

# 機械製品の動作音による半自動検査手法の提案

石田 繁巳<sup>1,a)</sup> 松井 威<sup>2</sup> 金井 貴浩<sup>3</sup>

**概要:** 近年、機械製品の動作音による検査の重要性が増している。これまでの熟練の検査員による官能検査が多く行われてきたが、検査員の負荷や技術伝承が問題となっており、自動化に向けた研究が進められている。自動化に向けては定量的な評価指標の決定に向けた分析が必要であり、手間が大きいことから、本稿では、製品ごとの特徴を人間が分析することなく半自動的に異常を検知する D-SAPI (Difficulty-based Semi-Automatic Product Inspection) システムを示す。D-SAPI システムは、機械の動作状態を推定する教師あり学習モデルを構築し、異常音では動作状態推定が難しくなることを利用して異常を検知する。実際の工場環境で収集した音声データを用いて評価を行った結果、最大 ROC-AUC 0.818 で異常を検知できることを確認した。

**キーワード:** 異常検知, 教師あり学習, 音声信号

## 1. はじめに

静音環境で用いられる機械製品には静けさが要求されることから、動作する際に発生する作動音の品質検査がますます重要となっている。動作音による検査は音の品質を保証するだけでなく、製品内部の異常による動作異常も検知できることから、静音環境向けの機械製品以外でも利用されている。

動作音による検査は、これまで熟練の検査員の聴覚に頼る官能検査が多く行われてきた。官能検査は工数や検査員の負荷、技術伝承の問題があることから、動作音を分析して定量的な評価を行う手法が報告されている [1], [2]。

しかしながら、定量的な評価指標の決定には多くの人手が必要であり、手間が大きい。評価指標の決定では大量の音データを分析し、音の周波数成分やその変化、高調波成分の割合や音の安定性、音量変化など、様々な特徴を可視化する。その上で、正常品と異常品の差を分析し、大きな差が見られる特徴に基づいて正常・異常を判定する評価指標を定める。一般に、このような分析の結果は製品ごと、異常ごとに異なるものとなるため、製品ごとの分析が必須となる。このような分析を自動化する手法や、体系化された分析手法はこれまでに報告されておらず、大きなコストを要する。

これに対し、本稿では、製品ごとの特徴を人間が分析する

ことなく半自動的に異常を検知可能な手法として、D-SAPI (Difficulty-based Semi-Automatic Product Inspection) を示す。機械製品の動作状態、例えば、回転する機械製品では回転速度を指定する「モード」を動作音から推定しようとしたとき、異常がある製品では推定が難しくなる。そこで、「推定の難しさ」に基づいて異常度を算出する。具体的には、正常音データを用いて動作音から機械製品の動作状態を推定する機械学習モデルを構築し、動作音から動作状態を推定したときの推定確信度に基づいて異常を検知する。

提案する D-SAPI の性能を検証するため、機械製品の検査時に収集したデータを用いて実証評価を行った。その結果、最大 ROC-AUC 0.818 で異常を検知できることを確認した。

本稿の構成は以下の通りである。2 では異常検知手法に関する関連研究を示し、3 で提案する D-SAPI を示す。4 で実証評価を行い、最後に 5 でまとめとする。

## 2. 関連研究

音による異常検知、すなわち異常音検知は広い分野で利用されている。環境音における異常音検知として、銃声や悲鳴、爆発音、ガラス破裂音、車の衝突音、高齢者の転倒音、いびき検知、心音の異常検知、嚙下音の異常検知など、日常生活の中での異常を検知する手法として、様々な分野で広く研究開発が進んでいる [3]。

しかしながら、筆者らの調査した範囲では、機械製品の動作音による検査において製品ごとの特徴を分析することなく異常を検知する手法は、これまでのところ報告されて

<sup>1</sup> 公立はこだて未来大学システム情報科学部  
<sup>2</sup> 公立はこだて未来大学大学院システム情報科学研究科  
<sup>3</sup> アイシン・ソフトウェア (株)  
a) ish@fun.ac.jp

いない。本節では、機械などの異常検知に関する関連研究を俯瞰する。正常音データと異常音データの両方を用いる教師あり学習に基づく手法も報告されているが [4]、本稿では、正常音データのみ、あるいは正常音データに加えて少数の異常データを用いる半教師あり学習手法を中心として関連研究について述べる。

