

物理モデルに基づくエレキベース奏法変換のための励起信号解析と特徴写像

守矢龍

法政大学情報科学部デジタルメディア学科

ryu.moriya.4u@stu.hosei.ac.jp

概要

本研究では、エレキベース演奏におけるピック弾き音を指弾き音へと変換するための音響信号処理手法を提案する。まず、指弾きおよびピック弾き音源から励起信号を抽出し、主成分分析を用いて上位3次元の特徴ベクトルへ変換する。得られた指弾き音の特徴ベクトルに対してガウス混合モデルを用いたクラスタリングを行い、音色特性の類似性に基づき5つのクラスタに分類した。

次に、指弾き音およびピック弾き音の対応関係を学習するため、主成分空間上で両者を結合したガウス混合モデルを構築し、条件付き期待値に基づく特徴変換を行った。具体的には、ピック弾き音の特徴が与えられたときの指弾き音特徴の条件付き期待値を推定し、これを変換後特徴として用いることで、統計的に最も尤もらしい指弾き音特徴への写像を実現した。この変換をクラスタごとに適用することで、演奏強度や音色分布のばらつきを考慮した変換を可能とした。

提案手法の有効性を評価するため、変換後特徴が指弾き音の特徴分布にどの程度近づいたかを示す指標として、PCA空間上で算出したz-scoreを用いた距離評価を行った。その結果、クラスタ2ではz-scoreが 2.16 ± 0.57 から 0.83 ± 0.54 へと減少し、約61%の改善が確認された。クラスタ3では 3.60 ± 0.62 から 1.89 ± 0.98 (約48%減少)、クラスタ4では 2.99 ± 0.90 から 0.33 ± 0.24 (約89%減少)、クラスタ5では 4.00 ± 0.27 から 0.79 ± 0.32 (約80%減少)と、すべてのクラスタにおいて顕著な距離低減が得られた。

これらの結果から、ガウス混合モデルと条件付き期待値を用いたクラスタ適応型特徴変換により、ピック弾き音を指弾き音の特徴分布へ写像できることが示された。

1 はじめに

エレキベースは現代音楽において欠かせない楽器であり、ポピュラー音楽やロック、ジャズ、ファンクなど幅広いジャンルで使用されている。演奏方法には指弾きやピック弾き、スラップ、タッピングなど様々なスタイルがあるが、特にFig. 1の指弾きとFig. 2のピック弾きは曲中で最も頻繁に用いられる主要な奏法である。弦を弾く手段が指かピックかによって、接触する材質が異なるため、音色にも大きな差が生まれる。一般に指弾きでは柔らかく温かみのある音が得られるのに対し、ピック弾きではアタックが強く高域成分が豊かで明瞭な音を得られる。この違いは、同じ弦や同じ位置で演奏しても音の立ち上がりや減衰、倍音構成に顕著に現れ、楽曲における雰囲気やミックス上の役割に大きな影響を与える。

このような奏法差は制作現場において音楽表現の重要な要素である一方、後処理によって柔軟に制御することが難しい。イコライザやマルチバンドコンプレッサを用いた音色変換は広く用いられているが、これらは主に周波数特性やダイナミクスの調整に基づくものであり、奏法に起因する音の時間的構造や励起特性の違いを十分に再現することは困難で



Fig.1: 指弾き



Fig.2: ピック弾き

ある。例えばアタックの立ち上がりや弦と指先・ピックの接触による微細なノイズ成分、さらには倍音の時間変化といった要素は単純なフィルタリングや圧縮だけでは本質的に変換できないことが多い。そのため、演奏スタイルを変更する場合には、同一フレーズを異なる奏法で再録音する必要が生じることが多く、制作の自由度や作業効率を低下させる一因となっている。

