

# ポンプ動作音の聴覚閾値を考慮した 機械学習に基づく音質評価モデル構築

## Construction of Sound Quality Evaluation Model Based on Machine Learning in Consideration of Auditory Threshold in Pump Operating Sound

精密工学専攻 24号 後藤 悠太  
Yuta Goto

### 1. はじめに

機械や製品を快音設計する際に、音質評価モデル(以下、モデル)の構築が欠かせない。また、近年機械学習やAIによる手法が注目されているが、適切な特徴量作成や抽出が難しい。従来、特徴量抽出には熟練者による特徴量決定や評価が行われてきたが、選択ミスや取りこぼしがあり、理解しづらい、もしくは設計には反映しにくい特徴量が含まれることが多い。そのため音の性質を反映し、分かりやすい特徴量選択が必要である<sup>(1)</sup>。

そこで本研究では、機械動作音の周波数特性、時間特性に対する指標化を行い、モデル構築を行った。周波数特性において、機械動作音の標準的な音源聴取時の純音聴覚閾値を測定および定義した。時間特性においては、音源の各周波数帯における変調周波数と変調度の閾値を測定および定義した。次に、顕著な周波数帯、ピーク、変調周波数や変調度の数値化を行った。そして、機械学習で快適感と関係する特徴量を抽出し、モデルを構築した。さらに精度評価や編集音源を用いた検証を行った。本研究の対象機器としてドライポンプ(以下、ポンプ)を使用した。なお、本研究の評価実験では、被験者の十分なインフォームドコンセントを得ている。

### 2. 印象把握実験

本章では、評価者が音源に対してどのような印象を持つか、印象の詳細について知るため、また、音源に対する評価得点を獲得するために実験を行う。

#### 2.1 音源について

使用したポンプ動作音は11音源である。ポンプはすべて同一製品で、運転状況も同一である。音源の中でも平均的な音源を決定するために、1/3 Octave Band(以下、OB)、A特性をかけた周波数特性を数値化し、平均を算出した。平均に最も近い音源を基準音源と設定し、音源名をPump-0とした。以降の実験で、相対評価実験を行う際、Pump-0を基準とした。残り10音源を評価音源とし、Pump-Aのように表現し、A-Jの名を付けた。

#### 2.2 実験概要

実験状況はFig. 1に示す。SD法による相対評価を用いて音質評価実験を行った。被験者は20代9名で、無響室にてヘッドホンで音源を提示した。被験者に基準音源を5秒、空隙時間1秒、評価音源を5秒という手順で提示した。提示音源セットの順序は無作為で聞かせた。

#### 2.3 評価語について

SD法では、形容詞対を用いるが、より幅広く適切な表現を目指すべく、形容詞のみならず、動詞を含む感情表現語をも加えた評価語を使用した。機械音質評価や官能評価に関する先行研究<sup>(2)(3)</sup>で使用された語句に、ポンプ動作音のイメージと重なる語句を加えた約300語収集した。



Fig. 1 The condition of the experiment

Table 1 Factor loadings

			Metal	Comfort	Powerful
Rough	-	Smooth	0.797	-0.301	0.235
Metallic	-	Moist	0.760	-0.200	0.290
Prickly	-	Roundness	0.738	-0.387	0.351
Dry	-	Wet	0.702	-0.227	0.201
Sharp	-	Dull	0.630	-0.308	0.176
Comfortable	-	Harsh	0.387	0.813	-0.354
Want to stay	-	Want to leave	0.304	0.756	-0.326
Want to listen-		Hate	0.423	0.741	-0.448
Not care	-	Annoying	0.384	0.714	-0.463
Wild	-	Mild	0.334	-0.413	0.774
Fierce	-	Soft	0.381	-0.420	0.704
Loud	-	Quiet	0.298	-0.490	0.629
Sum of sq.			3.562	3.303	2.471
% of variance			29.68	27.53	20.60
Cumulative %			29.68	57.21	77.80

被験者にポンプ動作音を提示し、イメージと合う評価語を選択させた。集計したのち、選択数が多かった上位評価語を抽出した。抽出する際には、感性プロセスにおける階層構造<sup>(4)(5)</sup>を考慮して、基礎・快適・嗜好・情緒感性に入る評価語を中心的に選択した。再度SD法で評価させた。

評価値に因子分析を用いて、音色の基本因子である「快適(美的)」「迫力」「金属性」の3因子を抽出するための適切な評価語を得るために、因子負荷量や寄与率など鑑みて、最終的な評価語を12対選定した。

#### 2.4 実験結果と考察

評価語や、印象把握実験の因子分析における因子負荷量や寄与率をTable 1に示す。2.3節において適切な評価語句の選定を行ったので、各々の因子負荷量、累積寄与率も高い。以降の実験や分析では、因子得点は分析ごとに値が変動するため、各因子に属する評価語群の平均評価点を基準化した数値を用いた。各評価語が1-7点で評価されるため、最大値

7, 最小値 1 として平均評価点の基準化を行った. 各因子の基準化平均評価点で表した各音源の快適得点, 金属得点を Fig. 2 に示す.