熟練の検査員の聴覚に頼る官能検査に代わる手法として、異常音を分析した上で行われるヒューリスティックな検知手法が報告されている。文献 [1] では、CD チェンジャのピックアップの動作音を分析し、5種類の特徴量を用いたルールベースの異常検知手法を示している。文献 [2] は、自動車部品の動作音検査において動作音の周波数が変化する異常音の検知方法として、ケプストラム解析とリッジ解析を組み合わせた手法を示している。これらの手法は非常に高い性能を示すものの、対象となる機械の異常音の分析が必須となる。

異常音を用いないアプローチとして、オートエンコーダの再構成誤差に基づく異常音検知手法が報告されている [5]。畳み込みオートエンコーダを正常音のスペクトログラムによって事前に学習しておき、テスト対象となる音を入力したときの再構成誤差をしきい値判定することで異常音を検知する。畳み込みオートエンコーダと One-class SVM (Support Vector Machine) を組み合わせることで性能を向上できることも示されている。

オートエンコーダではなく、分類器を用いるアプローチも報告されている。文献 [6] では、条件付き MAF (Masked Autoregressive Flow) を用いた生成モデルによって同一機械の同定や機械の種類を分類する分類器を構成し、推定時の対数尤度を用いて異常検知する手法が示されている。文献 [7] では、プロキシ異常データと呼ばれる異常音と類似性を有する疑似的な異常音データを用いて正常・異常を分類する 2 クラス分類器を学習する手法が示されている。いずれの手法も異常音データを用いていないが、異常音の特性を用いた分類器や出力処理が行われている。少数の異常音データとして外れ値データを用い、クラス間の距離が離れるように深層距離学習を行った多クラス分類器を用いる手法も報告されているが [8]、高い性能を得るためには多くの外れ値データや異常データが必要となる。

分類器とオートエンコーダを組み合わせる手法も報告されている [9]。分類器の中間層出力をオートエンコーダに入力し、異常確率と距離、再構成誤差を併用して異常を検知している。しかしながら、分類器を用いるアプローチと同様に、学習時に正常音データに加えて疑似的な異常音データが必要となる。

### 3. D-SAPI

#### 3.1 キーアイデア

提案する D-SAPI (Difficulty-based Semi-Automatic

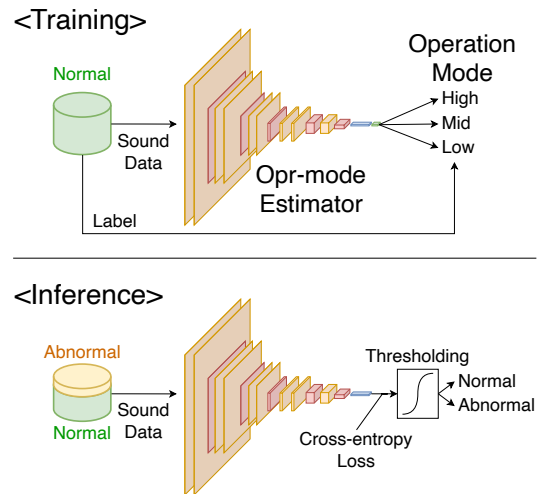


図 1 キーアイデア

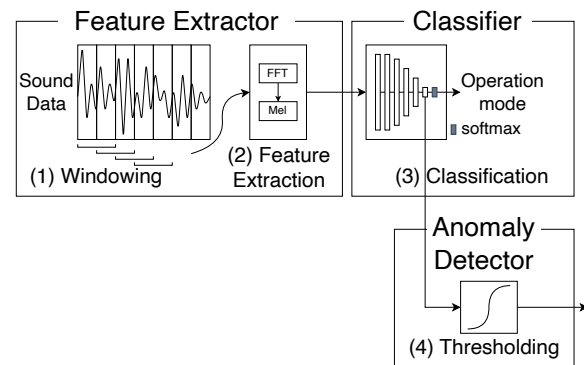


図 2 D-SAPI システムの概要

Product Inspection) のキーアイデアは、機械の動作状態の推定が正常音に比べて異常音では難しくなることを利用して、異常を検知することである。図 1 に示すように、学習フェーズでは、正常な機械の音声データを用いて動作状態、例えば回転速度などの「モード」を分類する機械学習モデルである動作状態推定器 (Opr-mode Estimator) を学習する。推論フェーズでは、推定対象の音声データに対する動作状態推定器の出力からクロスエントロピー損失、すなわち負の対数尤度を計算して、あらかじめ定められたしきい値を用いて異常判定を行う。

