J-POP 楽曲における歌声の声質を対象とした嗜好性の推定のための音響パラメータの提案

A Proposal of Acoustic Parameters to Estimate Listener's Preference for Voice Quality of Singing in J-POP Songs

桶　本　まどか　OKEMOTO Madoka　三　浦　雅　展　MIURA Masanobu

キーワード：MIR，音響パラメータ，歌声，嗜好性，機械学習

1. はじめに

近年，スマートフォンのような情報機器端末の普及にみられるようにIT（Information Technology）技術が急激な成長を遂げている．その中で音楽聴取の方法も変化している．2020年に行なわれた音楽の聴取方法に関するインターネット調査では，CD経由等での音楽聴取が減少傾向であるのに対し，定額制音楽サービス（いわゆる音楽サブスクリプションサービス．以下，音楽サブスク）での音楽聴取は増加傾向であると報告されている（一般社団法人レコード協会，2021a）．また，音楽の売り上げに関する2020年の調査では，CD等のオーディオレコードの売り上げが前年比85%と減少しているのに対し，音楽サブスクを含む音楽配信の売り上げが前年比111%であったと報告されており（一般社団法人レコード協会，2021b），このことからも音楽聴取の方法が変化してきていると考えられる．

このような，近年台頭している音楽サブスクは，その多くがスマートフォンアプリを中心にサービスが展開されている．例えば，Spotify Technology SAが提供するSpotify，Apple Inc.が提供するApple Music，Amazon.com, Inc.が提供するAmazon Music/Amazon Music Unlimited，Google LLCが提供するYouTube Music，LINE株式会社が提供するLINE MUSICなど多岐にわたる．各サービスでは，配信楽曲数の充実や音質など様々な独自性を打ち出している．そのうち，レコメンド機能は特に各サービスによってその独自性が示されるサービスであり，その仕様は明らかにしていない場合がほとんどである．例えば，SpotifyではSpotify APIによって取得可能な音響パラメータを用いて類似楽曲を推定するアルゴリズムがあり，そのアルゴリズムをレコメンド機能の一部として利用していると考えられる．YouTube Musicを利用可能なYouTubeでは，ビッグデータを利用しユーザーの視聴履歴の傾向が近い別のユーザーの視聴履歴に基づくレコメンドを行なうアルゴリズム，また，Googleアカウントのユーザー情報が近いユーザーの視聴履歴に基づくレコメンドを行なうアルゴリズムをレコメンド機能として利用しているとされている．しかし，いずれもレコメンド機能のアルゴリズムの全体像は明らかにされていない．

そのようなレコメンド機能では，邦楽，洋楽，K-POPといった音楽ジャンルだけではなく楽しい時に聴く曲，悲しい時に聴く曲などユーザーの状況応じたレコメンドなど様々なレコメンド機能が提供されているものの，レコメンド機能には，まだユーザーの音楽聴取状況に即していない部分も多数存

在する．その１つが楽曲を歌う歌手の歌声の嗜好性である．中学生，高校生を対象とした好きなアーティストに関する調査では，アーティストを好きな理由の１つとして，歌声に関する理由もいくつか挙げられている（高校生新聞ONLINE，2021）．しかし，このような歌声の嗜好性に着目したレコメンド機能に関しては，まだ十分に検討されていない．そこで，本論文では，歌声，特に声質の嗜好性を考慮した楽曲推薦を行なうための歌声の声質特徴を考慮した音響パラメータの提案を行なう．また，本論文では，前述のレコメンドアルゴリズムのうち，音響パラメータを用いた手法を提案する．本論文では，扱う歌声とは，J-POPなどのポップス楽曲における歌声を指す．

２．楽曲推薦技術

楽曲のレコメンドを行なうための音楽推薦技術は、音楽情報検索（MIR：Music Information Retrieval，以下MIR）と呼ばれる1990年代ごろに立ち上がった研究分野の一部である．本章では，楽曲推薦技術に関して先行研究などを踏まえて，整理を行なう．

2.1 推薦技術の概要

楽曲推薦技術については，音響パラメータを用いる推薦技術，視聴履歴などを用いる推薦技術の大きく2通りの手法，またその組み合わせによって行なっているとされている．そのうち，後者のあるユーザーの好みの楽曲を推薦するために，似た好みを持つ別のユーザーの視聴履歴を参考に推薦する技術は協調フィルタリングと呼ばれている．協調フィルタリングについては，楽曲推薦技術以外にも様々なサービスで利用されており，例えば，Amazonの商品レコメンドにおいてもアルゴリズムが採用されていると考えられている．この手法は，本，映画などコンテンツに問わず，行動履歴があれば採用できるため，多くの研究が行なわれた技術ではあるものの，こと楽曲推薦においてはユーザーの行動履歴の蓄積が十分でない楽曲が推薦されにくいという，コールドスタート問題があるとされている（吉井ら, 2009）．この問題を克服するために音響パラメータといった楽曲の内容に関する情報に基づいたフィルタリングによってユーザーの好みの楽曲と類似する楽曲を推薦する技術が盛んに研究されているものの，この手法は多くのユーザーが好むなどの楽曲の人気度合いを考慮することができないという問題があるとされている（吉井ら，2009）．このように，これらの２つの手法の問題点を解消するために，近年ではこれらを組み合わせたハイブリッド型の推薦技術に注目が集まっている．

本論文では，音響パラメータを用いた楽曲推薦技術に利用可能な提案パラメータの有効性検討を行なう．前述の通り，ハイブリッド型の推薦技術に近年注目が集まっているものの，協調フィルタリングのための大量の行動履歴，いわゆるビッグデータを取得するのが困難であるため，行動履歴を伴わない場合での有効性検討を行なう．

2.2 IT技術の発展におけるMIR

楽曲推薦技術はMIRの一部分と前述したが，このようなMIRの研究が盛んに行なわれている背景として，IT機器が小型化かつ大容量化したことによって，個人レベルで大量の音楽データを扱うことができるようになったこと（山田ら，2011），また，音楽配信サービスの開始により，ユーザーが聴取可

