに調査した主要な異常要因を基に発生音から生じる聴覚的な差異を聞き分け観点としてオノマトペで表現し，この表現を介して音質評価指標を対応づけた．その上で音質評価指標の算出要件を考察し，それを補完する短時間音響特徴量の考案にて異常音知覚に近づける音色関連特徴量実現までの検討結果を述べる．

3.1 聴知覚上の差異に基づく産業機器異常要因と音質評価指標との対応づけ

ものづくり工場では，原材料を加工するプレス機や部品を運搬するコンベヤなど，様々な機器が稼働している．これらの産業機器は担当する作業に適した機構を有しているが，その特性から下記 4 つに大別できると考えた．

1. 回転系
   送風機やモータなど，主要な構成要素（例えば，羽根や軸など）を回転させることで風を送ったり，モノを動かしたりする．

2. 摺動系
   直動スライドレールやコンベア，グラインダなど，レールの上の台車を滑らせてモノを運んだり，ヤスリなどを擦り付けてモノを削ったりする．

3. 打付け系
   鍛造プレスや電磁弁など，金属の板に型を打付けて成型したり，弁を壁に打付けながら経路を開閉することで物体を適切に運搬する．

4. 流体移送系
   給水ポンプや水車など，モータの力を使ったり，流路に流れを作ることにより流体を移送する．

各機器は継続的に発する動作音で稼働の状態を表しており，故障などで動作に異常をきたすとその要因に基づく個別の音の変化が生じる [69]．点検員（人）は

この動作音の変化を聞いた印象の差異として知覚（これを聴知覚上の差異とする）し，対象機器の機構上の知識や過去の診断経験などの事前知に基づいて非侵入型の機器診断に活用している．人は正常動作の音（正常音）と異常動作の音（異常音）の間を聞き分ける観点を音色としての印象に基づく聴知覚上の差異として捉えていると考えられる．そしてその差異は言語的な音色表現であるオノマトペと親和性が高いと考えられる [101]．過去に行った複数業種の工場で働く点検員を対象としたインタビューからは，この聴知覚上の差異をオノマトペによる音色変化として表現したり，点検者間で共有する傾向があることがわかった．

そこで点検員が有する事前知を想定しながら現場観察の知見を用いて，4 つの各機器群における異常要因を調査し，異常時の聴知覚上の差異をオノマトペ表現を介して音質評価指標に対応づけることとした．機械の異常音に対する差異のオノマトペ表現には，人が異常音から受ける印象に似せた発声を行うことに着目 [101] し，音声生成系における調音種（表 3.1 参照）との関連性も考察した [8]．

3.1.1 回転系について

異常要因

送風機（fan）のような回転体を有する機器においては，主な異常要因として回転子のバランス不良やベアリングの劣化や損傷が挙げられる．異常音が出ているにもかかわらず使い続けると機器の破損のみならず，コイルの焼損から火災に至る可能性もあり，異常音による早期の検出が求められる．回転子のバランス不良の場合，回転のエネルギーは振動に変換される．また，別の要因としてアラインメント不良（misalignment）や回転子のクラックなども起こり得る．保守には潤滑油が有効だが，補充が遅れた場合等の潤滑油不足が摩擦の増加などの問題を引き起こす [6]．

発生音と聴知覚上の差異

正常時の動作においては，回転子からの摺動音と羽根による風切り音が生じる．調音種の摩擦音に関連するヒュー（whir）音やビューン（whizz）音に近い音として知覚される．また，異常状態で生じる回転子のバランス不良や偏心に伴う音に対しては調音種の摩擦音と閉鎖音に関連するキーキー

（scratching）音やガタガタ（rattling）音が差異として知覚される．またベアリング不良に対しては，調音種における摩擦音のブーン音や鼻音系のウーン音が知覚される．

音質評価指標との対応

これらの聴知覚上の差異はオノマトペを用いた表現を介して，その音色としての印象から音質評価指標と対応づけることができる．正常音のヒュー音の印象からは明るさ感を示す brightness と鋭さ感と示す sharpness が対応する．異常音のキーキー音には鋭さ感の sharpness，ガタガタ音には粗さの roughness，ブーン音にはブーミング感の boominess，ウーン音には奥行感を示す depth を対応づけることができる．

3.1.2 摺動系

異常要因

代表的な機器として金属性のレールの上を金属性の台座がスライドすることで，台座に取り付けた物体のスライド動作を実現する直動スライドレール（slider）を考える．このような機器にはスムースなスライド動作のために多くのベアリングが使われている．このベアリングの典型的な異常原因としては潤滑油不足や破損が挙げられるが，回転疲労を生じて，最も厳しい異常として軌道面の表層部がうろこ状に剥がれるフレーキング（剥離）も生じる．そのため，潤滑油不足，破損やフレーキング，あるいはレールの損傷のいずれにおいても台座のスライド動作における摩擦の増加を生じる [102].

発生音と聴知覚上の差異

直動スライドレールの通常動作においては，台座のスライドに伴う摺動音が発生する．スライド時には調音種の摩擦音に関連するシュー（hiss）音と台座がレールの端で止まり折り返す際に調音種の閉鎖音に関連するカチカチ（click）音が知覚される． また，機器異常の状態では，主に軌道面の摩擦の増加により調音種の摩擦音に関連するキーキー（squeal）音やギーギー（creaking）音が，さらに台座の共振で発生する音が調音種の鼻音に関連するウーン音となり差異として知覚される．

音質評価指標との対応

正常音における滑らかな摺動音（摩擦音）の印象にはbrightness，sharpnessとroughnessが，カチカチ音にはsharpnessとbrightnessが対応する．異常音のキーキー音やギーギー音の印象にはsharpnessとroughnessが対応し，ウーン音にはdepthが対応する．

3.1.3 打付け系について

異常要因

打付け系の代表機種としての電磁弁（valve）は電磁石と弁を組合せた構造で，電気のON/OFFにより経路口を開閉することで流体を止めたり流したり，流れの方向を切り換える機器である．打付け動作の連続から，長時間動作において電磁弁の打付け面の摩耗や変形，さらには損傷が生じる [103].

発生音と聴知覚上の差異

正常時には弁の開閉に対応した打付け音が発生する．正常動作時は調音種の閉鎖音に関連する定期的なカチカチ音（click）が知覚される．異常時には，打付け面の劣化や疲労により打付け音の異常な振動（ビビり）による不明瞭感が調音種の閉鎖音に摩擦音が加わったビンビン音として，また部品の共振音が調音種の閉鎖音と鼻音系に関連するブンブン音として差異を感じる．

音質評価指標との対応

正常音のカチカチ音から受ける印象として，brightnessとsharpnessが対応する．異常音において打ち付け音が鈍く不明瞭になる印象にはroughnessとboominessが対応づけられ，ブンブン音にはdepthが対応する．

3.1.4 流体移送系について

異常要因

代表的機器は給水ポンプ（pump）である．その典型的な異常は，異物混入やキャビテーションによって生じる詰まりや水漏れの現象である．これら異常時には，流体を安定的に動かすことができず，不定期な吸引と排出が繰り

返し起こることになる [104]．また管内の急激な圧力変化が起こる場合があり，衝撃音を伴うウォータハンマ現象が発生する．

発生音と聴知覚上の差異

ポンプ自体から発する音はモータなどの回転体と同様だが，給水ポンプの主な活用シーンが水の供給（移送）となるため，主な正常時の動作音として流水音が生じる．ここでの流水音は排出された水がプールなどにためられた水の表面にあたって生じる音で，調音種の破擦音と閉鎖音に関連するピチャピチャ（splashing）音として知覚される．異常時はこの流水音が消失する．加えて配管の詰まりで生じる音が調音種の閉鎖音と摩擦音に関連するゴボゴボ音（gurgling）やウォータハンマで生じる音が調音種の閉鎖音と摩擦音に関連するブン音として差異を感じる．

音質評価指標との対応

正常音におけるピチャピチャ音からの印象として brightness と sharpness が対応する．異常音の配管の詰まりによるゴボゴボ音やウォータハンマによるブン音からの印象としては roughness と boominess が対応すると考える．

表 3.1: 音声生成系の調音種と発生音の対応 [8]

調音種	発生音
鼻音	マ行音，ナ行音，ン
閉鎖音	パ行/バ行音，タ行/ダ行音，キ，カ行/ガ行音
摩擦音	フ，サ，ザ，シ，ジ，ヒ，ハ，ヘ，ホ
接近音	ワ音，ヤ行音
破擦音	ツ，ヅ，チ，ヂ
流音	ラ行音

表 3.2: 聴知覚上の差異に基づく機器の異常要因と音質評価指標の対応づけ

	機器群（代表機種）	異常の要因	聴知覚上の差異	関連指標
1	回転系（fan）	回転子のバランス不良	回転軸のブレ（ガタガタ）音や摩擦（キーキー）音の知覚	boominess, brightness, depth, roughness, sharpness
		ベアリングの劣化や損傷	回転軸の摩擦（ブーン，ウーン）音の知覚	
2	摺動系（slider）	潤滑油不足，レールの損傷による摩擦の増加	摺動音に粗さ，軋み（ギーギー，キーキー）音，台座の共振によるウーン音の知覚	brightness, depth, roughness, sharpness
3	打付け系（valve）	打付け面の劣化や疲労，異物混入	打付け音の不明瞭感（ビビり）としてのビンビン音，打付け部品の共振（ブンブン）音の知覚	boominess, brightness, depth, roughness, sharpness
4	流体移送系（pump）	異物混入やキャビテーションによる配管の詰まりや水漏れ	流水（ピチャピチャ）音の消失	boominess, brightness, roughness, sharpness
			詰まり（ゴボゴボ）音やウォータハンマ（ブン）音の知覚	

3.2 音質評価指標の分析時間長の考察

音質評価指標の算出にはTimbral models（2.1.2項参照）を利用することとした．2.1.2項で述べた通り，本modelは入力ファイル単位でのマクロ的な指標値を算出する仕様だが，実運用を考慮すると連続動作のためのフレーム処理が必要であり適切な分析時間長の設定が必要となる．

そこで音色観点で産業機器動作音の変化を数値化しようとする音質評価指標には，全体的な印象を捉えつつ，その音響的な変化がどんな事象に関連づくのかを認知する [18] 必要性から，1,024 ms の分析時間長が適切だと考えた．この時間長は人が音の時間順序を判断するために最低700 ms を要するとする研究報告 [105] と，単一知覚の持続時間が最大で1.5 s までとする研究報告 [106] に基づき，700 ms〜1,500 ms のほぼ中間の時間長として1,024 ms を導いた．

音質評価指標としては適切な分析時間長を設定できたと考えるが，一方で音響符号化などは80 ms〜数百 ms オーダーで音響信号の細かい時間変化を捉えて圧縮を行っている [107]．このことからマクロ的な音質評価指標を補助して，産業機器動作音の数十から数百 ms オーダの細かい時間変化を捉えるためのミクロ的な特徴量分析を組合せることが有効だと考え，2種類の短時間音響特徴量の導入を検討した．

3.3 音質評価指標を補助する短時間音響特徴量の検討

3.3.1 波形振幅変動特徴量

まず1つ目の特徴量は，動作音の波形の変動を捉える振幅強調 shimmer（amplified shimmer: AS）である．この特徴量は動作音のデジタル系列の隣り合うサンプルの振幅値の変化量を評価する音響特徴量である shimmer を基に考案した．shimmer は式 (3.1) にて算出される [66]．

$$\text{shimmer (relative)} = \frac{\frac{1}{N-1}\sum_{n=1}^{N-1}|A_n - A_{n-1}|}{\frac{1}{N}\sum_{n=1}^{N}|A_n|} \tag{3.1}$$

ここで，A_n は入力信号の n 番目の振幅値で，N は算出フレームのサンプル数である．

2.2.1 項に記載した通り，shimmer は GRBAS スケールにおける粗そう性（rough）に強い相関を持つことが報告されている．そのため，音質評価指標の roughness をより高い時間分解能で分析できる特徴量として導入を検討した．

実際に直動スライドレール（slider）の動作音に対して，256 ms フレーム（2,048 サンプル）ごとに式 (3.1) で算出した shimmer と同じ区間の振幅平均値を並べた図を図 3.1 に示す．

直動スライドレールはレールの上を台座が往復運動する機器であるため，レール上を動く速度に応じて摩擦が強まり，速度と音の大きさが対応する．この音の波形に shimmer を重ねると，図の通り，台座の動きに合わせて音と shimmer の動きが重なり，shimmer の上下動が動作音から聞こえる（摩擦による）粗さの知覚に対応する．さらに，この知覚の度合いは信号の振幅に大きく依存することを考慮して，振幅強調 shimmer（amplified shimmer: AS）を特徴量として新たに考案した．

$$
\begin{aligned}
\text{amplified shimmer} &= \frac{\frac{1}{N-1}\sum_{n=1}^{N-1}|A_n - A_{n-1}|}{\frac{1}{N}\sum_{n=1}^{N}|A_n|} \times \frac{1}{N}\sum_{n=1}^{N}|A_n| \\
&= \frac{1}{N-1}\sum_{n=1}^{N-1}|A_n - A_{n-1}|
\end{aligned}
\tag{3.2}
$$

ここで，A_n は入力信号の n 番目の振幅値で，N は算出フレームのサンプル数である．

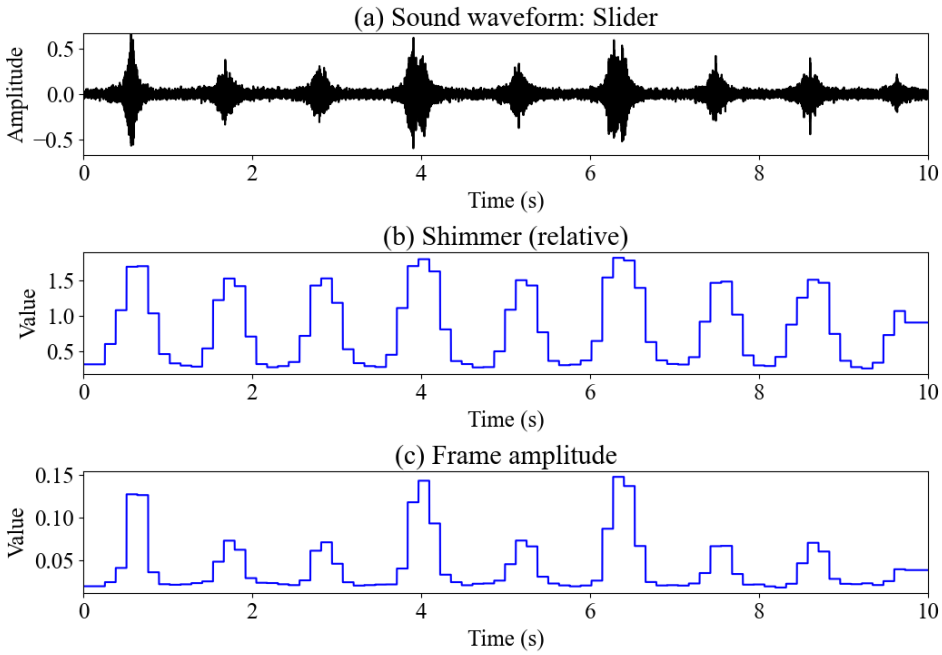


図 3.1: （a）正常状態の直動スライドレールの動作音波形,（b）動作音波形から計算した shimmer 値,（c）動作音波形における 256ms 毎のフレーム振幅値

3.3.2 優勢周波数変動特徴量

2つ目の特徴量は，動作音の音の高さの変動を捉える振幅強調優勢周波数（amplified predominant frequency: APF）とした．この特徴量は複合音のバーチャルピッチに関連が強いハーモニクスを基に，音の高さの感覚に寄与すると考えられている優勢な周波数（predominant frequency: PF）に基づく特徴量として考案した．PF は，サブハーモニクス加算（sub harmonics summation: SHS）法 [48] を適用し，SHS 値が最大となる基本周波数として求める．そのプロセスは下記の通り．

1. 振幅スペクトルの算出
   2,048 ポイント FFT を機器動作音入力に適用して，振幅スペクトルを算出する．具体的には，2,048 サンプル（16 kHz サンプリング信号の 128 ms に相当）長の窓を 1,024 サンプル（16 kHz サンプリング信号の 64 ms に相当）長ごとにシフトしながら，1,024 ポイントの振幅スペクトルを得る．

2. 振幅スペクトルからの基本周波数候補の抽出
   1,024 ポイントの振幅スペクトラムに対して，2,000 Hz 以下の周波数領域における振幅値の大きい方から 3 つの周波数を基本周波数候補として抽出する．周波数候補は f_i で表し，その振幅値を $A_{mp}(f_i)$ で表すこととする．なお，ここで i は 1〜3 となる．

3. サブハーモニクス加算値の算出
   i 個の基本周波数候補に対して，それぞれサブハーモニクス加算（SHS）値を算出する．基本周波数の k 倍の周波数における振幅値を足し合わせていく．その際，高い周波数に離れていくにつれて寄与度を下げるために，$h(k)$ を振幅値に乗じた．f_i におけるサブハーモニクス加算値 $SHS(f_i)$ は式 (3.3) で求めた．
   $$\mathrm{SHS}(f_i) = \sum_{k=1}^{K} h(k)\, A_{mp}(kf_i) \tag{3.3}$$
   ここで，$A_{mp}(kf_i)$ は，f_i Hz の k 倍の周波数 kf_i の振幅値, K は加算の次数で 5 に設定した．また $h(k)$ は 0.84^{k-1} に設定した．

4. 優勢周波数の算出

SHS (f_i) の最大を与える周波数 f_i Hz を優勢周波数 f_{pdf} とする．

$$f_{\mathrm{pdf}} = \underset{f_i}{\mathrm{argmax}}\,(\mathrm{SHS}(f_i)) \tag{3.4}$$

この優勢周波数 f_{pdf} を複合音としての産業機器動作音の高さを示す特徴量として抽出した．このサブハーモニクス加算値自体が知覚の度合いに対応することから強調因子として乗じることとし，式 (3.5) で示す振幅強調優勢周波数（amplified predominant frequency: APF）を新たに考案した．

$$\mathrm{amplified\,predominant\,frequency} = f_{\mathrm{pdf}} \times \mathrm{SHS}(f_{\mathrm{pdf}}) \tag{3.5}$$

3.4 音質評価指標と短時間音響特徴量の組合せによる音色関連特徴量の実現

前節までの検討を踏まえ，産業機器動作異常時の動作音の変化を音色観点から捉える音響特徴量として，音質評価指標と独自に考案した短時間音響特徴量の組合せにて構成される音色関連特徴量を定義した．具体的には 4 つに大別した産業機器の機器群ごとに 3.1 節にて対応づけを行った音質評価指標の 4〜5 種と 3.3 節で考案した 2 種の短時間音響特徴量とを組合せる．音質評価指標は機器群ごとに選択するが，短時間音響特徴量は全ての機器群に割り当てる．

産業機器は多種多様に存在するが，機構上の特性から一番近い機器群を当てはめることができると考える．また機構原理が全く異なる機器への拡張にも，同様のアプローチにて音質評価指標の割当変更や追加を行うことで対応可能と考える．

4 機器群ごとの音質評価指標と短時間音響特徴量の割当を表 3.3 にまとめる．

ここで音色関連特徴量を構成する要素と音色因子（2.1 節参照）がどのように関連するかを考察してみると，表 3.4 に示す通り，5 種の指標と 2 種の特徴量が 3 つの因子にバランス良く対応づくことがわかる．

表 3.3: 4 機器群に対する音色関連特徴量の割当

No.	機器群	代表機種	Bm	Br	Dp	Rg	Sp	APF	AS
1	回転系	Fan	○	○	○	○	○	○	○
2	摺動系	Slider	-	○	○	○	○	○	○
3	打付け系	Valve	○	○	○	○	○	○	○
4	流体移送系	Pump	○	○	-	○	○	○	○

ここで，Bm = boominess，Br = brightness，Dp = depth，
Rg = roughness，Sp = sharpness，APF = 振幅強調優勢周波数，
AS = 振幅強調 shimmer を示す.