本研究はこのような課題に対し、録音済みのエレキベース音源に対して奏法の違いを再現可能とすることを目的とする。

2 関連研究

2.1 演奏音の生成メカニズムと共鳴・励起の分離

弦楽器の演奏音は、楽器固有の共鳴特性と演奏者の奏法に依存する弦の励起過程の相互作用によって生成される。エレキベースにおいては、指弾きやピック弾きといった奏法の違いが、音の立ち上がり形状や高周波成分の分布に顕著な差

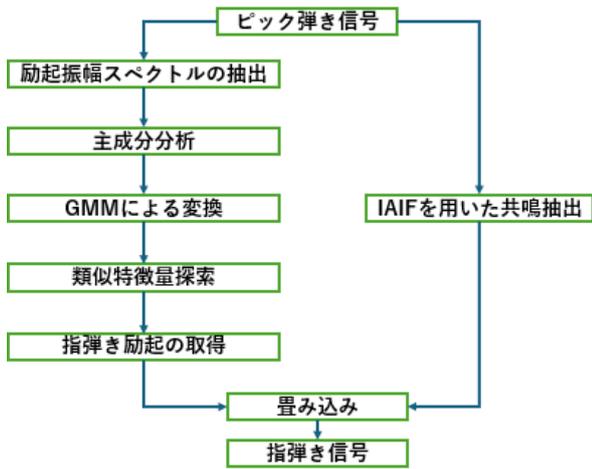


Fig.3: ピックから指弾きへの変換手法

として現れることが知られており、これらは演奏表現を特徴づける重要な要素である。

奏法差を信号処理的に扱うためには、音響信号を共鳴特性と励起信号に分離して考える枠組みが有効である。共鳴特性の推定手法としては、Iterative Adaptive Inverse Filtering (IAIF) [1] が広く用いられており、音声信号のみならず楽音信号に対しても、系の伝達特性を近似的に推定する手法として知られている。IAIF は音響信号から共鳴成分を段階的に除去することで、楽器や演奏条件に依存した共鳴特性をモデル化することが可能である。

一方、励起信号の扱いに関しては必ずしも統一的な方法論が確立されているわけではない。逆フィルタリングに基づく励起信号推定は、共鳴特性の推定には有効である一方、モデル仮定や推定誤差の影響を受けやすく、特に弦楽器音における発音直後の非定常な時間構造を十分に保持することが難しい。エレキベースの奏法差は、発音直後の短時間波形に強く現れることから、時間構造を直接保持した励起表現を扱うことが重要である。

2.2 励起信号を用いた既存研究と本研究の位置づけ

励起信号を用いた楽音合成に関する研究としては、Digital Waveguide モデルに励起信号を入力することで弦楽器音を生成する手法が報告されている [2]。また、励起信号を特徴量として表現し、ギター演奏における奏法の違いを解析する試みもあり、励起信号に奏法情報が含まれることが示唆されている [3]。これらの研究は、奏法と音響特徴の関係を理解する上で重要な知見を提供している。

しかしながら、既存研究の多くは奏法の分析や合成音の制御を主目的としており、録音済みの実音音源を対象に、ある奏法から別の奏法へ変換することを目的とした信号処理的枠組みは十分に検討されていない。特に、発音直後の励起波形を対象とし、その時間構造を保持したまま低次元特徴空間上に表現・操作する手法は、限られた検討にとどまっている。

以上の背景を踏まえ、奏法差を離散的なカテゴリとして扱うのではなく、励起信号を低次元特徴空間として表現し、その空間内で写像として奏法変換を定式化する枠組みが求められている。本研究では、発音直後の時間波形を直接切り出した励起信号を対象とし、共鳴特性とは独立に特徴空間化することで、録音済みエレキベース音源に対する奏法変換を実現することを目指す。

3 提案手法

本研究では、Fig. 3 に示すように録音済みのエレキベース音源に対してピック弾きから指弾きへの奏法変換を行う手法を提案する。弦楽器の演奏音は、楽器固有の共鳴特性と演奏者の奏法に依存する弦の励起の相互作用によって生成される。奏法差を共鳴成分とは独立した励起信号の違いとして捉え、これを低次元特徴空間上で写像することで奏法変換を実現する。本手法は、(1) 共鳴特性の抽出、(2) 励起信号の抽出と前処理、(3) 励起信号の特徴量化、(4) GMM による奏法間特徴写像、(5) 励起信号選択と音の再合成の 5 段階から構成される。

