音響マルチパラメータを用いたニューラルネットワークによる自動車走行音の特徴量抽出

Feature Extraction of Car Driving Sound Using Neural Network with Acoustic Multi-Parameters

精密工学専攻　14 号　大島　遥汰
Yota Oshima

1. はじめに

近年，企業において音の特徴を活用したサウンド・ブランディングに対するニーズが高まっている．現状では，主観に基づきサウンドデザインされることが多く，音響物理特性に基づくサウンドデザインがなされた事例は少ない．その中でも，自動車の開発においてエンジン，プラットフォームや部品などが共通化される場合があり，車種が異なる場合でも共通する音の要素が存在する可能性がある．各メーカで設計思想が同一の製品間の音に含まれる共通点はメーカのサウンドアイデンティティの確立に繋がる．

既報[1]では，時間変動成分の識別精度の高いニューラルネットワークとして，畳み込みニューラルネットワーク（CNN：Convolutional Neural Network）に準再帰型ニューラルネットワーク（QRNN：Quasi-Recurrent Neural Network）を組み合わせた「CNN-QRNN」を構築し，U-Net を用いて暗騒音を除去した疑似エンジン近接音の時間周波数解析画像を識別した．そして，時間周波数解析の画像に共通する特徴を抽出する手法として，勾配加重クラス活性化マッピング（Grad-CAM：Gradient-weighted Class Activation Mapping）に構造的類似性指数（SSIM：Structural SIMilarity）を適用している．これにより，時間周波数解析画像を分類し，共通する特徴量と相違する特徴量を抽出できることが示された．しかし，周波数特性だけでの特徴量抽出は，エンジンの気筒数が異なる同一メーカの車種間の共通特徴量抽出は困難である．

そこで既報[2][3]では，音響物理特性であるスペクトログラム，キャリア周波数と振幅変調周波数の相関，心理音響メトリクスを音響マルチパラメータと定義し，それらの解析画像を CNN に入力し，各パラメータの解析結果を統合してデータを分類する機械学習モデルを複数構築し，性能検証を行った．次に，最も性能の良いモデルを用いて複数の自動車走行音の分類を行い，協力ゲーム理論のシャープレイ値（Shapley Value）を機械学習に応用した SHapley Additive exPlanations（以下，SHAP）[4]を用いて，パラメータごとの分類貢献度の合計値を算出した．その結果に基づき，貢献度が高かったパラメータに対して各要素の SHAP を確認することで，各音源の特に特徴的な情報を抽出可能な XAI（Explainable Artificial Intelligence）を構築した．しかし，スペクトログラム画像内の特徴量を SHAP で可視化した際に，元画像内に回転次数成分，共振成分，暗騒音が混在しており，AI が重要と捉えた特徴量が音のどの成分であるか解釈が容易ではないという問題があった．

そこで本研究では，ニューラルネットワークが回転次数成分以外の特徴量を抽出するため，機械学習を用いてスペクトログラム画像から回転次数成分を除去した画像生成手法を検討する．次に，スペクトログラム画像の代わりに，回転次数成分除去スペクトログラム画像と時間-回転次数画像を入力データとし，既報において音の特徴量抽出に最適であると認められた機械学習モデルを用いて，再度機械学習モデルの性能検証を実施する[5]．最後に，自動車走行音の分類を行い，主成分分析による特徴量の次元削減に基づくサウンドマップの作成および車種間において相違する特徴量の抽出を実施する[6]．

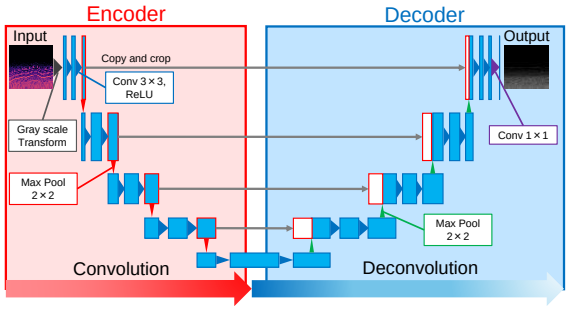


Fig. 1　U-Net architecture

2. U-Net を用いた回転次数成分除去

2.1 U-Net の概要

近年，ノイズが多くある画像や拡大により分解能が低下した画像，落書きされた画像などにおいて，CNN アーキテクチャを用いて復元を試みる事例が多数ある．そのようなニューラルネットワークの 1 つである U-Net は，CNN の中でも高い復元精度を誇る．

