研究背景および実施項目

◆機器の状態を音響で診断する音響診断技術は、振動診断のデメリットであるセンサ設置といった複雑な作業がないため、効率的かつ可搬的に診断できるといったメリットがある。しかしながら、収音されるデータに環境音といったノイズが含まれることなどから状態診断の精度は高くない。

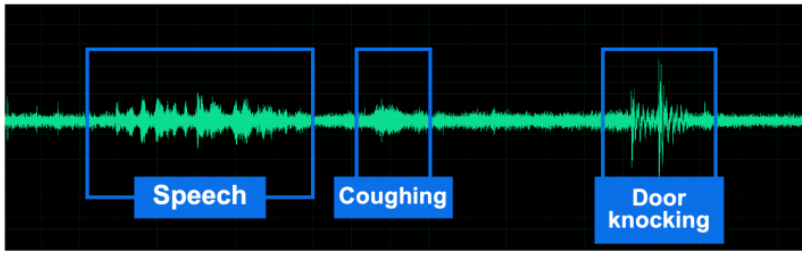
◆これまで取り組み（3Dマイクによる音源定位や機械学習・LSTMによる異常検知）を踏まえて、これまでに高度化できていなかった異常診断について、よりディープなネットワーク構造を持つAIモデル（例えば、環境音分類AIなどを参照する）を用いた異常診断を検討する。

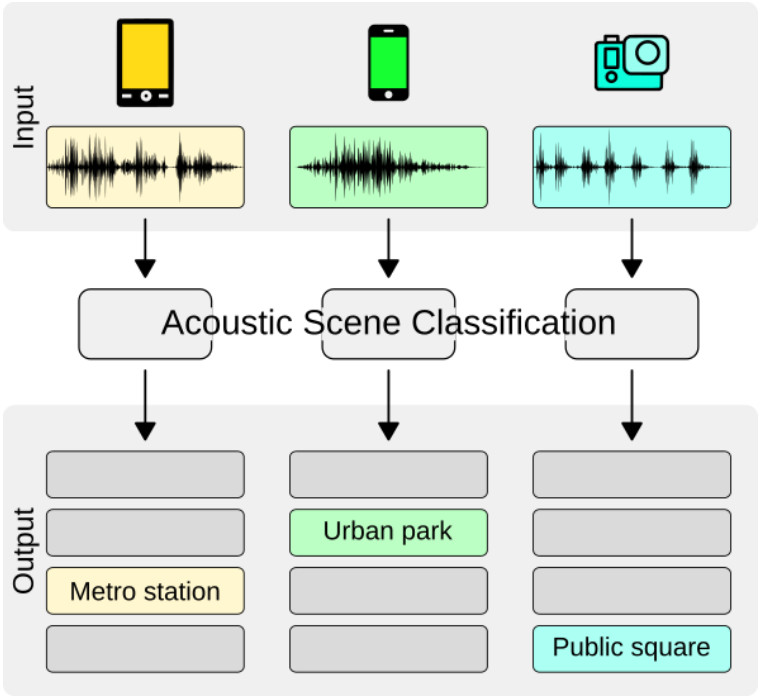
<u>実施項目</u>

- 時系列データを取り扱えるディープなネットワークを有するAIモデルの調査と異常検知の可能性の検討
- 時系列データ（音）を用いて、調査で得られたAIモデルの性能評価