3.1 Iterative Adaptive Inverse Filtering を用いた共鳴特性の抽出

単音で録音したエレキベース音源に対して、Iterative Adaptive Inverse Filtering (IAIF) [1] を適用することで、演奏方法に依存しない楽器固有の共鳴特性を抽出する。IAIF は、励起成分と共鳴成分が畳み込み構造で観測される音響信号に対して、それぞれを段階的かつ反復的に分離する手法であり、音声信号処理分野において広く用いられている。Fig. 4 に、本研究で用いた IAIF の処理フローを示す。

まず、入力信号 $s(n)$ に対して一次の線形予測分析 (LPC) を適用し、弦の放射特性や低周波数帯域の包絡を近似する低次フィルタ $H_{g1}(z)$ を推定する。

$$H_{g1}(z) = 1 + a_1 z^{-1}. \quad (1)$$

この低次フィルタは、励起信号に由来する緩やかなスペクトル傾斜をモデル化する役割を持つ。推定された $H_{g1}(z)$ を用いて入力信号を逆フィルタリングすることで、励起成分の影響を部分的に除去し、共鳴成分を強調した信号 $s_1(n)$ を得る。

$$s_1(n) = s(n) * H_{g1}(z). \quad (2)$$

次に、 $s_1(n)$ に対して次数 r の高次数 LPC 分析を行い、楽器ボディやピックアップ系に起因する共鳴特性を表すフィルタ $H_{vt1}(z)$ を推定する。

$$H_{vt1}(z) = 1 + \sum_{k=1}^r a_k z^{-k}. \quad (3)$$

このフィルタは、弦振動による励起成分が抑圧された信号に基づいて推定されるため、楽器固有の共鳴応答をより明瞭に近似できる。

さらに IAIF の反復的枠組みに従い、推定された $H_{vt1}(z)$ を用いて入力信号を再度逆フィルタリングし、その結果に対して低次 LPC 分析を適用することで、励起成分を表す補助的な低次フィルタ $H_{g2}(z)$ を推定する。この処理により、初期段階では十分に除去しきれなかった励起成分が補正される。最後に、 $H_{g2}(z)$ により補正された信号に対して高次数 LPC 分析を行い、高い周波数分解能を持つ最終的な共鳴フィルタ $H_{vt2}(z)$ を得る。

以上の処理により、演奏強度や奏法による励起の違いの影響を抑えつつ、エレキベース固有の安定した共鳴特性を抽出することが可能となる。本研究では、このようにして得られた共鳴特性を後段の奏法変換処理に用いる。

