日本保全学会 第18回学術講演会 C-1-4-5

2022年7月14日
京都大学

AIを用いた音源分離手法による音響診断の高度化

○中筋 俊樹、園部 経彰、角皆 学、宮 達夫（株式会社IIU）

※本講演は、
日本保全学会 状態監視技術高度化に関する調査検討分科会(CMT分科会)
第8フェーズ：2020/2021年度
（広域モニタリングによる異常の早期検知技術の開発及びAIの保全への活用の調査・検討）
の検討成果です。

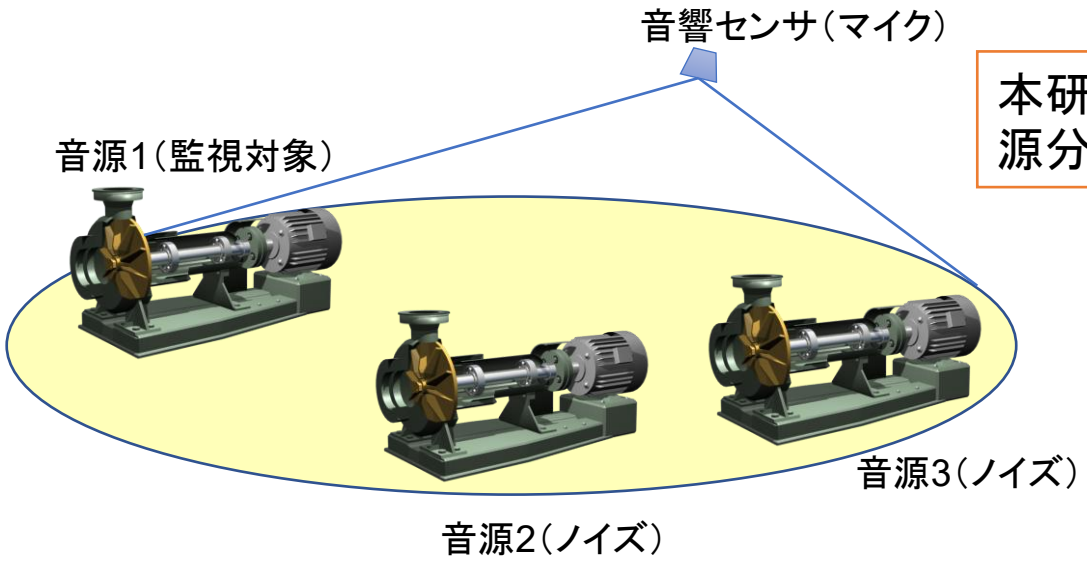
株式会社IIU
〒110-0008
東京都台東区池之端2-7-17井門池之端ビル7F
TEL : 03-5814-5350
FAX : 03-3827-0682
URL : http://www.iiu.co.jp

保全学会CMT分科会第8フェーズにおけるIIUの研究概要

これまで、設備の効率的な監視技術として音響診断の適用性を検討してきた。

第8フェーズ（2020年度 - 2021年度）では、音響診断の実機環境での適用性を向上させることを目的とし、正常時に様々なばらつきやノイズが含まれる場合において、AI・機械学習を用いることで機器の異常検知精度の向上を試みる。

特に、近年進歩が著しいAIによる音源分離技術について、設備の音響診断への適用性を検証し、音響診断の高度化を検討する。



本研究では、マイクで収録した音から音源分離AIによるノイズ除去を検討した

実施項目

1. AIモデルによる音源分離手法を調査
2. 選定された手法を用いたノイズ分離の効果検証
3. 異常診断の高度化の検討（LSTMによる異常検知）

音源分離AIモデルの調査結果

Défossez, Alexandre, et al. "Music source separation in the waveform domain." arXiv preprint arXiv:1911.13254 (2019).

Source to Distortion Ratio → 大きいほど分離精度が良い

Test data:
MusDB dataset (150 songs)