環境音分類AIについて

環境音認識とは、その場で発生している音から、状況を認識することである。近年AIを用いた環境音認識が行われている。





この技術を利用して、音による

- ✓ 機器の異常検知
- ✓ 現場の状況把握

の可能性を検討した

音の分類AIモデルについて

主な手法：

1. 時系列データから短時間フーリエ変換（STFT）して得られるスペクトログラムを画像出力し、画像分類AIモデルを用いて分類する。

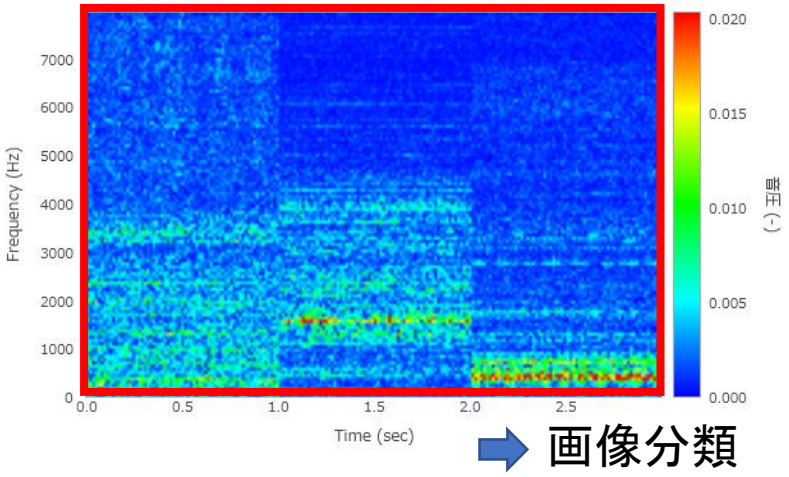
   ※画像分類：写っているものを特定して分類する

   ☞ 特徴量に周波数特性を用いる

2. 時系列データそのものをCNNなどによって特徴量抽出を行い、特徴量を用いて分類する。

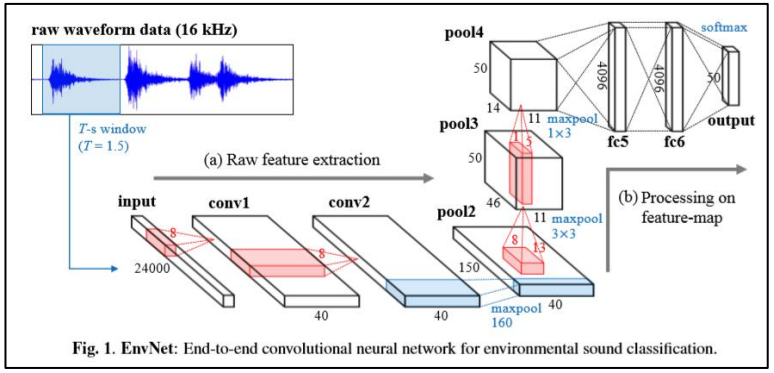
   ☞ 特徴量抽出からAIを用いる

スペクトログラム



➡ 画像分類

LEARNING ENVIRONMENTAL SOUNDS WITH END-TO-END CONVOLUTIONAL NEURAL NETWORK



EnvNet

本研究では、1. と2. を組み合わせた手法によるAIモデルを利用する

PANNsによる環境音分類について

Large-Scale Pretrained Audio Neural Networks for Audio Pattern Recognition

Kong, Qiuqiang, et al. "Panns: Large-scale pretrained audio neural networks for audio pattern recognition." IEEE/ACM Transactions on Audio, Speech, and Language Processing 28 (2020): 2880-2894.