Fig. 1 に示すように，U-Net はエンコーダとデコーダからなるアーキテクチャであり，エンコーダ部分においては，入力された画像を何度か畳み込み，その画像の特徴量を抽出する．そして，デコーダ部分において，エンコーダによって抽出された特徴量を受け取り，通常の畳み込みの逆処理（逆畳み込み）を行い，入力画像と同じサイズのデータを出力する[7]．本研究では，U-Net をエンジンの回転次数成分除去に適用する．予め暗騒音と共振成分を学習させた U-Net に疑似車室内音のスペクトログラム画像を入力する場合，エンコーダ部分で共振成分と暗騒音の特徴量のみを抽出し，デコーダ部分で画像を元のサイズに復元することで，回転次数成分を除去し，共振成分と暗騒音のみの画像を推定する．

2.2 学習に使用するデータ

本研究では，Fig. 2 (a)のような擬似車室内音のスペクトログラム画像，および Fig. 2 (b) のような回転次数成分を除去した擬似音画像を用意し，その画像を U-Net に学習させることで，U-Net が暗騒音領域を予測し，元の画像から回転次数成分除去を実施できるようにする．

学習用データは，直列 4 気筒および直列 6 気筒のエンジンを想定して顕著な回転次数成分を設定し，回転次数成分の音圧や暗騒音の音圧および周波数勾配，共振成分の音圧および周波数はランダムに設定する．

2.3 結果および考察

Table 1 に訓練用データと検証用データの学習終了時の損

失を示す．ここで，訓練用データと検証用データとは，学習に使用するデータを分割して機械学習の訓練と検証に使用するデータであり，損失とは画像の予測値と真値とのずれを示す．本解析では，損失が非常に小さいことから，高い精度で画像の復元が実施されていると考えられる．

また，U-Net による出力画像の音圧レベルの正確性を評価するため，学習用データにおける，全出力画像と疑似音画像とのヒストグラム相関を算出する．ただし，Fig. 3(a)に示すように，出力画像には回転次数成分が存在していた領域において，回転次数成分の影が残存するため，Fig. 3(b)に示すように，画像の時間軸方向に微小なガウスフィルタをかけた画像を用いて，相関を求めることとする．

全出力画像と疑似音画像のヒストグラム相関の平均値は 99.9 %となり，対応するピクセルのほとんどで画素値が一致することが確認できる．また，Fig. 4 に学習用データ以外のスペクトログラム画像を入力した際の出力結果を示す．回転次数成分の影は残存しているものの，その音圧レベルが周辺の暗騒音の音圧レベルと同等となっていることから，回転次数成分を正確に除去できていると考えられる．

3. 機械学習モデルの特徴量抽出精度の検証

本章では，既報にて音の特徴量抽出に最適と示された機械学習モデルを用いて，入力する音響パラメータを 2 章で生成した画像に変更した場合においても，十分な特徴量抽出精度を有するか検証する．

3.1 使用するニューラルネットワーク

本研究で使用した音響マルチパラメータ機械学習モデルの概要を Fig. 5 に示す．解析して得られるデータをヒートマップ画像として CNN に入力することで，各データの特徴を際立たせ，得られた多角的な特徴量を基に音源を分類する．

CNN 部については既報と同様に，Modulation Spectrum の連続画像においては 3D ResNet を使用し，それ以外のパラメータ画像においては RepLKNet を使用する．