3.2 励起信号の抽出方法

録音信号から奏法差を反映する励起信号を抽出するため、切り出し、正規化、および時間整列からなる前処理を行う。

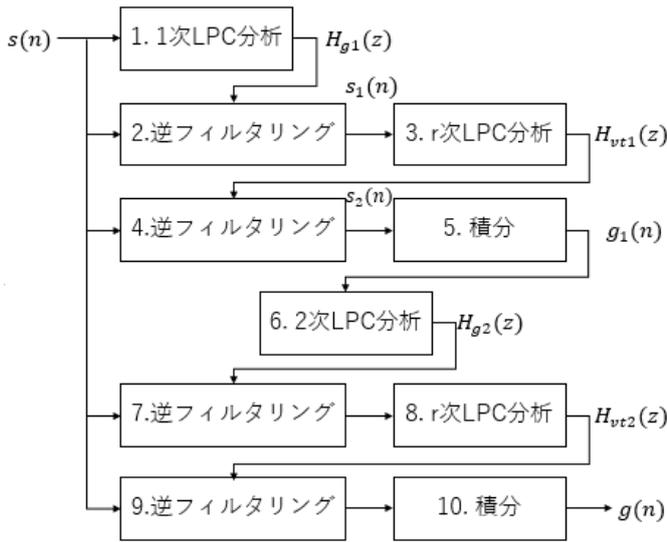


Fig.4: IAIF の処理フロー

本前処理の目的は、弦の励起直後に生じる非定常かつ高エネルギーな過渡成分を安定して抽出し、演奏方法に起因するスペクトルの違いを明確に反映させることである。

3.2.1 励起の切り出し

入力された WAV 音源の波形を時間軸上に可視化し、励起開始付近と判断される基準時刻を手動で指定する。この基準時刻は、弦が外力によって変位し、振動が開始される瞬間に対応する急峻な立ち上がり成分を含む位置として選択される。自動検出では奏法や演奏条件によるばらつきの影響を受けやすいため、本研究では再現性と信頼性を優先し、視覚的判断に基づく手動指定を採用した。

指定された基準時刻を中心として、前方向に 15 ms の区間を含むように開始点を設定する。

終了点については、基準時刻以降において信号が負から正へ遷移する最初のゼロクロス点を検出する。検出されたゼロクロス点の前後サンプルに対して線形補間を行い、振幅が 0 となる時刻を推定することで、サンプリング周期に依存しない精度で終了時刻を決定する。これにより、励起信号の主要な立ち上がり成分を安定して含む区間を抽出することが可能となる。

3.2.2 正規化と波形整列

切り出しを行い得た励起信号に対して、まず振幅方向の正規化を行う。各信号 $x[n]$ に対して

$$x_{\text{norm}}[n] = \frac{x[n]}{\max |x[n]|} \quad (4)$$

によりピーク正規化を施し、最大振幅が ± 1 となるように統一する。これにより、演奏強度の違いや録音条件のばらつきに起因する振幅スケールの影響を除去し、波形形状そのものの違いに着目した解析を可能とする。

なお、励起信号の時間長は最低でも 4 弦ベースの最低音 E (41.2 Hz) の 2 周期分を確保するように切り出す。すなわちその時間長は約 $2 \times \frac{1}{41.2} \approx 0.0485$ 秒であり、サンプリング周波数 f_s の下ではサンプル数

$$N = \left\lceil 2 \cdot \frac{1}{41.2} f_s \right\rceil$$

を満たすように切り出す。ここでは $f_s = 192 \text{ kHz}$ を想定すると、 $N \approx \lceil 0.048543689 \times 192000 \rceil = \lceil 9323.999 \dots \rceil = 9324$ サンプルを確保することになる。