$$\mathrm{SDR} = 10\log_{10}\frac{\|s\|^2}{\|\hat{s}-s\|^2}$$

s: 実際の音源信号
$\hat{s}$: 推定した音源信号

Architecture	Wav?	Extra?	Test SDR in dB All	Drums	Bass	Other	Vocals
IRM oracle	✗	N/A	8.22	8.45	7.12	7.85	9.43
Wave-U-Net	✓	✗	3.23	4.22	3.21	2.25	3.25
Open-Unmix	✗	✗	5.33	5.73	5.23	4.02	6.32
Meta-Tasnet	✓	✗	5.52	5.91	5.58	4.19	6.40
Conv-Tasnet†	✓	✗	5.73 ±.10	6.02 ±.08	6.20 ±.15	4.27 ±.03	6.43 ±.16
DPRNN	✓	✗	5.82	6.15	5.88	4.32	6.92
D3Net	✗	✗	6.01	**7.01**	5.25	**4.53**	**7.24**
Demucs†	✓	✗	6.28 ±.03	6.86 ±.05	**7.01** ±.19	4.42 ±.06	6.84 ±.10
Spleeter	✗	∼ 25k*	5.91	6.71	5.51	4.55	6.86
TasNet	✓	∼ 2.5k	6.01	7.01	5.25	4.53	7.24
MMDenseLSTM	✗	804	6.04	6.81	5.40	4.80	7.16
Conv-Tasnet††	✓	150	6.32 ±.04	7.11 ±.13	7.00 ±.05	4.44±.03	6.74 ±.06
D3Net	✗	1.5k	6.68	7.36	6.20	**5.37**	**7.80**
Demucs†	✓	150	**6.79** ±.02	**7.58** ±.02	**7.60** ±.13	4.69 ±.04	7.29 ±.06

*: each track is only 30 seconds, †: from current work, ††: trained without pitch/tempo augmentation, as it deteriorates performance.