#### 3.2 システム概要

図 2 に、提案する D-SAPI システムの概要を示す。D-SAPI システムは、特徴量抽出ブロック、分類ブロック、異常検知ブロックの 3 つのブロックで構成される。まず、特徴量抽出ブロックにおいて機械の動作音データからメルスペクトログラムを取得する。分類ブロックでは、あらかじめ学習された教師あり学習モデルである動作状態推定器を用いて、メルスペクトログラムから機械の動作状態を推定する。異常検知ブロックは、分類ブロックの動作状態推定

器の学習時に用いる損失関数の出力を取り出して異常スコアとし、しきい値判定によって異常を検知する。図2は教師あり学習モデルとしてニューラルネットワークを用いる場合を描いており、損失関数としてクロスエントロピー損失を取り出して異常を検知している。

以降では各ブロックについて詳述する。

### 3.3 特徴量抽出ブロック

特徴量抽出ブロックでは、機械動作音の音声データから特徴量としてメルスペクトログラムを抽出する。メルスペクトログラムは、スペクトログラムの周波数軸をメル尺度変換したメルスペクトログラムであり、音の特徴を示す特徴量として音声認識タスクで広く利用されている。

まず、音声データを固定幅のウィンドウで区切り、FFT (Fast Fourier Transform) を施す。FFT で得られる周波数成分  $X$  は、全ウィンドウの数  $N_{\text{win}}$ 、FFT ウィンドウサイズ  $N_{\text{fft}}$  として、 $N_{\text{win}} \times N_{\text{fft}}$  の行列である。次に、メルフィルタバンクを適用してウィンドウ毎のメルスペクトログラムを得る。メル尺度のチャンネル数を  $N_{\text{mel}}$  とするとメルフィルタバンク  $M$  は  $N_{\text{fft}} \times N_{\text{mel}}$  の行列として表され、メルスペクトログラム  $Y$  は、

$$Y = XM \quad (1)$$

として得られる。

対象とする音声データの周波数成分に大きな偏りがある場合には、メルスペクトログラムを対数スケールに変換し、ログメルスペクトログラムを得て使用する。ログメルスペクトログラム  $S$  は、 $Y$  の各成分に対数スケールへの変換、すなわち dB (デシベル) 変換を施して得られる。

$$S = 10 \log_{10} Y \quad (2)$$

### 3.4 分類ブロック

分類ブロックでは、学習済みの動作状態推定器を用いて特徴量抽出ブロックで得たメルスペクトログラム  $S$  から動作状態を推定し、推定時の損失関数出力を得る。本稿で提案する D-SAPI では、動作状態推定器として使用する機械学習モデルを限定しない。学習時に使用される損失関数を出力として利用できれば D-SAPI に組み込むことが可能である。

学習フェーズでは、正常音データを用いて動作状態推定器を学習する。損失関数の出力が最小となるように各動作状態の正常音を用いて学習を行う。本稿では、動作状態推定器の一例として、音声データによる機械の異常判定手法 [7] を参考に、ResNet を用いる。

推定フェーズでは、動作状態推定器にテスト対象の音データを入力し、学習フェーズの学習で用いていた損失関数の出力を異常スコアとして取り出す。ResNet を用いた

動作状態推定ではクロスエントロピー損失から softmax 層を経て動作状態の推定結果が出力されるため、softmax 層に入力しているクロスエントロピー損失を異常スコアとする。

短いウィンドウで計算したメルスペクトログラムでは動作状態推定・異常検知に必要な情報が含まれないことも想定されるため、動作状態推定器には複数ウィンドウのメルスペクトログラムをまとめた「フレーム」を入力する。フレーム長、すなわちフレームに含まれるウィンドウ数を  $L_{\text{fr}}$  とすると、動作状態推定器の入力は、 $N_{\text{mel}} \times L_{\text{fr}}$  の行列となる。フレームはウィンドウを1つずつずらして定義する。動作状態推定器からは  $N_{\text{win}} - L_{\text{fr}} + 1$  個の異常スコアが得られる。

### 3.5 異常検知ブロック

異常検知ブロックでは、分類ブロックの推定フェーズで得られる損失関数出力に対して、しきい値判定を施して異常を検知する。D-SAPI は、正常音データを用いたときと異常音データを用いたときで動作状態推定の難易度が異なることを前提としている。動作状態推定器は教師あり学習モデルであり、異常音データを入力したときは正常音データを入力したときと比べて損失関数の出力、すなわち異常スコアが大きくなる。そこで、異常スコアがあらかじめ定められたしきい値を上回る場合に異常と判定する。