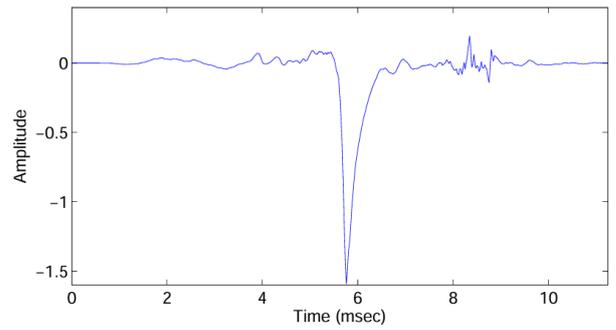


Fig.5: 正規化した励起信号

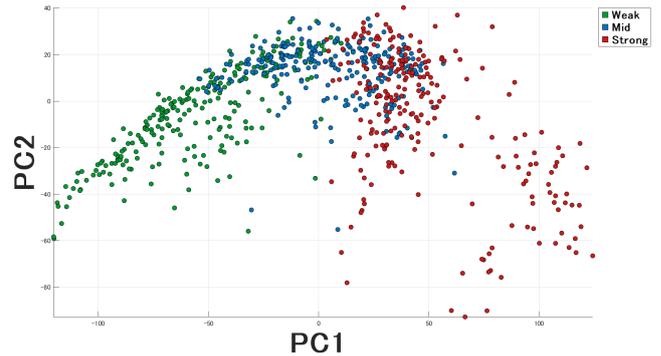


Fig.6: 強度ごとの励起信号の分布

さらに、切り出しおよび正規化後の励起信号群に対して、時間軸方向のばらつきを補正するための整列処理を行う。まず、整列の基準となる励起信号を 1 つ選択し、基準信号および整列対象信号の双方に対して振幅の 3 乗を施す。この非線形変換により、微小振幅成分の影響を抑制しつつ、励起初期における大振幅成分を強調することで、立ち上がり付近の特徴が相互相関計算においてより支配的となる。

次に、立方化した信号同士の相互相関を計算し、相関値が最大となるラグを推定する。得られたラグを用いて元の信号をシフトすることで、各励起信号の時間位置を基準信号に整列させる。

以上の処理により、振幅スケールおよび時間位置の影響が低減された励起信号が得られる。Fig. 5 に示すように、正規化後の励起信号は発音直後の極めて短い時間区間において一度急峻に下方向へ変化した後に 0 に向かう、下に凸な時間包絡を示しており、弦が瞬間的に励起された際の力学的応答が時間波形として明瞭に観測される。このように整列・正規化された信号群は、時間的に整合した初期過渡成分と統一された振幅スケールを持つため、主成分分析やクラスタリングによる特徴抽出が安定して行え、奏法差に起因する違いを捉えることが可能となる。

3.3 振幅スペクトルに対する主成分分析

抽出した励起信号に対して、まず振幅スペクトルを算出する。振幅スペクトルは、励起信号に含まれる立ち上がり形状や高周波成分の分布など、奏法差を反映する特徴を表す。

得られた振幅スペクトルは高次元であり、直接的な解析や比較が困難であるため、PCA を用いて低次元の潜在特徴空間に写像する。PCA は、スペクトル間の線形相関構造を反映した直交基底を構築し、寄与率の高い主成分のみを選択することで次元圧縮を行う。

Fig. 6 に励起信号の振幅スペクトルを PCA により低次元

空間へ写像した結果を示す。第1主成分を横軸、第2主成分を縦軸として特徴ベクトルをプロットしている。弱・中・強の演奏強度に対応する特徴ベクトルは、第1主成分の値が小さい側から大きい側へと順に配置されており、第1主成分が演奏強度を反映した軸であることが確認できる。

このように、振幅スペクトルに対する主成分分析は、励起信号に含まれる演奏強度を低次元の潜在特徴として表現することを可能にする。

3.4 ガウス混合モデルによる特徴量モデリング

ピック弾きおよび指弾きの振幅スペクトル特徴ベクトルを統計的に表現するため、本研究では結合ベクトルを用いたガウス混合モデル (Gaussian Mixture Model: GMM) による特徴量モデリングを行う。GMM を用いることで、単一の平均・分散では表現できない奏法内のばらつきや非線形な分布構造を、複数の局所的な確率分布の重ね合わせとして表現することが可能となる。

特に、同一演奏強度内であっても、弦の触れ方の違いにより励起信号およびスペクトル特徴にはばらつきが生じる。GMM はこのような確率の変動を自然に扱うことができるため、奏法変換における統計的対応関係をモデル化する手法として適している。

3.4.1 結合ベクトルの GMM 学習

低次元化された振幅スペクトル特徴ベクトルは、ピック弾きと指弾きの奏法差を反映した情報を含んでおり、これらを統計的に対応付けることで奏法変換が可能となる。本研究では、録音済み音源間の対応関係を直接モデル化するため、ピック弾き特徴と指弾き特徴を結合した特徴ベクトルを用いた GMM 学習を行う。