生波形に対して畳み込みニューラルネットワークを通すことで、次元を削減する

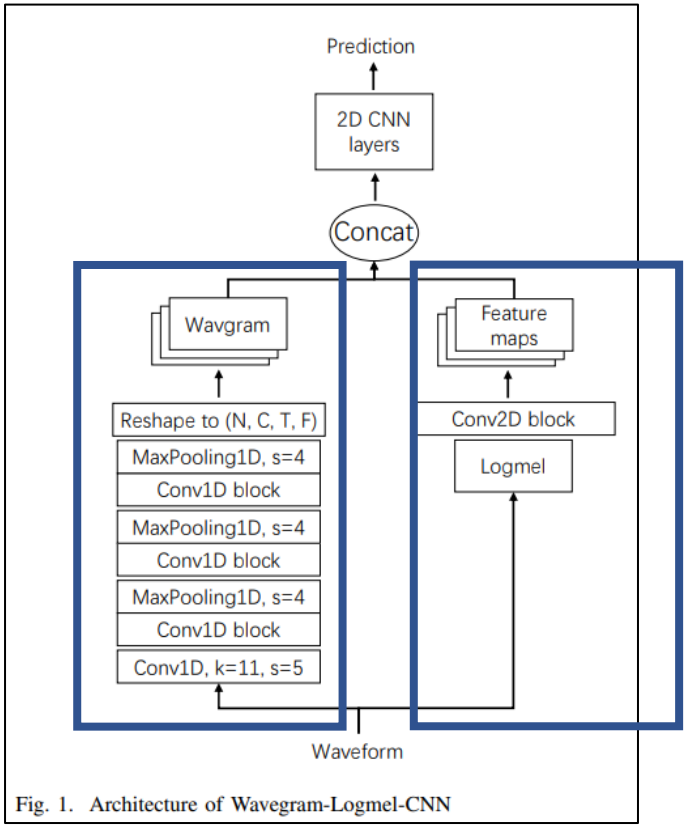


Fig. 1. Architecture of Wavegram-Logmel-CNN

スペクトログラム（Log-mel スペクトログラム）を用いて画像分類を行う

Youtubeから収集した1,934,187のオーディオファイルを用いて、632種類に分類するように学習。

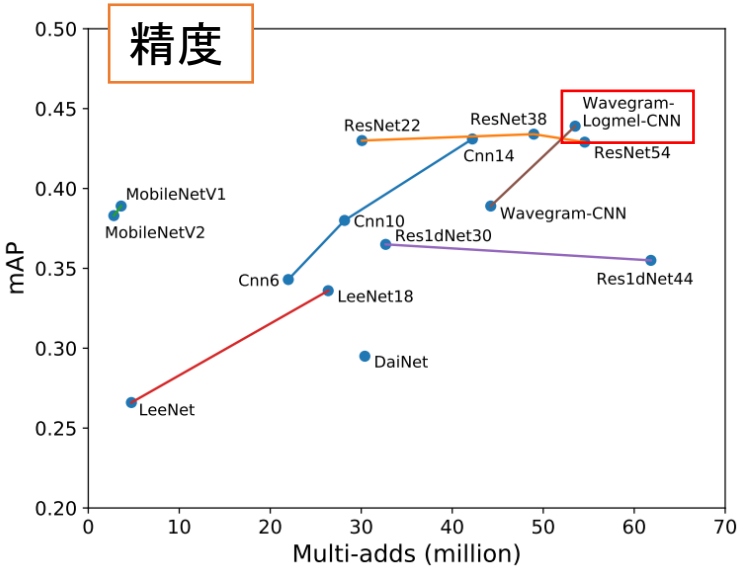
PANNsによる環境音分類について

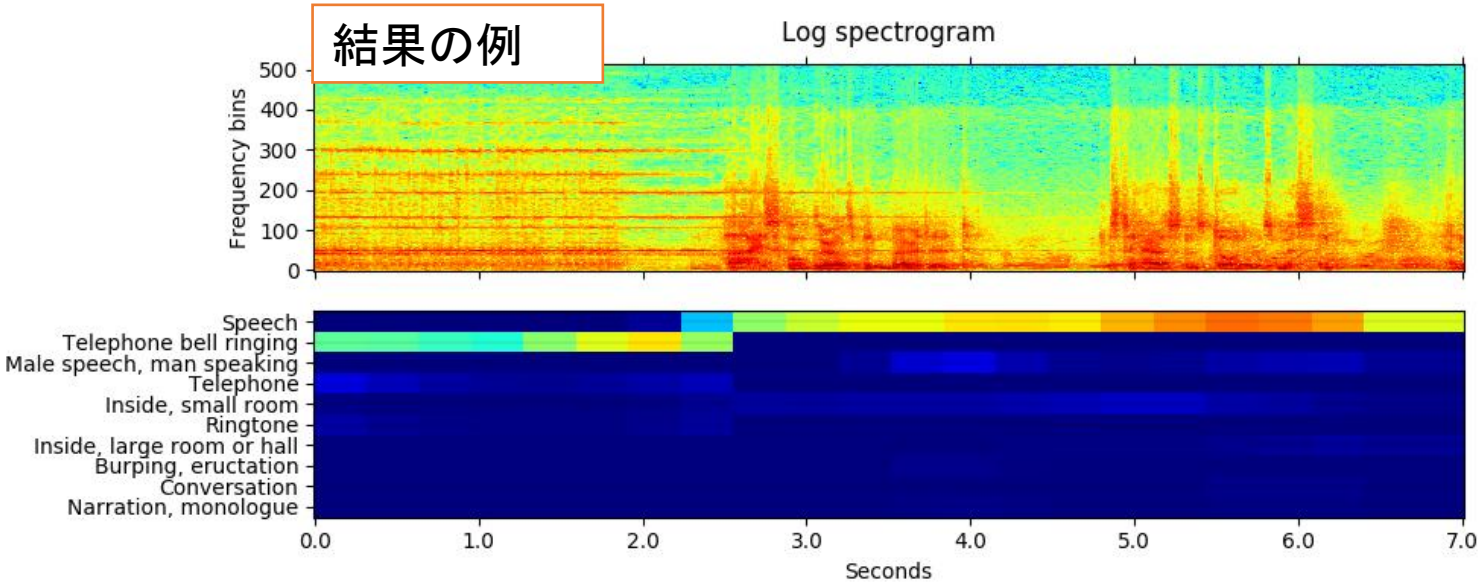
Large-Scale Pretrained Audio Neural Networks for Audio Pattern Recognition