動作状態推定器からは  $N_{\text{win}} - L_{\text{fr}} + 1$  個の異常スコアが得られることから、異常検知ブロックではこれらを集約処理して1つの値に集約した上で、しきい値判定をして異常を検知する。集約処理としては、平均や中央値、最大値、ピークツーピーク値、四分位範囲などを想定しており、評価において適切な集約処理を検討する。

判定に使用するしきい値は、分類ブロックの動作状態推定器を正常音データで学習した際の異常スコアの分布から定める。学習済みの動作状態推定器に正常音データを入力した際の異常スコアの集合を  $A$ 、ある値  $x$  以下の要素からなる  $A$  の部分集合

$$A_s(x) = \{a \in A | a \leq x\} \quad (3)$$

とする。許容する誤検知率を  $p_{\text{max}}$  とするとき、しきい値  $a_{\text{th}}$  は以下の式を解いて得られる。

$$\frac{|A_s(a_{\text{th}})|}{|A|} = p_{\text{max}} \quad (4)$$

実際に使用する際には異常音データの異常スコアも考慮してしきい値  $a_{\text{th}}$  を定めることが望ましい。 $a_{\text{th}}$  を大きくすると偽陽性 (FP: False Positive) の検知ミスは減少するが、偽陰性 (FN: False Negative) の検知ミスが増加する可能性がある。このため、少数の異常音データを用いてしきい値  $a_{\text{th}}$  を検討する必要がある。本稿の評価では、しき

表 1 評価データセットのデータ数

Anomaly	OPEN/CLOSE	OPEN	CLOSE
None	964	36	36
A	0	2	2
B	6	0	0

い値  $a_{th}$  を変化させたときの ROC 曲線から ROC-AUC を得て性能を検証する。

## 4. 評価

提案する D-SAPI システムの基本性能を検証するため、実際の工場で収集された機械製品検査時の音声データを用いて評価を行った。

### 4.1 評価データセット

評価には、モータによって往復動作を行う製品の製造時検査における音声データを用いた。往路/復路動作をそれぞれ OPEN/CLOSE とし、OPEN/CLOSE のそれぞれについて防音室内で音声データサンプルを収集した。OPEN/CLOSE はいずれも 5 秒程度を要することから、動作開始から終了までの約 5 秒間の音声をサンプリング周波数 51.2kHz、量子化ビット数 24bit で取得し、1 サンプルの音声データとした。なお、対象とする製品は単一のものであるが、複数の品種が含まれている。

表 1 に、評価に使用したデータセットのデータ数を示す。表では、機械 1 個体に対して OPEN/CLOSE の両方のサンプルが揃っているもの、OPEN または CLOSE のみのサンプルがあるもののそれぞれでデータ数を示している。熟練した検査員が音声データを聴いてアノテーションを行い、正常 (Anomaly None) と、2 種類の異常 (Anomaly A, Anomaly B) のラベル付けを行った。OPEN, CLOSE のそれぞれで、異常のない正常データは  $964 + 36 = 1000$  サンプル、異常 A のデータは 2 サンプル、異常 B のデータは 6 サンプルであった。

事前処理として、サンプルのデータ長を揃えてメルスペクトログラムを得た。各サンプルはデータ長が異なることから、もっとも短いデータの長さに合わせて先頭から切り出した。そして、ウィンドウサイズ  $N_{fft} = 1024$  で 50% オーバーラップさせながらウィンドウに分割し、各ウィンドウで 128 次元のログメルスペクトログラムを算出した。

### 4.2 評価環境

評価データセットを学習データとテストデータに分割し、学習データを用いて動作状態推定器を学習した上で、テストデータを用いて異常検知性能を評価した。表 1 に示すように異常データが極めて少ないことから、テストデータは異常データ及び異常データと同数の正常データとした。OPEN/CLOSE の両方が揃っている正常データから 8

サンプルをランダムに取得し、異常データと合わせてテストデータとした。テストデータを除いたあとに残ったデータを学習データとして動作状態推定器を学習した。

動作状態推定器は ResNet-18 を用い、入力サイズに合わせて最初の畳み込み層、MaxPool 層を変更した。動作状態として OPEN/CLOSE を推定するため、出力層は 1 つの出力に Sigmoid 関数を適用して 2 クラスに分類する形に変更した。そして、推定器の未変更部分に ImageNet で事前学習された ResNet-18 の重みを適用し、Binary Cross Entropy を損失として動作状態推定器全体の学習を行った。学習データは、トレーニングデータ : バリデーションデータ = 8 : 2 となるように分割し、最大エポック数を 100 として、バリデーションデータに対する損失が 3 エポック連続で減少しなかった場合に Early Stop させて学習した。