U-Net:
左右対称の
Encoder–Decoder構造

Demucs:
Wave-U-Netの構造をベースとして、LSTMを追加

・Wavファイルが扱える
・追加学習により精度が向上

→ Wave-U-Net、Conv-Tasnet、Demucsを利用する

検討の対象とする機器の測定音のスペクトルグラム

1) ターボポンプ

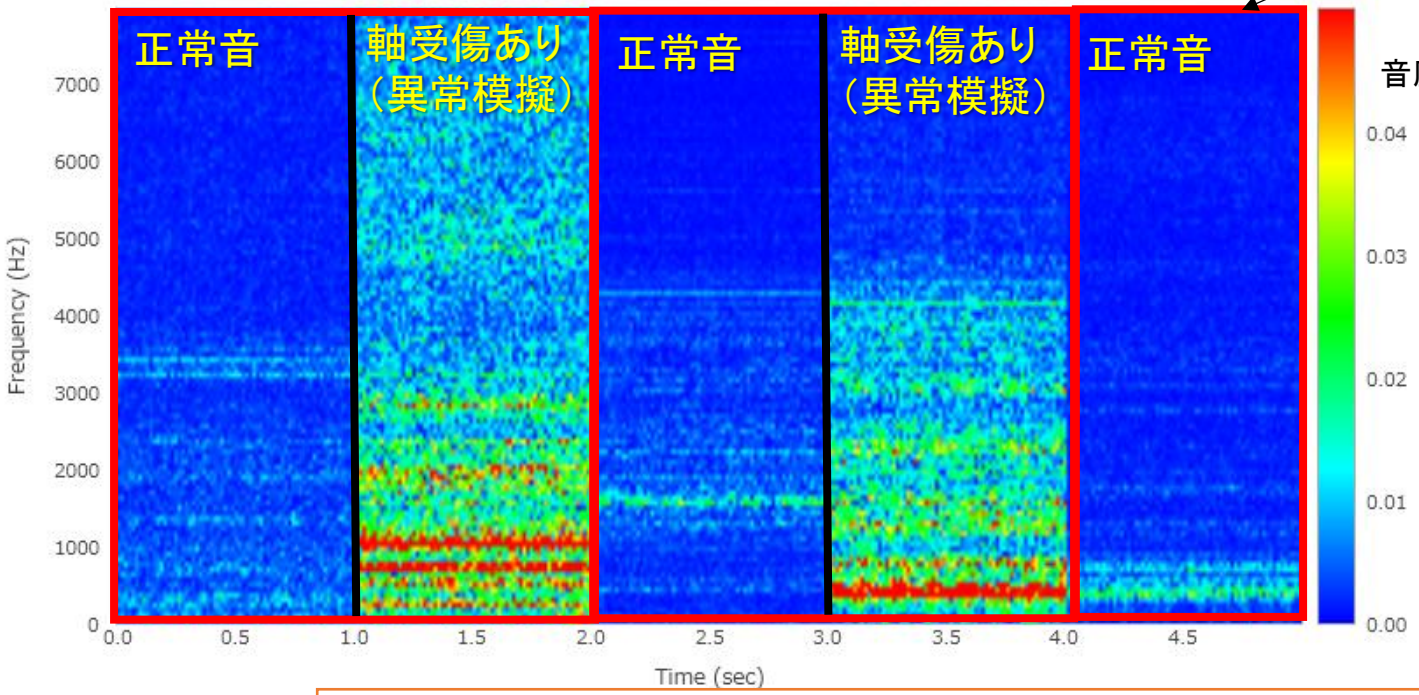


2) 軸受試験機



3) 容積式ポンプ





- 特徴的な周波数が存在する。
- 軸受傷なしと傷ありで、周波数の特徴が変わる
- 2)の傷ありと3)の傷なしは周波数特性が似ている

個々の機器の測定音を合成し、環境音として取り扱った

音源分離AIによるノイズ除去

監視対象の音を分離するように正常時のデータで学習すると、監視対象に異常が発生した際に正しく分離できないという問題がある。

そこで、監視対象ではなくノイズを学習し、ノイズを抽出・除去するという目的で音源分離AIを使用した場合の有効性を検証する。

監視対象の機器と、その近傍にノイズ源となる機器があるような環境を考える。

監視対象



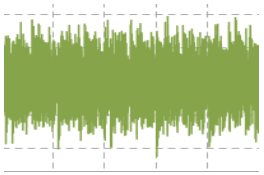
＋

ノイズ源となる機器



＝

監視対象＋ノイズ音　:環境音



測定データから1s間のデータをランダムにを抽出

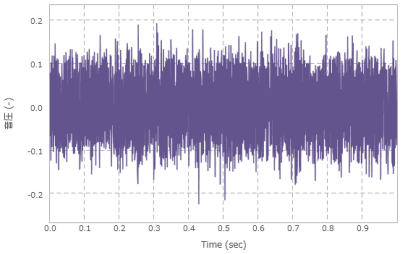
学習データ数　900
評価データ数　100

環境音からノイズ音を分離するよう学習

AIによるノイズ分離と状態診断の手順

監視対象＋ノイズ源の環境音

m

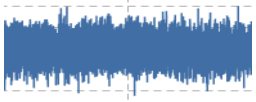


監視対象（正常音+ランダムノイズ）
※ノイズは異常状態を模擬するため

学習済みAIによる
ノイズ分離

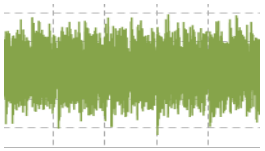
分離されたノイズ

$\hat{s}_2$



$\hat{s}_1 = m - \hat{s}_2$

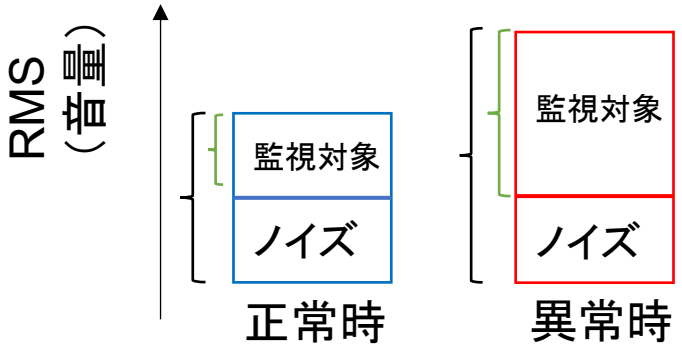
ノイズを除去した監視対象音$\hat{s}_1$



状態診断

状態診断

ここでは、音のRMS（二乗平均平方根）の正常時と異常時の比を指標として評価した。



正常時と異常時のRMS比について、
ノイズ分離前（合成音）
と
各AIモデルによる分離後（ノイズ除去後）
を比較した

音源分離AIによるノイズ除去効果検証①　ターボポンプ内輪傷

監視対象：ターボポンプ



ノイズ源： 容積式ポンプ



学習条件

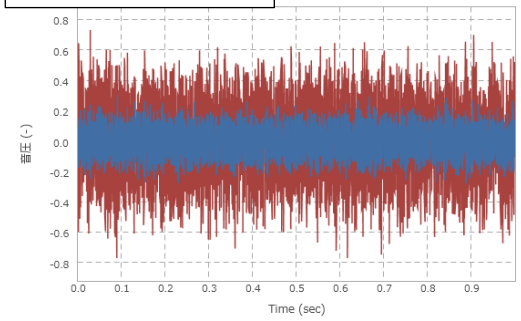
- 監視対象とノイズ源の単独運転音を信号レベル比1：1で混合
- 監視対象は正常時のデータのみを学習に使用（ランダムノイズを加える）
- ノイズ源の音を分離するよう学習