また, 快適得点が高い Pump-E と快適得点が高い Pump-J の周波数特性を Fig. 3, 4000 Hz (1/1 OB) における変調周波数特性を Fig. 4 に示す. 一部例示したが, 各ポンプ動作音には周波数特性, 変調に違いがあることがわかっている. Fig. 3 より 1000 Hz 周辺, 3850 Hz のピーク, 5000 Hz 付近での音圧レベルの違いや, Fig. 4 より変調周波数や変調度の違いがわかる. このような違いが快適得点に影響していると考えられる.

### 3. 聴覚閾値実験

Zwicker のマスキング理論<sup>6)</sup>では, 周波数特性が平坦なマスキャーを使用している. そのため, マスキング理論を周波数特性を有する機械動作音には適用できないと考えた.

そこで本章では, ポンプ動作音の特徴を捉えた周波数特性を持つマスキャーに対して, 純音認知の閾値を測定し, 聴覚閾値ラインを作成した. これを TH line とし, TH line を超えた周波数や突出量などを特徴量として使用する.

#### 3.1 純音と平滑化ポンプ音源について

実験音源を Fig. 5 に示す. 実験では, 純音とマスキャーの 2 種類を用意した. 純音は, より正確な TH line を作成するために, 100, 200, 400, 800, 1000, 2000, 2700, 4000, 6000, 8000, 9000 Hz の 11 音源を用意した. マスキャーとして, ポンプの個体差による影響を減らすために音源を合成し, さらに顕著なピークを削除して平滑化した音源を作成した. この平滑化ポンプ音源を Pump-SMTH とする. 各純音の音圧は, Pump-SMTH の各周波数の音圧レベルに対応するように調整を行った.

#### 3.2 聴覚閾値の推定

Pump-SMTH と 1 つの純音を同時に提示し, 純音のみ音圧レベルを上昇させ, 被験者が純音を認知できた音量を記録した. 上昇幅は 2 dB とし, 曖昧な判定がある場合は, その音量を記録し, その地点で音量増減を繰り返し, 適切なレベルを記録した. 測定した上昇分の音圧レベルを  $\Delta$ TH とする. 1 種類の純音が終わり次第, 純音を次の周波数に変更し, 繰り返し行った. 被験者は 20 代 11 名で, 実験音源は無響室においてヘッドホンで提示した. Fig. 5 の 4000 Hz の点線は, Pump-SMTH の 4000 Hz に当たる SPL から  $\Delta$ TH 分だけ上昇させた例であり, 10 dB となる.

#### 3.3 測定結果と考察

各純音における被験者閾値を Fig. 6 に示す. 高周波になるにつれ, 音の認知に必要な音量が上昇して行くことがわかった. これは Zwicker のマスキング理論と同様に高周波数ほど  $\Delta$ TH は大きくなる. しかし, 低周波数での  $\Delta$ TH の増減や, 2000 Hz が 1000 Hz, 2700 Hz と比較して音圧レベルが高いことはマスキング理論では表現できない. そのため, 各純音における  $\Delta$ TH はポンプ動作音における聴覚閾値の特徴を表現できる指標となる.

また, 2000 Hz における  $\Delta$ TH の音圧レベルが大きい理由は, Pump-SMTH の音圧レベルが 2000 Hz で減少し, 相対的に周囲の周波数の音圧レベルが高く, 聴覚的に優位になるためと考えられる.