- パラメータ数は多いが、判定の精度はほかのモデルよりも良い。
- 音声を1秒程度に分割して検知することで、その場の状況の時間進展が把握できる。

本研究では、このモデルを用いて検討を行う

学習は、文献のモデルを初期値として取り扱った。





対象とする機器と音測定

3台の回転機器それぞれを単独で運転し、6種類の運転時の音を録音した。

1) ターボポンプ



状態：正常、異常

↑ 内輪きずに模擬き裂を付与

2) 軸受試験機



状態：正常

3) 容積式ポンプ

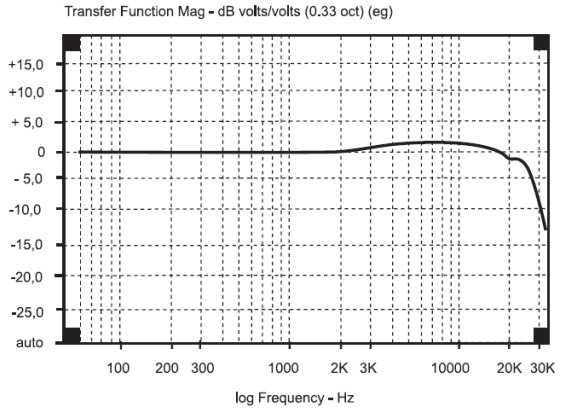


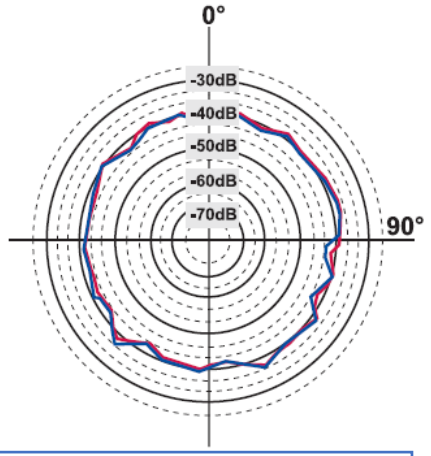
状態：正常（10Hz、20Hz、30Hz）

↑ 運転条件を変更

サンプリングレート 44.1kHz
マイク：無指向性マイク







それぞれの機器の音を学習させ、分類が可能かどうかを検討した。

2台稼働時（正常のみ）の音も含めて学習する

入力データを10秒間とする

カテゴリーとラベル

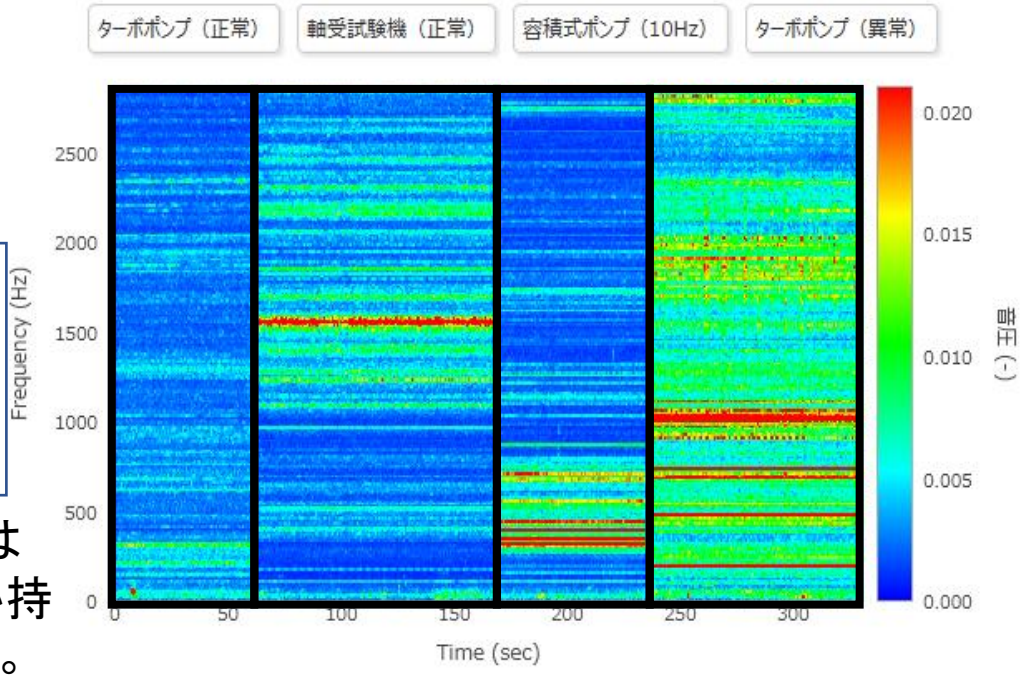
1. ターボポンプ（正常）：00
2. 軸受試験機（正常）：01
3. 容積式ポンプ（10Hz）：02
4. 容積式ポンプ（20Hz）：03
5. 容積式ポンプ（30Hz）：04
6. ターボポンプ（異常）：05