ピック弾き特徴 \mathbf{y} と、それに対応する指弾き特徴 \mathbf{x} を縦方向に結合し、結合ベクトル

$$\mathbf{z} = [\mathbf{y}; \mathbf{x}] \quad (5)$$

を定義する。GMM はこの結合ベクトル \mathbf{z} の分布を

$$p(\mathbf{z}) = \sum_{k=1}^K \pi_k \mathcal{N}(\mathbf{z}; \boldsymbol{\mu}_k, \Sigma_k) \quad (6)$$

として近似する。ここで K は混合成分数、 π_k は各成分の混合係数を表し、 $\mathcal{N}(\mathbf{z}; \boldsymbol{\mu}_k, \Sigma_k)$ は平均 $\boldsymbol{\mu}_k$ 、共分散行列 Σ_k を持つ多変量正規分布である。

この結合 GMM の枠組みにより、ピック弾きと指弾きの特徴空間を個別にモデル化するのではなく、両者の相関構造を含めた同時分布として表現することが可能となる。すなわち、各混合成分は、特定の演奏ニュアンスや音色傾向に対応した局所的な奏法対応関係を表していると解釈できる。

GMM の学習には期待値最大化法 (Expectation-Maximization: EM アルゴリズム) を用い、対数尤度が最大となるように各混合成分の平均 $\boldsymbol{\mu}_k$ 、共分散行列 Σ_k 、および混合係数 π_k を推定する。EM アルゴリズムを用いることで、潜在変数としての混合成分割当を考慮しながら安定したパラメータ推定が可能となる。

3.4.2 条件付き期待値による奏法変換

学習済みの GMM を用いて、入力となるピック弾き特徴 \mathbf{y} から対応する指弾き特徴 \mathbf{x} を推定する。本研究では、結合ベクトルの条件付き分布に基づく条件付き期待値を用いて奏法変換を定式化する。これは、確率論的に最も尤もらしい指弾き特徴を推定する手法であり、連続的かつ安定した特徴変換を実現できる。

各混合成分 k に対して、結合ベクトルの平均および共分散行列を

$$\boldsymbol{\mu}_k = \begin{bmatrix} \mu_{y,k} \\ \mu_{x,k} \end{bmatrix}, \quad \Sigma_k = \begin{bmatrix} \Sigma_{yy,k} & \Sigma_{yx,k} \\ \Sigma_{xy,k} & \Sigma_{xx,k} \end{bmatrix} \quad (7)$$

のように分割する。ここで $\mu_{y,k}$ 、 $\Sigma_{yy,k}$ はピック弾き特徴に対応する平均および共分散を表し、 $\mu_{x,k}$ 、 $\Sigma_{xx,k}$ は指弾き特徴に対応する。

入力ピック弾き特徴 \mathbf{y} に対する指弾き特徴の条件付き期待値は、次式で与えられる。

$$\hat{\mathbf{x}} = \sum_{k=1}^K \gamma_k(\mathbf{y}) \left[\mu_{x,k} + \Sigma_{xy,k} \Sigma_{yy,k}^{-1} (\mathbf{y} - \mu_{y,k}) \right] \quad (8)$$

ここで $\gamma_k(\mathbf{y})$ は、入力 \mathbf{y} が混合成分 k に属する事後確率であり、

$$\gamma_k(\mathbf{y}) = \frac{\pi_k \mathcal{N}(\mathbf{y}; \mu_{y,k}, \Sigma_{yy,k})}{\sum_{j=1}^K \pi_j \mathcal{N}(\mathbf{y}; \mu_{y,j}, \Sigma_{yy,j})} \quad (9)$$

として計算される。

この条件付き期待値による推定は、各混合成分における線形回帰を事後確率で重み付けした形と解釈でき、非線形な奏法対応関係を滑らかに近似する効果を持つ。これにより、入力特徴が複数の演奏ニュアンスにまたがる場合でも、それらを統合した連続的な特徴変換が可能となる。

以上の手法により、入力ピック弾き特徴に対して統計的に最も適合する指弾き特徴を推定することができる。確率的対応付けに基づく奏法変換であるため、同一演奏強度内での微細なニュアンスのばらつきを保持しつつ、平均的な指弾き音色への変換が実現される。