#### 3.4 聴覚閾値ラインについて

各音源の認知ピークや音圧レベルが顕著な周波数帯を検知するために, 各周波数における  $\Delta$ TH をラインになるように繋げ, 聴覚閾値ライン TH line とした. 9000 Hz より高い周波数は実験で計測していないため, Zwicker のマスキング理論の閾値は 10 dB/decades で上昇することに従って, 9000 Hz

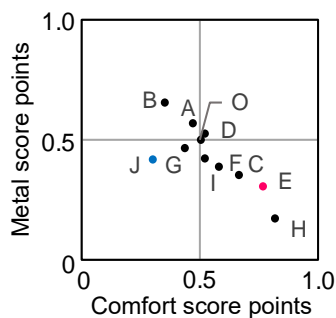


Fig. 2 Plot of the pump's score

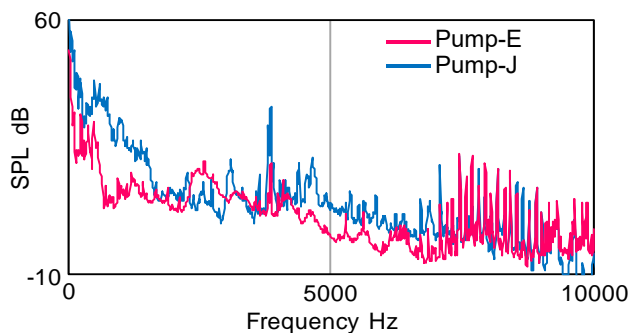


Fig. 3 Frequency spectrum of the pumps

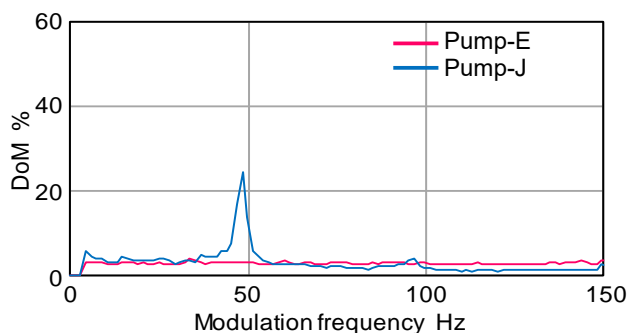


Fig. 4 Modulation frequency spectrum of the pumps

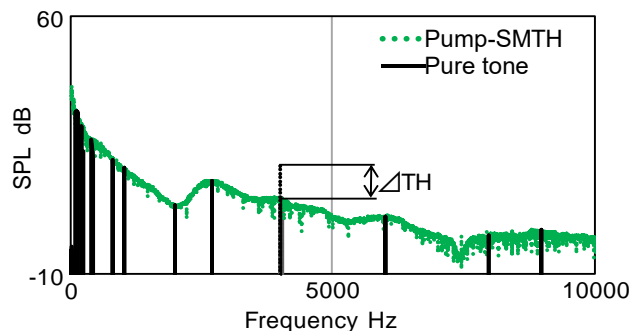


Fig. 5 Frequency spectrum of the test sounds

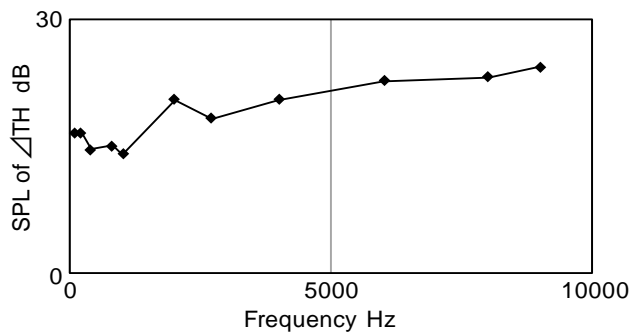


Fig. 6  $\Delta$ TH for the test sounds

を基に 10000, 11000, 12000 Hz の値を算出した。接続方法は、 $\Delta$ TH の値を 3 次スプライン曲線で各点間を補間した。

### 3.5 聴覚閾値ラインに基づく特徴量抽出

例として Pump-0 の周波数特性に TH line を適用し、認知されるピークを検出した例を Fig. 7 に示す。TH line を用いて、各ポンプ動作音の認知ピークの周波数や音圧レベルを数値化した。

## 4. 変調度閾値実験

機械動作音は定常音であるが、各周波数帯における時間特性によって変動を感じる。実際、ポンプ動作音にも各周波数バンドにおいて支配的な変調周波数が存在する。

そこで本章では、Pump-SMTH の一部周波数バンドの変調周波数、変調度を変更し、変調度閾値 TH-M を測定することで、各周波数バンドにおける変調度閾値ライン TH-M line を作成する。