＋

7. 00+01	11. 01+02	14. 02+03
8. 00+02	12. 01+03	15. 02+04
9. 00+03	13. 01+04	16. 03+04
10. 00+04		

※単独音の合成により生成

ラベル：　00　01　02　05



- 検討当初、単独音のみでの学習では異常音などの特徴周波数をいくつか持つ音の判定が難しいことが分かった。
- 複雑な音（特徴周波数が複数含まれる）でも判断できるように合成音も使って学習する。

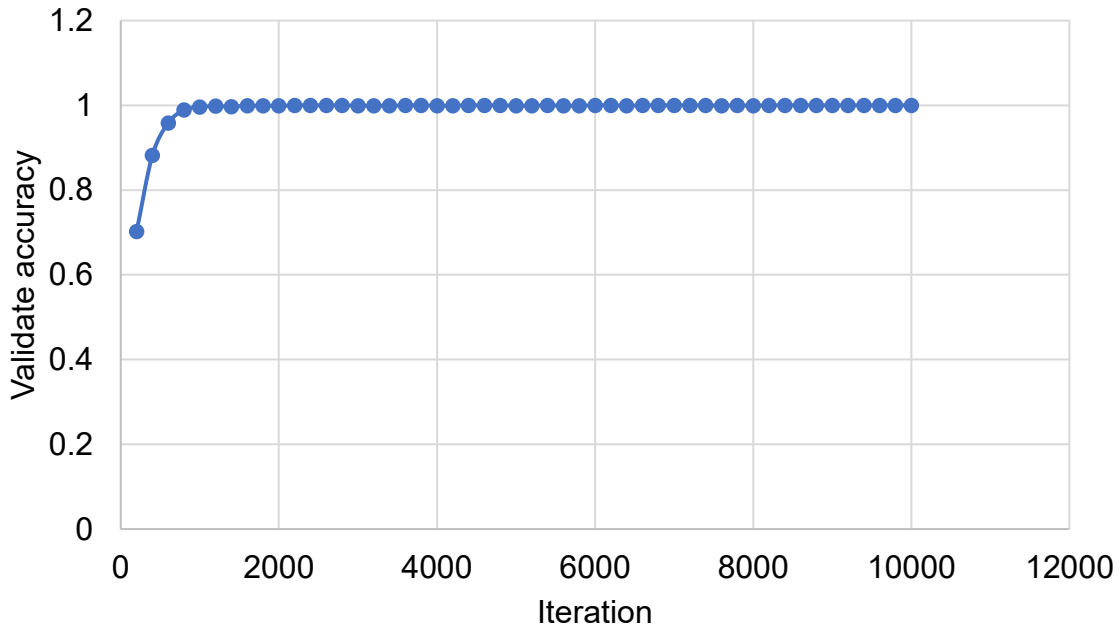
学習・検証に用いる

それぞれのカテゴリーの音について、10秒間のデータをランダムに1000個を切り出す

学習および検証結果

検証結果

反復回数が1000を超えると、検証用データに対する正解率は1となる。



各機器の単独音と2台稼働時の音を学習させたモデルについても、学習させた条件では状況の識別は精度良く可能である

3台稼働時の音の検知結果

ランダムに切り取った4個のデータに対して、判定を行った

入力合成音	判定結果①	判定結果②	判定結果③	判定結果④
・ターボポンプ（正常）:00 ・軸受試験機（正常）:01 ・容積式ポンプ（10Hz）:02	00+01: 0.971 00+02: 0.546 01+02: 0.008	00+01: 0.995 00+02: 0.068 01+02: 0.006	00+01: 0.924 00+02: 0.636 01+02: 0.014	00+01: 0.976 00+02: 0.243 01+02: 0.014
・ターボポンプ（正常）:00 ・軸受試験機（正常）:01 ・容積式ポンプ（20Hz）:03	00+01: 0.990 00+03: 0.450 01+03: 0.001	00+01: 0.898 00+03: 0.816 01+03: 0.001	00+03: 0.820 00+01: 0.812 01+03: 0.003	00+01: 0.995 00+03: 0.053 01+03: 0.001
・ターボポンプ（正常）:00 ・軸受試験機（正常）:01 ・容積式ポンプ（30Hz）:04	00+04: 0.906 00+01: 0.710 01+04: 0.002	00+04: 0.812 00+01: 0.338 01+04: 0.068	00+01: 0.907 00+04: 0.730 01+04: 0.010	00+01: 0.983 00+04: 0.311 01+04: 0.041