表 3.4: 音色関連特徴量の構成要素と音色因子の関係について

音色因子	音質評価指標	短時間音響特徴量
金属性因子	brightness，sharpness	APF
迫力因子	boominess，depth	APF
美的因子	roughness	AS

ここで，APF = 振幅強調優勢周波数，
AS = 振幅強調 shimmer を示す.

3.5 異常音検知システム化検討

音色関連特徴量を用いた異常音検知システムを実現するために図 1.2 の構成に沿って，特徴量抽出の分析時間長と異常検出の学習および判別法を検討した．

3.5.1 音色関連特徴量の分析時間長の検討

検討対象は音質評価指標と短時間音響特徴量である．まず，3.2 節における考察結果から音質評価指標の分析時間長は 1,024 ms とした．

次にこの音質評価指標に対して，短時間音響特徴量は高い時間分解能にて音響分析する位置づけであるため，分析時間長を音質評価指標の 1,024 ms に対して 1/4 から 1/8 の範囲で考えた．音質評価指標に対する時間長比率を少なくすればより高い時間分解能で分析が行えることになるが，ベースとなる音質評価指標との相関を維持し難くなると考えた．そこで音の高さの変動を捉える APF は音響符号化のフレーム長（80 ms）を参考に 128 ms とし，波形振幅の変動を捉える AS は粗さの知覚に関連することから比率を下げて 256 ms とした．

本論文では上記の考えから分析時間長を設定したが，分析時間長は特徴抽出の重要なパラメータであるため，捉える事象の特性を信号処理分析および心理聴覚の両面から解析することで常に最適化していく必要があると考える．

3.5.2 機械学習による異常分類仕様の検討

観測された音データが異常動作音に属するか否か，つまり異常性の有無の判別には機械学習技術を利用した．本論文では産業機器の動作音の変化を捉える音色関連特徴量の実現を主眼としているため，予め正常と異常のラベル付けが行われているデータを想定し，教師あり学習による二値判別器として，サポートベクターマシン（support vector machine:SVM）を適用することとした [108].

サポートベクターマシンは 2 つのクラスにマージンを伴う境界を設定し，そのマージンを最大化する規範にそって最適な分類を行う（図 3.2 参照）.

SVM は他の二値分類器であるロジスティック分類等と比較して，高次元領域の分類に優れた正確性を有することを特長としている．

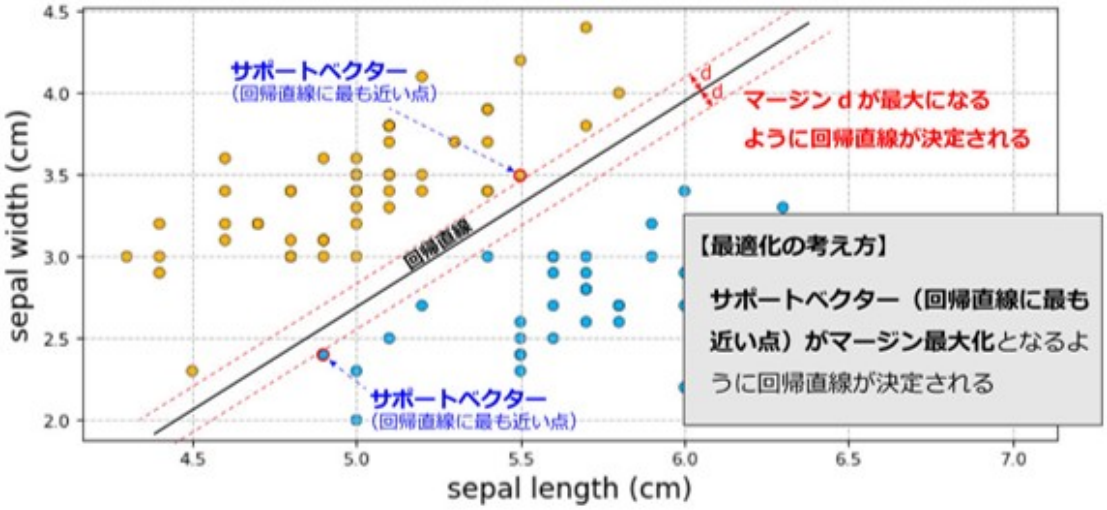


図 3.2: サポートベクターマシン （出典元： https://kit.socinno.com/2_3_ai/）

本章では4つに大別した産業機器の機器群において，機構上の異常の要因から発生する音に対する聴知覚の差異をオノマトペを交えて表現し，関連する音質評価指標との対応づけを表3.2としてまとめた．音質評価指標より高い時間分解能で機器動作音の波形振幅変動と音の高さに着目した周波数変動を捉える2種類の短時間音響特徴量を考案し，音質評価指標と組合せることで音色関連特徴量を実現した（表3.3参照）．そして，この音色関連特徴量に対してサポートベクトルマシンで正常／異常を分類する仕様を検討し，異常音検知システム実現の筋道を立てた．

第 4章

音質評価指標に基づく音響特徴量の有効性検証

本章では，異常音の聴知覚上の差異に関連づけた音質評価指標 5 種と短時間音響特徴量 2 種の組合せによる音色関連特徴量について，産業機器動作音データセットの収録音を入力とした特徴量分布の評価および分析により異常音検知に対する有効性を検証する．

本章以降の検証にあたっては，実際の産業機器動作音を収録したデータセットとして，2.3.1 項で紹介した産業機器動作音データセット（MIMII）を使用した [90]. 本データセットには実運用環境を想定した実機動作音が多数収録されており，正常時の動作音を “正常音”，異常時の動作音を “異常音” としてラベル付きで公開されている．Kaggle のデータセットなども含めて産業機器動作音のデータセットを幅広く探索したが，現時点でここまで大量の実機動作音が適切な条件でラベル付けされて公開されているデータセットは他にはない．

本データセットに収録されている機種は送風機（fan），給水ポンプ（pump），直動スライドレール（slider），電磁弁（valve）の 4 機種である（表 2.9 参照）．本章以降はこの 4 機種全てを対象に検証を進めていく．また本データセットは実運用環境を想定し，別収録した工場内騒音を 3 種の SNR 条件（6 dB,0 dB,−6 dB）で重畳しているが，本論文では環境騒音の影響を最小に抑えるために SNR 条件としては 6 dB を採用した．

4.1 産業機器動作音スペクトラムを用いた聴知覚上の差異の考察

産業機器動作音データセットに収録の 4 機種の各々について，正常時動作音（正常音）と異常時動作音（異常音）の聴知覚上の差異をスペクトラムを図示しながら考察し，音質評価指標との関連づけ（表 3.2）を検証する．

送風機 （fan） 図 4.1 参照

正常音 ブレード（羽根）の安定した回転による風切り音と回転子からの微かな摺動音が生じる．これらの音がスペクトラム（図 4.1）においては 1,000 Hz 以下に水平に並んだ線（調波構造）として観察される．音色としては主に明るさの boominess と鋭さの sharpness が知覚される．

異常音 回転軸の劣化やアンバランスで摩擦が増え，強い摺動音や振動音が生じる．スペクトルには 1,000 Hz 以下の調波構造に摺動音に対応する高振幅成分が見られ，500 Hz や 200 Hz 付近に振動音に対応する散発的な高振幅成分（濃い赤い点）が観察されている．音色としては強い風切り音に対応するブーミング感の boominess，振動音と振動音に粗さの roughness や奥行感の depth，さらに鋭さの sharpness が強く知覚される．

給水ポンプ （pump） 図 4.2 参照

正常音 機器自体の動作音とともに収録音からは安定した流水音（ピチャピチャ音）が聞こえる．スペクトラム（図 4.2）には 4,000 Hz 以上の領域に排出された水が水面にあたる音に相当する赤の線が観測される．この水面への衝突音が複数無秩序に発生している様子がランダムな振幅変化（赤の領域の点在）で確認できる．音色としては brightness や sharpness が知覚される．

異常音 配管の詰まりによる流水音の消失と詰まり音（ゴボゴボ音）が聞こえる．スペクトルでは流水音に対応する 4,000 Hz 以上のランダムな振幅変化が無くなり，500 Hz 付近に強い振幅（濃い赤の領域）が確認できる．音色としては boominess や低域の粗さ（roughness）が知覚される．

直動スライドレール （slider） 図 4.3 参照

正常音 金属レールの上を台座が往復動作する動きに伴う定期的に連続する薄い摺動音が生じる．スペクトラム（図 4.3）には 500 Hz 以下と 5,000 Hz 付近に高い振幅（濃い赤）が観察される．音色としては brightness，sharpness と roughness が知覚される．

異常音 摩擦増加による粗い摺動音（ビービー音）が生じる．スペクトルでは 250 Hz 付近の高い振幅（濃い赤）と台座がレール上を滑っている区間のビビり音に対応する 8,000 Hz までの均一の高い振幅（濃い赤）が観察される．音色としては主に粗さの roughness に加えて共振に対応する奥行感の depth が知覚される．

電磁弁　（valve）　図 4.4 参照

正常音 弁が壁に当たった際の打付け音が定期的に聞こえる（カチカチ音）．スペクトラム（図 4.4）には 2,000 Hz までの小さい打付け音と 8,000 Hz までの振幅が大きい打付け音（赤い棒状の線）が規則正しく並んで観察される．音色としては brightness，sharpness が知覚される．

異常音 劣化により打付け動作が不安定となり，8,000 Hz の高周波域までの成分を持つ打付け音（メイン）の濁りが聞こえ，2,000 Hz までの成分を持つ打付け音（サブ）が消失する．スペクトルではサブの打付け音の線が見えなくなり，メインの打付け音の細り（余韻領域の減少）と 500 Hz 以下の振幅増（濃い赤色）が観察される．音色としては粗さの roughnes に低周波域のビビりやブンブン音に対応する boominess と depth が知覚される．

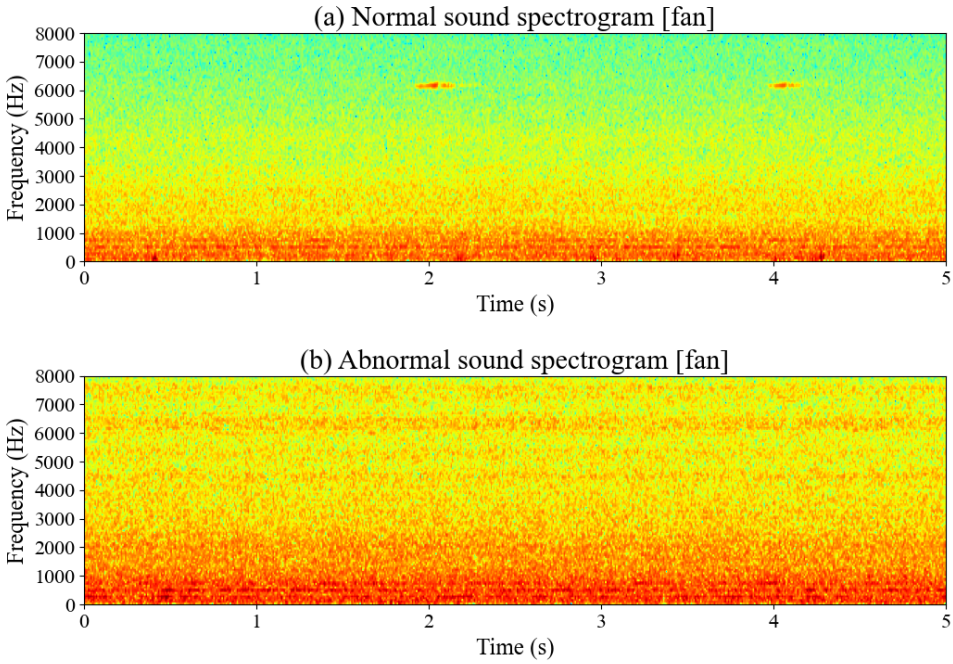


図 4.1: 送風機（fan）の動作音スペクトラム：(a) 正常動作音，(b) 異常動作音

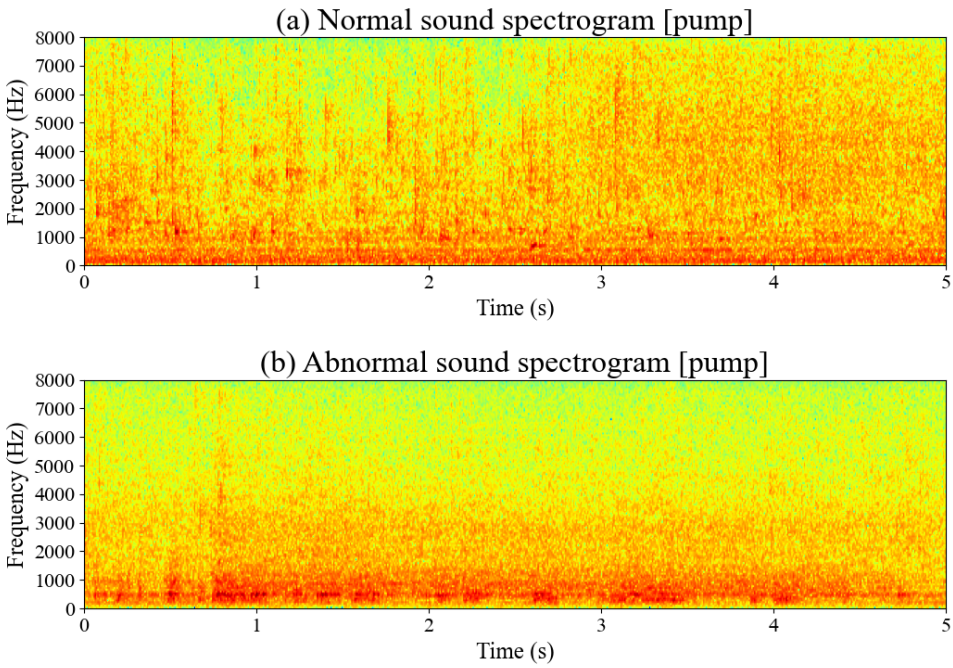


図 4.2: 給水ポンプ（pump）の動作音スペクトラム：(a) 正常動作音，(b) 異常動作音

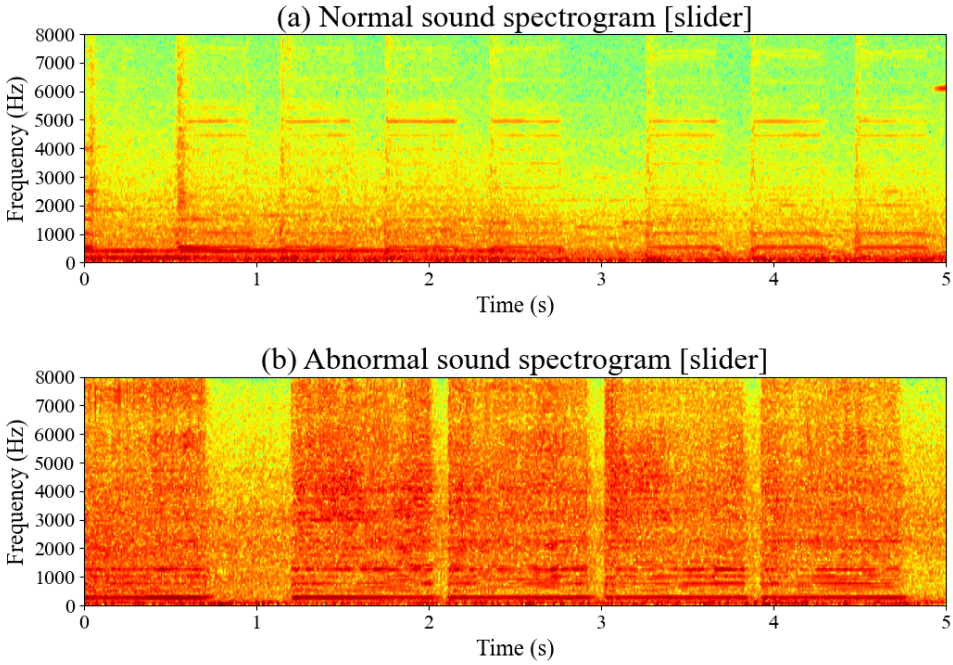


図 4.3: 直動スライドレール（slider）の動作音スペクトラム：(a) 正常動作音，(b) 異常動作音

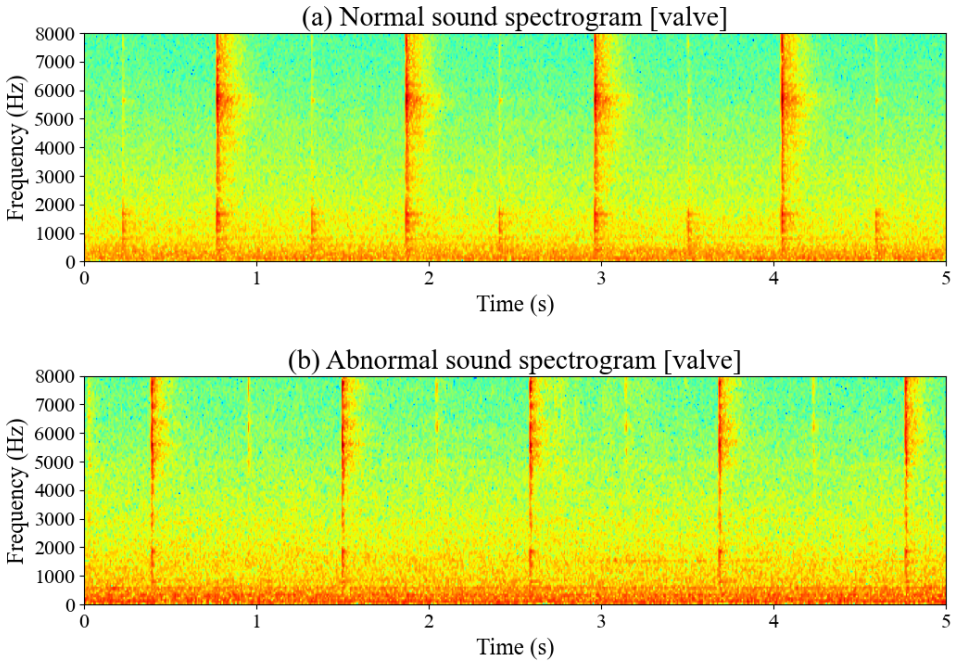


図 4.4: 電磁弁（valve）の動作音スペクトラム：(a) 正常動作音，(b) 異常動作音

4.2 短時間音響特徴量と音質評価指標の関連性検証

比較的長い時間長で音色属性を数値化する音質評価指標に対して，高い時間分解能で分析する短時間音響特徴量を 2 種類考案したが (3.3 節参照)，ここで産業機器動作音データセットの正常音を用いて短時間音響特徴量と音質評価指標の関連性を検証する．具体的には，短時間音響特徴量（x）と同一フレーム信号から算出したターゲット音質評価指標値（y）の相関値（$r = r_{xy}$）を評価した．