監視対象＋ノイズ源

監視対象の状態
正常
軸受内輪傷

正常時と異常時の
信号RMS比
（検証用100データの平均）

2.64



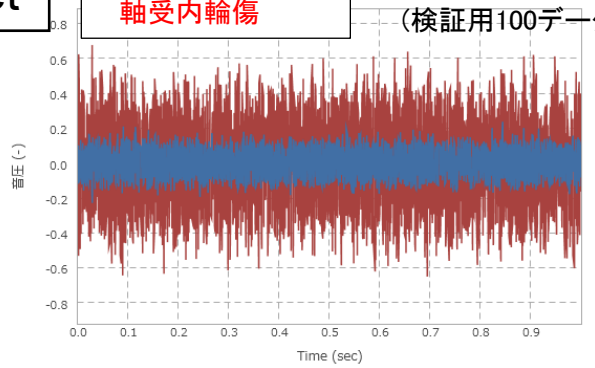
AIによる
ノイズ除去

Wave-U-Net

監視対象の状態
正常
軸受内輪傷

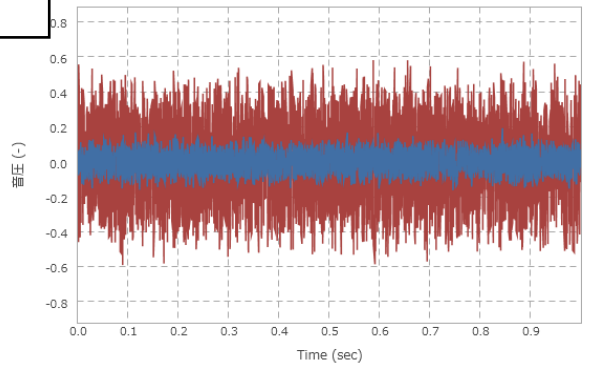
正常時と異常時の
信号RMS比
（検証用100データの平均）

3.08

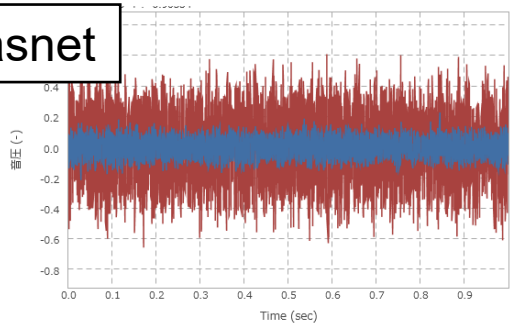


Demucs

4.01

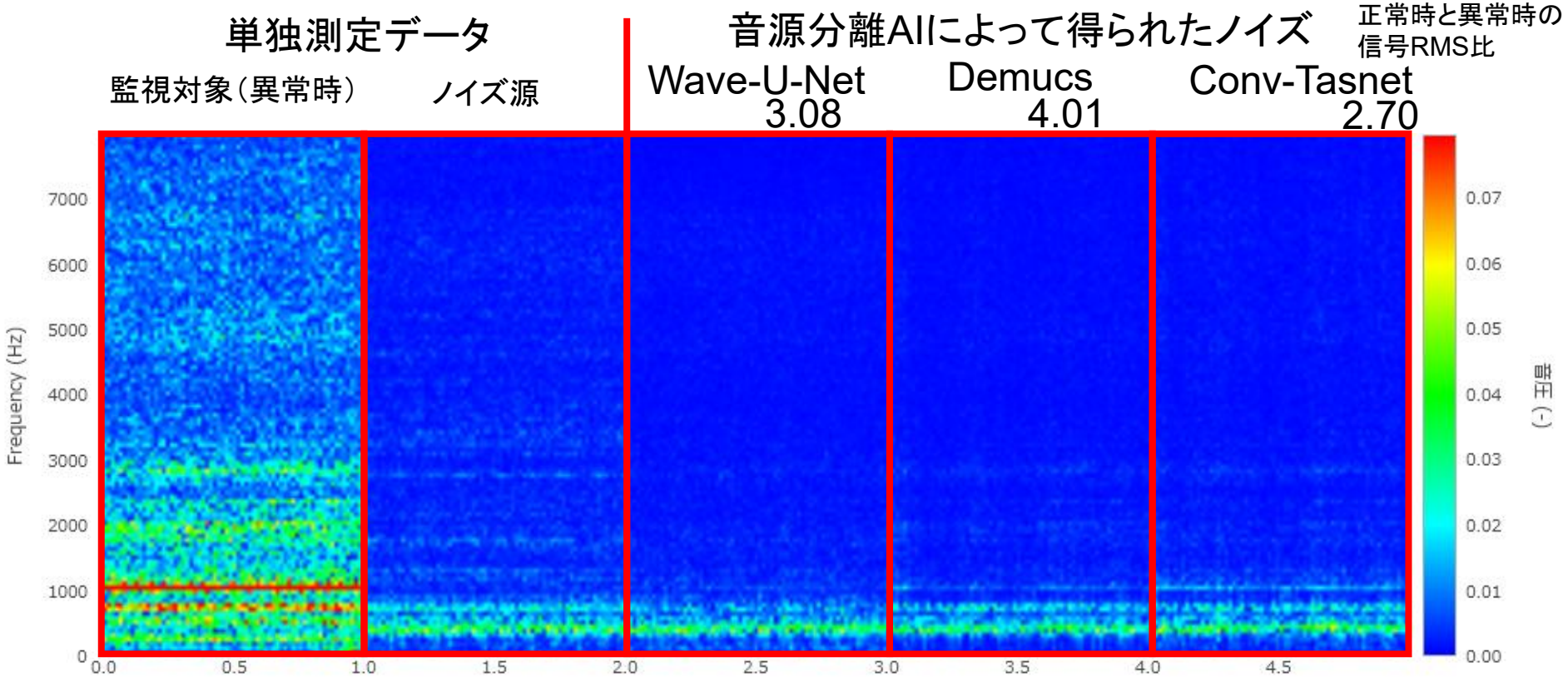


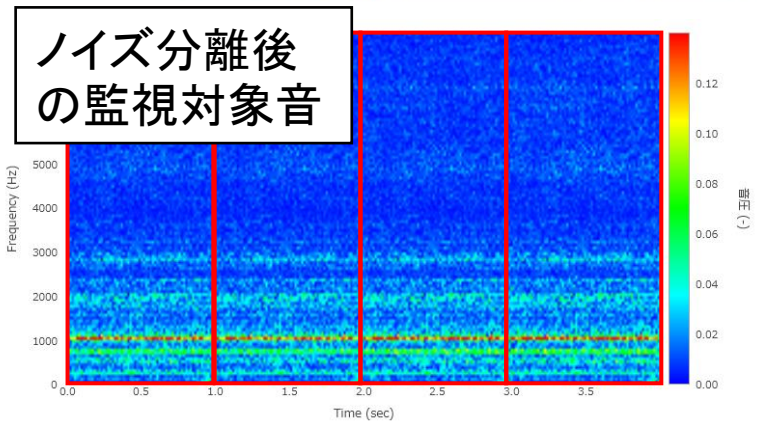
Conv-Tasnet



2.77

音源分離AIによるノイズ除去効果検証①　ターボポンプ内輪傷





- ノイズについて、Wave-U-Netよりも広い周波数帯域で分離に成功した。
- Conv-Tasnetは、異常音をノイズとして認識してしまっている。
  - → マスク処理による分離であるため、学習データ（正常音+乱数）にない帯域の音はマスクされずに出力されてしまう可能性がある。