05：ターボポンプ（異常）

数値は判定の確率を示す

3台稼働させた時について、ラベルを切り分けて考える必要があるが、検知は可能である

異常音の検知結果

ランダムに切り取った4個のデータに対して、判定を行った

03：容積式ポンプ（20Hz） 04：容積式ポンプ（30Hz）

入力合成音	判定結果①	判定結果②	判定結果③	判定結果④
・ターボポンプ（正常）：00 ・ターボポンプ（異常）：05	05: 1.000 01+03: 0.000 00+01: 0.000	05: 1.000 01+03: 0.000 00+03: 0.000	05: 1.000 01+03: 0.000 01+02: 0.000	05: 1.000 01+03: 0.000 01: 0.000
・軸受試験機（正常）：01 ・ターボポンプ（異常）：05	03+04: 0.923 05: 0.771 00+01: 0.000	03+04: 0.919 05: 0.587 00+01: 0.000	03+04: 0.977 05: 0.408 00+01: 0.000	03+04: 0.904 05: 0.623 00+01: 0.000
・容積式ポンプ（10Hz）：02 ・ターボポンプ（異常）：05	05: 0.998 00+02: 0.099 02+04: 0.000	05: 0.999 00+02: 0.211 02+04: 0.000	05: 0.994 00+02: 0.082 02+04: 0.002	03+04: 0.997 05: 0.052 00+01: 0.000

数値は判定の確率を示す

- 軸受試験機（正常）が動いている場合、容積式ポンプが動いていると検知されてしまう。
  - → 軸受試験機（正常）とターボポンプ（異常）の合成音が容積式ポンプに似ているため
- ラベル05は高い確率で検知されているので、2台稼働しても異常の検知は可能。