能な楽曲の幅が増えたことがある．音楽再生機器でいえば，1979年に登場したSONYのウォークマンではカセットテープで録音可能な数曲がいつでもどこでも聴取できるようになり，2001年に発表されたiPodでは1000曲もの楽曲がいつでもどこでも聴取できるようになり，2020年代では，音楽サブスク等によって数万曲以上がいつでもどこでも聴取できるようになった．このように，人々がいつでもどこでも膨大な数の楽曲の中から選択的に音楽を聴取できるようになっていった．しかし，聴取可能な楽曲が増えたからといって，すべての楽曲をある１ユーザーが聴取することは不可能であり，膨大な楽曲の中からユーザーが聴取したい楽曲を選択する必要がある．MIRの研究分野の技術は，その楽曲選択の一助となるものであり，このように大量の楽曲を手軽に聴取することが可能な時代においては，ニーズの高い分野である．

2.3 楽曲推薦に関する諸研究

楽曲推薦に関する研究は，MIRの研究分野の発展と共に盛んに研究が行なわれている．例えば，ユーザーの気分に適した楽曲推薦技術，歌詞情報を考慮した楽曲推薦技術など，様々な目的に対し，様々な手段によって楽曲推薦技術の提案が行なわれている．なお，ここでは，協調フィルタリングに関する技術を含まない楽曲推薦技術について触れる．また，ここでは楽曲の印象などの推定に関する研究も楽曲推薦技術の関連技術として取り扱う．

楽曲推薦技術において，盛んに取り上げられるテーマの１つは，楽曲の印象に基づいた推薦技術がある．この技術は楽曲に印象に関するメタ情報が付与されていなければ実現が難しい．また，その際に，すべての印象を人の手によって付与することも困難である．そこで，音響分析によって音響パラメータを取得し，印象評価実験等で得られた楽曲に対する印象とのマッチングを取り，そのマッチングの結果を基に未知の楽曲に対し，印象のメタ情報を付与する楽曲推薦の基礎技術となり得る研究が行なわれている（伊藤ら，2011；西川ら，2011；など）．これらの研究では，印象評価実験に用いた印象評価の基準に沿った形式の印象が推定される．例えば，音楽ゆらぎ特徴を用いた楽曲印象値の推定手法を提案している研究では，SD法（Osgood, et al., 1957）に基づいた印象評価，楽曲の印象軌跡に関する研究では，RussellのValence – Arousalの感情平面（Russell, 1980）に基づいた印象評価を行なっており，同じ印象を推定する研究といっても，その推定の対象となる印象の形式は異なる．印象を求める際には，楽曲の印象軌跡に関する研究では，音響パラメータだけでなく歌詞情報も用いており，既に利用可能となっているメタ情報を利用する場合もある．

その他にも，楽曲の時代ごとの特徴に関する研究（Serrà, et al., 2012；岡田ら，2018；など），動画像に適した楽曲に関する研究（桐本ら，2008；追木ら，2018；など）など，一重に楽曲推薦技術といっても，その目的，手法，対象は多種多様となっている．

2.4 歌声の特徴を考慮した楽曲推薦に関する研究

歌声の特徴を考慮した楽曲推薦に関する研究では，声質の類似度に基づく楽曲推薦に関する研究（藤原ら，2007）がある．この研究では，伴奏音つきの音源から伴奏音抑制技術を用いて取得した音響パラメータを取得する手法，また相互情報量を用いて２つの特徴ベクトルの類似度を算出する手法が

提案され，これらを組み合わせることで歌声の類似度に基づく楽曲推薦を行なっている．また，得られた推薦結果を被験者に評価させた結果，比較手法に対する，提案手法の有効性が示されている．

声質の類似度に基づく楽曲推薦では，用いる音響信号は伴奏音つきの音源であるため，その音響信号に対し，伴奏音抑制を行なっている．まず，歌声における基本周波数（以下，F0）(重野，2006）をPreFEst（Goto，2005）によって推定し，次に推定されたF0に基づいて歌声の調波構造を抽出，最後に正弦波重畳モデル（Moorer，1977）に基づいて音響信号の再合成を行なう，伴奏音抑制技術が用いられている．この伴奏音抑制技術によって再合成された音響信号を用いて，LPCメルケプストラム係数（徳田ら，1988）とΔF0（Ohishi, et al.，2005）の２つの音響パラメータを算出している．なお，この時，PreFEstの特性上，楽曲の間奏部分，また伴奏音のダイナミクスが大きい場合については，再合成された音響信号が歌声の特徴を表す音響パラメータに相応しくない音響信号となっている場合があるため，分析に適さない区間に対し算出された音響パラメータを類似度計算から除外する処理を加え，その後に，類似度計算を行なっている．この研究では，音響パラメータを算出する際に，楽曲を歌唱する際の旋律情報として変化するF0が考慮されていないという問題があり，歌声の類似度において，その声質の類似度と，歌い方の類似度のいずれがどの程度の割合で考慮されているのかが明確にされていない．

３．音響パラメータ

歌声の類似度に関する研究（藤原ら，2007）では，歌い方など声質以外の要素を排除したパラメータとなっているかということについて十分に議論されていない．そこで，本論文では，歌声の声質のみに着目した音響パラメータを提案する．また，音響パラメータの有効性検証として，提案パラメータ群を説明変数，歌声の声質の好みに対する主観評価結果を目的変数とした機械学習を行なう．

3.1 歌声の抽出

歌声の声質を考慮した音響パラメータを設計するにあたり，音響パラメータの算出対象とする音響信号は歌声のみの音響信号が望ましい．しかし，今後，大量の楽曲を対象とした楽曲推薦システムに用いることを考慮すると，歌声のみの楽曲というのはマイノリティであり，マジョリティである伴奏音を伴った楽曲を対象とした音響パラメータを考慮するのが妥当である．そこで，まず，伴奏音を伴う楽曲から歌声を抽出する必要がある．