3.5 類似特徴量の探索

機械学習によって得られた指弾き特徴量に基づき、録音済みの指弾きオリジナル音源データベースの中から、最も類似する励起信号を探索する。本処理の目的は、GMM によって推定された連続値の指弾き特徴を、実在する励起信号にすることで、物理的に妥当かつ自然な励起波形を用いた音生成を可能とすることである。すなわち、統計モデルによる特徴変換と、実データに基づく波形選択を組み合わせることで、過度な平滑化や不自然な励起形状の発生を抑制する。

探索は、GMM によって出力された n 次元の指弾き特徴量 $\mathbf{y} = [y_1, \dots, y_n]$ と、指弾き励起信号のデータベース \mathcal{D} に含まれる各励起信号に対応する n 次元特徴量 $\mathbf{x}_i = [x_{i,1}, \dots, x_{i,n}]$ との距離を計算することで行う。本研究では、特徴空間上での類似度指標としてユークリッド距離を採用し、距離が最小となる特徴量に対応する励起信号を類似信号として選択する。

距離計算においては、各主成分が全体の分散に寄与する割合の違いを考慮するため、主成分分析により得られた寄与率 w_j を用いた重み付き距離を定義する。これにより、寄与率の高い主成分、すなわち音色差や奏法差を強く反映する成分が、探索結果により大きな影響を与えるように調整される。

$$\mathbf{x}_{\text{sim}} = \mathbf{x} \in \mathcal{D} \sqrt{\sum_{j=1}^n w_j (y_j - x_{i,j})^2} \quad (10)$$

この重み付き距離は特徴空間の統計的構造を反映した類似度評価を実現する。特に、低寄与率成分に含まれる微小なノイズ成分や個体差の影響を抑制しつつ、主要な音色・励起特性に基づく安定した探索が可能となる。

以上の類似特徴量探索により、GMMによって推定された指弾き特徴と最も整合する実在の励起信号を選択することができる。

3.6 音の再合成

得られた指弾き特徴量に対応する励起信号 x_{sim} に対して、前述の IAIF により推定した楽器共鳴フィルタ $H_{\text{res}}(z)$ を畳み込むことで、実際の演奏音に必要な音程・減衰特性・共鳴構造を付与する。時間波形 $x_{\text{sim}}[n]$ に対して線形畳み込みを行い、最終的な出力信号 $s_{\text{out}}[n]$ を得る。

$$s_{\text{out}}[n] = x_{\text{sim}}[n] * h_{\text{res}}[n] \quad (11)$$

ここで $h_{\text{res}}[n]$ は $H_{\text{res}}(z)$ のインパルス応答を表す。

4 実験条件

4.1 評価方法

本研究では、提案手法によって変換された特徴量が、指弾き演奏に由来する特徴量分布への程度近づいたかを定量的に評価することを目的とする。評価は、特徴量空間を事前に複数のクラスタに分割した上で、クラスタごとに独立して行った。これにより、演奏強度や音色特性の異なる領域を分離した評価が可能となる。

まず、各クラスタに属する指弾き演奏の特徴量集合を基準分布とし、その平均ベクトル μ_k および標準偏差ベクトル σ_k を算出した。ここで k はクラスタ番号を表す。次に、ピック弾き演奏および変換後特徴量 \mathbf{x} に対して、対応するクラスタの基準分布を用いて正規化を行い、次式で定義される正規化特徴量 \mathbf{z} を求めた。

$$\mathbf{z} = \frac{\mathbf{x} - \mu_k}{\sigma_k} \quad (12)$$

この正規化により、各特徴次元のスケール差を除去し、次元間で均等に距離を評価できるようにしている。さらに、正規化特徴量に対してユークリッドノルムを計算し、基準分布からの乖離度を表す正規化距離 D_z を次式で定義した。