本稿の評価では「いずれかの状態で異常がある機械個体を異常と判定」するものとし、OPEN, CLOSE のいずれかで異常スコアがしきい値を超える場合を「異常」とした。このため、機械の各個体に対して動作状態推定器から OPEN, CLOSE のそれぞれの異常スコアを得て、小さくない方の異常スコアをその個体の異常スコアとした。異常 A については機械の各個体について OPEN または CLOSE のみのサンプルしか存在しないことから、存在するサンプルから得られた異常スコアをその機械個体の異常スコアとした。例えば、OPEN のみのサンプルがある機械個体の場合には、OPEN のサンプルから得られた異常スコアをその機械個体の異常スコアとした。

3.5 で示したように D-SAPI はしきい値を用いて異常検知を行うことから、異常検知の性能は ROC-AUC (Receiver Operating Characteristic Curve-Area Under Curve) で評価した。しきい値を変化させたときの真陽性 (TP: True Positive) 率、偽陽性 (FP) 率を計算して ROC 曲線を描き、曲線下部の面積比率を算出した。学習データとテストデータの分割から ROC-AUC の算出までを 1 試行とし、50 試行の ROC-AUC を得てその平均値を算出した。

### 4.3 異常検知性能

表 2 に、異常検知結果の平均 ROC-AUC を示す。表は、フレーム長  $L_{fr}$ 、異常スコアの集約処理の組み合わせごとの結果を示している。各フレーム長、すなわち各列において、太字は平均 ROC-AUC が最大のもの、下線は 2 番目に

表 2 異常検知結果の平均 ROC-AUC

Aggregation	$L_{fr} = 1$	$L_{fr} = 2$	$L_{fr} = 4$	$L_{fr} = 8$
mean	0.796	0.782	<u>0.760</u>	<u>0.727</u>
median	0.743	0.752	0.747	<b>0.743</b>
std	<b>0.818</b>	0.796	<b>0.761</b>	0.719
max	<u>0.807</u>	<u>0.809</u>	0.756	0.700
iqr	0.771	0.756	0.743	0.725
p2p	<u>0.807</u>	<b>0.810</b>	0.756	0.700

表 3 異常の種類ごとの異常検知結果の平均 ROC-AUC  
(a) 異常 A

Aggregation	$L_{fr} = 1$	$L_{fr} = 2$	$L_{fr} = 4$	$L_{fr} = 8$
mean	0.588	0.578	0.511	0.495
median	0.459	0.482	0.458	0.476
std	<u>0.674</u>	<u>0.632</u>	<u>0.534</u>	<b>0.500</b>
max	<b>0.753</b>	<b>0.709</b>	<b>0.553</b>	<u>0.498</u>
iqr	0.535	0.507	0.468	0.470
p2p	<b>0.753</b>	<b>0.709</b>	<b>0.553</b>	<u>0.498</u>

(b) 異常 B

Aggregation	$L_{fr} = 1$	$L_{fr} = 2$	$L_{fr} = 4$	$L_{fr} = 8$
mean	<b>0.935</b>	0.918	<u>0.927</u>	0.882
median	<u>0.933</u>	<b>0.933</b>	<b>0.939</b>	<b>0.917</b>
std	0.914	0.906	0.912	0.865
max	0.842	0.876	0.892	0.836
iqr	0.928	<u>0.922</u>	0.926	<u>0.895</u>
p2p	0.842	0.877	0.892	0.836

大きいものを示している。表 2 より、以下のことが分かる。

- フレーム長  $L_{fr} = 1$ , 集約処理が標準偏差のときに最大の ROC-AUC = 0.818 が得られた。異常スコアはフレームごとに異なる値となるが、異常がある機械個体では異常スコアがフレームごとに大きくなったり小さくなったりすることでバラツキが大きくなり、その変化を捉える標準偏差によって高い異常検知性能を得られたと考えられる。
- フレーム長  $L_{fr} = 4, 8$  のときと比較して、フレーム長が  $L_{fr} = 1, 2$  のときの方が高い ROC-AUC を得られた。サンプリング周波数 51.2 kHz で取得したデータをウィンドウサイズ  $N_{fft} = 1024$  で分割しているため、1 つのウィンドウは  $1024 / (51.2 \times 10^3) = 0.02$  秒、すなわち 20 ミリ秒に相当する。ウィンドウを 50% オーバーラップで取り出していることに注意すると、フレーム長  $L_{fr}$  は  $20 + 10(L_{fr} - 1) = 10(L_{fr} + 1)$  ミリ秒に相当する。 $L_{fr} = 1, 2$  はそれぞれ 20 ミリ秒、30 ミリ秒に相当することから、本評価で用いたデータセットの異常音が 20 ミリ秒、30 ミリ秒程度の長さで捉えられるものであったと考えられる。