相関値はピアソンの積率相関係数（以下，ピアソン相関係数）を用いて，式 (4.1) にて算出した [109].

$$r_{xy} = \frac{\sum_{n=1}^{N} (x_n - \bar{x})(y_n - \bar{y})}{\sqrt{\sum_{n=1}^{N} (x_n - \bar{x})^2}\sqrt{\sum_{n=1}^{N} (y_n - \bar{y})^2}} \tag{4.1}$$

ここで，N は相関計算に用いたデータサンプル数，$\bar{x}$，$\bar{y}$ はそれぞれ x と y の平均値を表す.

4.2.1 振幅強調 shimmer（AS）と roughness の相関について

振幅強調 shimmer（AS）についてターゲット音質評価指標である roughness 算出値と組合せてプロットした散布図を図 4.5 に示す．両者の数値の相関をピアソン相関係数にて評価したところ，$r = 0.906$ と高い相関を示すことがわかった．このことから産業機械の動作音データに対しては，考案した振幅強調 shimmer（AS）が roughness の音質評価指標値と相関を持ちながら，4 倍の高い時間分解能で分析できていることが確認できた.

4.2.2 振幅強調優勢周波数（APF）と boominess の相関について

同様に，振幅強調優勢周波数（APF）についてもターゲット音質評価指標である boominess 値との散布図を図 4.6 に示す．両者の数値間のピアソン相関係数値も $r = -0.856$ と負の高い相関を示すことがわかった．2.1.2 項に記載の通り，音の高さに関連する優勢ピッチが低いほど boominess 値は高くなる原理であるため，負の相関を持つことになる．このことから前項同様，産業機械の動作音データに対し

ては，考案した振幅強調優勢周波数（APF）が boominess の音質評価指標値と相関を持ちながら，8 倍のより高い時間分解能で分析できていることが確認できた．

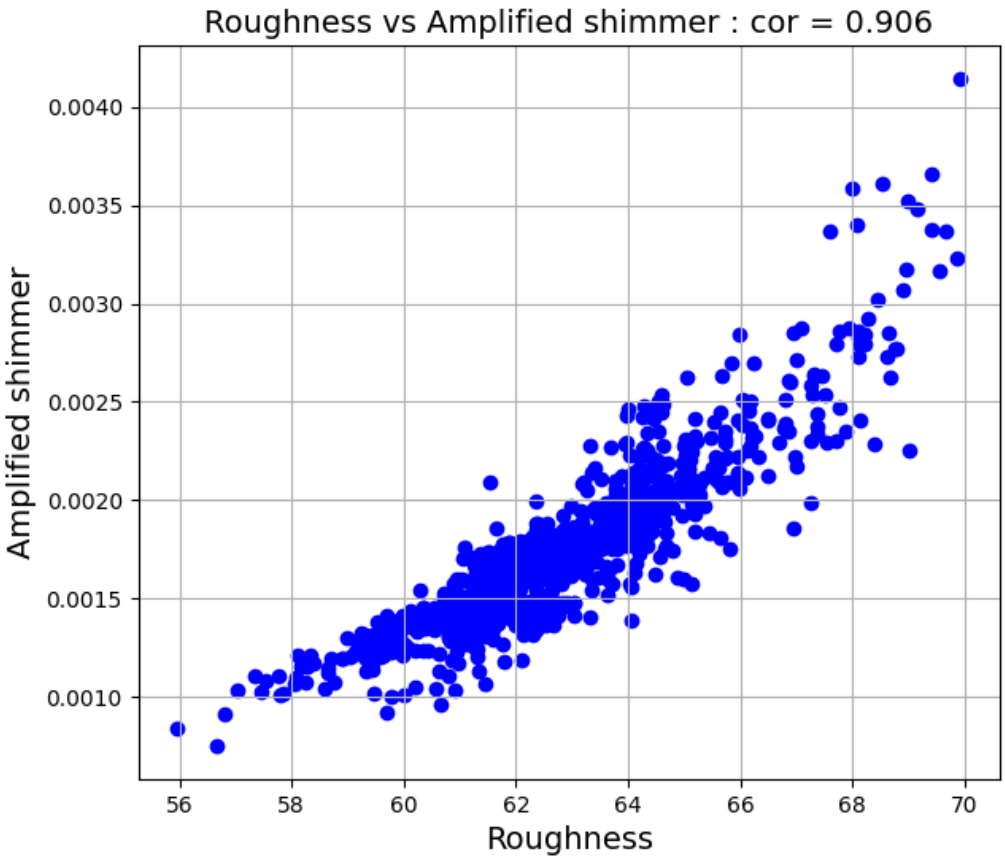


図 4.5: 振幅強調 shimmer（AS）と roughness の相関評価

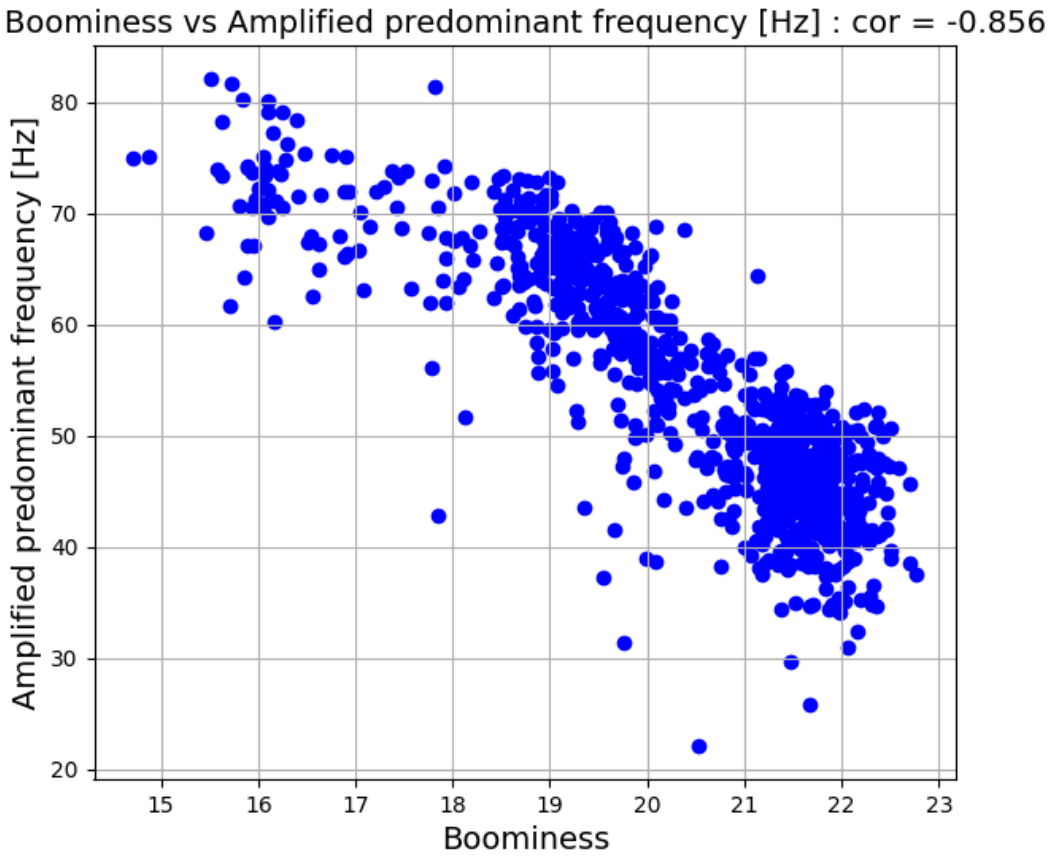


図 4.6: 振幅強調優勢周波数（APF）と boominess の相関評価

4.3 音色関連特徴量の確率統計分析による有効性検証：単次元正規分布仮定

音質評価指標と短時間音響特徴量の組合せで実現した音色関連特徴量を対象に，産業機器動作音データセットの異常検知に対してどの程度の有効性を有するのかを本節（4.3 節）と次節（4.4 節）で検証する．音色関連特徴量が産業機器動作音における異常音を外れ値として適切に分離できる特徴量となっているか否かを統計的に検証する．具体的には音質評価指標 5 種と短時間音響特徴量 2 種の各々について，産業機器の正常音における特徴量の分布が正規分布に従うとの仮定の下で尤度に基づく異常度スコアを定義する．その上で正常音の異常度スコアに対して，異常音の異常度スコアを外れ値と判別できるほどの高い数値として示せるかを考察した．

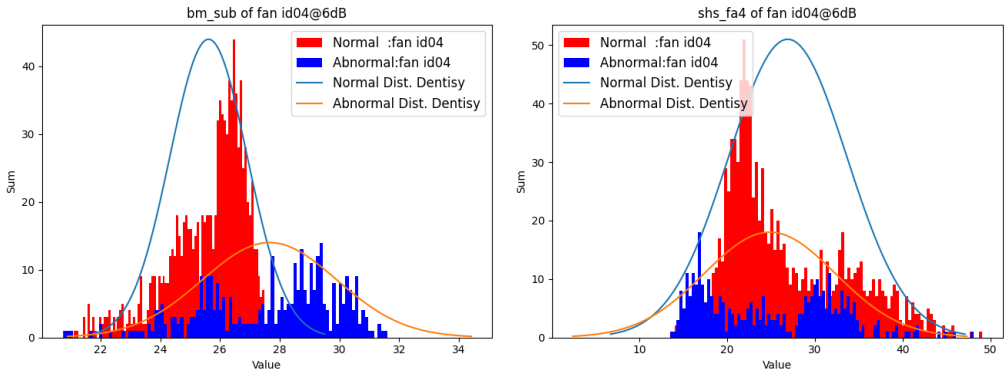
4.3.1 異常度スコアの定義

産業機器動作音データセット収録の 4 機種× 4 個体の 16 種の個々において，正常音の特徴量分布が正規分布に従う条件下で異常度スコアを定義する．

正規分布は式 (4.2) にてモデル化できる．

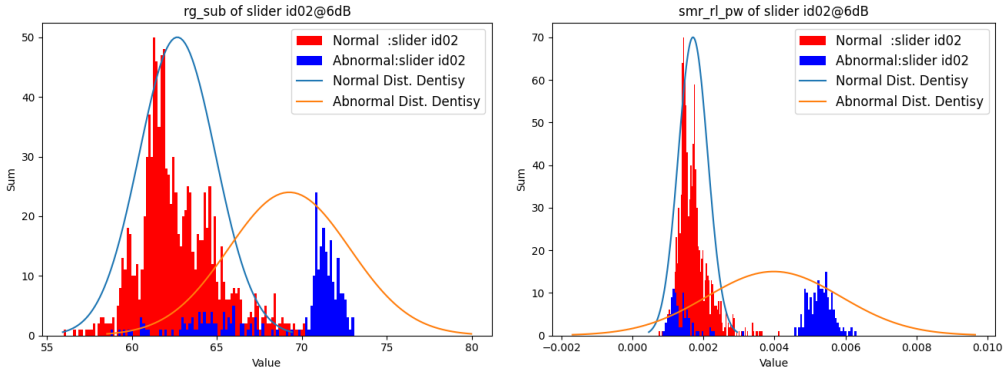
$$\mathscr{N}(x|\mu,\sigma^2) \triangleq \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}\ (x-\mu)^2\right\} \tag{4.2}$$

参考までに fan（id04）と slider（id02）における音質評価指標 1 種と短時間音響特徴量 1 種のヒストグラムを図 4.7 と図 4.8 に示す．赤のバーが正常音の特徴量分布（ヒストグラム）を示し，この分布を正規分布でモデル化したものを青の包絡線で示している.（青のバーとオレンジの包絡線は異常音に対する特徴量分布とその特徴量分布を正規分布でモデル化したものをそれぞれ表す）



(a) boominess の分布比較　　**(b)** APF の分布比較

図 4.7: 正常音／異常音の特徴量数値分布の比較 [fan id04]



(a) roughness の分布比較　　**(b)** AS の分布比較

図 4.8: 正常音／異常音の特徴量数値分布の比較 [slider id02]

正規分布を仮定して推定した正常音の分布に対して，観測されたデータがどの程度稀であるかを尤度を使って定量化する．まず正常音分布における入力データ x の尤度を $p(x|\mathrm{Normal})$ とする．ここで，“Normal” は正常音から抽出された音色関連特徴量の確率変数であるとの意である．観測データ x' に対して求めた尤度の負の対数値を，単次元正規分布を仮定した異常度スコア $A_{\mathrm{score_mono}}(x')$ として式 (4.3) で定義する [110].

$$A_{\mathrm{score_mono}}(x') = -\ln p(x'|\mathrm{Normal}) \tag{4.3}$$

ここでの異常度スコアは正常音の分布に対して，観測されたデータが確率統計上どの程度稀であるかを示す数値（スコア）として算出され，稀であればあるほど数値が大きく評価される．この異常度スコアは式 (4.2) のパラメータを使って算出することができる．本式のパラメータである平均ベクトル μ と分散 σ^2 は最尤推定法により，学習用の正常音データ系列 $x(n)$ から式 (4.4)，式 (4.5) により求められる．

$$\hat{\mu} = \frac{1}{N}\sum_{n=1}^{N} x(n) \tag{4.4}$$

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{n=1}^{N} (x(n) - \hat{\mu})^2 \tag{4.5}$$

ここで，N は正常音データ系列のサンプル数である．

この推定パラメータ値を用いて，観測データ x' に対する異常度スコアは式 (4.6) にて算出される．

$$A_{\mathrm{score_mono}}(x') \triangleq \frac{1}{\hat{\sigma}^2}(x' - \hat{\mu})^2 = \left(\frac{x' - \hat{\mu}}{\hat{\sigma}}\right)^2 \tag{4.6}$$

4.3.2 異常度スコアの分布分離度評価

式 (4.6) で求めた異常度スコア $A_{\mathrm{score_mono}}(x')$ を用いて，音質評価指標 5 種と短時間音響特徴量 2 種の有効性を検証した．正常音から算出された異常度スコアに対して異常音から算出された異常度スコアの分布が互いに十分に離れているかど

うかを評価することで，異常音を外れ値として的確に判別するための特徴量としての貢献度が数値化できると考えた．

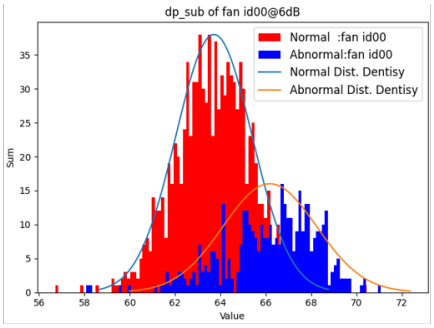
具体的には産業機器動作音データセットの 16 種（4 機種× 4 個体）各々において，正常音データを学習用と評価用に分け，学習用正常音データ（以下，学習データ）の音色関連特徴量で分布を推定（実際は推定パラメータを式 (4.4)，式 (4.5) にて算出）し，推定パラメータに基づいて評価用正常音データの異常度スコア（以下，正常音スコア）と異常音データの異常度スコア（以下，異常音スコア）を算出する．

正常音スコアと異常音スコアの分布分離度は，“positive” or “negative” としての分類を想定した ROC-AUC（area under the receiver operating curve. 以下，AUC）にて評価した．正常音スコアと異常音スコアの分布が十分分離されていれば，AUC は 1.0 に近づく．

図 4.9 に分布分離性の検証例を示す．図中の青で示す異常音の異常度スコアの分布が 0 から離れた数値大の領域に分布すればするほど分離性が高い．この分離性を ROC-AUC（AUC）値で数値化し 1.0 に近いほど貢献度の高いと評価する．

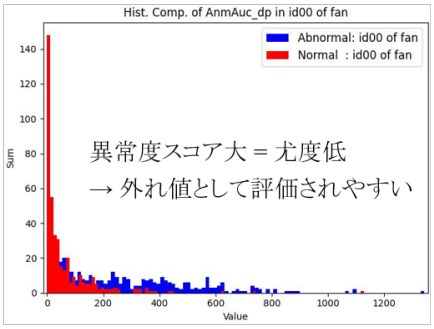
この仕様で音質評価指標 5 種と短時間音響特徴量 2 種の各々について，異常度スコアの AUC 値を比較した (表 4.1)．

また各機種ごとの平均値を表 4.2 に示す．本表において機種ごとに選択した音質評価指標値を黄色の網掛けで示したが平均値で 0.647 から 0.882 までの高い数値を示しており，異常度分類への貢献度からもほぼ適切な選択であったことが確認できた．一方，valve における音質評価指標のうち brihtness や depth で低い数値を示す個体も見受けられる．これらについては，対象個体における数値を対象指標の観点から異常音の聴感上の差異の傾向と突合せていくことで解析が可能となる．



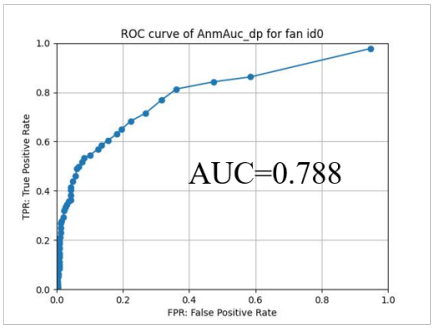
depthの指標分布 [fan id00]
[正常音の指標:赤, 異常音の指標:青]

異常度スコア
(正常音分布に
対する負の対数尤度)



depthの異常度スコア分布[fan id00]
[正常音スコア(赤), 異常音スコア(青)]
※正常音スコアが0近傍に, 異常音スコアが
数値大の領域に分布することが貢献大.