学習データに異常データを含めない場合の検討

カテゴリーとラベル
1. ターボポンプ（正常）：00
2. 軸受試験機（正常）：01
3. 容積式ポンプ（10Hz）：02
4. 容積式ポンプ（20Hz）：03

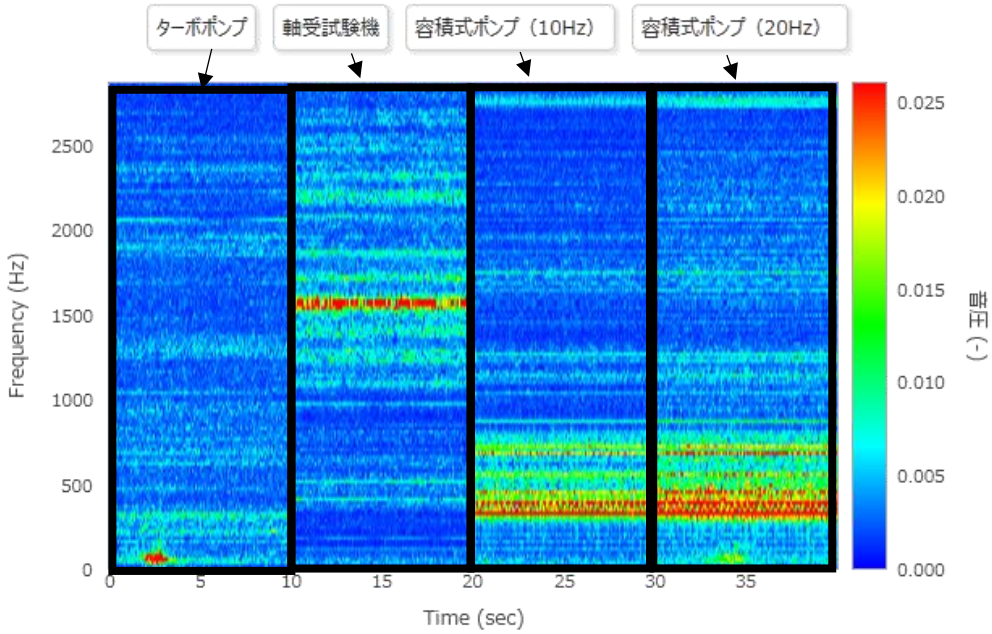
＋

5. 00+01　8. 01+02
6. 00+02　9. 01+03
7. 00+03　10.02+03

※単独音の合成により生成

入力データを10秒間とする

ラベル：　00　01　02　03



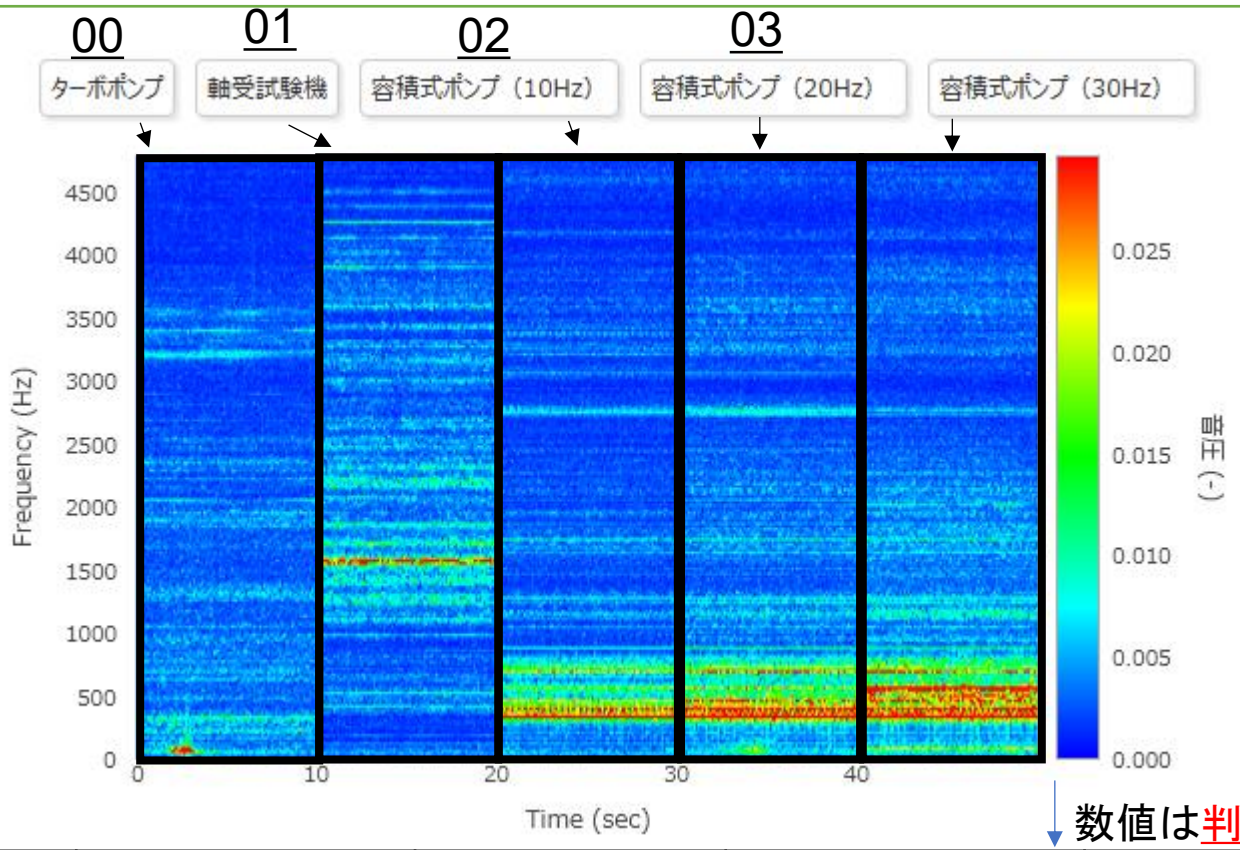
それぞれのカテゴリーの音について、10秒間のデータをランダムに1000個を切り出す