以上より、提案する D-SAPI システムによって正常データのみを用いた教師あり学習で異常を検知できることが確認された。

#### 4.4 異常の種類ごとの異常検知性能

異常 A, B のそれぞれに対する性能を検証するため、試行ごとに異常 A, B のそれぞれと正常データを取り出して ROC-AUC 得て、その平均値を算出した。表 3 に、異常 A, B のそれぞれに対する平均 ROC-AUC を示す。各フレーム長において、太字、下線はそれぞれ平均 ROC-AUC が最大、2 番目のものを示している。表 3 より、以下のこ

とが分かる。

- 最大の ROC-AUC は、異常の種類ごとに異なるフレーム長で得られた。異常 A, B に対してそれぞれ  $L_{fr} = 1, 4$  のときに ROC-AUC は最大の 0.753, 0.939 となった。異常の種類ごとに異常を観測しやすい時間幅が異なるためと考えられる。
- 異常 A では、標準偏差や最大値、ピークツーピーク値など、フレームごとの異常スコアのバラツキを評価する集約処理のときに比較的高い ROC-AUC が得られた。異常 A は短いフレーム長で異常が観測され、異常が観測されるフレームとされないフレームが存在することでフレームごとの異常スコアのバラツキが大きくなるものと考えられる。
- 異常 B では、平均や中央値など、フレームごとの異常スコアの平均的な値を評価する集約処理のときに比較的高い ROC-AUC が得られた。異常 B が長い時間で観測される異常であり、多くのフレームで異常スコアが大きくなったためと考えられる。
- 異常 A では、フレーム長  $L_{fr}$  の増加とともに ROC-AUC が大幅に低下した。長いフレーム長では異常 A の異常を観測しづらいためと考えられる。
- 異常 B では、フレーム長  $L_{fr}$  を変化させても ROC-AUC は大きく変化しなかった。異常 B では短時間の観測であっても平均的に異常スコアが大きくなる傾向があるためと考えられる。

以上より、異常の種類ごとに適切なフレーム長、適切な集約処理を行うことで異常を検知できることが確認された。一方で、異常の種類によってはフレーム長の設定が異常検知性能に大きな影響を与えることも確認された。

図 3 に、フレームごとの異常スコアのヒストグラムを示す。図は、フレーム長ごとに、全試行、全フレームの異常スコアのヒストグラムを正常、異常 A, 異常 B のそれぞれについて描いている。図より、以下のことが分かる。

- 分布のピークの異常スコアを比較すると、異常 A, 異常 B ともに正常のときと比べて大きな異常スコアにピークが存在する。すなわち、異常の種類に関わらず、異常スコアが大きくなるフレームが正常の場合よりも多い傾向にある。
- 異常 A, 異常 B の分布は、正常な場合の異常スコアの分布と多くの部分で重なっている。フレームごとに異常スコアを見て異常を判定することは難しいと言える。
- 異常 B は、フレーム長が長いほど大きな異常スコアを示すフレーム数が増加している。異常 B が長い時間で観測される異常であるためと考えられる。
- 異常に関係なく、異常スコア  $10^{-6}$  辺りで分布が不連続に変化している。評価用の音声データは機械の動作が停止してから記録を停止しており、ほぼ無音の区間が含まれる。このような無音区間のフレームの異常ス