異常度スコアの分布差を
ROC-AUC値で数値化



ROC-AUC(AUC)値
※AUC値が1.0に近いほど, 分類貢献度大と評価.

図 4.9: 異常度スコアにおける分布分離性の検証例 1

表 4.1: 異常度スコアの AUC 比較：単次元正規分布仮定

Machine type	id	AUC						
		boominess	brightness	depth	roughness	sharpness	APF	AS
fan	00	0.680	0.635	0.788	0.668	0.531	0.772	0.533
	02	0.665	0.967	0.718	0.963	0.799	0.674	0.894
	04	0.784	0.798	0.756	0.827	0.615	0.615	0.683
	06	0.686	0.898	0.749	0.774	0.759	0.922	0.954
pump	00	0.817	0.810	0.687	0.833	0.787	0.775	0.960
	02	0.718	0.866	0.764	0.719	0.886	0.763	0.747
	04	0.585	0.642	0.599	0.655	0.851	0.628	0.670
	06	0.887	0.702	0.874	0.705	0.601	0.804	0.768
slider	00	0.984	0.982	0.971	0.925	0.990	0.674	1.000
	02	0.670	0.777	0.836	0.902	0.784	0.730	0.915
	04	0.719	0.856	0.750	0.952	0.833	0.708	0.785
	06	0.644	0.655	0.634	0.750	0.599	0.609	0.637
valve	00	0.882	0.599	0.775	0.856	0.810	0.601	0.988
	02	0.633	0.621	0.650	0.676	0.623	0.588	0.683
	04	0.653	0.632	0.579	0.734	0.643	0.829	0.789
	06	0.627	0.738	0.619	0.657	0.757	0.585	0.665
Average on each machine type								
fan		0.704	0.825	0.753	0.808	0.676	0.746	0.766
pump		0.752	0.755	0.731	0.728	0.781	0.742	0.786
slider		0.754	0.817	0.798	0.882	0.801	0.680	0.834
valve		0.699	0.647	0.656	0.731	0.708	0.651	0.781
Average in total		0.727	0.761	0.734	0.787	0.742	0.705	0.792

表 4.2: 異常度スコアの AUC 比較（機種平均値）：単次元正規分布仮定

Machine type	AUC						
	boominess	brightness	depth	roughness	sharpness	APF	AS
Average on each machine type							
fan	0.704	0.825	0.753	0.808	0.676	0.746	0.766
pump	0.752	0.755	0.731	0.728	0.781	0.742	0.786
slider	0.754	0.817	0.798	0.882	0.801	0.680	0.834
valve	0.699	0.647	0.656	0.731	0.708	0.651	0.781
Average in total	0.727	0.761	0.734	0.787	0.742	0.705	0.792

※黄色の網掛けは機種ごとに選択した音質評価指標を示す.

4.4 音色関連特徴量の確率統計分析による有効性検証：多次元正規分布仮定

本節では音質評価指標と短時間音響特徴量の組合せによる音色関連特徴量を多次元特徴量として捉えて，4.3 節の考えを多次元に拡張して検証する．具体的には前節同様，産業機器の正常音における各特徴量数値の分布が正規分布に従うとの仮定の下で，正規分布を多次元に拡張した上で対数尤度から異常度スコアを定義して，異常音における異常度スコアが外れ値と判別できる十分な数値として表せるかどうかで考察を行った．

4.4.1 異常度スコアの定義

4.3 節と同様に産業機器動作音データセット収録の 4 機種× 4 個体の 16 種の個々について，正常音の特徴量分布が多次元正規分布に従う条件下で異常度スコアを定義する．多次元の正規分布は式 (4.7) にてモデル化できる．

$$\mathscr{N}(\boldsymbol{x}|\boldsymbol{\mu},\Sigma) \triangleq \frac{|\Sigma|^{1/2}}{(2\pi)^{M/2}} \exp\left\{-\frac{1}{2}\,(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\} \tag{4.7}$$

本式のパラメータである平均ベクトル $\boldsymbol{\mu}$ と共分散行列 Σ は最尤推定法により，学習用の正常音ベクトル系列 $\boldsymbol{x}(n)$ から式 (4.8)，式 (4.9) により求められる．

$$\hat{\boldsymbol{\mu}} = \frac{1}{N}\sum_{n=1}^{N}\boldsymbol{x}(n) \tag{4.8}$$

$$\hat{\Sigma} \triangleq \frac{1}{N}\sum_{n=1}^{N}(\boldsymbol{x}(n)-\hat{\boldsymbol{\mu}})(\boldsymbol{x}(n)-\hat{\boldsymbol{\mu}})^{\mathrm{T}} \tag{4.9}$$

前節同様，この分布に対する負の対数尤度 $-\ln\mathscr{N}(\boldsymbol{x}'|\hat{\boldsymbol{\mu}},\hat{\Sigma})$ を異常度スコアとして式 (4.10) で定義する．

$$A_{\mathrm{score_multi}}(\boldsymbol{x}') \triangleq (\boldsymbol{x}'-\hat{\boldsymbol{\mu}})^{\mathrm{T}}\hat{\Sigma}^{-1}(\boldsymbol{x}'-\hat{\boldsymbol{\mu}}) \tag{4.10}$$

この異常度スコアは入力音 (=観測音) から求めた音色関連特徴量の多次元ベクトル $\boldsymbol{x}'$ が，正規分布でモデル化した正常音の音色関連特徴量分布に対して，どの

程度の尤度を有するかを数値化したもので，稀であればあるほど大きな値を取ることになる．

4.4.2 異常度スコアの分布分離度評価

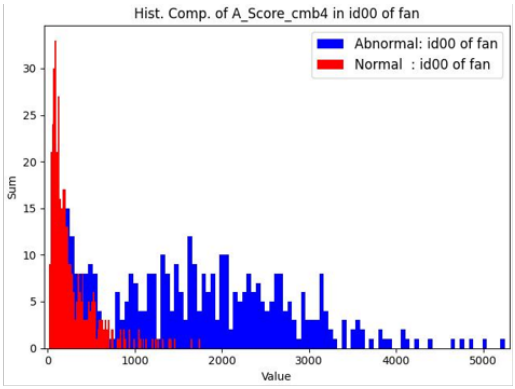
式 (4.10) で求めた異常度スコア $A_{\mathrm{score_multi}}(\boldsymbol{x}')$ は，入力データ $\boldsymbol{x}'$ が標本平均 $\hat{\boldsymbol{\mu}}$ からどの程度離れているかを表しており，マハラノビス距離（の二乗）と対応している [110]．このことは多次元特徴量空間における観測データの確率的な異常度が距離尺度として評価できることを意味する．

この利点を活用し，多次元正規分布を仮定した各機種の正常音の音色関連特徴量分布に対して，正常音データの異常度スコアと異常音の異常度スコアのそれぞれの数値範囲（＝分布）が互いに十分に離れているかどうかを評価する．この分布分離度が高ければ高いほど異常性を判断しやすくなる特徴量であると言え，特徴量としての有効性が高いと考えた．

検証にあたっては 4.3.2 項と同様に，まず産業機器動作音データセットの 16 種（4 機種× 4 個体）各々において，正常音データを学習用と評価用に分け，学習用正常音データ（以下，学習データ）の音色関連特徴量で分布を推定（実際は推定パラメータを式 (4.8)，式 (4.9) にて算出）し，推定パラメータに基づいて評価用正常音データの異常度スコア（以下，正常音スコア）と異常音データの異常度スコア（以下，異常音スコア）を算出する．

正常音スコアと異常音スコアの分布分離度は，同じく “positive” or “negative” としての分類を想定した ROC-AUC（以下，AUC）にて評価した．正常音スコアと異常音スコアの分布が十分に分離されていれば，AUC は 1.0 に近づく．

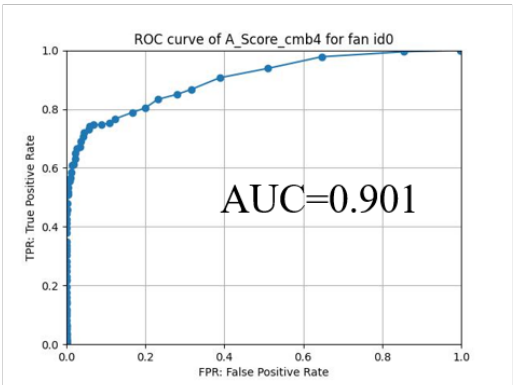
図 4.10 に 7 次元の音色特徴量を対象とした分布分離性の検証例を示す．前節同様に図中の青で示す異常音の異常度スコアの分布が 0 から離れた数値大の領域に分布すればするほど分離性が高く，この分離性から AUC 値が 1.0 に近いかどうかで貢献度の高さが評価できる．図 4.9 との比較により青で示された異常音の異常度スコア分布が 0 から離れた数値大の範囲に延びていることがわかる．これに応じて AUC 値も 0.901 と高い値を示しており，このことから特徴量としての貢献度が向上していることが確認できる．



音色関連特徴量の異常度スコア分布[fan id00]

[正常音スコア(赤), 異常音スコア(青)]

異常度スコアの分布差を
ROC-AUC値で数値化



ROC-AUC(AUC)値

※AUC値が1.0に近いほど, 分類貢献度大と評価.

図 4.10: 異常度スコアにおける分布分離性の検証例 2

この異常度スコアの AUC 値を音質評価指標のみ（SQM），SQM と振幅強調優勢周波数（APF）の組合せ（SQM+APF），SQM と振幅強調 shimmer（AS）の組合せ（SQM+AS），そして SQM と APF と AS の組合せ（SQM+APF+AS），APF と AS の組合せ（APF+AS）の 5 パターンで比較した（表 4.3）.

本結果から，全 16 種平均値においては音色関連特徴量とした SQM+APF+AS が AUC 値で 0.856 と最も分離貢献度が高く，最も有効性が高いことが示された.

また，表 4.3 の数値から下記の点も抽出できた.

機種平均の傾向

- fan と pump は SQM+APF+AS の組合せが最も数値が高いことを確認した. SQM のみでも 0.864 以上の高い数値を示しており，短時間音響特徴量の追加で数値が上がっていることから両者がバランス良く貢献していると考えられる.
- slider は SQM+AS の組合せが最も高い数値を示している．摺動音の差異が有効に捉えられていると考えられる.
- valve は SQM+APF+AS の組合せが最も数値が高いものの 0.756 と他機種に比べて数値が低い．弁の打付け（アタック）音が聴知覚上の差異の着目点となるため，分析長の短縮など時間分解能のさらなる検討が必要と考える.

機種や個体別の傾向

- fan は一部個体（id00）で APF 追加による向上（0.087）が見られ，他個体でも僅かではあるが短時間音響特徴量追加による向上で SQM+APF+AS の組合せが最高値を示している.
- pump は全般的に短時間音響特徴量追加で僅かではあるが数値が向上しており，SQM+APF+AS の組合せが最高値を示している.
- slider は短時間音響特徴量の追加で数値が向上する個体（id00，id02，id06）と追加では変わらない個体（id04）が混在している．特に id06 は数値が 0.6 台と低い傾向にある.

表 4.3: 異常度スコアの AUC 比較：多次元正規分布仮定

Machine type	ID	AUC				
		SQM	SQM+APF	SQM+AS	SQM+APF+AS	APF+AS
fan	00	0.814	0.901	0.812	0.901	0.788
	02	0.959	0.959	0.960	0.960	0.914
	04	0.843	0.843	0.846	0.845	0.743
	06	0.958	0.959	0.969	0.969	0.963
pump	00	0.888	0.885	0.898	0.898	0.929
	02	0.934	0.934	0.933	0.936	0.775
	04	0.822	0.822	0.822	0.822	0.665
	06	0.813	0.812	0.814	0.819	0.756
slider	00	0.985	0.984	0.999	0.999	1.000
	02	0.860	0.831	0.891	0.858	0.803
	04	0.979	0.978	0.977	0.978	0.751
	06	0.669	0.681	0.669	0.680	0.590
valve	00	0.969	0.960	0.975	0.966	0.854
	02	0.645	0.636	0.649	0.640	0.659
	04	0.748	0.868	0.751	0.870	0.851
	06	0.573	0.550	0.575	0.548	0.558
Average on each machine type						
fan		0.894	0.916	0.897	0.919	0.852
pump		0.864	0.863	0.867	0.869	0.781
slider		0.873	0.868	0.884	0.879	0.786
valve		0.734	0.754	0.738	0.756	0.731
Average in total		0.841	0.850	0.846	0.856	0.787

- valve も短時間音響特徴量の追加で数値が向上する個体（id00，id04）と追加では変わらない個体（id00，id06）が混在している．特に id06 は数値が 0.5 台と極めて低い傾向にある．

本章では，異常音の聴知覚上の差異に関連づけた音質評価指標 5 種と短時間音響特徴量 2 種の組合せによる音色関連特徴量に関して，産業機器動作音データセットを用いた有効性検証を行った．機器異常音と聴知覚との対応を産業機器 4 機種に各々についてスペクトラムを用いて考察し，短時間音響特徴量の音質評価指標への高い関連性を相関係数値にて確認した．その上で音色関連特徴量の異常音検知に対する貢献度を多次元正規分布を仮定した異常度スコアにより数値化し，機種によりばらつきがあるものの AUC の全機種平均値が 0.856 と高い数値をもって音色関連特徴量（SQM+APF+AS）の有効性を示した．

第 5章

音色関連特徴量に基づく異常音検知システム：Timbre-feature-based ASD

本章では，音色関連特徴量に基づく異常音検知手法の実現性を検証するため，サポートベクターマシンによる学習および分類部を組合せた異常音検知システムを構成し，産業機器動作音データセットを用いた分類性能の評価を行う．

5.1 基本構成

図 5.1 に音色関連特徴量に基づく異常音検知システムのブロックダイヤグラムを示す．本システムは音質評価指標算出部，短時間音響特徴量算出部，音質評価指標と短時間音響特徴量の組合せによる音色関連特徴量生成部，学習および分類部で構成される．音質評価指標算出部では，機器動作音の入力信号に対して 5 種の指標（boominess，brightness，depth，roughness，sharpness）を算出する．短時間特徴量算出部は，機器動作音の入力信号に対して波形振幅の変動を捉える amplified shimmer（AS）と音の高さの変動を捉える amplified pre-dominant frequency（APF）の 2 種の特徴量を算出する．その後，入力音の機種ごとに適切な音質評価指標を選択し，短時間特徴量と組合せて音色関連特徴量が生成され，本特徴量を入力として学習および分類が行われる．

5.2 特徴量算出部

音質評価指標算出部は 2.1.2 項で紹介した英国 Surry 大が開発の Timbral models を利用して，5 種の指標（boominess，brightness，depth，roughness，sharpness）を算出する．分析フレーム長は 3.5.1 項における検討結果から 1,024 ms とした．時間変化への対応も考慮し，分析フレーム間のシフト時間をフレーム長の半分の 512 ms としている．産業機器動作音データセットの収録ファイルは 16 kHz サンプリングの 10 s 長なので，1 ファイルあたり 18 フレーム分の指標値が算出されるが，本論文ではファイルあたりの平均値を求めて利用した．

短時間音響特徴量の APF と AS の算出部は，それぞれ 3.3.2 項と 3.3.1 項に記載の仕様で算出する．分析フレーム長は上記同様 3.5.1 項における検討結果から，APF を 128 ms，AS を 256 ms とした．上記同様分析フレーム間のシフト量は分析フレー