歌声抽出に関する技術を含む音源分離技術については，2007年から開催されてきたSiSEC（Signal Separation Evaluation Campaign）などで盛んに報告されている．近年では，音響信号そのものを用いた機械学習が可能となり，その結果精度の高い歌声抽出が行なわれるようになった．例えば，Tensorflowを利用したSpleeter（Prétet, et al., 2019）などがある．

本論文では，ソースが公開されており利用可能な歌声抽出として，GitHubにて公開されているvocal-remover v4.0.0（Turumeso，2019）(以下，vocal-remover）を用いる．vocal-removerを採用した理由としては，歌声抽出技術の精度が他のものと比較し優位だと，第一著者と第二著者で判断したためである．vocal-removerによって抽出された歌声に対し，音響パラメータの算出を行なう．

3.2 提案する音響パラメータ

本論文では，歌声の声質のみに着目した音響パラメータ，つまり，歌い方といった歌声における声質以外の特徴が影響しない音響パラメータの設計を行なう．歌声の声質に着目した音響パラメータとして，18種類の音響パラメータを設計した．この18種類の音響パラメータは，周波数とメル周波数の側面から算出された２つの倍音構造パラメータを算出し，それらの傾向を表す９つの傾向パラメータを組み合わせることによって算出を行なう．また，提案パラメータの他に2つの基礎的な信号分析に用いられるパラメータ（以下，基礎分析パラメータと呼称）を組み合わせた，計20個の音響パラメータ群を歌声の声質の特徴として提案する．表1に提案パラメータ群の一覧を示す．

表１：提案パラメータ

	H_1	H_2	B_1	B_2
T_1	○	○		
T_2	○	○		
T_3	○	○		
T_4	○	○		
T_5	○	○	○	○
T_6	○	○		
T_7	○	○		
T_8	○	○		
T_9	○	○		

3.2.1 倍音構造パラメータ（H）

本論文では，倍音構造パラメータとして，F0を考慮したパラメータとメル尺度（古井，1992；日本音響学会，2011；など）を考慮したパラメータの２種類を提案する．

3.2.1.1 F0を考慮した倍音構造パラメータ（H1）

このパラメータは，F0から３オクターブ分の倍音構造を，分析フレーム毎に算出するパラメータであり，音高に影響されない歌声の倍音を取得することができる．算出方法は，歌声抽出音源の無音区間を除く．続いて，YIN（Cheveigné, et al., 2002）に基づいて作成されたF0算出プログラムを用いて，F0カーブを取得する．次に，取得したF0から３オクターブのパワースペクトルを取得するバンドパスフィルタを通過させることでF0を考慮した倍音構造パラメータを算出している．

3.2.1.2 メル尺度を考慮した倍音構造パラメータ（H2）

3.2.1.1で述べたパラメータと異なり，こちらにはバンドパスフィルタを通過させず，またF0を考慮せず，バンドパスフィルタ外のスペクトル成分が取得可能となっており，具体的には，F0以下の周波数帯域，またF0から３オクターブより上の周波数帯域の成分も取得可能である．また，人間の聴覚構

造を考慮したメル尺度を適用したパラメータとすることで，高周波数の成分の抑制などが行なわれている．

3.2.2 傾向パラメータ（T）

3.2.1で取得した倍音構造パラメータに対し，９つの傾向パラメータを算出し，歌声の声質に着目した音響パラメータを算出する．

3.2.2.1 最大値（T_1）

倍音構造パラメータの最大値を算出する．

3.2.2.2 セントロイド（T_2）

倍音構造パラメータの重心を算出する．

3.2.2.3 周波数軸方向の差分の平均（T_3）

倍音構造パラメータの周波数軸方向の差分を取得し，その平均値を算出する．

3.2.2.4周波数軸方向の差分の標準偏差（T_4）

倍音構造パラメータの周波数軸方向の差分を取得し，その標準偏差（東京大学教養学部統計学教室，1991；小島，2006；など）を算出する．

3.2.2.5 時間数軸方向の差分の平均（T_5）

倍音構造パラメータの時間軸方向の差分を取得し，その平均値を算出する．

3.2.2.6 時間軸方向の差分の標準偏差（T_6）

倍音構造パラメータの周波数軸方向の差分を取得し，その標準偏差を算出する．

3.2.2.7 2次の曲線近似におけるx^2の係数（T_7）

倍音構造パラメータに対し２次の曲線近似を行ない，x^2の係数を取得する．

3.2.2.8 2次の曲線近似におけるxの係数（T_8）

倍音構造パラメータに対し２次の曲線近似を行ない，xの係数を取得する．

3.2.2.9 2次の近似曲線との差分の平均（T_9）

倍音構造パラメータに対し２次の曲線近似を行ない，得られた近似式より，近似曲線を算出する．算出された近似曲線と倍音構造パラメータとの差分の平均値を算出する．

3.2.3 基礎分析パラメータ（B）

Zerocrossings（以下，ゼロクロス）(山田ら，2014）とSpectral Skewness（以下，スペクトルスキューネス）(Peeters，2004）を算出する．音楽音響信号を対象とした楽曲の年代を推定した研究では，RMS（Root Means Square）などの音響パラメータが用いられている（岡田ら，2018）ものの，声質のみに着目したパラメータにおいては不適切と判断した，その理由として，異なる楽曲つまり，異なるリズム構造をしていることによるRMSの変化が，歌声の声質に対する個人差に対し，非常に大きな変化となり，声質に関するパラメータとして，ノイズの多いパラメータとなってしまうためである．本論文では，ある１つの楽曲を対象とした場合における歌声の声質の好みではなく，様々な楽曲を対象とした場合における歌声の声質の好みを推定することを目的としているためである．さらに，3.2.2で提案するパラメータと近い機能を有するパラメータについても除外した．