<u>モデル構造の違いによる影響</u>

音源分離AIによるノイズ除去効果検証② 軸受試験機内輪傷

Wave-U-Netでのノイズ除去効果があまり見られなかった条件（ノイズ源と異常音の特徴周波数が近い）

監視対象：軸受試験機



ノイズ源： 容積式ポンプ



学習条件
- 監視対象とノイズ源の単独運転音を信号レベル比1：1で混合
- 監視対象が正常時のデータのみ使用
- ノイズ源の音を分離するよう学習

AIによるノイズ除去

ノイズ分離前の混合音における正常時と異常時の信号RMS比（検証用100データすべて）

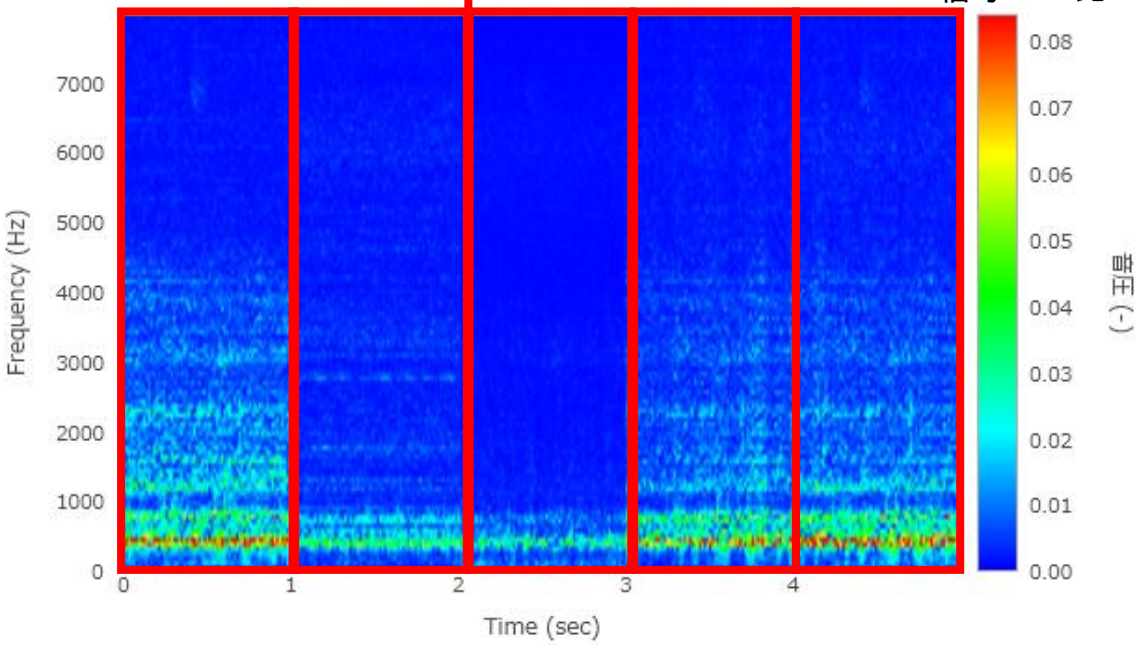
ノイズ除去前 1.55

単独測定データ

監視対象（異常時） ノイズ源

音源分離AIによって得られたノイズ

Wave-U-Net 1.91　Demucs 1.13　Conv-Tasnet 1.47　信号RMS比

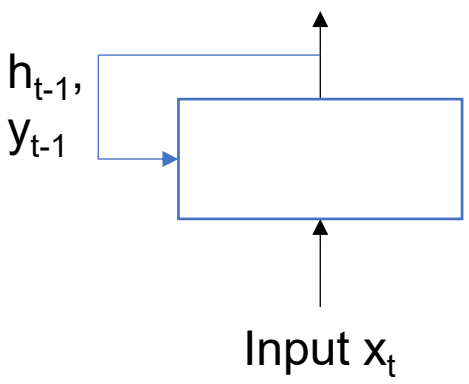


- DemucsおよびConv-Tasnetでは、信号RMS比がノイズ除去前より下がった。
- ノイズ源と異常音の特徴周波数が近いため、異常音がノイズとして除去されてしまったと考えられる。

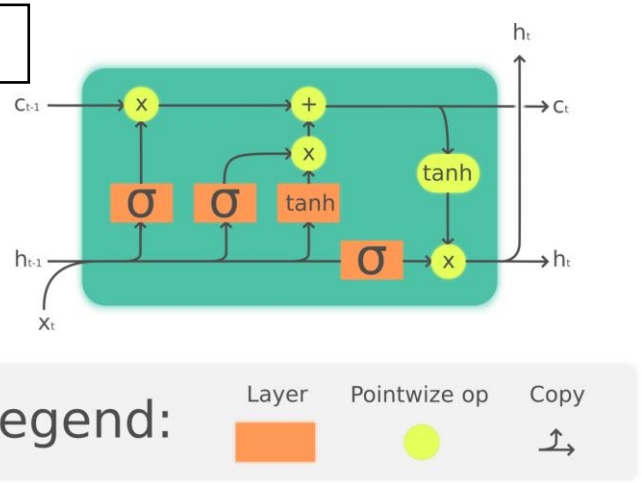
異常診断の高度化 ～LSTMの活用～

- Recurrent Neural Network（RNN, 再起型ニューラルネットワーク）はニューラルネットワークの出力層や中間層の出力を次のステップの時系列データの処理時にも利用するネットワーク。時系列データの取り扱いに有効。
- LSTM（ Long Short Term Memory、長・短期記憶）ネットワークは、RNNで不得意であった長期の記憶を、忘却ゲートを追加して取り扱えるようにしたものである。