3.2 精度の検証に使用するデータ

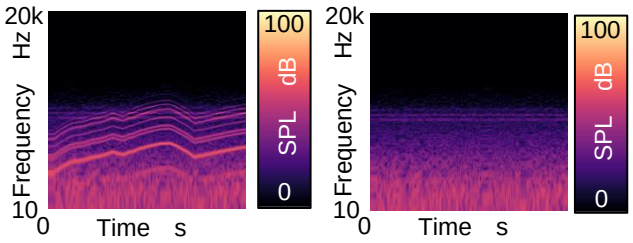
Table 2 に示すように，広帯域音，回転次数成分，共振成分の各条件を変更した 7 種類の疑似自動車加速音を使用する．各音源を解析し，Fig. 5 に示す 8 種類のパラメータ情報を得る．Sound 1 は基準となる回転次数成分が異なるため Time-order map が，Sound 2 は周波数勾配が異なるため Order-removed spectrogram が，Sound 3 は共振成分が存在するため Order-removed spectrogram と Prominence ratio が，Sound 4 は音圧が大きいため Loudness が，Sound 5, Sound 6, Sound7 は広帯域音が振幅変調を伴うため Modulation spectrum や Roughness, Fluctuation strength, Impulsiveness が特徴的なデータになるように，各パラメータを設定している．なお，画像や連続画像のデータをかさ増しするために，ランダムにガウシアンノイズを加えた画像を作成している．

3.3 学習・貢献度算出結果および考察

Table 3 に訓練用データと検証用データの学習終了時の損失と正解率を示す．損失は非常に小さく，正解率は訓練用データが 99.9 %，検証用データが 100 %と，高い精度で画像を識別していることが確認できる．

次に，Fig. 6 に音源ごとに正規化した，各音源における固有の特徴的なパラメータの貢献度を示す．この結果より，各音源において特徴付けしたパラメータの貢献度が，特徴的でないパラメータよりも大きい傾向があると認められた．

そして，単一パラメータ画像の SHAP について，2 音源間での比較を行う．音響パラメータ画像の SHAP について，2 音源間での比較を行う．Fig. 7 に，Fluctuation strength の貢献度が低かった Sound 4，Fig. 8 に，特徴的なパラメータが

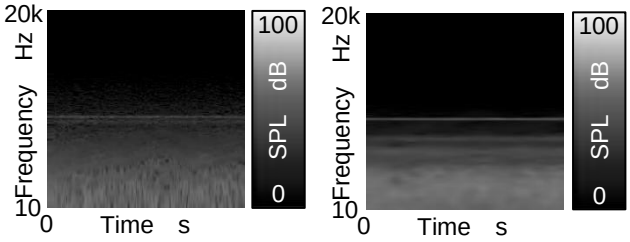


(a) Before order removed　(b) After order removed

Fig. 2 Spectrogram images for learning

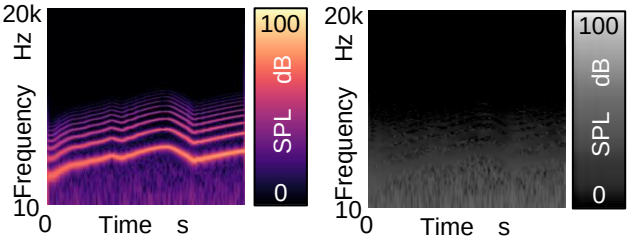
Table 1 Loss of train and validate

	Train	Validate
Loss	5.19×10^{-4}	5.32×10^{-4}



(a) Before Gaussian filtered　(b) After Gaussian filtered

Fig. 3 Order-removed spectrogram



(a) Before order removed　(b) After order removed

Fig. 4 Spectrogram images not for learning

Fluctuation strength であり，貢献度が音源内で 1 位であった Sound 6 の特徴量抽出結果を示す．Sound 4 においては，Fluctuation strength 画像全体における SHAP 値が小さいが，Sound 6 においては，広帯域音を振幅変調させているため，回転次数成分が存在しない領域における SHAP 値が顕著に大きいことが確認できる．

同様に，Fig. 9 に Sound 3 の Order-removed spectrogram の SHAP を示す．U-Net による回転次数成分除去前との画像比較より，回転次数成分が除去され，共振成分と暗騒音領域の SHAP 値が大きいことから，2 章での U-Net における回転次数成分除去による画像分離が有効であると考えられる．

以上より，音響マルチパラメータ機械学習モデルによる音源分類と SHAP による貢献度算出により，パラメータごとの貢献度算出の有効性，および単一画像内の特徴量抽出の有効性が認められた．また，U-Net における回転次数成分除去により，特徴量を解釈しやすくなったと認められた．