3.2.3.1 ゼロクロス（B_1）

音響信号の振幅値が0と交差する回数を算出する．

3.2.3.2 スペクトルスキューネス（B_2）

音響信号のスペクトル歪度を算出する．

４．評価実験

３章で述べた音響パラメータの妥当性を評価するためには，歌声の好みに対する印象評価結果が必要である．この印象評価結果を取得するために，印象評価実験を行なった．

4.1 概要

印象評価実験では，５秒の伴奏音を伴う楽曲を連続聴取させ，その歌声の声質が「好みである/好みでない」を２件法で回答いただいた．なお，本実験は，国立音楽大学研究倫理委員会の（主に）人を対象とする研究に関する研究計画等審査で承認されている．

4.2 聴取者

聴取者は，第一著者を含む国立音楽大学，もしくは，国立音楽大学大学院に所属する10名であった．以後，L1からL10と表記する．聴取者には，ヘルシンキ宣言に基づく実験参加同意をいただいた．

4.3 実験刺激

前述の通り，伴奏音を伴う楽曲を聴取させ，その歌声の声質について主観評価をさせた．この時，聴取させた楽曲のアーティストのイメージが評価に影響を与える可能性があると考え，VOCALOID等の歌声合成ソフトによるオリジナル楽曲を人がカバー歌唱している楽曲，計70通りを用いた．その70通りの楽曲からサビ部分を５秒抜粋し，正規化を行ない，冒頭と終端に0.5秒のテーパーを付与した実験刺激を作成した．そして，刺激番号を英語で読み上げた音声，歌唱音，読み上げ音声，実験刺激

が交互になるような実験刺激を合成した．実験開始前に，テスト音源を聴取し音量調整を行なっていただき，以後は音量調整を行なわないように指示した．

4.4 評価法

評価は，歌唱の声質が「好みである/好みでない」の 2 件法とした．注意事項として，歌い方などではなくあくまで声質にのみに着目し回答するように，また，回答に迷った際には，直感で判断を行なうように指示をした．さらに，今回の好みに対する判断において何らかの基準があった場合はその基準を訪ねる自由記述欄を回答用紙の最下部に設定した．

4.5 回答方法

本実験は，ファイル共有サービスにて，実験刺激，回答用紙を配布して，回答を行なっていただいた．回答終了後，回答用紙は再度ファイル共有サービスにアップロードしていただいた．

4.6 回答の分布

図 1 に10人の聴取者の回答の分布を表す積み上げ折れ線グラフで示す．また，表 2 に聴取者の回答の相関係数を示す．なお，相関係数の算出には，回答をそれぞれ，「好みである」を 1，「好みでない」を 0 に置き換えることで算出を行なった．図 1 の横軸は刺激IDを示し，縦軸は各聴取者が好みであると回答した数を示している．例えば，刺激ID S01については，L1，L3，L4，L7，L10の計5名が好みであると回答していることを表している．

表 2 より，聴取者の回答における相関の絶対値の平均は0.20であった．図 1 と表 2 より，相関係数は全体として低い傾向にあることより，歌声の声質の好みにおける非共通性が確認された．一方，4 通りの実験刺激については，全ての聴取者が好みであると回答しており，歌声の声質の好みに対する共通性も確認された．これより，歌声の声質の好みについては，一貫性は見出せないものの，その一部においては何らかの共通認識を有する可能性が示唆された．