RNN



LSTM



自己回帰（AR, autoregressive）モデル：

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_n y_n$$

を高度化したもの

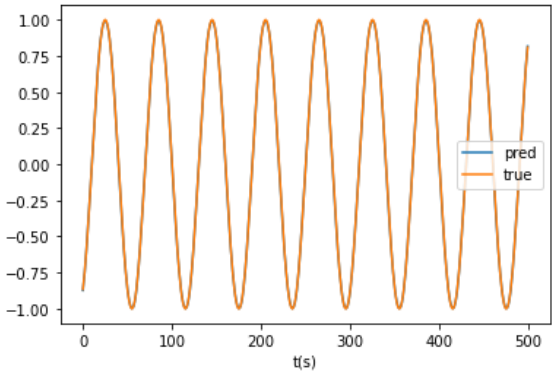
ϕはパラメータ

LSTMによる時系列データの変化に対する異常検知の試行計算

学習

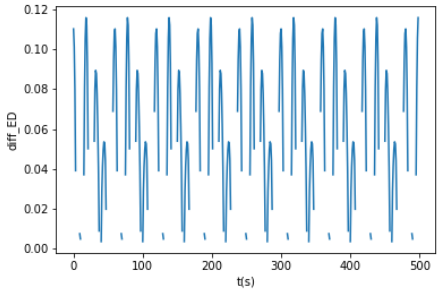
SIN波（freq=60[Hz]）のみを学習（データ長は50点）→LSTMにより51点目を予測