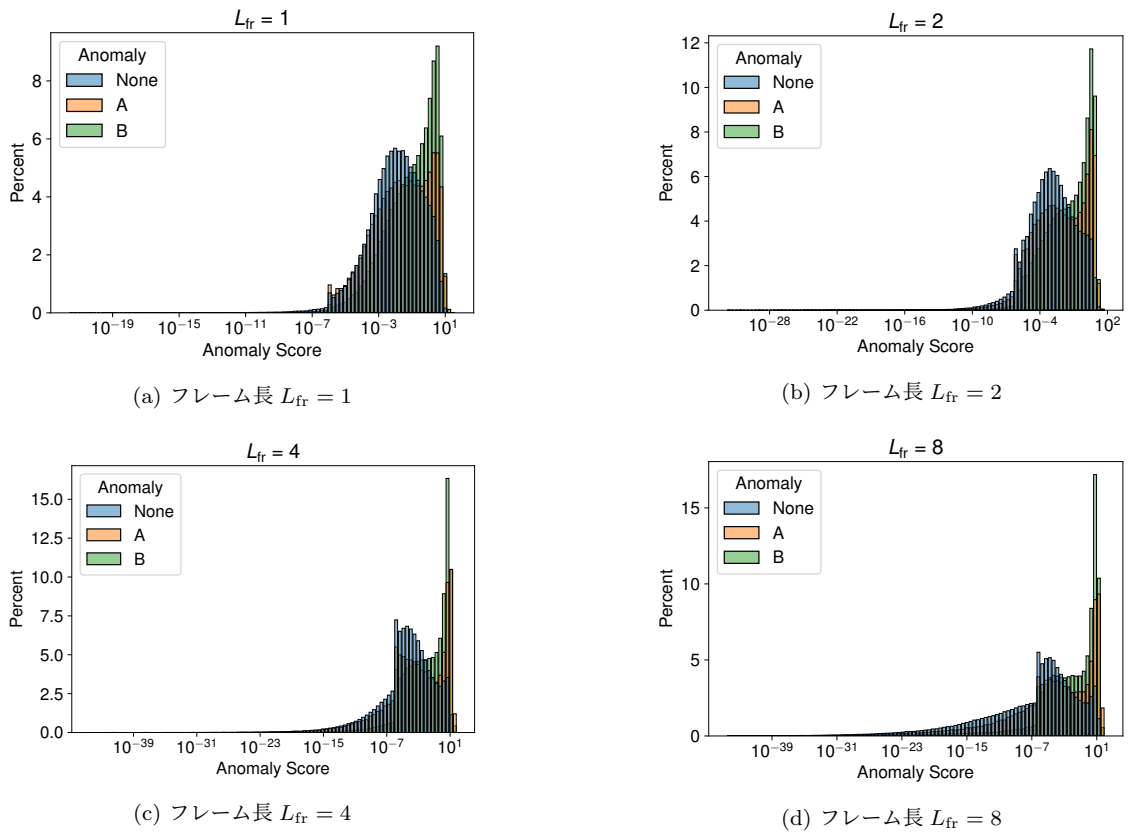


図 3 フレームごとの異常スコアのヒストグラム

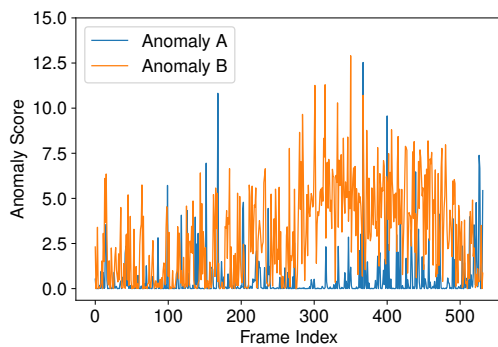


図 4 フレームごとの異常スコアの例

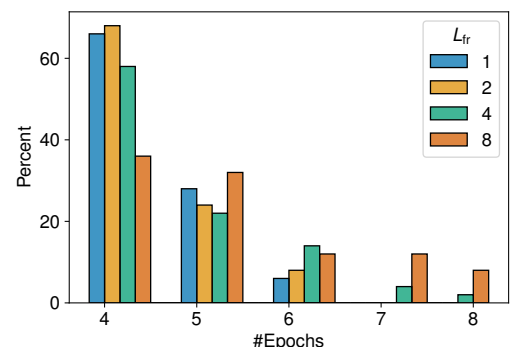


図 5 学習エポック数のヒストグラム

コアが一律に小さくなったものと考えられる。

図 4 に、異常 A、B それぞれのある試行におけるフレームごとの異常スコアの例を示す。図 4 より、異常 A は最大の異常スコアは大きいものの、異常スコアが小さいフレームも多く存在することが分かる。これに対し、異常 B では異常スコアが大きいフレームが多く、平均異常スコアが大きくなる。

以上より、単独のフレームで異常を検知することは難しく、フレームごとの異常スコアの統計処理が必須であることが確認された。異常 A と異常 B でフレームごとの異常スコアの傾向が異なるため、統一した検知基準を定めることが難しいことも確認された。