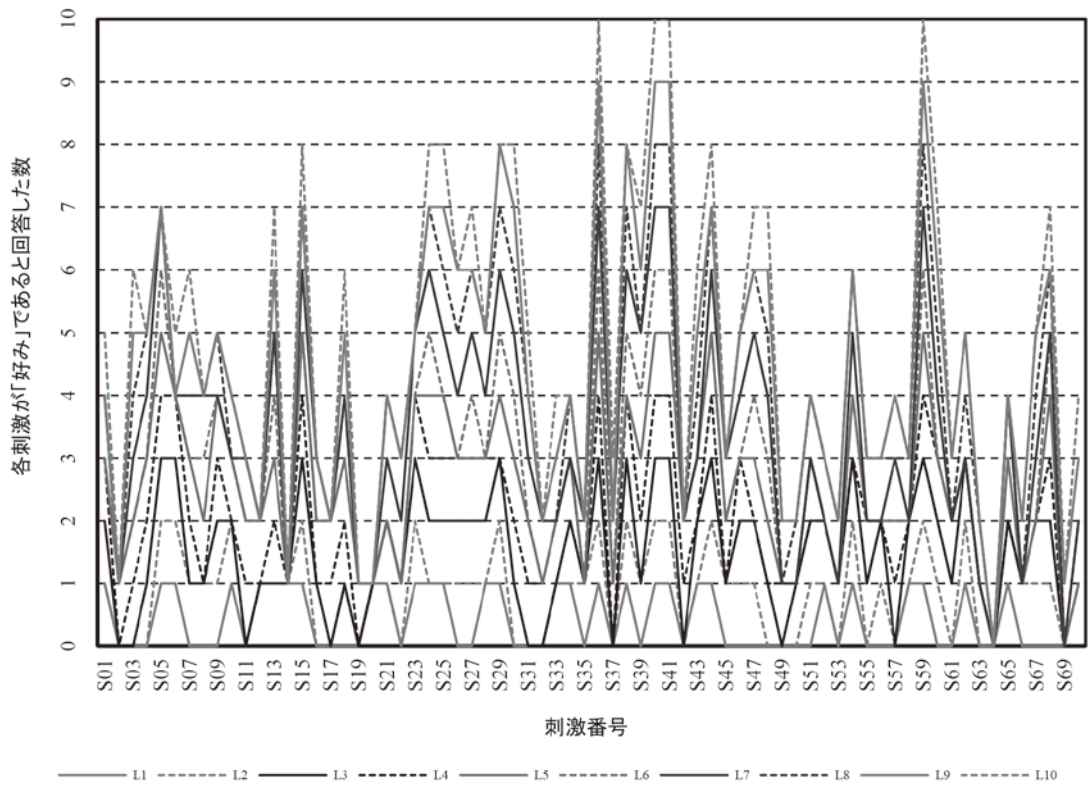

図１：聴取者の回答の分布

表２：回答結果の相関係数

	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10
L1	1.00									
L2	0.10	1.00								
L3	0.27	0.22	1.00							
L4	0.07	0.09	0.13	1.00						
L5	-0.01	0.13	-0.20	0.11	1.00					
L6	0.20	0.35	0.19	0.19	0.37	1.00				
L7	0.00	0.25	0.14	0.33	0.30	0.34	1.00			
L8	0.13	0.15	0.77	0.08	-0.07	0.19	0.16	1.00		
L9	0.05	0.16	-0.07	0.20	0.12	0.24	0.19	0.02	1.00	
L10	0.13	0.13	0.08	0.41	0.39	0.39	0.41	0.17	0.36	1.00

５. 機械学習による有効性検証

有効性検証として，提案パラメータ群を説明変数，歌声の声質の好みに対する主観評価結果を目的変数とした機械学習を行なう．

5.1 検証の概要

本論文では，３章で述べたパラメータ群をすべて使ったALL条件，HとTの組み合わせからなるパラメータ群を用いたHT条件，Bのみを用いたB条件でそれぞれ機械学習を行ないパラメータの有効性を検討する．この時，Closedテスト（クローズドテスト）と，Openテストとして10 fold-CV（荒木，2018）による検証を行なう．機械学習のアルゴリズムとしてランダムフォレスト（random forest）（Breiman，2001）にて分類を行ない，F-measure（荒木，2018）によって推定精度の評価を行なう．

5.2 結果

図２に３条件に対するクローズドテストの推定結果，図３に３条件に対する10 fold-CVの推定結果を示す．図２，３はそれぞれ横軸が聴取者，縦軸がF-measureであり，各聴取者の回答を３条件のパラメータでランダムフォレストを行なった際のF-measureが表されている．また，グラフ中のCLはチャンスレベルを指しており，歌声の声質に対し，「好みである／好みでない」の２件法で回答させたので，チャンスレベルは50%であることを表している．そして，図２，３のグラフ右側のAvg. については，条件ごとにF-measureの値を平均したものである．

図２より，クローズドテストの場合は，いずれの場合においても，F-measureの値は，1.00であった．図３より，L8は最もF-measureの値が高い条件HTの場合においても0.47であるものの，その他の聴取者においては，提案パラメータを含む条件でF-measureの値はいずれもチャンスレベルを超えていることが確認できる．また，聴取者ごとで最もF-measureの値が高い条件は，ALL条件が３名，HT条件が６名，ALLとHT条件が同等であるのが１名であった．これより，提案パラメータを用いた方が，F-measureの値が向上する傾向が確認できた．F-measureの値が高い音響パラメータの組み合わせが異なるのは，歌声の声質に対する評価基準の差が影響していると考えられる．さらに，Avg.については，条件Bの場合はF-measureの値は0.49であり，チャンスレベルを下回っていたのに対し，最もF-measureの値が高いHT条件ではF-measureの値は0.58であり，F-measureの値が大幅に向上していることからも提案パラメータの有効性が示唆される．