4. 自動車走行音の特徴量抽出

本章では，3 章にて十分な性能を有すると認められた機械学習モデルを用いて，自動車走行音の分類を行い，主成分分析による特徴量の次元削減に基づくサウンドマップの作成および車種間において相違する特徴量の抽出を実施する．

4.1 学習に使用するデータ

本研究では，自動車音源 CD[8]内のディーゼルエンジン車の

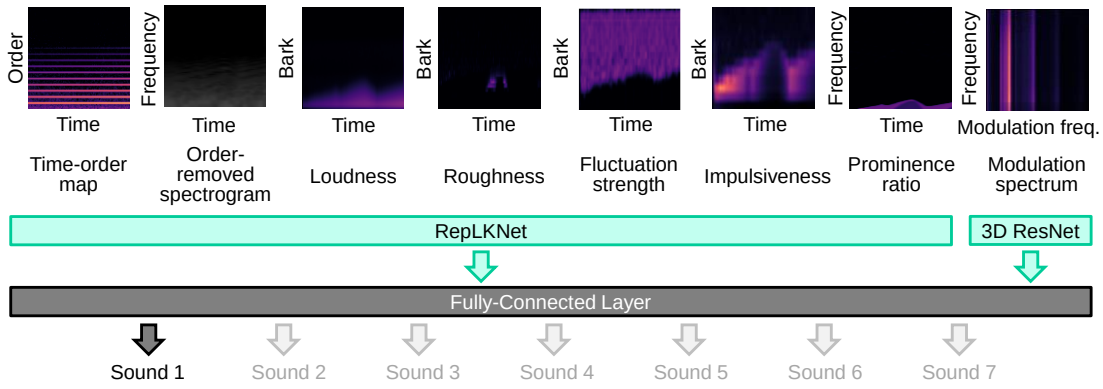


Fig. 5　Overview of multi modal neural network architecture

Table 2　Each sound information

Sound number		1	2	3	4	5	6	7
Broadband sound	Freq. gradient　dB / octave (Range : 10 - 20000 Hz)	-6	-8	-6				
	SPL　dB	80	80	80	100	80 (Before amplitude modulated)		
	Modulation wave	None	None	None	None	Sine	Sine	Sawtooth
	Modulation degree	None	None	None	None	1	1	1
	Modulation freq.　Hz	None	None	None	None	70	4	10
Rotational order	Base order number	3rd	2nd	2nd	2nd	2nd	2nd	2nd
	Base order max SPL　dB	80	80	80	80	80	80	80
Resonance component	Freq.　Hz / SPL dif.　dB	None	None	200 / +20	None	None	None	None

■ Characterized parameter

Table 3　Loss of train and validate

	Train	Validate
Loss	0.146	0.146
Accuracy　%	99.9	100

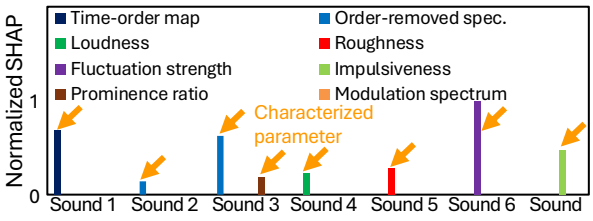
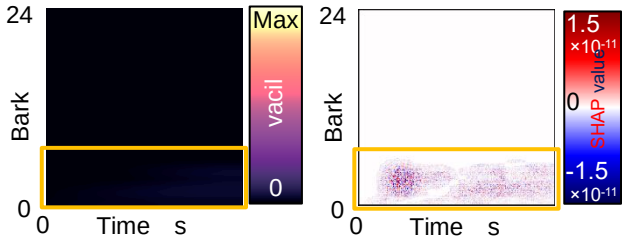
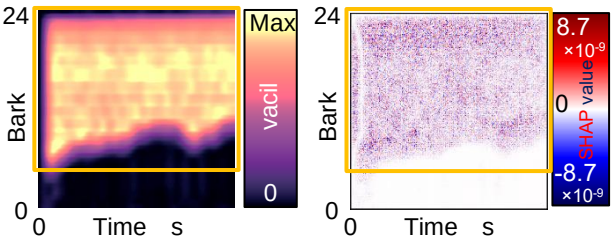