### 4.1 音源作成について

Pump-SMTH を中心周波数帯が 250, 500, 1000, 2000, 4000, 8000 Hz の 1/1 OB に分割した音源を作成した。Pump-SMTH の変調度は低く一様である。今回実験のために編集した分割音源は中心周波数が 250, 500, 4000, 8000 Hz の音源である。1000, 2000 Hz は変調周波数において全ポンプ動作音の変調度が低く、特徴がないため、実験対象とはしていない。分割音源の一つを 4, 8, 10, 20, 40, 70, 80, 100 Hz の変調周波数、約 0 - 80% で変調度を調整し、変調した。変調周波数 100 Hz 以降では、変調の音の聞き分けが難化するため、100 Hz までとした。高変動しているほど、ホワイトノイズの粗さがある音源に近くなる。編集分割音源を、他の未編集分割音源と合成することで、変調音源を完成させた。

音源の一例として、Fig. 8 に 8000 Hz (1/1 OB) の分割音源を変調周波数 100 Hz、変調度 20% で変調させ、作成した変調音源の 8000 Hz (1/1 OB) における変調度を示す。

### 4.2 変調度閾値の推定

基準音として Pump-SMTH を用い、評価音には変調音源を用いて測定実験を行った。

測定方法は、被験者が基準音源と比較して変動や粗さなどに違いを認められる際の変調周波数と変調度を記録した。各 1/1 OB における各変調周波数の変調度を上昇させ、測定結果を TH-M とした。被験者は 20 代 6 名で行った。実験環境は無響室においてヘッドホンで提示した。

### 4.3 測定結果と考察

各 1/1 OB における変調周波数ごとの TH-M を Fig. 9 に示す。3.4 節と同様に各変調周波数における TH-M から 3 次スプライン補間を行い、TH-M line を作成した。100 Hz より大きい変調周波数は未測定かつ、変調音の聞き分けが難いため、100 Hz と同一値で扱った。

250, 500 Hz (1/1 OB) のような低周波帯が変動している音源は変調周波数が高くなるにつれて、TH-M がより上昇する。低周波帯が高変動しているものは、他周波数帯の粗さに変調音が埋もれてしまうため、認知に必要な変調度がより大きくなったと考えられる。4000, 8000 Hz (1/1 OB) のような高周波帯が変動している音源は変調周波数の高低に関わらず、TH-M は低く推移している。これは、高周波帯の変調音は耳に残りやすく、認知しやすいたことが原因だと考えられる。

### 4.4 特徴量抽出

Fig. 10 に一部ポンプ動作音の 8000 Hz (1/1 OB) における変動解析を行い、TH-M line を適用し、認知される変調度を検出した例を示す。TH-M line を用いて、各ポンプ動作音の各 1/1 OB の変調周波数、認知変調度を数値化した。