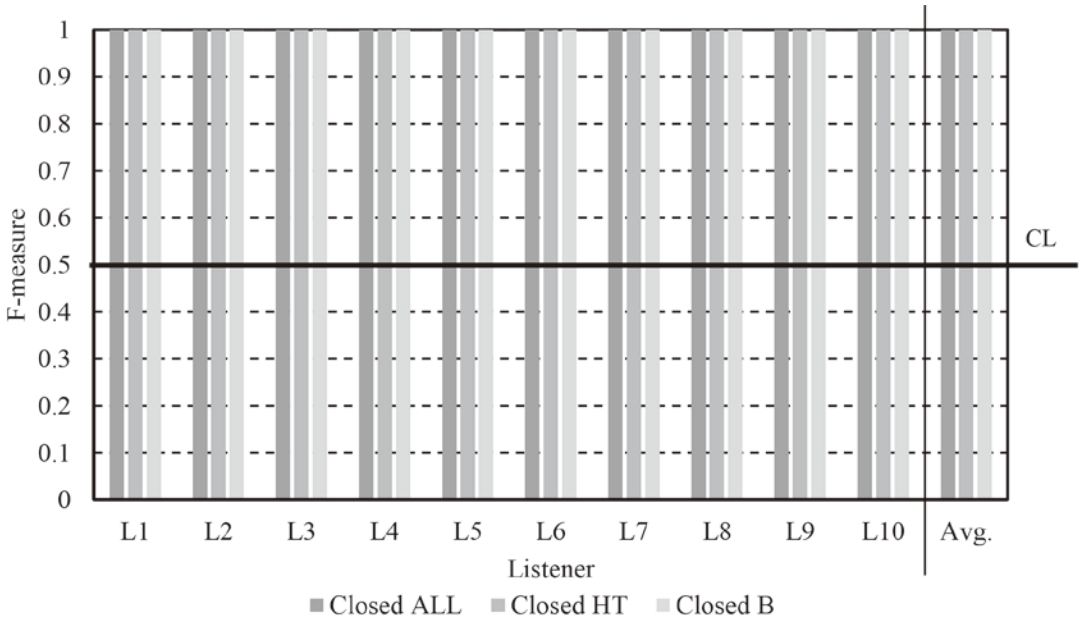

図２：クローズドテストの推定結果

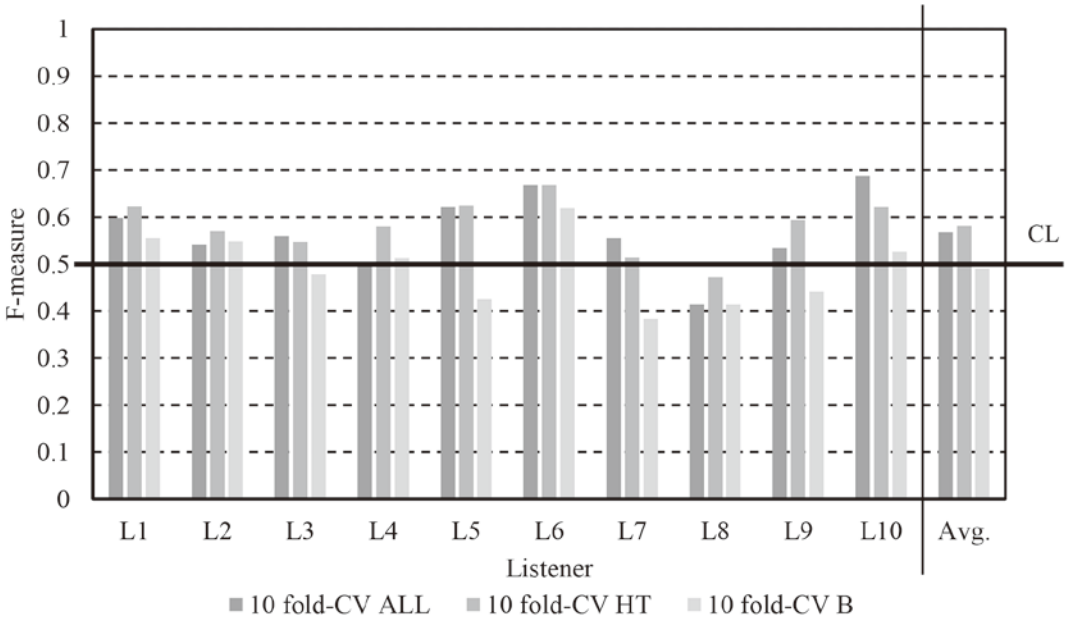

図３：10 fold-CVの推定結果

6．考察

5.2で述べたように，提案パラメータの有効性は示唆されたものの，聴取者ごとに，最もF-measureの値が高い条件も異なっていたことについて更なる検証を行なうため，各聴取者の回答と20個の音響パラメータの相関係数を確認する．このとき，表２の相関係数を算出した際と同様に，各被験者の回答については「好みである」を１，「好みでない」を０に置き換えた．図４に各音響パラメータに対する聴取者ごとの回答の相関係数を示す．図４は，横軸が相関係数，縦軸が音響パラメータ，棒が各音響パラメータとの比較対象となった聴取者IDを表している．例えば，H1T1の音響パラメータに対し，聴取者L1の回答の相関係数は0.29であることが表されている．図４より，H1T1の音響パラメータを確認しても，L3の相関係数は0.35であるのに対し，L2については -0.03であり，その相関係数の傾向が異なることが確認できる．同様に，H1T3などの音響パラメータについても，相関係数の傾向が異なる音響パラメータが確認できる．これより，歌声の声質の好みについては，聴取者ごとにパラメータ群を調整することが望ましいことが示唆される．このことは，4.3で説明した自由記述欄の回答でも示唆された．回答欄に記述のあった10名中４名の回答を次に示す．「無理なく出ている声，作っていない声が好み」，「中性的な声が好み」，「聴いていて高音が痛く感じないか，ストレートな声か」，「喉ががらがらした音を含んでいる，また高くきんきんしている，中性的な声は苦手」というような記述があった．これをみても，「高音が痛い，きんきんしている」といった声の倍音成分に関係するような特徴を着目している場合，また「喉ががらがらした音を含んだ声」，「ストレートな声」といった歌声のラフネス（Terhardt, 1974）やFluctuation Strength（Fastl, 1990）に関係がありそうな特徴に着目するなど，様々な基準で好みの評価を行なっていることが確認できる．これより，更に音響パラメータを拡充することによって，F-measureの値がより高い推定を行なうことができる可能性が示唆された．

本論文では，歌声の声質の嗜好性を対象とした音響パラメータの設計を試みた．表２に示した通り，好みについて被験者間の回答に対する相関係数が低い傾向にあり，その要因としてその判断基準が多様であるからだと結論を述べた．このような好みと被験者間の回答の傾向に対しては，顔の好みに関する報告（中山ら，2012）でも，被験者間の相関係数は0.27であったと報告されており，本論文の好みに対する評価の被験者間相関が低いということと同様の傾向を示している．また，布の好みの個人差について分析を行なった研究（市原，1996）では，布に対し20の印象評定項目で7段階評価を行なわせ，その結果から評定の個人差について言及している．その結果，若々しいやさわやかといった印象評定については比較的個人差が少なかったものの，好きの印象評定については強い個人差がみられたと報告されている．更に，この研究では，印象評定値を用いて因子分析を行なうことで布の印象を構成する因子を取得し，布の印象評価モデルを構築，検討を行なっている．これによって，人が布に対し印象評価を行なう際の認知過程の一端が明らかになっている．歌声の声質の好みについても，その認知過程が明らかになれば，音響パラメータの設計に対し有用性が期待されると考えられ，歌声の声質における好みの推定についても適用可能であると考える．

以上，個人差の更なる検討は必要であるものの，歌声の好みに対して第一次近似的な観点から言って「F0からの倍音構造」や「メル尺度の倍音構造」に関する音響パラメータの有効性が示された．こ

れは，歌声の好みという複雑な現象において，そのスペクトルに含まれる倍音という成分が少なくとも好みに寄与している可能性が示され，今後の研究へと適用される可能性が見出された.