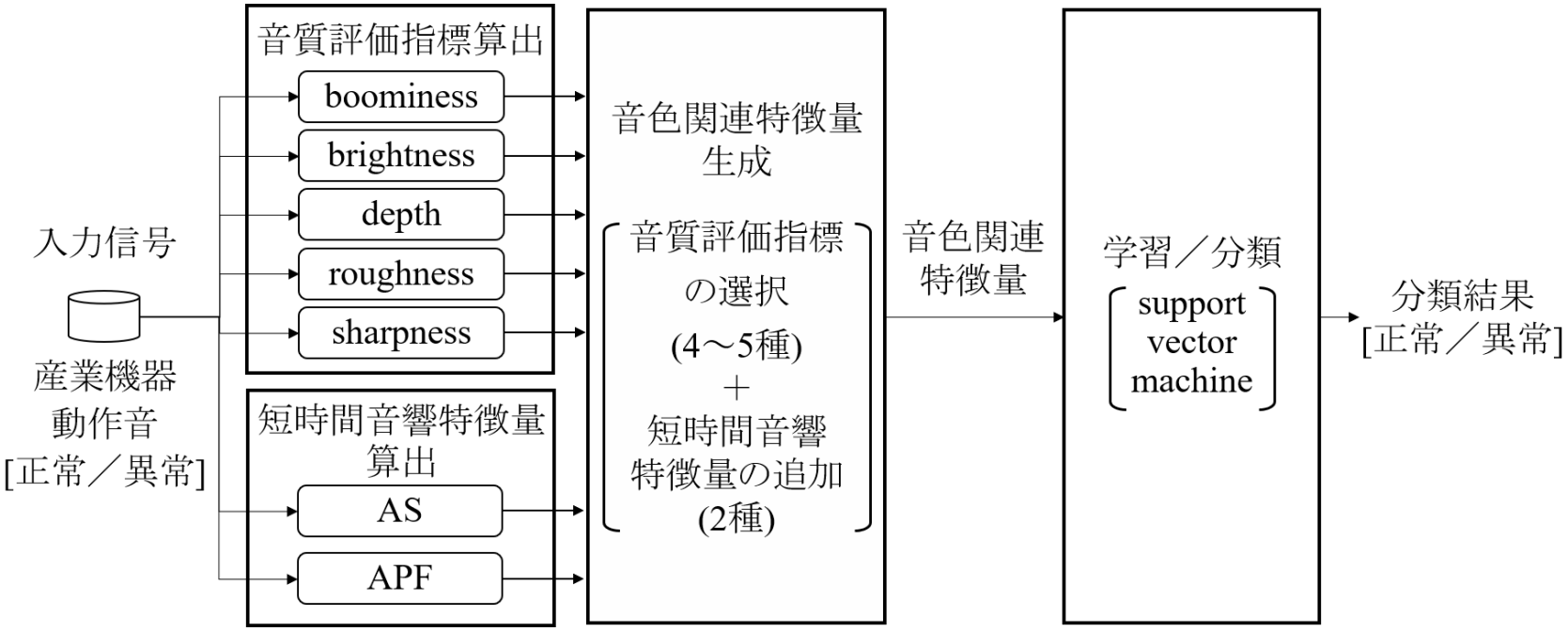


図 5.1: TF-ASD のブロックダイヤグラム

ム長の半分として，それぞれ 64 ms と 128 ms とした．この 2 つの算出特徴量についてもファイル単位で平均値を求めて利用した．

5.3 音色関連特徴量生成部

特徴量算出部で算出された 5 種の音質評価指標値と 2 種の短時間音響特徴量を対象となる機種ごとに組合せて，音色関連特徴量を生成する．機種ごとの音質評価指標の選択と短時間音響特徴量との組合せは表 3.3 に記載の仕様に沿って行った．

5.4 学習および分類部

学習および分類部には，サポートベクターマシン（SVM）の二値分類を適用した（3.5.2 項参照）．産業機械動作音データセットの 4 機種 4 個体（合計 16 種）のそれぞれにおいて，正常音ファイルと異常音ファイルの各 7 割を学習用データとして学習を行った．学習は付与されているラベルを基に異常音データを “true”，正常音データを “false ’’ として実行し，16 種の学習モデルを生成している．評価には学習用データを除いた残りの各 3 割を学習外評価用データとして用いて，各機種各個体ごとに SVM 分類性能を評価した．7 割の学習用データと 3 割の評価用データはランダムに分割し，5 回の分割パターンによる評価の平均値を採用している．

システムは python を使って実装しており，SVM については機械学習ライブラリの scikit-learn を利用した [111]．ハイパーパラメータはカーネルに rbf カーネルを指定した以外はデフォルトで適用した．

5.5 分類性能の評価尺度

分類性能の評価尺度としては正解率（accuracy）, F 値（F-measure）, マシューズ相関係数（matthews correlation coefficient: MCC）の 3 種を用いて，SVM 分類性能を評価した．

正解率（accuracy）は，全てのデータインスタンスに対して正しく分類されたインスタンスの割合を示す尺度で，式 (5.1) で算出される．

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{5.1}$$

ここで，TP は真陽性（true positive），TN は真陰性（true negative），FP は偽陽性（false positive），FN は偽陰性（false negative）を表す．数値範囲は 0.0〜1.0 で数値が 1.0 に近いほど良い．

F 値は適合率（precision）と再現率（recall）の調和平均による誤差評価尺度で式 (5.2) で算出される．

$$\text{F-measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{5.2}$$

ここで適合率（precision）は陽性クラスと予測したインスタンスの中からどのくらい正しく陽性と分類できたかの比率を評価する尺度で，式 (5.3) で算出される．

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{5.3}$$

再現率（recall）は陽性と分類したインスタンスの中でどれくらい陽性と予測されていたかの比率を評価する尺度で，式 (5.4) で算出される．

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{5.4}$$

F 値も数値範囲は 0.0〜1.0 で数値が 1.0 に近いほど良く，正常と異常のサンプル数が非対称の関係にあるデータの評価に有効である．

マシューズ相関係数（MCC）は式 (5.5) により計算される．

$$\text{MCC} = \frac{\text{TP} \times \text{TN} - \text{FP} \times \text{FN}}{\sqrt{(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})}} \tag{5.5}$$

マシューズ相関係数も非対称データに有効で 1.0 に近いほど良い尺度だが，数値範囲が −1.0〜1.0 と負の相関を持つ上に分母が 0.0 を取る場合に例外処理が必要となるなど少々扱い難い性質を持つ．

5.6 評価結果

4 機種の各 4 個体で 16 種類の産業機器個体別に SVM による分類性能を評価した．産業機器動作音データセットの SNR 条件が 6dB のデータセットを入力とした．音色関連特徴量として，音質評価指標のみ（SQM），SQM に APF を加えた組合せ（SQM+APF），SQM に AS を加えた組合せ（SQM+AS），SQM に APF と AS を加えた組合せ（SQM+APF+AS）の 4 種類の組合せごとに 5.5 節で説明した正解率（accuracy），F 値（F-measure），マシューズ相関係数（MCC）の 3 種の尺度で評価した．

表 5.1 に示す通り，異常音の聴覚上差異に対応づけた音色関連特徴量に基づく異常音検知システムが F 値で 0.920 と優れた分類性能を示すことが確認できた．また，F 値のみならず正解率やマシューズ相関係数においても，音色関連特徴量の数値が最大を示し，その上で音質評価指標（SQM）のみに対して短時間音響特徴量（APF，AS）の追加が分類性能の向上に貢献していることが確認できた．

本結果はこの手法が 2.3.2 項で取り上げた最新の競合手法 [98] に対して，機種や異常要因を限定しない幅広い機器群に汎用的に適用できるだけでなく，その異常音分類性能が正解率（accuracy）において競合手法の 0.984（低ノイズ条件の両耳平均値）と同等の 0.984 と優れた性能を示す優位な技術であることを示している．

本章では音色関連特徴量に基づく異常音検知手法の実現性を検証するため，サポートベクターマシンによる学習および分類部を組合せた異常音検知システムを構成し，産業機器動作音データセットを用いた分類性能の評価を行った．その結果，異常音の聴知覚上差異に対応づけた音色関連特徴量に基づく異常音検知システムが F 値で 0.920 と優れた分類性能を示すことが実証できた．構成特徴量を組合せた比較評価も行ったところ，音質評価指標のみ（SQM）と比べて短時間音響特徴量の APF と AS の追加により分類性能が向上し，音質評価指標に短時間音響特徴量を組合せることの有効性も確認できた．これにより，提案する音色関連特徴量が人の聴知覚上の差異に対応していることでシステムの分類結果の解析や性能向上が行いやすい利点を持ちながら，異常音検知向け特徴量としても優れていることが実証できた．

表 5.1: SVM 分類性能評価結果

Machine type	id	Accuracy				F-measure				MCC			
		SQM	SQM +APF	SQM +AS	SQM +APF +AS	SQM	SQM +APF	SQM +AS	SQM +APF +AS	SQM	SQM +APF	SQM +AS	SQM +APF +AS
fan	00	0.926	0.953	0.932	0.977	0.862	0.913	0.872	0.958	0.792	0.873	0.827	0.932
	02	0.994	0.994	0.995	0.994	0.989	0.988	0.991	0.988	0.980	0.985	0.986	0.990
	04	0.975	0.979	0.996	0.997	0.949	0.958	0.991	0.993	0.927	0.928	0.986	0.981
	06	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
pump	00	0.980	0.981	0.988	0.988	0.908	0.916	0.950	0.950	0.901	0.901	0.946	0.932
	02	0.993	0.993	0.993	0.991	0.965	0.964	0.965	0.954	0.947	0.957	0.930	0.915
	04	0.996	0.996	0.998	0.996	0.983	0.983	0.992	0.983	0.981	0.981	0.990	0.981
	06	0.982	0.991	0.986	0.991	0.901	0.948	0.918	0.949	0.878	0.933	0.920	0.923
slider	00	0.998	0.996	0.999	1.000	0.995	0.992	0.998	1.000	0.988	0.988	0.997	1.000
	02	0.979	0.987	0.987	0.987	0.945	0.967	0.966	0.966	0.912	0.947	0.957	0.977
	04	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.994	0.994
	06	0.903	0.924	0.949	0.947	0.517	0.663	0.786	0.781	0.558	0.641	0.808	0.739
valve	00	1.000	0.999	1.000	1.000	1.000	0.993	1.000	1.000	1.000	0.992	1.000	1.000
	02	0.956	0.952	0.988	0.983	0.819	0.800	0.957	0.938	0.745	0.730	0.920	0.917
	04	0.972	0.985	0.978	0.989	0.852	0.927	0.886	0.949	0.799	0.890	0.832	0.914
	06	0.901	0.907	0.904	0.911	0.141	0.277	0.198	0.306	0.079	0.128	0.000	0.128
Average on each machine type													
fan		0.974	0.981	0.981	0.992	0.950	0.965	0.964	0.985	0.925	0.947	0.950	0.976
pump		0.988	0.990	0.991	0.992	0.939	0.953	0.956	0.959	0.927	0.943	0.947	0.938
slider		0.970	0.977	0.984	0.983	0.865	0.905	0.937	0.937	0.865	0.894	0.939	0.927
valve		0.957	0.961	0.968	0.971	0.703	0.749	0.760	0.798	0.656	0.685	0.688	0.740
Average in total		0.972	0.977	0.981	0.984	0.864	0.893	0.904	0.920	0.843	0.867	0.881	0.895

第 6章

統合考察

本章では，共に産業機器動作音データセットを用いた 4 章における音色関連特徴量の検証結果と 5 章における音色関連特徴量に基づく異常音検知システム性能の評価結果を統合した検証を行う．

6.1 音色関連特徴量に基づく異常音検知システム性能評価

本節では音色関連特徴量を基に，図 5.1 に示す構成で実現した異常音検知システムの性能（SVM 分類性能）評価結果を F 値にフォーカスして考察する．

5 章の評価結果（表 5.1）の F 値の平均値部分を表 6.1 に示す．この結果より，音質評価指標（SQM）のみの平均値 0.864 に対して，APF と AS の追加により F 値が 0.920 に向上していることが確認できる．機種別平均値では fan と pump では SQM のみで 0.939 以上と高い分類性能が実現され，slider では APF と AS の追加効果が高いことがわかる．一方，valve は比較的 SQM のみから数値が低く，APF と AS の追加でも 0.798 と頭打ちとなっている．

次に 4 機種（fan，pump，slider，valve）の各 4 個体（id00,02,04,06）の全 16 種個別の F 値性能値（表 6.2）の傾向が下記の 3 つに大別できる．

(1) 音質評価指標のみ（SQM）で F 値 0.95 以上の高い性能を示す個体
　→ 9 個体/16 個体

(2) 短時間音響特徴量（APF，AS）の追加で性能が向上している個体
　→ 5 個体/16 個体

(3) 全ての組合せ（SQM+APF+AS）でも分類性能が低い個体
　→ 2 個体/16 個体

上記 (2) に分類した 5 個体については，追加特徴量と数値変化幅から下記の点が示唆される．

(i) fan id00

結果 0.096 の向上幅．APF 追加の効果大．

分析 回転機器としての摩擦増加による音の高さの変化を捉えた効果が示唆される．

(ii) pump id00

結果 0.042 の向上幅．AS 追加の効果大．

分析 配管の詰まり（ゴボゴボ音）による音の粗さの変化を捉えた効果が示唆される．

(iii) pump id06

結果 0.048 の向上幅．APF 追加の効果大．

分析 流水音の消失による音の高さの変化を捉えた効果が示唆される．

(iv) valve id02

結果 0.119 の向上幅．AS 追加の効果大．

分析 金属弁の劣化による打付け音の濁りにおける音の粗さの変化を捉えた効果が示唆される．

(v) valve id04

結果 0.097 の向上幅．APF 追加の効果大．

分析 金属弁の打付け頻度の減少に伴う打付け音の高さの変化を捉えた効果が示唆される．

さらに上記 (3) で指摘した分類性能が低い 2 個体については，その数値結果から下記の点が示唆される．

(i) slider id06

結果 SQM のみの F 値が 0.517．AS+APF の追加により 0.263 も数値が向上．AS 追加の効果大．

分析 SQM のみの数値が低い理由として人の聴知覚では捉えられない差異か，新たな音質評価指標（音色属性）追加の必要性が示唆される．AS が機能していることから波形変動，例えば変調度などを捉える特徴量の追加による性能向上効果が期待できる．

(ii) valve id06

結果 SQM のみの F 値が 0.141．AS+APF の追加でも 0.165 の数値向上．

分析 SQM のみの数値が低い理由として同じく人の聴知覚では捉えられない差異か，新たな音質評価指標（音色属性）追加の必要性が示唆される．valve は金属弁から発生する打付け音の音色変化をより細かく捉える必要があることから，新たな属性としては hardness（固さ）の追加による性能向上効果が期待できる．

これら推察に対しては観測データの主観評価との突合せなどのフィードバック検証が必要だが，このように対象機器の知識を前提に評価結果の数値分析から効果の推測や問題点の抽出が行える点が音色関連特徴量を用いた異常音検知アプローチの最も有望な利点の一つであると考える．

表 6.1: SVM 分類性能評価結果（F 値平均値）

Machine type	F-measure			
	SQM	SQM +APF	SQM +AS	SQM +APF +AS
Average on each machine type				
fan	0.950	0.965	0.964	0.985
pump	0.939	0.953	0.956	0.959
slider	0.865	0.905	0.937	0.937
valve	0.703	0.749	0.760	0.798
Average in total	0.864	0.893	0.904	0.920

表 6.2: SVM 分類性能評価結果（F 値機種個別）

Machine type	id	F-measure			
		SQM	SQM +APF	SQM +AS	SQM +APF +AS
fan	00	0.862	0.913	0.872	0.958
	02	0.989	0.988	0.991	0.988
	04	0.949	0.958	0.991	0.993
	06	1.000	1.000	1.000	1.000
pump	00	0.908	0.916	0.950	0.950
	02	0.965	0.964	0.965	0.954
	04	0.983	0.983	0.992	0.983
	06	0.901	0.948	0.918	0.949
slider	00	0.995	0.992	0.998	1.000
	02	0.945	0.967	0.966	0.966
	04	1.000	1.000	1.000	1.000
	06	0.517	0.663	0.786	0.781
valve	00	1.000	0.993	1.000	1.000
	02	0.819	0.800	0.957	0.938
	04	0.852	0.927	0.886	0.949
	06	0.141	0.277	0.198	0.306

6.2 音色関連特徴量のSVM分類性能への貢献と異常度スコア分布との関係

本節では6.1節で検証したSVM分類性能結果と異常度スコア分布との関係を考察する．

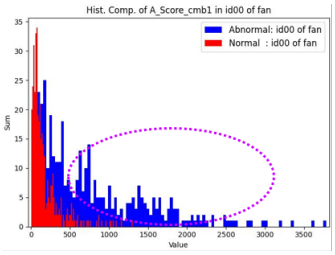
まず，6.1節の(2)で指摘したfan id00における短時間音響特徴量追加の効果に着目し，異常度スコア分布（AUC値）とSVM分類性能（F値）の関係を図示する．図6.1(a)と図6.1(c)を比較すると，短時間音響特徴量の追加に伴い異常音スコア（異常音の異常度スコア）の分布（青色のヒストグラム）が数値大方向（図では右方向）にシフトしていることがわかる．このシフトが特徴量の追加による分布分離度向上の効果であり，図6.1(b)と図6.1(d)の比較でAUC値の増加として表れている．これに伴ってF値も向上しているため，特徴量の分離貢献度（AUC値）の向上がSVM異常音分類性能（F値）の向上に寄与していることが推測できる．

<u>音質評価指標のみ</u>

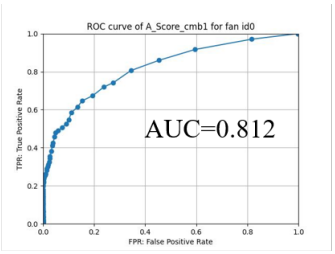
(SQM)

<u>SVM分類性能</u>

F値=0.862



(a) 異常度スコア分布(SQMのみ)
[赤:正常音スコア、青:異常音スコア]



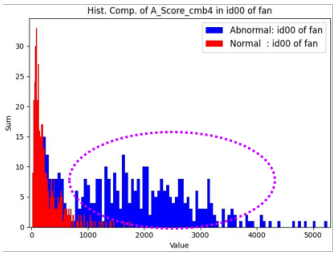
(b) 異常度スコアのROC-AUC
(SQMのみ)

<u>音色関連特徴量</u>

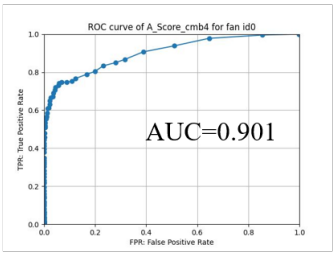
(SQM+APF+AS)