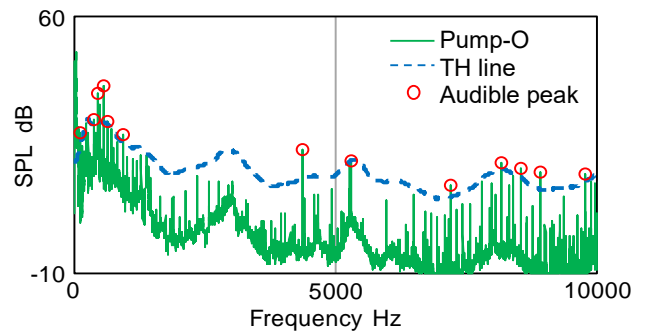


Fig. 7 Audible peak of Pump-O

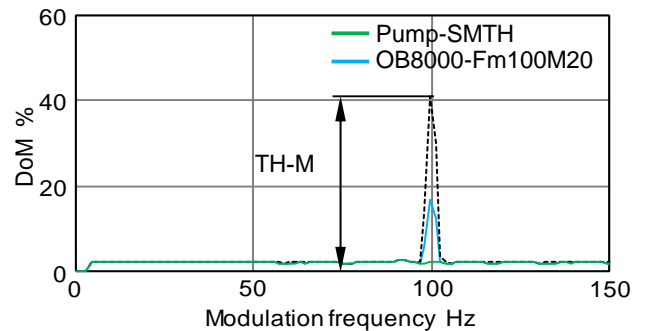


Fig. 8 Modulation frequency spectrum of the test sound

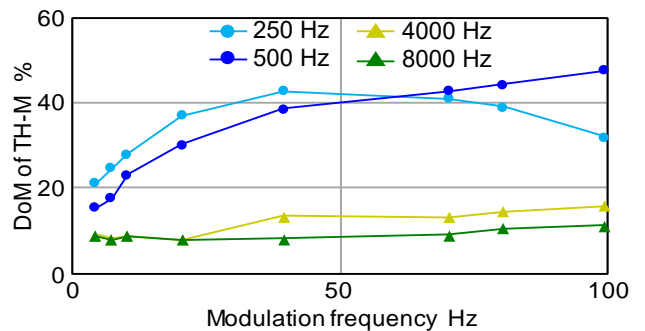


Fig. 9 TH-M each 1/1 Octave Band

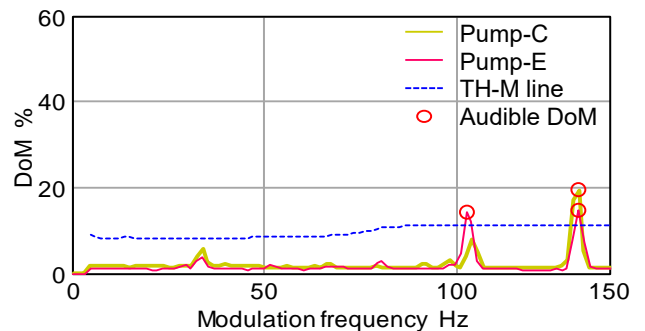


Fig. 10 Audible DoM of the pumps

## 5. モデル構築

本章では、機械学習の一つであるランダムフォレストを用いて、特徴量抽出を行い、重回帰モデルの作成を行う。

### 5.1 ランダムフォレスト

ランダムフォレストは機械学習手法の一つで分類や回帰に用いることができる手法である<sup>(7)</sup>。目的変数に対し説明変数の影響度の評価をする変数重要度という指標がある。

本研究では、変数重要度を使用し、重回帰分析で用いる特徴量を抽出するためにランダムフォレストを用いた。

### 5.2 特徴量抽出

3章における周波数の検知ピークはデータが大量にあるた

め、検出した周波数特徴量の抽出のみを対象にランダムフォレストの重要度を使用し、特徴量抽出を行った。目的変数を快適因子に属する形容詞の基準化平均評価点とした。説明変数は、認知ピークがある周波数の有無や出現回数とした。全音源の0 - 5000 Hzのうち1つでもピークがある周波数を並べ、ピークがあるものは1、ないものは0としたカテゴリ変数に変換した。5000 Hz以降はピークが多く乱立しているため、中心周波数6300, 8000, 10000 Hzの1/3 OB内のピーク出現回数をカウントし、カテゴリ変数とした。

抽出方法として、まずモデル作成条件は変更せず、重要度の出力を10回行い、各解析で重要度が高い上位10項目順に点数化し、10回分の合計値を算出し、重要度得点とした。次に、重要度得点が10点以上の周波数を特徴量として抽出した。

### 5.3 モデル構築と精度

抽出特徴量の各周波数の音圧レベル、1/3 OBの音圧レベル、4章で検出した各1/1 OBの各変調周波数における変調度を説明変数とした。目的変数として快適因子に属する形容詞の基準化平均評価点を用いて、重回帰分析を行った。抽出した特徴量の中でも、分散や修正決定係数を基に説明変数を選択し、モデル構築した式(1)に示す。

$$\begin{aligned}
 y = & +0.026 * [SPL \text{ at } 367 \text{ Hz}] \\
 & - 0.008 * [SPL \text{ at } 3850 \text{ Hz}] \\
 & - 0.054 * [DoM \text{ at } 4 \text{ Hz}(4000 \text{ Hz } 1/1 \text{ OB})] \\
 & - 0.019 * [SPL \text{ at } 8000 \text{ Hz } 1/3 \text{ OB}] \\
 & + 0.021 * [DoM \text{ at } 103 \text{ Hz}(8000 \text{ Hz } 1/1 \text{ OB})] \\
 & + 0.686 \qquad \qquad \qquad \dots (1)
 \end{aligned}$$

式(1)の修正決定係数 $R_f^2$ は0.862となった。快適感の評価が低い音源の傾向の一つとして、ポンプ回転周波数(91.6 Hz)に依存する3850 Hzの音圧レベルが顕著であることが挙げられる。そのため快適感に悪影響を及ぼしている3850 Hzの音圧レベルの低減で快適感の向上を式(1)で示せた。

## 6. 検証

本章では、モデル構築時に使用していない評価音源を用いて、モデルの未知データに対する予測精度の検証を行う。

### 6.1 音源について

Fig. 11に精度検証データとして使用した評価音源の一例を示す。D-3850はPump-Dの3850 Hzのピークを減少させた。他評価音源も式(1)を基に各ポンプ動作音を編集した。編集方法は各周波数部位の音圧レベルの増減を行い、評価音源を作成した。各ポンプ動作音は、固有の変調周波数があるため、周波数特性のみ変化させた。