Fig. 6　Parameter contribution



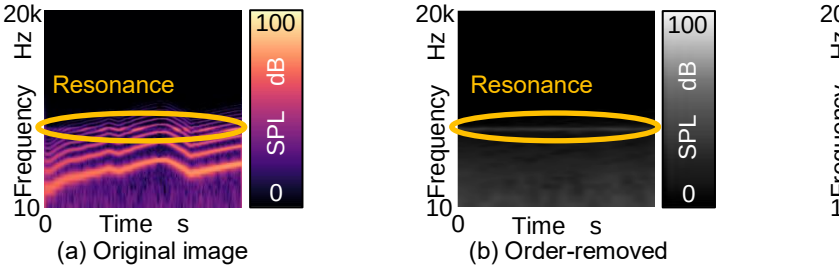
(a)　Original image　(b)　SHAP

Fig. 7　Fluctuation strength of sound 4



(a)　Original image　(b)　SHAP

Fig. 8　Fluctuation strength of sound 6



(a) Original image　(b) Order-removed　(c) SHAP

Fig. 9　Spectrogram of sound 3

加速音を，車格ごと（L_truck，M_truck，H_truck，Sedan）に分類する．加速時データは各車格 4 車種ずつ，車種ごとに 1 音源のみであることから，画像や連続画像のデータをかさ増しするために，ランダムにガウシアンノイズを加えた画像を作成する．

4.2　学習結果

Table 4 に訓練用データと検証用データの学習終了時の損失と正解率を示す．本解析では，正解率は訓練用データ，検

Table 4　Loss of train and validate

	Train	Valid.
Loss	0.047	0.047
Accuracy %	99.9	100

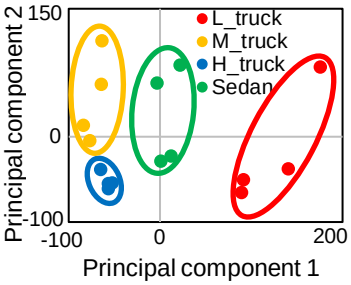


Fig. 10　Sound feature map by SHAP-PCA

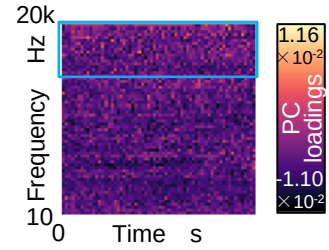


Fig. 11　Principal component 1 loadings of order-removed spectrogram

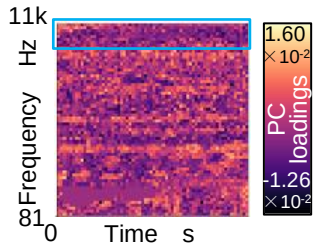
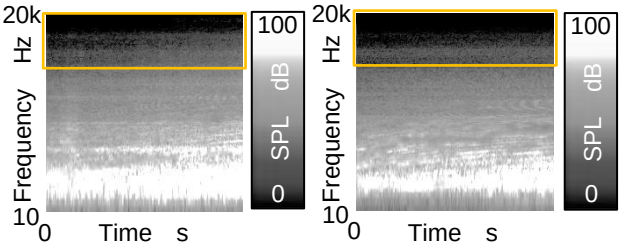
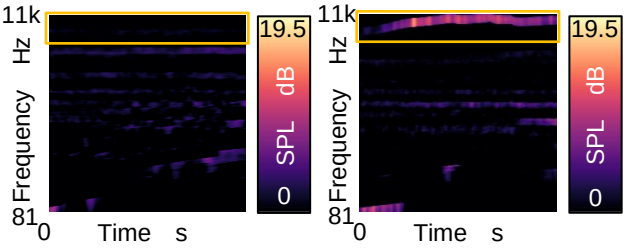