学習・検証に用いる

➡ 異常データを学習データに含めずに学習を行って、異常検知する。

学習結果：検証用データ（正常音）に対する正解率　→100%

学習データにない容積式ポンプ（30Hz）の検知結果

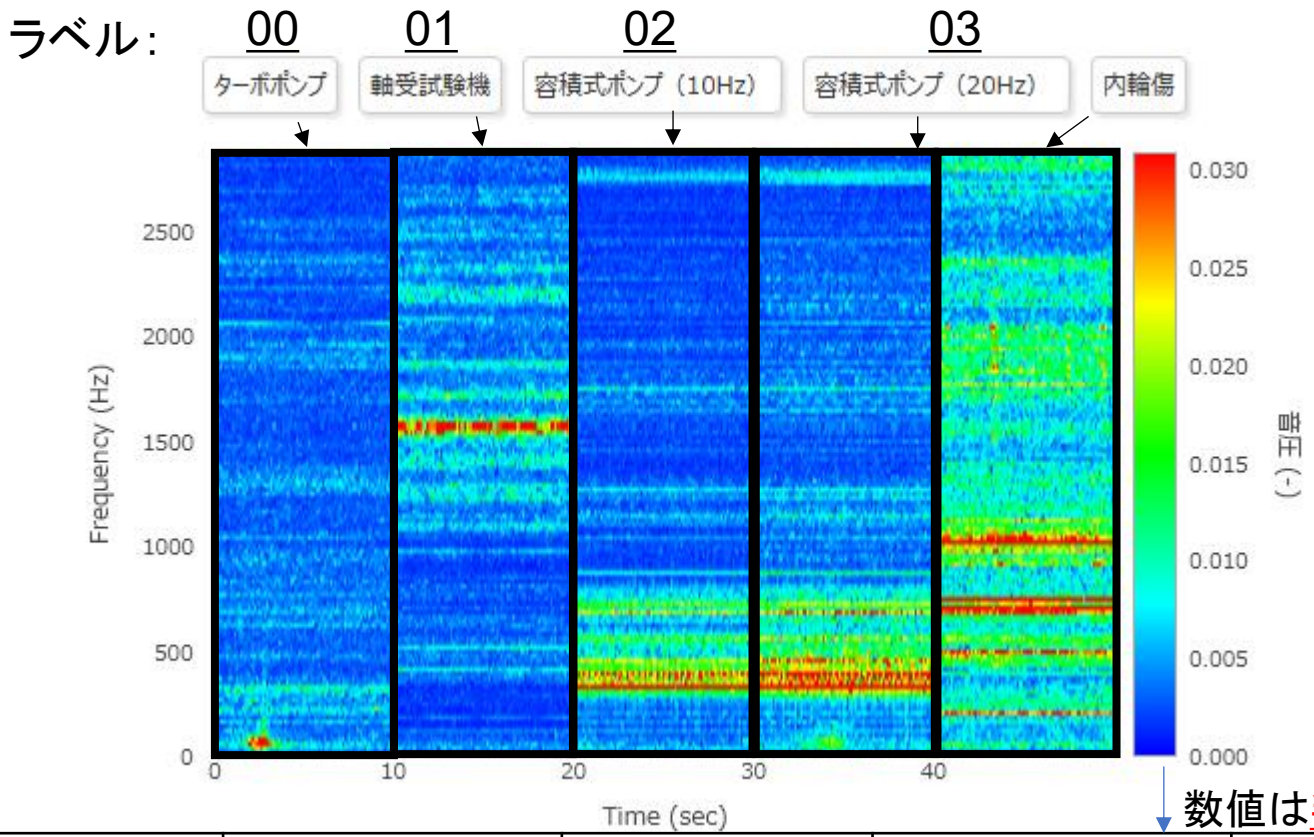


数値は判定の確率を示す

判定結果①	判定結果②	判定結果③	判定結果④	判定結果⑤
03: 0.993 02+03: 0.377 01+03: 0.000	03: 0.933 02+03: 0.602 01+03: 0.000	03: 0.823 02+03: 0.802 01+03: 0.000	03: 0.921 02+03: 0.639 01+03: 0.000	03: 0.968 02+03: 0.398 01+03: 0.000

◆ 似ている正常音（容積式ポンプ20Hz）のラベルが出力される。
◆ 容積式ポンプ（10Hz）とは区別できている。

異常音（ターボポンプ内輪傷）の検知結果



数値は判定の確率を示す

判定結果①	判定結果②	判定結果③	判定結果④	判定結果⑤
01+02: 0.029 02+03: 0.001 00+02: 0.000	01+02: 0.038 02+03: 0.000 00+02: 0.000	01+02: 0.026 02+03: 0.000 00+02: 0.000	01+02: 0.038 02+03: 0.000 00+02: 0.000	01+02: 0.020 02+03: 0.000 00+02: 0.000

- どのラベルにも該当しないという結果となった。
- これにより異常（正常にない状態）の検知が可能である。

まとめ

以下の項目の検討を進めた。得られた知見を示す。

- 時系列データを取り扱えるディープなネットワークを有するAIモデルの調査と異常検知の可能性の検討
  - ✓ スペクトルグラム画像と生波形の両方から特徴を抽出して、精度よく状況を分類するモデルを利用した。
- 時系列データ（音）を用いて、調査で得られたAIモデルの性能評価
  - ✓ 単独音のみを学習したモデルについて、単独音のみが鳴っている状況の識別は精度よく可能である。
  - ✓ 単独音とそれらの組合せ音（2種類）を学習させることで、複数の機器が稼働する状況の識別の精度が向上した。
  - ✓ 異常音が鳴っているときの異常検知は精度よく可能である。
  - ✓ 学習データに異常音を含めない場合、異常音が正常音と似ていなければ、どのラベルにも該当しないという結果が得られる。

どのラベルにも該当しないという結果を用いて異常検知が可能。
異常音データが蓄積されたら、学習を実施してラベル付けを行う。