学習結果



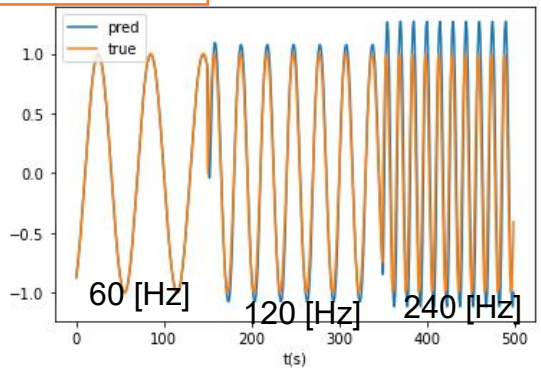
正解値と予測値の差の絶対値　を
異常度　とする

異常度

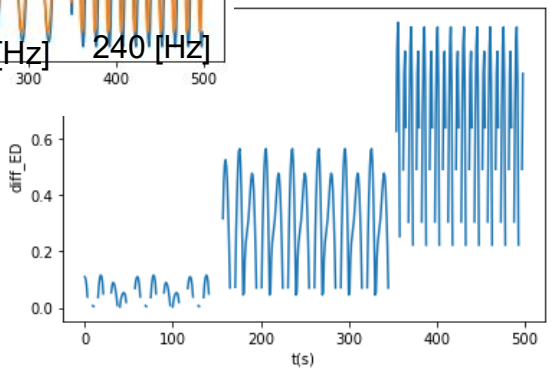


異常模擬

周波数が段階的に変化

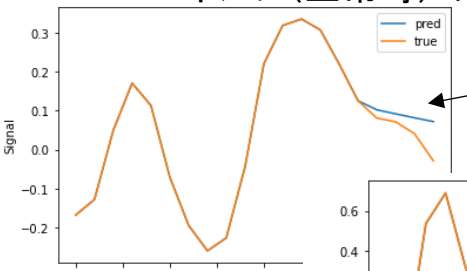


異常度

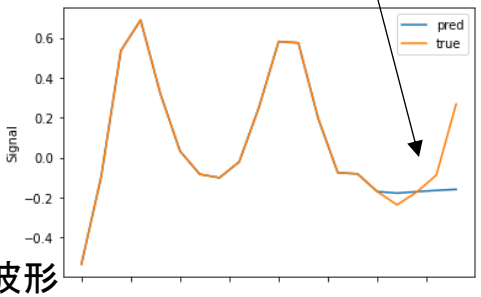


正常時のデータのみで学習した
「LSTMによる予測」と「実測」の乖離具合による異常検知を検討した

ポンプ（正常時）の音波形



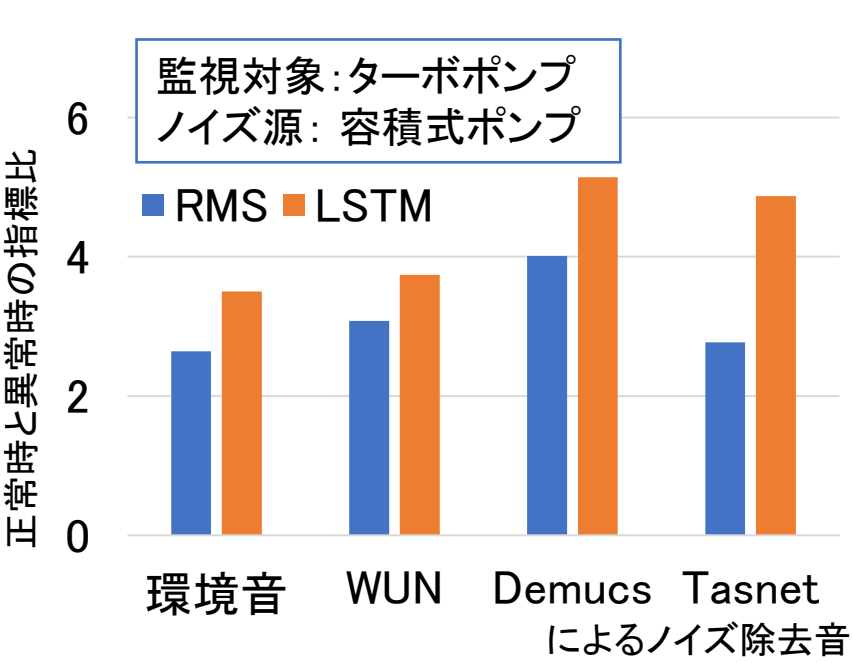
青：LSTMにより3点先まで予測
オレンジ：実測



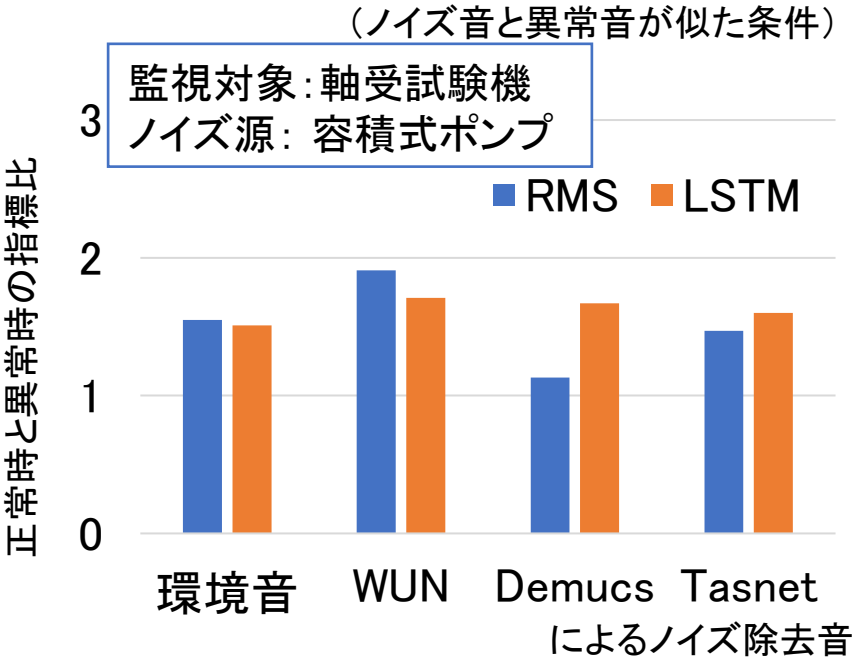
ポンプ（異常時）の波形

LSTMによる異常検知の検討

- ✓ 4000点の時系列データから次の1点のデータを予測するモデルをLSTMを用いて構築した。(学習データは正常時のデータのみ)
- ✓ 異常検知の指標は、「LSTMによる予測」と「実測」の差の絶対値とした。



(ノイズ音と異常音が似た条件)