<u>SVM分類性能</u>

F値=0.958



(c) 異常度スコア分布(音色関連特徴量)
[赤:正常音スコア、青:異常音スコア]



(d) 異常度スコアのROC-AUC
(音色関連特徴量)

図 6.1: 特徴量追加に伴う異常度スコア分布と SVM 分類性能値の比較（fan id00）

次に，多次元異常度スコアに単次元異常度スコアを組合せた分布の遷移を図 6.2 に示す．図 6.2(a) の SQM のみの異常度スコア分布に APF のみの異常度スコア分布（図 6.2(b)）が加わることで，図 6.2(c) の SQM+APF では青で示す異常音スコア（異常音の異常度スコア）が数値大の方向にシフトされ，特徴量分布上の分離性能が向上していることが確認できる．さらに AS を加えた SQM+APF+AS（図 6.2(d)）の条件では異常音スコア（青部分）の分布に若干の変化が見れるものの AUC 数値としては変化がなく，AS 追加の影響が限定的であることもわかる．

このように異常度スコアの分布遷移を特徴量の各要素に分解して解析することで，どの音色関連特徴量が異常音検知に寄与しているかが分析できる．この分析に事前知識と組合せることで機器の異常要因の推定が行えたり，点検員の主観評価結果と突合せることで自身の聞き分けポイントの数値化（可視化）が行える可能性があると考える．

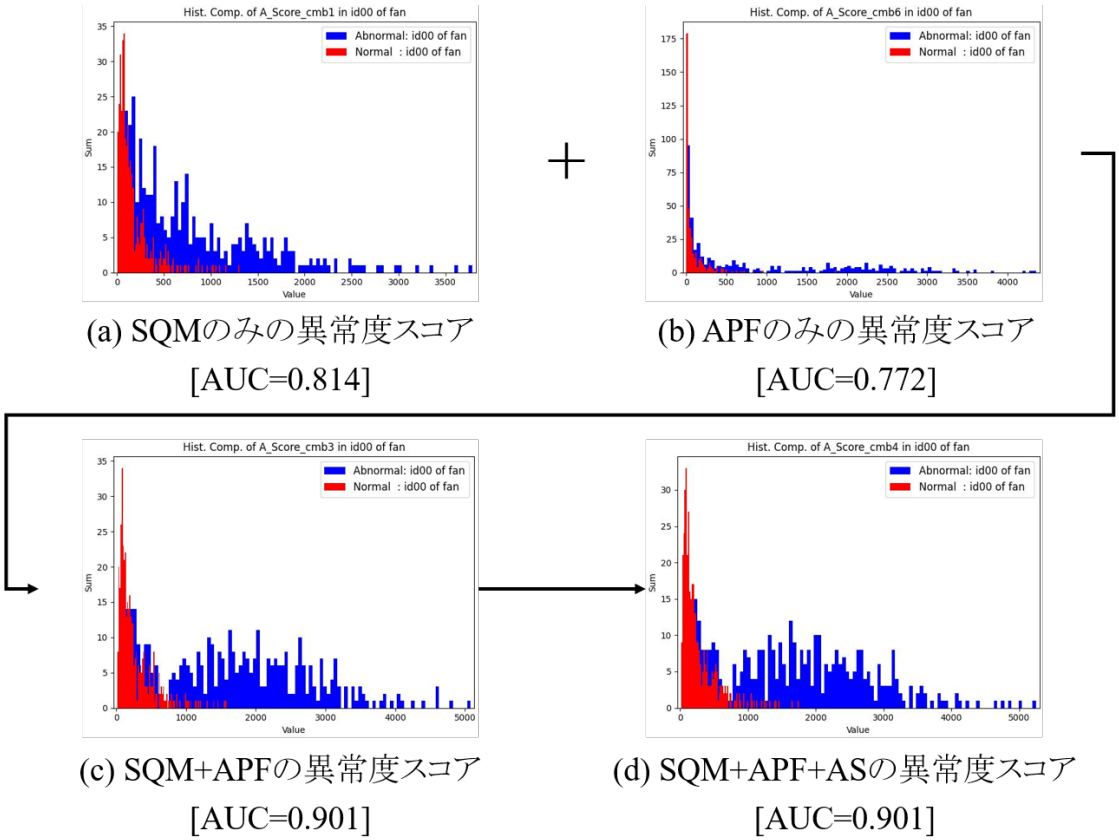


図 6.2: 多次元／単次元異常度スコア分布の遷移解析（fan id00）

6.3 産業機器動作音データセットにおける正常音の正規分布性についての考察

多くの産業機器では通常稼働時に複数の動作仕様で運用を行っている．モータなどの回転系であればその回転数，物品の運搬を行う摺動系であればその動作スピードなどが作業の要求に合わせて制御されている．そのため正常時動作であっても発生する音には動作仕様に応じた種類の状態が生じ，主観的な音色の違いとして知覚される場合があり得る．つまり正常動作音（正常音）であっても音色関連特徴量としては複数の状態が存在することで，特徴量分布の正規分布性が必ずしも保証されない場合も考えられる．そのため，本節では産業機器動作音データセットにおける正常音の正規分布性について考察を加える．

6.3.1 音色関連特徴量分布の正規分布性の調査

産業機器動作音データセットの正常音に対する 5 種の音質評価指標と 2 種の短時間音響特徴量の数値分布の正規分布性を調査した．4 種の機種，各々4 個の個体において shapiro-wilk 法（SW 法） [112] による検定を行った．

結果を表 6.3 に示す．算出値（p 値）が 5％（0.05）以上を示すとき，有意水準 5％の確率で帰無仮説を採択＝正規分布性があるとされる．

表中，0.05（5.0E−02）以上を示す条件をハイライトした．この結果より SW 法による検定からは産業機器動作音データセットの正常音においては，音色関連特徴量の数値分布に厳密な正規分布性が認められないことがわかった．

表 6.3: 正規分布性評価（Shapiro-wilk test）：MIMII データ全体

Machine type	id	Shapiro-wilk test						
		boominess	brightness	depth	roughness	sharpness	APF	AS
fan	00	2.55E-09	4.64E-23	1.09E-01	6.60E-16	2.30E-20	4.24E-04	5.13E-29
	02	6.97E-27	1.10E-37	5.26E-08	8.92E-21	7.20E-32	2.96E-21	3.37E-36
	04	4.02E-22	1.47E-35	8.14E-04	2.75E-18	2.28E-28	4.34E-24	7.38E-37
	06	3.42E-25	1.20E-34	1.52E-05	1.12E-23	1.55E-26	3.99E-16	5.79E-35
pump	00	1.82E-11	2.51E-15	3.20E-07	6.88E-15	3.34E-12	9.20E-19	5.31E-23
	02	5.96E-06	2.17E-19	2.60E-04	1.50E-14	1.62E-20	1.85E-01	2.07E-18
	04	4.85E-13	1.00E-20	6.86E-07	4.35E-23	2.38E-07	2.08E-24	3.35E-07
	06	5.65E-03	1.31E-05	2.17E-06	3.79E-05	1.56E-05	3.22E-24	1.17E-19
slider	00	2.20E-12	4.95E-19	2.66E-20	1.72E-13	2.48E-15	6.18E-35	5.40E-32
	02	2.82E-19	4.93E-22	9.09E-22	6.06E-15	1.84E-17	8.59E-23	5.23E-26
	04	5.78E-03	7.45E-06	6.35E-03	5.02E-11	2.79E-04	5.42E-08	4.94E-03
	06	4.47E-03	2.96E-04	5.48E-02	1.59E-01	4.27E-05	5.66E-22	8.89E-07
valve	00	1.01E-06	5.03E-31	7.11E-26	3.90E-08	3.64E-28	9.13E-32	9.39E-14
	02	1.43E-09	3.55E-24	2.85E-18	4.34E-05	6.87E-23	1.28E-20	1.01E-16
	04	5.19E-05	2.50E-30	9.07E-28	2.12E-14	4.19E-27	1.41E-21	5.97E-31
	06	4.86E-03	1.37E-23	3.58E-11	3.69E-18	1.30E-16	3.75E-18	9.40E-21

※赤色枠は数値が 5.0E−02 以上の条件

6.3.2　正規分布性検定値を向上させたサブセットの抽出

正規分布性の要件の一つとして，産業機器動作音の正常音に含まれる複数の動作仕様ごとの音であれば特徴量分布に正規分布性が仮定できるのではないかと考えた．これは例えば回転系の正常動作音ではその回転数の多寡が音色の違い，すなわち音の高さや鋭さなどの違いとして知覚できることから，同じ回転数から生じる音の音色関連特徴量の分布は正規分布と見なせるのではないかとの仮説に基づく．この仮説にそって，産業機器動作音データセットにおける 4 機種各 4 個体の 16 種のそれぞれにおける特徴量分布から正規分布性が仮定できる部分集合（サブセット）を抽出した．

まず各機種ごとに動作仕様の違いが表れると推測される音質評価指標を表 6.4 の通り選定した．

そして選定した音質評価指標において，SW 法検定値（p 値）の向上で正規分布性が確認できる特徴量分布のサブセットを 16 種全てから抽出した．本作業の一例を図 6.3 に示す．

図 6.3 で例示したサブセット抽出は，点検保守の現場における機器の動作仕様ごとの正常データ収集に対応すると考えている．実運用においては機器操作の動作仕様ごとに動作音の収集および特徴量算出を行うことで，ある程度の正規分布性が仮定できる学習用データが入手できると想定している．

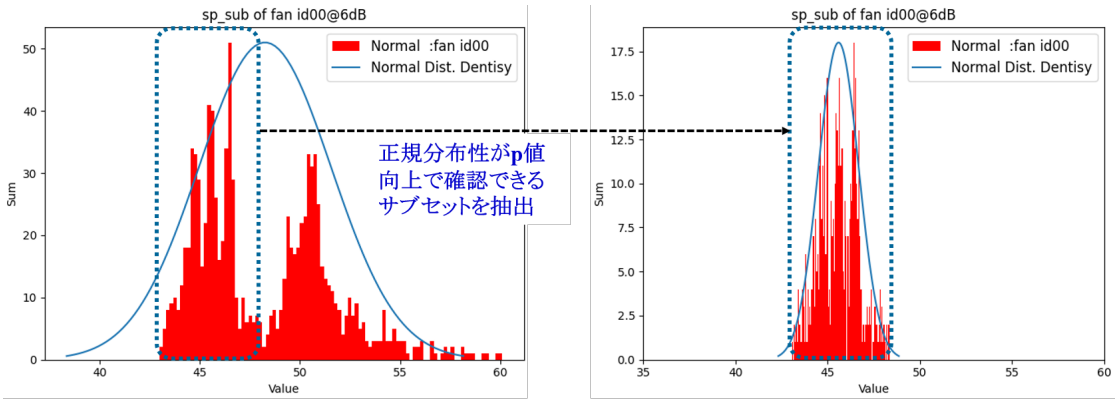
次に各個体ごとに抽出した 16 種のサブセットに対して SW 法による検定を実施し，表 6.5 に整理した．

表 6.5 に示す通り，選択した音質評価指標におけるサブセット抽出により SW 法による検定値（p 値）が大幅に向上していることが確認できる．

さらにこのサブセット抽出により対象とした音質評価指標以外の指標や特徴量の p 値も向上していることがわかった．ただし，ここで示す選択指標以外の特徴量に対する影響については波及効果の詳細な調査および解析が必要であり，今後の課題である．

表 6.4: 動作仕様の違いが表れると推測される音質評価指標

機種	音質評価指標	理由
fan	sharpness	回転数の変化が音の鋭さに表れるとの仮説
pump	brightness	流水音の変化が音の明るさに表れるとの仮説
slider	roughness	往復動作速度の変化が音の粗さに表れるとの仮説
valve	brightness	打付け音の変化が音の明るさに表れるとの仮説



p値=2.30E-20
[データサンプル数 = 1011]

p値=1.59E-03
[データサンプル数 = 539]

図 6.3: サブセット抽出例（fan id00 の sharpness 分布）

表 6.5: 正規分布性評価（Shapiro-wilk test）：正規分布性向上サブセット

Machine type	id	Shapiro-wilk test						
		boominess	brightness	depth	roughness	sharpness	APF	AS
fan	00	1.66E-10	7.99E-10	2.06E-03	5.75E-07	1.59E-03	2.79E-01	6.05E-09
	02	3.12E-24	1.48E-01	3.62E-02	7.38E-05	1.02E-12	8.87E-21	6.35E-03
	04	3.87E-05	1.92E-10	2.01E-01	1.45E-07	2.40E-02	2.60E-11	1.24E-12
	06	3.92E-11	4.57E-04	1.11E-03	8.81E-06	2.59E-01	3.92E-01	2.69E-02
pump	00	9.66E-08	2.55E-01	1.01E-05	1.17E-02	2.78E-04	9.20E-14	1.73E-02
	02	4.92E-04	1.70E-04	5.96E-04	2.45E-17	9.11E-08	2.50E-01	2.86E-13
	04	1.05E-09	3.32E-01	6.43E-03	8.41E-02	3.61E-03	1.24E-16	1.05E-11
	06	3.19E-02	9.82E-03	1.53E-01	5.80E-03	1.14E-03	2.65E-03	4.17E-15
slider	00	6.73E-14	3.32E-07	1.17E-19	1.15E-02	1.16E-09	1.28E-27	1.91E-19
	02	1.11E-20	1.19E-23	5.99E-18	2.38E-04	7.38E-15	9.44E-07	1.36E-23
	04	2.72E-02	2.57E-08	2.57E-02	5.19E-01	3.16E-06	5.92E-01	4.52E-06
	06	6.54E-03	3.17E-04	5.29E-02	6.94E-01	4.43E-05	3.56E-22	5.24E-07
valve	00	2.05E-13	7.84E-03	4.12E-02	3.45E-15	5.09E-03	8.63E-16	1.38E-13
	02	2.21E-07	4.80E-01	2.59E-03	8.04E-16	9.30E-02	1.91E-12	1.26E-09
	04	8.19E-08	1.71E-02	1.93E-14	2.88E-16	4.04E-03	8.23E-17	1.29E-32
	06	2.39E-03	1.67E-02	1.89E-03	1.60E-25	9.37E-04	8.68E-16	1.31E-13

※赤色枠は数値が 5.0E−02 以上の条件

6.3.3 正規分布性向上サブセットに基づく音色関連特徴量の有効性検証

前節までに行ったSW法検定値に基づく正規分布性を向上させたサブセットを対象に，4.3節で行った異常度スコアによる有効性検証を再度実施した．各指標と特徴量ごとの貢献度をROC-AUC（AUC）値で評価した機種ごとの平均値を表6.6に示す．

表6.6に示す通り，黄色の網掛けで表した機種ごとに選択した音質評価指標が上位の数値を示すことが改めて実証された．正常音の特徴量分布の正規分布性が向上した点に加えて，相対的に正常音と異常音のデータ数の差が縮まり非対称性が緩和された点も本来の性能が見えてきた好材料と考えている．

ただし，本結果は産業機器動作音データセットの単一サブセットにおける試行に過ぎない点に留意する必要がある．正規分布を仮定できる単一サブセットでも音質評価指標の有効性が示せた点は想定通りの望ましい結果だが，一方で個々に見ると一部の条件（例えばvalveにおけるboominessのAUC値）で貢献度が十分でない点など想定外の問題点も判明した．

大量の正常音データが入手できることが見込まれる実運用条件を見据えて，特徴量分布の正規分布性の仮定の妥当性を再考し，現産業機器動作音データセット（MIMII）の仕様限界と捉えるか，MIMIIデータセットの特徴量分布に対する新たな定義を模索するかなどをMIMIIデータを題材とした今後の検討課題に対する有益な情報が本試行で得られた．

一方，実運用下においては通常稼働音である正常音の特徴量分布をいかに的確かつ迅速に捉えるかが最も重要な課題となるはずである．これに対して，本6.3節で議論した正規分布性の仮定の下での機器の稼働条件との関連づけが運用における学習データの選別（クレンジング等）や学習状態の検証に役立つと考えられる．

表 6.6: 異常度スコアの AUC 比較（機種平均値）：正規分布性向上サブセット

Machine type	AUC						
	boominess	brightness	depth	roughness	sharpness	APF	AS
Average on each machine type							
fan	0.869	0.936	0.791	1.000	0.999	0.842	0.934
pump	0.801	0.840	0.794	1.000	0.998	0.758	0.850
slider	0.832	0.866	0.833	0.933	0.871	0.612	0.834
valve	0.751	0.942	0.947	0.960	0.814	0.669	0.884
Average in total	0.813	0.896	0.841	0.973	0.921	0.720	0.876

※黄色の網掛けは機種ごとに選択した音質評価指標を示す.