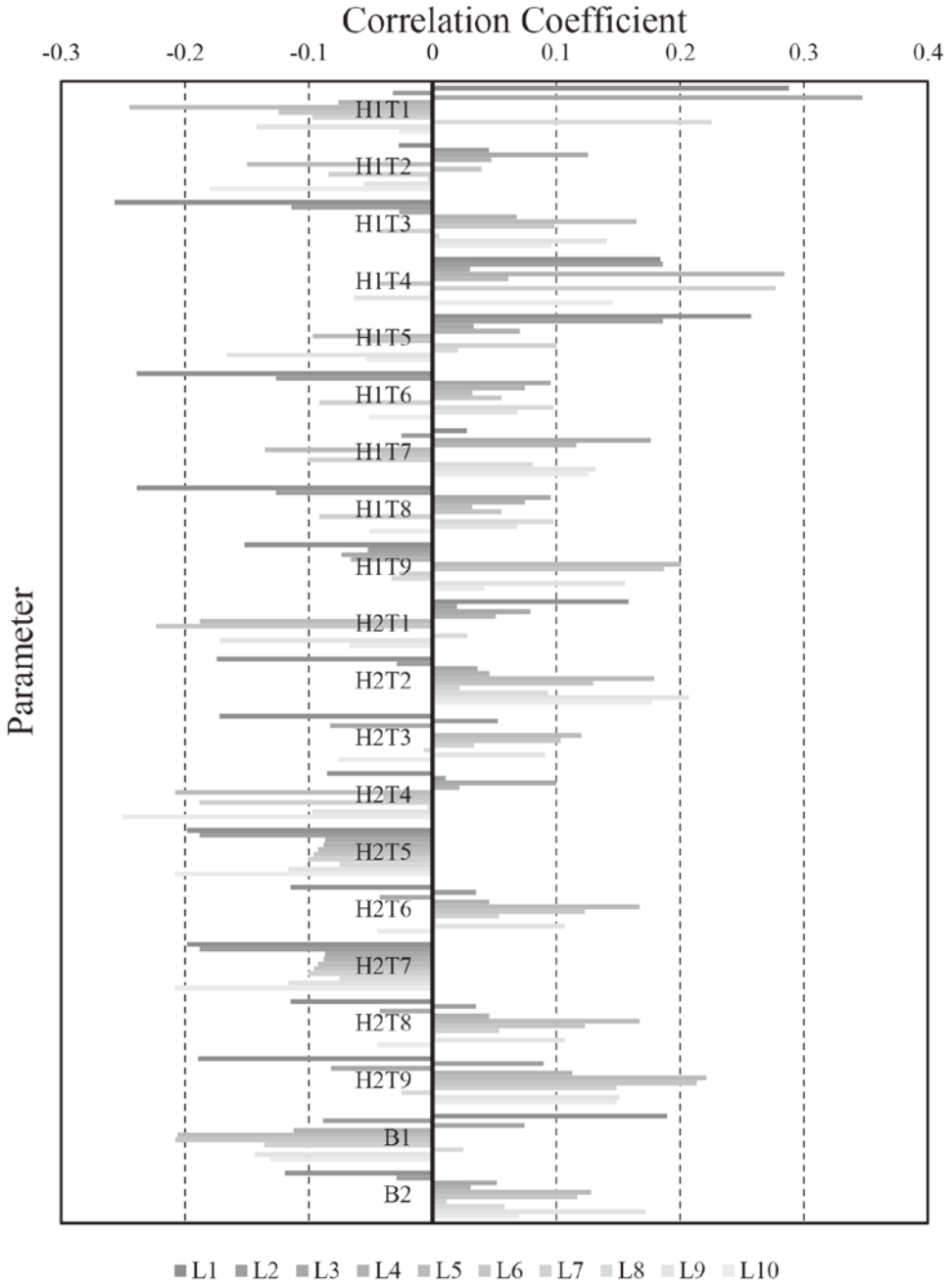

図４：各音響パラメータに対する聴取者ごとの回答の相関係数

7．まとめと今後の課題

本論文では，歌声の声質の嗜好性を推定するための音響パラメータの設計を試みた．10人の聴取者による歌声の声質に好みに関する評価に対して，各聴取者の回答を目的変数，提案する音響パラメータを説明変数とし，ランダムフォレストによる分類を行なった結果，提案する倍音構造パラメータによる分類において，F-measureの値の平均はチャンスレベルである50％より高い0.58であることが確認され，提案する音響パラメータの有効性が示唆された．つまり，歌声に含まれる倍音という特徴が好みを決定づける要因であることが第一次近似的に示された．一方，聴取者の判断基準のいくつかが音響パラメータに反映できていない可能性が高く，それらの音響パラメータを設計することでよりF-measureの値が高い分類を行なうことが可能であることが示唆された．更に，歌声の声質の好みについて，その判断基準を明文化できない聴取者も存在するので，SD法などの印象評価の側面から好みの認知過程を明らかにすることによって音響パラメータの設計に活かすことができる可能性も示唆された．なお，本論文内の聴取実験については，聴取者がいずれも国立音楽大学，もしくは，国立音楽大学大学院に所属する学生であり，回答にもそのバイアスが影響した可能性が示唆される．例えば，ピアノ演奏の評価基準については，ピアノ演奏の経験者と非経験者では異なる評価を示す傾向にあることが報告されており（宮脇ら，2016），今後は異なる聴取者群を対象とした聴取実験を行なうことで，歌声の声質の好みに対する知見が拡充されることが示唆される．

今後の展望として，様々なジャンルの楽曲を対象として，本論文と同様の実験を行なうことで，楽曲のジャンルと歌声の声質の好みにおける重要視するパラメータの差を明らかにすることができることが期待される．また，本論文の結果については，歌声の声質の好みと音響パラメータに対する一次近似的な結果を示すものであり，今後更なる検討を行なうことで，二次，三次近似的な結果をもたらすことも示唆される．

謝辞

聴取実験に参加いただいた方々に深く感謝いたします．本論文は桶本が立案，データの収集，原稿の執筆を行ない，三浦が研究全体に対する指導と助言を担当した．本研究の一部はJSPS科研費 JP21J13539の助成を受けた．

引用文献

荒木雅弘，“フリーソフトではじめる機械学習入門: Python/Weka で実践する理論とアルゴリズム 第２版”，森北出版（2018）．

Breiman, L.,“Random forests”, Machine learning, 45(1), pp.5-32(2001).