### 6.2 検証実験概要

編集音源を評価音源として、基準音源 Pump-0 との相対評価によるSD法で評価させた。音源は10個用意し、実験条件や使用語句は2章と同様である。

### 6.3 検証

実測値と予測値の誤差を計算し、検証を行った。実測値には実験結果の快適感得点、予測値にはモデルと使用音源特徴量で算出した得点を使用した。予測値と実測値のプロットをFig. 12に示す。回帰モデルに使用される精度評価指標の誤差を平均的に評価するMAE、外れ値を大きく評価するRMSE、決定係数 $R^2$ を算出すると、MAE : 0.118, RMSE : 0.137,  $R^2$  : 0.640となった。MAE, RMSEはどちらも予測精度は0に近いほど高い。 $R^2$ は1に近いほど適合性が高く、今回構築したモデルは、学習データだけでなく、精度検証データに対する当てはまりが良いことから精度が高いといえる。

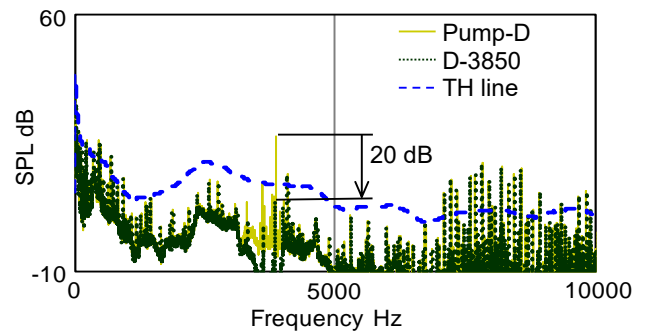


Fig. 11 Frequency spectrum of the edited Pump-D

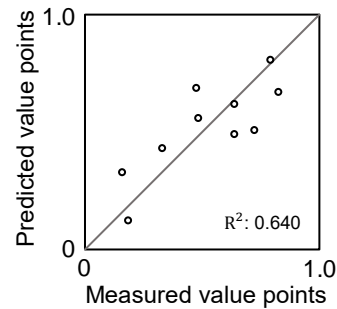


Fig. 12 Plot of predicted vs. measured value

## 7. 研究成果

- (1) ポンプ動作音を標準化した平滑化ポンプ音源における純音認知の閾値を測定、聴覚閾値ライン TH line を定義し、各ポンプ動作音における認知ピークの周波数帯や音圧レベルを特徴量化した。
- (2) 平滑化ポンプ音源の一部周波数帯を変調させた音源を用いて、各1/1 OBにおける変調度閾値 TH-M を測定、変調度閾値ライン TH-M line を定義し、認知変調度や変調周波数を特徴量化した。
- (3) 大量の認知ピークを特徴量とし、機械学習ランダムフォレストの重要度算出を複数回繰り返すことで、特徴量の重要度を定量化し、特徴量を抽出した。
- (4) 認知ピークや認知変調度からなる特徴量を用いて、重回帰分析を行い、モデルを構築した。
- (5) モデル構築に使用していない検証データを作成し、検証実験を行い、モデル精度が高いことを確認した。

### 参考文献

- (1) 後藤悠太, 戸井武司, 機械学習による音源の特徴量選択に基づいた快音化モデルの構築, 日本音響学会講演論文集(春)(2021)pp. 1185-1186.
- (2) 難波精一郎, 音色の定義を巡って, 日本音響学会誌, 49-11(1993)pp. 823-831.
- (3) 安倍幸治, 小澤賢司, 鈴木陽一, 曾根敏夫, 音色表現語, 感情表現語及び音情報関連語による環境音評価, 日本音響学会誌, 54-5(1998)pp. 343-350.
- (4) 小澤賢治, 松本恵, 感性プロセスにおける階層構造の検討-(1)モデルの提案と語彙調査, 第5回日本感性工学会大会予稿集(2003)pp. 44.
- (5) 松本恵, 小澤賢治, 感性プロセスにおける階層構造の検討-(2)階層構造の検証実験, 第5回日本感性工学会大会予稿集(2003)pp. 45.
- (6) Fastl, H. and Zwicker, E., PSYCHO-ACOUSTICS(2006) pp. 61-66.
- (7) Cutler, A., Cutler, D. R. and Stevens, J.R., Ensemble Machine Learning(2011)pp. 157-176.