本章では，異常音検知システム性能結果と異常度スコア分布の分析結果と統合した検証を行うことで，SVM 性能向上が異常度スコアの分布分離性と対応づく可能性が示唆された．さらに多次元と単次元異常度スコア分布評価を組合せることで，音色関連特徴量の異常音検知への寄与度が要素ごとに分解できる可用性が示され，異常要因の推定や点検員の聞き分けポイントの数値化に活用できる可能性も示唆された．

前提としていた正常音の正規分布性については評価対象とした産業機器動作音データセット全体では動作仕様の多様性から保証されない可能性を指摘し，Shapiro-wilk 法の p 値で正常分布性を担保したサブセットにて音質評価指標の選定に対する有効性を検証した．本検証は一試行に過ぎず正規分布性仮定の妥当性についての再考は必須であるが，一方で本仮定に基づく機器の稼働条件との対応づけが実運用下での検証に役立つと考えられる．

第 7章

結論

本論文では，本論文で得られた成果を要約し，明らかにしたことと今後の展望について述べる．

7.1 本論文における成果

本研究における成果は下記 3 点となる

1. 産業機器動作音の異常要因に基づいた聴覚上の変化と音色（音質評価指標）との対応づけ.
   産業機器動作音データセットに収録されている 4 機種（送風機，給水ポンプ，直動スライドレール，電磁弁）を対象に機種ごとに典型的な異常要因を調査し，発生音の聴覚上の変化（差異）をオノマトペで表現した．このオノマトペを参考に異常音における聴知覚上の変化（差異）に関連する音色として 5 種の音質評価指標（boominess，brightness, depth, roughness, sharpness）を対応づけた．本対応づけは異常検知特徴量化の妥当性根拠になっただけでなく，音色表現語やオノマトペとの関連づけを整備することで，現場点検員の異常検出観点（ノウハウ）の言語化にも役立てられる．

2. 音質評価指標に基づいた音色関連特徴量の考案
   前項で対応づけた音質評価指標と高時間分解能分析で本指標を補助する短時間音響特徴量の組合せで音色関連特徴量を考案した．この聴覚上の差異に対応づけた特徴量の実現により，シンプルな分類手法との連携で異常音検知の判別ロジックの理解しやすさに繋がる．さらに音色関連特徴量の分析や個別音色要素への分解により，音色観点からの機器状態の推定や産業機器自体の騒音評価や快音化にも活用できる．

3. 音色関連特徴量を適用した異常音検知システムで優れた異常音分類性能を実現
   各機種ごとに適切な組合せを行った音色関連特徴量に対してサポートベクターマシン（SVM）の二値分類にて異常性を検知するシステムを構成し，産業機器動作音データセットによる性能評価にて F 値で 0.920 と優れた分類性能を実現した．本性能により音質評価指標を特徴量とする既存技術と比べ，

機種や障害種を限定せずに同等の分類性能を実現できる優位な方式であることが示せた．本論文では適用現場にて自らの経験・知見に照らした結果分析や性能調整等が行える利用者にやさしい Human-centric な異常音検知技術を具現化した．

これら成果により，下記 2 点の研究課題への具体的な対応が得られた．

課題 1 人の異常音聞き分け観点と機器の異常要因が音色関連特徴量を介して定量的に対応づけられるか?

対応 1 産業機器の異常音の聞き分け観点を異常要因に基づいた聴覚上の差異としてオノマトペ表現し，音色観点で音質評価指標と対応づけることで音色関連特徴量による定量化を行った．産業機器の実動作音データにて本特徴量の有効性も検証した．

課題 2 機器の動作音の変化と関連づけた音色関連特徴量により優れた異常検知を実現できるか?

対応 2 機器の動作音から算出した音色関連特徴量に対して機械学習手法（SVM）の二値分類にて異常音を判別する異常音検知システムを構成し，実動作音データに対して優れた分類性能を実証した．

このように産業機器動作音データセットの活用により，本研究で産業機器の動作音に関して，異常音と音色との関連性を明らかにすることができた．本論文における研究アプローチと成果の関係を図 7.1 に示す．

本論文では産業機器を対象に点検員を想定した人の知覚を音色の観点で音質評価指標に対応づけるアプローチにて人間基点の Human-centric な異常音検知の実現を目指した．一般の産業機器を幅広くカバーする設計思想の下で有効かつ既存技術より優位な汎用技術として具現化できたが，生産現場の音環境への適用性や音色関連特徴量による異常音検知の網羅性については現場データによる評価と検証が今後必要と考える．さらに機械学習モデルの組合せについては SVM 二値分類による一例を提示した初期の段階であるため，auto encoder や convolutional neural network などの機械学習の最新手法との組合せによるポテンシャル評価にも取り組みたいと考えている．

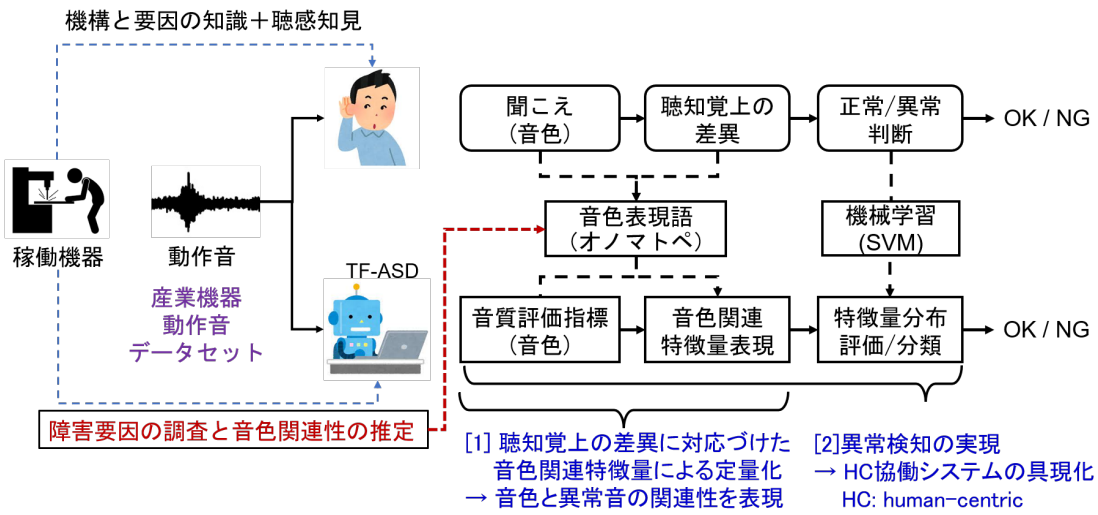


図 7.1: 研究アプローチと本論文における成果

7.2 今後の展望

本論文における成果に基づいた本研究の今後の展望を述べる.（図 7.2）

1. 音色関連特徴量（異常音分類結果／異常度スコア）分析による機器異常要因推定

   本論文成果の異常音検知システム（Timbre-feature based anomalous sound detection：TF-ASD）を工場等の生産現場に導入することで，点検員（人）の異常音診断（聴感判別）結果に音色関連特徴量数値を対応づけた数値データとしての蓄積が期待できる．この蓄積に対して機器の障害要因との対応づけを解析（学習）することで，異常音から抽出した特徴量を通じて対象機器の異常要因が推定できる可能性がある．また 6.2 節で示した通り，本研究アプローチには音色指標ごとに分析が行える利点があるため，機器異常の要因推定の結果を音色指標ごとに分解することで特定の異常要因にフォーカスした TF-ASD のカスタマイズ化も考えられる.

2. 点検員自身の主観評価結果との突合せによる聴感知見の数値化，および言語化

   点検員（人）の聴感判別知見には属人性があり互いに共有できれば集合知としての発展活用が期待できる．実際には各人が有する知見の客観化が課題であるため，機器動作音における各指標の関連性が明確になれば，オノマトペを介した言語化や特徴量を介した数値化が可能となり，聴感判別知見の顕在知化，さらには熟練者知見の共有化や技術継承等に向けた有効な技術として展開できる.

3. 機器動作音に対する聴感知見数値化分析による展開

   日常の保守点検が適切に行われていれば，機器の異常が発生する頻度は極めて低い．この通常稼働時の動作音も音色の観点から評価できる利点を活用して，機器の騒音評価や快音化，サウンドデザインへの展開が考えられる [113]. さらに音質評価指標のモデル改良などへのフィードバックや人が周囲音から状況を判別するためのサウンドスケープや情景分析への応用など幅広い展開も期待できる.

これら新たな価値の顕在化には点検業務に携わる点検員自身の聴感知見の数値化が前提となるが，現状ではSD（semantic differential）法などの手法は確立しているものの主観評価のための音源サンプルのセットアップや評価時間の確保など運用上の難しさがある．これに対しては日常の点検作業における診断結果にオノマトペ表現した聴感印象も入力対象にするなど，現場作業と連携したシステム運用策で効率的なデータ活用サイクルの構築が期待できる．

このように本論文における成果は，様々なものづくり現場において人と協業しながら育っていくHuman-centric IT技術として幅広い可能性を有していると考える．人の聴感のデジタル化（数値化・言語化）によるビジネス貢献を目的に，今後はクラウド連携システム等を想定した本成果の社会実装を目指していきたい．

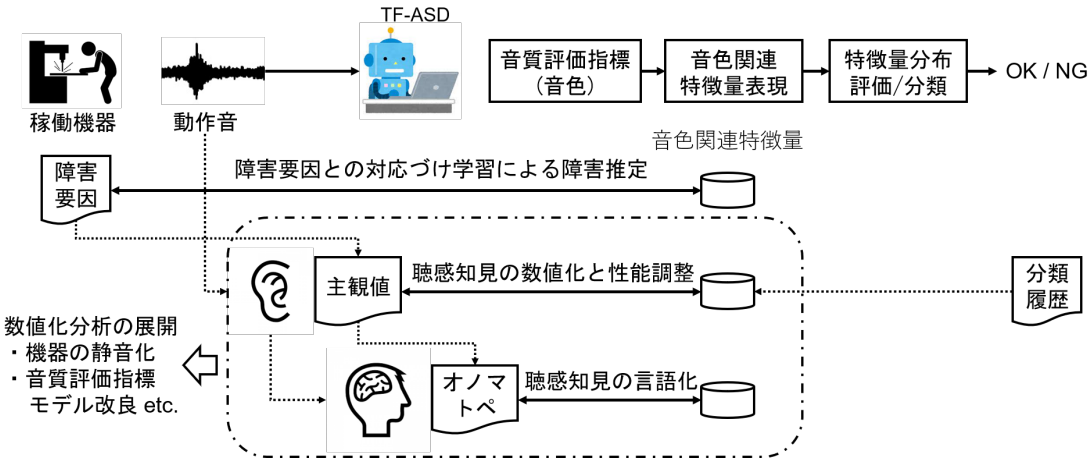

TF-ASD
稼働機器
動作音
音質評価指標
（音色）
音色関連
特徴量表現
特徴量分布
評価/分類
OK / NG
音色関連特徴量
障害
要因
障害要因との対応づけ学習による障害推定
主観値
聴感知見の数値化と性能調整
分類
履歴
数値化分析の展開
・機器の静音化
・音質評価指標
モデル改良 etc.
オノマ
トペ
聴感知見の言語化

図 7.2: 今後の展望

謝辞

本研究を遂行するにあたって，北陸先端科学技術大学院大学 先端科学技術研究科 鵜木祐史教授には，指導教員として終始熱心なご指導を頂きました．心から感謝いたします．特に赤木正人名誉教授とともに本学社会人コース入学時から貴重なご助言をいただき，素晴らしい研究環境に導いていただいたことに対して，深い感謝の意を申し上げます．また、北陸先端科学技術大学院大学 先端科学技術研究科 岡田将吾准教授には主テーマの立ち位置をご理解いただきながら，副テーマとして懇切丁寧なご指導をいただきました．また予備審査の場では，新たな着眼点からのご助言にて研究の発展を導いてくださいました．深く感謝申し上げます．

九州大学 芸術工学研究院 音響設計部門 高田 正幸准教授には音響設計学の専門家として，同志社大学 理工学部 情報システムデザイン学科 井本 桂右准教授には環境音分析と異常音検知の専門家として予備審査の場で大変貴重な示唆に富んだご意見とご助言をいただきました．また北陸先端科学技術大学院大学 先端科学技術研究科 長谷川忍教授，および吉高淳夫准教授のお二方からも同じく予備審査の場で研究の視点や取組みの甘さを指摘いただく貴重なご助言をいただきました．貴重なご意見やご助言により一層見識が深まり，博士論文の厚みが増しました．諸先生方に深く感謝の意を申し上げます。

鵜木研で日々研究を共にさせていただいた木谷俊介講師，上江洲安史特任助教，磯山拓都さんには日頃から貴重なご助言をいただき，研究遂行に対して知識面から意欲面に渡る幅広いご支援いただきました．深く感謝申し上げます．また鵜木研の卒業生である倉誠吾さんには，ご自身の異常音検知研究の資産にて多大なご支援をいただきました．特に提供いただいた計算機資産はシミュレーション環境の基盤として円滑な研究遂行の推進力となりました．深く感謝いたします．

最後に暖かい支援で励まし続けてくれた妻，百香に感謝の意を示します。

参考文献

[1] 経済産業省, 環境共創イニシアチブ, 野村総合研究所. スマート保安事例集. Technical report, 経済産業省, 2020.

[2] 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.

[3] William J Davies, Mags D Adams, Neil S Bruce, Rebecca Cain, Angus Carlyle, Peter Cusack, Deborah A Hall, Ken I Hume, Amy Irwin, Paul Jennings, et al. Perception of soundscapes: An interdisciplinary approach. *Applied acoustics*, Vol. 74, No. 2, pp. 224–231, 2013.

[4] 山西健司, 久野遼平, 島田敬士, 峰松翼, 井手剛. 異常検知からリスク管理へ. サイエンス社, 2022.

[5] 大串健吾. 音色および音質の研究. テレビジョン学会誌, Vol. 34, No. 12, pp. 1079–1087, 1980.

[6] N Tandon and A Parey. Condition monitoring of rotary machines. *Condition monitoring and control for intelligent manufacturing*, pp. 109–136, 2006.

[7] Susana Vaz Freitas, Pedro Melo Pestana, Vítor Almeida, and Aníbal Ferreira. Integrating voice evaluation: correlation between acoustic and audio-perceptual measures. *Journal of Voice*, Vol. 29, No. 3, pp. 390–e1, 2015.

[8] 長町三生. 言葉の響きに関する感性工学. 日本音響学会誌, Vol. 49, No. 9, pp. 638–644, 1993.

[9] 井熊均, 木通秀樹. 大胆予測！ＩｏＴが生み出すモノづくり市場２０２５. 日刊工業新聞, 2018.

[10] 富士電機製品コラム. スマートファクトリーとは何か？Technical report, 富士電機, 2023. https://www.fujielectric.co.jp/products/column/fa/fa_10.html.

[11] スマート保安官民協議会. スマート保安推進のための基本方針. Technical report, 経済産業省, 2020.

[12] Public-Private Sector Council on Smart Industrial Safety High Pressure Gas Safety Committee, Ministry of Economy, Trade and Industry Japan. *"High Pressure Gas Safety: Smart Industrial Safety Action Plan,"*, 2020. [ONLINE] Accessed on: Mar. 16, 2023. Available:https://www.meti.go.jp/english/press/2020/pdf/0710_007a.pdf.

[13] Shikha Agrawal and Jitendra Agrawal. Survey on anomaly detection using data mining techniques. *Procedia Computer Science*, Vol. 60, pp. 708–713, 2015.

[14] 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.

[15] 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.

[16] 岩宮眞一郎. 音質評価指標 –入門とその応用–. 日本音響学会誌, Vol. 66, No. 12, pp. 603–609, 2010.

[17] 大串健吾. 音響聴覚心理学. 誠信書房, 2019.

[18] 難波精一郎. 音色の定義を巡って. 日本音響学会誌, Vol. 49, No. 11, pp. 823–831, 1993.

[19] 厨川守, 八尋博司, 柏木成豪. 音質評価のための 7 属性. 日本音響学会誌, Vol. 34, No. 9, pp. 493–500, 1978.

[20] Masayuki Takada, Kazuhiko Tanaka, and Shin-ichiro Iwamiya. Relationships between auditory impressions and onomatopoeic features for environmental sounds. *Acoustical science and technology*, Vol. 27, No. 2, pp. 67–79, 2006.

[21] Nozomu Fujisawa, Shin-ichiro Iwamiya, and Masayuki Takada. Auditory imagery associated with japanese onomatopoeic representation. *Journal of physiological anthropology and applied human science*, Vol. 23, No. 6, pp. 351–355, 2004.

[22] Masayuki Takada, Nozomu Fujisawa, Fumino Obata, and Shin-ichiro Iwamiya. Comparisons of auditory impressions and auditory imagery associated with onomatopoeic representation for environmental sounds. *EURASIP Journal on Audio, Speech, and Music Processing*, Vol. 2010, pp. 1–8, 2010.

[23] 桑野園子. 機械音騒音の音質評価方法. 日本音響学会誌, Vol. 53, No. 6, pp. 456–461, 1997.

[24] 脇田敏裕. 自動車の音色評価. 豊田中央研究所 R&D レビュー, Vol. 27, No. 3, pp. 23–32, 1992.

[25] 田中俊光. 機械音の音色改善設計. 日本機械学会誌, Vol. 98, No. 917, pp. 295–299, 1995.

[26] Kristoffer Jensen. The timbre model. *Journal of the Acoustical Society of America*, Vol. 112, No. 5, pp. 2238–2238, 2002.

[27] Geoffroy Peeters. A large set of audio features for sound description (similarity and classification) in the cuidado project. 01 2004.

[28] James Mcdermott, Niall Griffith, and Michael O'Neill. Timbral, perceptual, and statistical attributes for synthesized sound. 01 2006.

[29] Andy Pearce, Tim Brookes, and Russell Mason. Timbral attributes for sound effect library searching. In *Audio Engineering Society Conference: 2017 AES International Conference on Semantic Audio*. Audio Engineering Society, 2017.

[30] Andy Pearce, Tim Brookes, and Russell Mason. First prototype of timbral characterisation tools for semantically annotating non-musical content. Deliverable d5.2, AudioCommons, 2017.

[31] Andy Pearce, Saeid Safavi, Tim Brookes, Russell Mason, Wenwu Wang, and Mark Plumbley. Release of timbral characterisation tools for semantically annotating non-musical content. Deliverable d5.8, AudioCommons, 2019.

[32] H. Fastl and E. Zwicker. *Psycho-Acoustics Facts and Models. Third Edition.* Springer, Berlin Heidelberg, 2007.

[33] Peter Daniel and Reinhard Weber. Psychoacoustical roughness: Implementation of an optimized model. *Acta Acustica united with Acustica*, Vol. 83, No. 1, pp. 113–123, 1997.

[34] Ernst Terhardt. On the perception of periodic sound fluctuations (roughness). *Acta Acustica united with Acustica*, Vol. 30, No. 4, pp. 201–213, 1974.

[35] V Vencovskỳ. Roughness prediction based on a model of cochlear hydrodynamics. *Archives of Acoustics*, Vol. 41, No. 2, pp. 189–201, 2016.

[36] Roland Sottek, Julian Becker, and Thiago Lobato. Progress in roughness calculation. In *INTER-NOISE and NOISE-CON Congress and Conference Proceedings, InterNoise20*, pp. 2835–2846. Institute of Noise Control Engineering, 2020.

[37] Pantelis N Vassilakis and K Fitz. Sra: A web-based research tool for spectral and roughness analysis of sound signals. In *Proceedings of the 4th Sound and Music Computing (SMC) Conference*, pp. 319–325, 2007.