#### 4.5 学習エポック数

D-SAPI では動作状態推定器を正常データで学習することが必須であることから、動作状態推定器の学習にどの程度のコストがかかるのかを見積もるため、学習時のエポック数を確認した。4.2 で述べたように、バリデーションデータに対する損失が 3 エポック連続で減少しなかった場合に Early Stop させた。このため、最短では 4 エポックで学習が終了する。

図 5 に、学習エポック数のヒストグラムを示す。図は、フレーム長ごとに全試行の学習が終了したエポック数を集計した結果を示している。図より、以下のことが分かる。

- フレーム長に関わらず、60%以上の試行で 5 エポック以内に学習が終了した。学習に要したエポック数は最

大でも 8 である。本稿の評価で対象とした機械製品では、データ動作音に基づいて動作状態を推定することが比較的容易であり、少ないエポック数で学習が終了したと考えられる。

- フレーム長が長いほど、大きな学習エポック数の試行が増加した。3.4 で示したように、動作状態推定器の入力は  $N_{\text{mel}} \times L_{\text{fr}}$  の行列であり、フレーム長が長いほど大きくなるため、学習エポック数が増加したと考えられる。

以上より、動作状態の識別が比較的容易である場合には、動作状態推定器の学習に要するコストは小さいことが確認された。動作状態の推定が容易である場合には動作状態推定器としてより小さなモデルを利用することも可能と考えられるが、動作状態推定器の性能と異常検知性能の関係性については検討の余地が残る。

## 5. おわりに

本稿では、動作音による機械製品の検査において製品ごとの特徴を分析することなく半自動的に異常を検知する手法として、D-SAPI システムを示した。D-SAPI システムでは正常な機械の動作音から機械の動作状態を推定する教師あり学習モデルを構築した上で、動作状態推定が異常によって困難になることを利用して異常検知を行う。実際の工場環境で収集した機械製品検査時の音声データを用いて評価を行った結果、最大 ROC-AUC 0.818 で異常を検知できることを確認した。

## 参考文献

- [1] 大瀧淳司, 小菅 太, 三譚 勉, 植原孝之, 藤江龍一: 異音を定量的に検出する手法の開発, 富士通テン技報 (Fujitsu Ten Technical Report), Vol. 23, No. 2, pp. 37–46 (2005).
- [2] 笠原太郎, 坂本湧暉, 宗戸俊樹, 高部晃好, 伊藤秀浩: 製品動作音の新検査法の開発, アイシン技報 (AISIN Technical Review), Vol. 28, pp. 39–42 (2024).
- [3] 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・境界サイエンス Fundamentals Review, Vol. 15, No. 4, pp. 268–280 (2022).
- [4] 土坂祐太郎, 伊藤 徹, 酒井 涉, 森田克明: 深層学習モデルによる雑音環境下における機械音の異常分類, 人工知能学会全国大会論文集, Vol. 2Win5-65, pp. 1–4 (2025).
- [5] Duman, T. B., Bayram, B. and İnce, G.: Acoustic Anomaly Detection Using Convolutional Autoencoders in Industrial Processes, *14th International Conference on Soft Computing Models in Industrial and Environmental Applications (SOCO 2019)* (Martínez Álvarez, F., Troncoso Lora, A., Sáez Muñoz, J. A., Quintián, H. and Corchado, E., eds.), Seville, Spain, pp. 432–442 (2019).
- [6] Haunschmid, V. and Praher, P.: Anomalous Sound Detection with Masked Autoregressive Flows and Machine Type Dependent Postprocessing, *Detection and Classification of Acoustic Scenes and Events (DCASE) Challenge Technical Report*, Tokyo, Japan, pp. 1–4 (2020).
- [7] Primus, P., Haunschmid, V., Praher, P. and Widmer, G.: Anomalous Sound Detection as a Simple Binary Classification Problem with Careful Selection of Proxy Out-lier Examples, *Detection and Classification of Acoustic Scenes and Events (DCASE) Workshop*, Tokyo, Japan, pp. 170–174 (2020).
- [8] Kuroyanagi, I., Hayashi, T., Takeda, K. and Toda, T.: Anomalous Sound Detection Using a Binary Classification Model and Class Centroids, *2021 29th European Signal Processing Conference (EUSIPCO)*, Dublin, Ireland, pp. 1995–1999 (2021).
- [9] 松本啓汰, 山田武志, 牧野昭二: 多クラス分類器とその中間層出力の再構成器を用いた異常音検知の検討, 電子情報通信学会技術研究報告 音声 (SP) 研究会, Vol. 122, pp. 77–81 (2022).