De Cheveigné, A., & Kawahara, H.“YIN, a fundamental frequency estimator for speech and music”, The Journal of the Acoustical Society of America, 111(4), pp.1917-1930 (2002).

Fastl, H.,“The hearing sensation roughness and neuronal responses to AM-tones”, Hearing Research, 46(3), pp.293-295 (1990).

藤原弘将，後藤真孝，“Vocalfinder: 声質の類似度に基づく楽曲検索システム”，情報処理学会研究報告 音楽情報科学（MUS），2007（81（2007-MUS-071）），pp.27-32（2007）.

古井貞熙，“電子・情報工学入門シリーズ 2 音響・音声工学”，近代科学社，p.25（1992）.

Goto, M.,“PreFEst: A predominant-F0 estimation method for polyphonic musical audio signals”, Proceedings of the 2nd Music Information Retrieval Evaluation eXchange (2005).

市原茂，“布の好みの個人差の因果分析的研究”，人間工学，32（1），pp.21-27（1996）.

一般社団法人レコード協会，“音楽メディアユーザー実態調査　2020年度調査結果”，https://www.riaj.or.jp/f/pdf/report/mediauser/softuser2020.pdf（2021年公開，2021.9参照a）.

一般社団法人レコード協会，“生産実績・音楽配信売上実績　過去10年間　合計”，https://www.riaj.or.jp/f/data/annual/msdg_all.html（2021.9参照b）.

伊藤雄哉，山西良典，加藤昇平，“音楽ゆらぎ特徴を用いた楽曲印象の推定”，日本音響学会誌, 68（1）, pp. 11-18（2011）.

桐本篤，佐々木史織，清木康，“風景画像とサンプル楽曲を用いた環境状況コンテクスト対応型音楽推薦システムの実現”，情報処理学会研究報告データベースシステム（DBS），2008（88（2008-DBS-146））, pp.157-162（2008）.

高校新聞ONLINE，“高校生・中学生が選んだ「好きなアーティスト」ランキング”，https://www.koukouseishinbun.jp/articles/-/6818（2021.9参照）.

小島寛之，“完全独習 統計学入門”，ダイヤモンド社（2006）.

宮脇聡史，三浦雅展，“固有演奏を用いたピアノ熟達度の評価基準における多様性の可視化手法”，日本音響学会誌，72(10)，pp.617-626 (2016).

Moorer, J. A,“Signal processing aspects of computer music”, A survey. Proceedings of the IEEE, 65（8）, pp.1108-1137（1977）.

中山功一, 乗冨喜子, 大島千佳,“顔の好みの分析”, 人工知能学会全国大会論文集 第 26 回全国大会, 3B-2-R-2-7, 一般社団法人 人工知能学会（2012）.

日本音響学会（編），鈴木陽一，赤木正人，伊藤彰則，佐藤洋，苣木禎史，中村健太郎（共著），“音響入門シリーズ A-1 音響学入門”，コロナ社，p.40（2011）.

西川直毅，糸山克寿，藤原弘将，後藤真孝，尾形哲也，奥乃博，“歌詞と音響特徴量を用いた楽曲印象軌跡推定法の設計と評価”，研究報告 音楽情報科学（MUS）2011.7，pp.1-8（2011）.

Ohishi, Y., Goto, M., Itou, K., & Takeda, K,“Discrimination between singing and speaking voices”, Ninth European Conference on Speech Communication and Technology (Eurospeech 2005), pp.1141-1144(2005).

追木智明，欅惇志，宮崎純，“物体の色や表情情報を利用した画像の印象にあった音楽推薦手法の提案”，研究報告情報基礎とアクセス技術（IFAT），2018（25），pp.1-6（2018）.

岡田創太，山口翔也，三浦雅展，“女性アイドルポピュラ音楽を対象とした動的パラメータによる年代推定システムの構築”，日本音響学会誌, 74（7）, pp.363-371（2018）.

Osgood, C. E., Suci, G. J., and Tannenbaum, P. H.“The Measurement of Meaning”, University of Illinois

press (1957).

Peeters, G. “A Large Set of Audio Features for Sound Description (Similarity and Classification) in the CUIDADO Project”, Technical Report; IRCAM: Paris, France, (2004).

Prétet, L., Hennequin, R., Royo-Letelier, J., & Vaglio, A, “Singing voice separation: A study on training data”, ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 506-510, (2019).

Russell, J. A., “A circumplex model of affect”, Journal of Personality and Social Psychology, 39(6), pp.1161–1178 (1980).

Serrà, J., Corral, Á., Boguñá, M., Haro, M., & Arcos, J. L., “Measuring the evolution of contemporary western popular music,” Sci. Rep., 2, pp.1–6 (2012).

Terhardt, E., “On the perceptions of periodic sound fluctuation (Roughness)” Acustica 30, pp.201-213, (1974).

徳田恵一，小林隆夫，今井聖，“メル一般化ケプストラムの再帰的計算法”，電子情報通信学会論文誌 A, 71（1），pp.128-131（1988）.

東京大学教養学部統計学教室,“統計学入門（Vol. 1）”，東京大学出版会（1991）.

Turumeso，“vocal-remover v4.0.0”，https://github.com/tsurumeso/vocal-remover（2019年公開，2021.9参照）.

山田真司，三浦雅展，“音楽情報処理で用いられる音響パラメータによる音楽理解の可能性”，日本音響学会誌，70(8)，pp.440-445（2014）.

山田真司，西口磯春，“音楽はなぜ心に響くのか，音響サイエンスシリーズ 4, 日本音響学会編”，コロナ社（2011）.

吉井和佳，後藤真孝，“音楽情報処理技術の最前線:7. 音楽推薦システム”，情報処理 50.8 pp.751-755（2009）.