Fig. 12　Principal component 2 loadings of prominence ratio



(a)　L_truck_4　(b)　H_truck_4
Fig. 13　Order-removed spectrogram



(a)　M_truck_1　(b)　H_truck_2
Fig. 14　Prominence ratio

証用データともに約 100 %と，高い精度で画像を識別していることが確認できる．

4.3　SHAP 主成分分析結果および考察

SHAP により得られた自動車走行音の特徴は，一つの音響パラメータでは表現しきれず，特徴量の評価が難しいため，各音響パラメータの SHAP をダウンサンプリングしたデータを基に，主成分分析による特徴量の次元削減を実施する．参考文献[9]より，SHAP は機械学習の分類結果に基づく，各パラメータの大小に依らない貢献度を示す．そのため，SHAP 主成分分析により，第 1 主成分と第 2 主成分による特徴量マップとして，各車格の特徴を明示できる．

Fig.　10 に，SHAP 主成分分析による特徴量マップを示す．第 1 主成分と第 2 主成分の主成分得点により，各車格の特徴量を表現できている．次に，画像内の主成分負荷量の濃淡が明確であるパラメータの 1 つとして，Fig.　11 に Order-removed spectrogram の第 1 主成分負荷量，Fig.　12 に Prominance ratio の第 2 主成分負荷量を示す．また，第 1 主成分得点に差があった L_truck と H_truck から，L_truck_4 と H_truck_4 の Order-removed spectrogram を Fig.　13 に，第 2 主成分得点に差があった M_truck と H_truck から，M_truck_1 と H_truck_2 の Prominence ratio を Fig.　14 に示す．Order-removed spectrogram において，H_truck_4 と比較し，L_truck_4 は主成分負荷量の大きい高周波数の暗騒音の音圧レベルが大きい傾向にあると認められた．また，Prominence ratio において，M_truck_1 と比較し，H_truck_2 は主成分負荷量の小さい高周波数の突出成分が顕著であると認められた．

以上より，音響マルチパラメータ機械学習モデルにおける，高精度での車格ごとの音源分類が可能であると認められる．また，SHAP 主成分分析による特徴量の次元削減，および車格ごとの差が明確な特徴量の抽出が有効であると認められる．

5.　研究成果

(1)　エンジンの回転次数成分と暗騒音を別々に評価するための，U-Net を用いたスペクトログラム画像からの回転次数成分除去が有効であると認められた．

(2)　特定の音響パラメータが特徴的となる音を音響マルチパラメータ機械学習モデルに入力し，SHAP による貢献度算出を実施することで，本モデルが特徴量抽出に適すると認められた．

(3)　自動車走行音を機械学習モデルに入力し，SHAP に対して主成分分析を適用することで，特徴量の次元削減による特徴量マップ作成手法を確立した．また，主成分負荷量の大きい画像領域を可視化することで，車格ごとの差が明確な特徴量の抽出が可能であると認められた．

参考文献

(1)　大島遥汰 他，機械学習による時間変動を伴うエンジン音の特徴量抽出，自動車技術会論文集，**55**-6 (2024) pp. 1231-1237.

(2)　大島遥汰 他，機械学習を用いた音響物理特性に基づく特徴量抽出，第 34 回環境工学総合シンポジウム 2024 予稿集，**24**-9 (2024) pp. 133-136.

(3)　大島遥汰 他，音響マルチパラメータを用いたニューラルネットワークによる自動車走行音の特徴量抽出，自動車技術会論文集，**56**-2 (2025).

(4)　Scott Lundberg et al., A Unified Approach to Interpreting Model Predictions, NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems (2017) pp.4768-4777.

(5)　大島遥汰 他，音響マルチパラメータを用いた機械学習による自動車走行音の特徴量抽出－第 1 報 ニューラルネットワークを用いた回転次数成分と暗騒音分離による特徴量抽出の精度向上－，日本音響学会講演論文集(春) (2025).

(6)　大島遥汰 他，“音響マルチパラメータを用いた機械学習による自動車走行音の特徴量抽出－第 2 報 ニューラルネットワークを用いた車種間における共通特徴量抽出－”，日本音響学会講演論文集(春) (2025).

(7)　Olaf Ronneberger et al., U-net：Convolutional Networks for Biomedical Image Segmentation, International Conference on Medical image computing and computer assisted intervention (2015) pp.234-241.

(8)　自動車技術会，自動車音源 CD(DVD 版) (2004).

(9)　小山幸典, SHAP 主成分分析を用いた目的変数指向化合物マップの作成，第 68 回応用物理学会春季学術講演会 講演予稿集，19p-Z32-4 (2021).