[38] S. Hatano and T. Hashimoto. Booming index as a measure for evaluating booming sensation. In *The 29th International Congress and Exhibition on Noise Engineering*, 2000.

[39] 橋本竹夫. 非定常な機械音の音質評価. 日本音響学会誌, Vol. 60, No. 6, pp. 330–335, 2004.

[40] Emery Schubert and Joe Wolfe. Does timbral brightness scale with frequency and spectral centroid? *Acta acustica united with acustica*, Vol. 92, No. 5, pp. 820–825, 2006.

[41] 大串健吾. 音のピッチ知覚. コロナ社, 2016.

[42] 古井貞熙. ディジタル音声処理. 東海大学出版会, 1985.

[43] Anssi Klapuri, Tuomas Virtanen, and Jan-Markus Holm. Robust multipitch estimation for the analysis and manipulation of polyphonic musical signals. In *Proc. COST-G6 Conference on Digital Audio Effects*, pp. 233–236, 2000.

[44] 西一樹, 安部素嗣, 安藤繁. 聴覚情景分析のための多重ピッチ追跡と調波分離アルゴリズム. 計測自動制御学会論文集, Vol. 34, No. 6, pp. 483–490, 1998.

[45] Tuomas Virtanen and Anssi Klapuri. Separation of harmonic sounds using linear models for the overtone series. In *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 2, pp. II–1757–II–1760, 2002.

[46] Ernst Terhardt. Pitch, consonance, and harmony. *The Journal of the Acoustical Society of America*, Vol. 55, No. 5, pp. 1061–1069, 1974.

[47] Ernst Terhardt. Calculating virtual pitch. *Hearing research*, Vol. 1, No. 2, pp. 155–182, 1979.

[48] Dik J Hermes. Measurement of pitch by subharmonic summation. *The journal of the acoustical society of America*, Vol. 83, No. 1, pp. 257–264, 1988.

[49] 富士通株式会社. Iot を活用した「居住者見守りステーション」を提供開始. Press release, 富士通株式会社, 2016.

[50] 児島宏明, 橋本泰治, 佐土原健, 佐宗晃. 音響的素性記述に基づく環境音認識の検討. 日本音響学会講演論文集, 2011.

[51] 井本桂右, 野口賢一, 島内末廣, 大室仲, 羽田陽一. 複数の生活音の出現頻度に基づくユーザ行動の識別手法とコミュニケーションへの応用. 画像電子学会 VMA 研究会, 2012.

[52] Keisuke Imoto, Yasunori Ohishi, Hisashi Uematsu, and Hitoshi Ohmuro. Acoustic scene analysis based on latent acoustic topic and event allocation. In *2013 IEEE international workshop on machine learning for signal processing (MLSP)*, pp. 1–6. IEEE, 2013.

[53] 井本桂右, 大石康智, 植松尚, 大室仲, 小野順貴. 逐次的な観測のための音響シーン分析手法の提案. 日本音響学会誌, Vol. 72, No. 6, pp. 293–305, 2016.

[54] 林知樹, 戸田智基. 統計的手法による音響イベント検出. 日本音響学会誌, Vol. 75, No. 9, pp. 532–537, 2019.

[55] Stavros Ntalampiras, Ilyas Potamitis, and Nikos Fakotakis. An adaptive framework for acoustic monitoring of potential hazards. *EURASIP Journal on Audio, Speech, and Music Processing*, Vol. 2009, pp. 1–15, 2009.

[56] 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.

[57] 吉澤亜耶, 橋本洋一. 異常検知技術の概要と応用動向. INTEC Technical Journal 17, INTEC.

[58] 吉村隆, 浅野太, 麻生英樹, 北脇信彦ほか. 高次統計量の分布モデルを用いた音声・環境音識別法の検討. 情報処理学会研究報告音声言語情報処理 (SLP), Vol. 2004, No. 74 (2004-SLP-052), pp. 137–142, 2004.

[59] P.K. Atrey, N.C. Maddage, and M.S. Kankanhalli. Audio based event detection for multimedia surveillance. In *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, Vol. 5, pp. V–V, 2006.

[60] Akinori Ito, Akihito Aiba, Masashi Ito, and Shozo Makino. Detection of abnormal sound using multi-stage gmm for surveillance microphone. In

2009 Fifth International Conference on Information Assurance and Security, Vol. 1, pp. 733–736. IEEE, 2009.

[61] Erik Marchi, Fabio Vesperini, Florian Eyben, Stefano Squartini, and Björn Schuller. A novel approach for automatic acoustic novelty detection using a denoising autoencoder with bidirectional lstm neural networks. In *2015 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pp. 1996–2000. IEEE, 2015.

[62] B Uğur Töreyin, Yiğithan Dedeoğlu, and A Enis Çetin. Hmm based falling person detection using both audio and video. In *International workshop on human-computer interaction*, pp. 211–220. Springer, 2005.

[63] Yaniv Zigel, Dima Litvak, and Israel Gannot. A method for automatic fall detection of elderly people using floor vibrations and sound—proof of concept on human mimicking doll falls. *IEEE transactions on biomedical engineering*, Vol. 56, No. 12, pp. 2858–2867, 2009.

[64] Charalampos N Doukas and Ilias Maglogiannis. Emergency fall incidents detection in assisted living environments utilizing motion, sound, and visual perceptual components. *IEEE Transactions on Information Technology in Biomedicine*, Vol. 15, No. 2, pp. 277–289, 2010.

[65] 牧山清. 嗄声の聴覚心理的評価 (grbas 尺度). 日本耳鼻咽喉科学会会報, Vol. 115, No. 10, pp. 930–931, 2012.

[66] Aalto University. Introduction to speech processing: Jitter and shimmer. https://speechprocessingbook.aalto.fi/Representations/Jitter_and_shimmer.html.

[67] 細川清人, 岩橋利彦, 小川真, 加藤智絵里, 猪原秀典. 音響分析の概念と実際. 喉頭, Vol. 28, No. 2, pp. 78–87, 2016.

[68] 湯本英二. 嗄声度の定量的指標に関する研究. 耳鼻咽喉科臨床, Vol. 76, No. 9, pp. 2151–2168, 1983.

[69] 小林健二, 鈴木英男, 金井浩, 大橋正尚. 音・振動による診断工学. コロナ社, 2000.

[70] Stephen McAdams. The perceptual representation of timbre. *Timbre: Acoustics, perception, and cognition*, pp. 23–57, 2019.

[71] Stephen McAdams, Antoine Chaigne, and Vincent Roussarie. The psychomechanics of simulated sound sources: Material properties of impacted bars. *The Journal of the Acoustical Society of America*, Vol. 115, No. 3, pp. 1306–1320, 2004.

[72] Daniel J Freed. Auditory correlates of perceived mallet hardness for a set of recorded percussive sound events. *The Journal of the Acoustical Society of America*, Vol. 87, No. 1, pp. 311–322, 1990.

[73] Roberta L. Klatzky, Dinesh K. Pai, and Eric P. Krotkov. Perception of material from contact sounds. *Presence*, Vol. 9, No. 4, pp. 399–410, 2000.

[74] 橘秀樹, 田中ひかり, 上野佳奈子, 横山栄, 船場ひさお. 音と生活. コロナ社, 2016.

[75] Margret Sibylle Engel, André Fiebig, Carmella Pfaffenbach, and Janina Fels. A review of the use of psychoacoustic indicators on soundscape studies. *Current Pollution Reports*, pp. 1–20, 2021.

[76] International Organization for Standardization. *ISO 7729: 2010. Acoustics – measurement of airbone noise emitted by information technology and telecommunication equipments*, 2010.

[77] Gordon R. Bienvenue and Matthew A. Nobile. Prominence ratio for noise spectra with discrete tones: A procedure based on zwicker's critical band research. In *INTER-NOISE and NOISE-CON Congress and Conference Proceedings, InterNoise91.*, pp. 53–56. Institute of Noise Control Engineering, 1991.

[78] J.D. Johnston. Transform coding of audio signals using perceptual noise criteria. *IEEE Journal on Selected Areas in Communications*, Vol. 6, No. 2, pp. 314–323, 1988.

[79] Julian Mathias Becker, Roland Sottek, and Thiago Henrique Gomes Lobato. Progress in tonality calculation. 2019.

[80] Yuma Koizumi, Shoichiro Saito, Hisashi Uematsu, and Noboru Harada. Optimizing acoustic feature extractor for anomalous sound detection based on neyman-pearson lemma. In *2017 25th European Signal Processing Conference (EUSIPCO)*, pp. 698–702. IEEE, 2017.

[81] Yuta Kawachi, Yuma Koizumi, and Noboru Harada. Complementary set variational autoencoder for supervised anomaly detection. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2366–2370. IEEE, 2018.

[82] Yohei Kawaguchi and Takashi Endo. How can we detect anomalies from subsampled audio signals? In *2017 IEEE 27th International Workshop on Machine Learning for Signal Processing (MLSP)*, pp. 1–6. IEEE, 2017.

[83] Robert M'uller, Steffen Illium, Fabian Ritz, and Kyrill Schmid. Analysis of feature representations for anomalous sound detection. *arXiv preprint arXiv:2012.06282*, 2020.

[84] Ji Wu, Fei Yang, and Wenkai Hu. Unsupervised anomalous sound detection for industrial monitoring based on arcface classifier and gaussian mixture model. *Applied Acoustics*, Vol. 203, p. 109188, 2023.

[85] 大石康智. 環境音分析の研究を促進させる競争型ワークショップ. 日本音響学会誌, Vol. 75, No. 9, pp. 519–524, 2019.

[86] Yuma Koizumi, Yohei Kawaguchi, Keisuke Imoto, Toshiki Nakamura, Yuki Nikaido, Ryo Tanabe, Harsh Purohit, Kaori Suefusa, Takashi Endo, Masahiro Yasuda, and Noboru Harada. Description and discussion on

dcase2020 challenge task2: Unsupervised anomalous sound detection for machine condition monitoring, 2020.

[87] Yohei Kawaguchi, Keisuke Imoto, Yuma Koizumi, Noboru Harada, Daisuke Niizumi, Kota Dohi, Ryo Tanabe, Harsh Purohit, and Takashi Endo. Description and discussion on dcase 2021 challenge task 2: Unsupervised anomalous sound detection for machine condition monitoring under domain shifted conditions, 2021.

[88] Kota Dohi, Keisuke Imoto, Noboru Harada, Daisuke Niizumi, Yuma Koizumi, Tomoya Nishida, Harsh Purohit, Takashi Endo, Masaaki Yamamoto, and Yohei Kawaguchi. Description and discussion on dcase 2022 challenge task 2: Unsupervised anomalous sound detection for machine condition monitoring applying domain generalization techniques, 2022.

[89] Kai Li, Quoc-Huy Nguyen, Yasuji Ota, and Masashi Unoki. Unsupervised anomalous sound detection for machine condition monitoring using temporal modulation features on gammatone auditory filterbank. In *Proceedings of the 7th Workshop on Detection and Classification of Acoustic Scenes and Events 2022, DCASE 2022, Nancy, France*, No. 8, Session1. Tampere University, 2022.

[90] Harsh Purohit, Ryo Tanabe, Kenji Ichige, Takashi Endo, Yuki Nikaido, Kaori Suefusa, and Yohei Kawaguchi. Mimii dataset: Sound dataset for malfunctioning industrial machine investigation and inspection. *arXiv preprint arXiv:1909.09347*, 2019.

[91] Yanxue Wang, Richard Markert, Jiawei Xiang, and Weiguang Zheng. Research on variational mode decomposition and its application in detecting rub-impact fault of the rotor system. *Mechanical Systems and Signal Processing*, Vol. 60-61, pp. 243–251, 2015.

[92] Deepam Goyal, Anurag Choudhary, BS Pabla, and SS Dhami. Support vector machines based non-contact fault diagnosis system for bearings. *Journal*

of Intelligent Manufacturing, Vol. 31, pp. 1275–1289, 2020.

[93] Rismaya Kumar Mishra, Anurag Choudhary, Amiya Ranjan Mohanty, and Shahab Fatima. Multi-domain bearing fault diagnosis using support vector machine. In *2021 IEEE 4th international conference on computing, power and communication technologies (GUCON)*, pp. 1–6. IEEE, 2021.

[94] Adam Glowacz, Ryszard Tadeusiewicz, Stanislaw Legutko, Wahyu Caesarendra, Muhammad Irfan, Hui Liu, Frantisek Brumercik, Miroslav Gutten, Maciej Sulowicz, Jose Alfonso Antonino Daviu, Thompson Sarkodie-Gyan, Pawel Fracz, Anil Kumar, and Jiawei Xiang. Fault diagnosis of angle grinders and electric impact drills using acoustic signals. *Applied Acoustics*, Vol. 179, p. 108070, 2021.

[95] Kesaaki Minemura, Tetsuji Ogawa, and Tetsunori Kobayashi. Acoustic feature representation based on timbre for fault detection of rotary machines. In *2018 International Conference on Sensing, Diagnostics, Prognostics, and Control (SDPC)*, pp. 302–305. IEEE, 2018.

[96] P.V. Kane and A.B. Andhare. Critical evaluation and comparison of psychoacoustics, acoustics and vibration features for gear fault correlation and classification. *Measurement*, Vol. 154, p. 107495, 2020.

[97] Sneha Singh, Sagar Potala, and Amiya R Mohanty. An improved method of detecting engine misfire by sound quality metrics of radiated sound. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, Vol. 233, No. 12, pp. 3112–3124, 2019.

[98] Tauheed Mian, Anurag Choudhary, and Shahab Fatima. An efficient diagnosis approach for bearing faults using sound quality metrics. *Applied Acoustics*, Vol. 195, No. 108839, pp. 1–11, 2022.

[99] Min Hu, Zhiwei Ji, Ke Yan, Ye Guo, Xiaowei Feng, Jiaheng Gong, Xin Zhao, and Ligang Dong. Detecting anomalies in time series data via a meta-feature

based approach. *Ieee Access*, Vol. 6, pp. 27760–27776, 2018.

[100] Zhiwei Ji, Yi Wang, Ke Yan, Xiaojun Xie, Yang Xiang, and Jing Huang. A space-embedding strategy for anomaly detection in multivariate time series. *Expert Systems with Applications*, Vol. 206, p. 117892, 2022.

[101] 田中基八郎, 松原謙一郎, 佐藤太一. 機械の異常音の擬音語表現. 日本音響学会誌, Vol. 53, No. 6, pp. 477–482, 1997.

[102] Bearing failure rca: Flaking. ONLINE Accessed on: Mar. 16, 2023., Oct. 2019. https://www.bearing-news.com/bearing-failure-rcaflaking/.

[103] Bürkert Fluid Control Systems GmbH & Co. KG. What is a solenoid valve and how does it work? Technical report, Bürkert Fluid Control Systems GmbH & Co. KG., 2023. https://www.burkert-usa.com/en/company-career/what-s-new/press/media/technical-reports/what-is-a-solenoid-valve-and-how-does-it-work.

[104] Four common causes of pump failure. ONLINE Accessed on: Mar. 16, 2023., Feb. 2020. https://www.pumpindustry.com.au/four-common-causes-of-pump-failure/.

[105] Richard M. Warren, Charles J. Obusek, Richard M. Farmer, and Roslyn P. Warren. Auditory sequence: Confusion of patterns other than speech or music. *Science*, Vol. 164, No. 3879, pp. 586–587, 1969.

[106] Paul Fraisse. Perception and estimation of time. *Annual Review of Psychology*, Vol. 35, No. 1, pp. 1–37, 1984.

[107] 守谷健弘, 鎌本優, 原田登, 杉浦亮介. 音声音響符号化技術の進展. 電子情報通信学会 基礎・境界ソサイエティ Fundamentals Review, Vol. 10, No. 4, pp. 246–256, 2017.

[108] 竹内一郎, 鳥山昌幸. サポートベクトルマシン. 講談社, 2015.

[109] 松原望, 縄田和満, 中井検裕. 統計学入門. 東京大学出版会, 1991.

[110] 井手剛. 入門 機械学習による異常検知. コロナ社, 2015.

[111] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, Vol. 12, pp. 2825–2830, 2011.

[112] S. S. Shapiro and M. B. Wilk. An analysis of variance test for normality.

[113] 田中基八郎（編）. 静音化&快音化設計技術ハンドブック. 科学情報出版, 2012.

本研究に関する発表論文

ジャーナル論文

[1] Yasuji Ota and Masashi Unoki, "Anomalous Sound Detection for Industrial Machines Using Acoustical Features Related to Timbral Metrics," *IEEE Access*, vol. 11, pp. 70884–70897, Jul. 2023.

国際会議

（査読有・ポスター発表）

[2] Yasuji OTA, Seigo KURA and Masashi UNOKI, "Anomalous sound detection using objective metrics related to timbral attributes," *24th International Congress on Acoustics*, ABS–0395, Oct. 2022.

（査読有・ポスター発表・共著）

[3] Kai Li, Quoc-Huy Nguyen, Yasuji OTA and Masashi Unoki, "Unsupervised Anomalous Sound Detection for Machine Condition Monitoring Using Temporal Modulation Features on Gammatone Auditory Filterbank," *DCASE2022 Workshop on Detection and Classification of Acoustic Scene and Events*, Nov. 2022.

国内発表

国内学会

[4] 大田恭士, 倉誠吾, 鵜木祐史, “音色属性の客観評価指標を活用した異常音検知の検討,” 日本音響学会 2022 年秋季研究発表会, 1-Q-14, 2022 年 9 月.

[5] 大田恭士, 鵜木祐史, “音色の客観指標と信号分析を組合せた異常音検知の一検討,” 第 37 回信号処理シンポジウム, 2022 年 12 月.

[6] 大田恭士, 鵜木祐史, “産業機器の異常音検知に向けた音色関連特徴量の検討,” 日本音響学会 2023 年秋季研究発表会, 1-R-15, 2023 年 9 月.

[7] 大田恭士, 鵜木祐史, “音色に関連づけた信号処理手法による異常音検知の一検討,” 第 38 回信号処理シンポジウム, 2023 年 11 月.