※LSTMは最も良い条件(予測点数、隠れ層数=5)の結果を表示している

- ◆ 環境音・ノイズ除去音の両方でLSTMを用いることで異常検知の感度は向上した。
- ◆ ノイズ音と異常音が似た条件について、RMSの比では異常検知の感度が低下する場合でも、LSTMであれば異常検知の感度は向上した。

まとめ

音源分離AIを用いた環境音からのノイズ除去

- 調査を行い、楽曲の音源分離性能がよく、使い勝手の良い、3種類（Wave-U-Net、Demucs、Conv-Tasnet）を選定した。
- 異常検知のためのノイズ除去効果を検証したところ、ノイズ源と異常音の周波数特性が似ていなければ、Demucsによるノイズ除去が有用であることが分かった。

LSTMによる異常検知

- 4000点の時系列データから次の時刻のデータを予測するモデルをLSTMを用いて構築した。（学習データは正常時のデータのみ）
- 異常検知の指標は、LSTMによる予測と実測の差の絶対値とした。
- 環境音（合成音）について、従来の指標（正常時と異常時のRMS値の比）を用いるよりも、LSTMによる異常検知指標比のほうが有効であることが確認された。
- 音源分離AIによるノイズ除去後の音に対して、RMS値の比による異常検知が困難な場合でも、LSTMを利用することで異常検知できることが明らかになった。