$$D_z = \|\mathbf{z}\|_2 \quad (13)$$

各クラスタにおける評価値は、対象となる全サンプルの D_z の平均および分散として集約した。これにより、個々のサンプルに依存しない統計的な比較を可能とした。

提案手法の効果は、ピック弾き演奏に対する正規化距離の平均値 D_z^{before} と、変換後特徴量に対する正規化距離の平均値 D_z^{after} を比較することで評価した。さらに、変換による改善度を相対的に表す指標として、次式で定義される正規化距離の減少率 R_z を導入した。

$$R_z = \frac{D_z^{\text{before}} - D_z^{\text{after}}}{D_z^{\text{before}}} \quad (14)$$

本研究では、 D_z が小さいほど特徴量が指弾き分布に近いことを示すものと解釈し、 R_z が正の値をとる場合、提案手法によって指弾き演奏の特徴量分布への接近が達成されたと判断する。

さらに、定量評価に加えて、変換後特徴量の分布構造を視覚的に検証するための評価を行う。PCAにより得られた低次元特徴空間上において、指弾き演奏、ピック弾き演奏、および変換後特徴量の散布図を作成し、それぞれの分布関係を比較する。

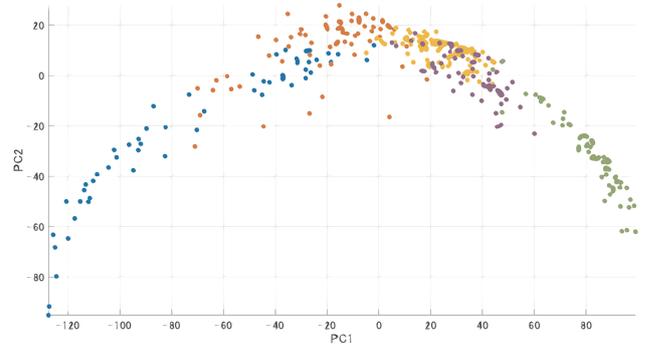


Fig.7: K=5 のピック弾き学習データの分布

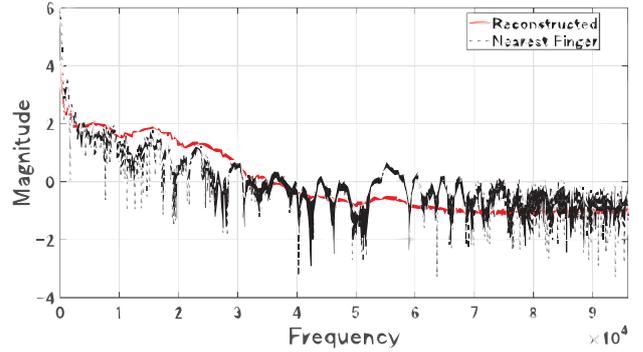


Fig.8: 3次元を用いた類似スペクトルの探索

この散布図により、変換後特徴量が指弾き分布に接近しているかどうかに加え、各クラスタにおける分布の外形や分散構造が変換前後でどの程度保持されているかを確認した。特に、クラスタ構造が大きく崩れることなく指弾き分布側へ移動している場合、奏法差を反映した統計的写像が適切に学習されていると解釈できる。

4.2 録音データ

ピック弾きと指弾きの音響的差異を明確に捉えるため、アコースティックベースを使用して音源を収集する。このベースには圧電素子を内蔵したピエゾピックアップが搭載されており、弦の圧力変化を高精度に検出できる。このピックアップはローパス特性を持ちにくく、励起成分が明瞭に記録されるため、演奏スタイルの違いを捉えるのに適している。

ピック弾きには、セルロイド製・厚さ 0.8mm・おにぎり型の Fender Medium Pick を使用する。録音環境は、Studio One 6 (Windows 環境) を使い、Rubix24 オーディオインターフェースを介してライン録音を行う。録音はモノラル入力・192kHz・32bit の条件で行い、各サンプルの録音時間は約 4 秒とする。

録音対象は、ベースの 1~4 弦の開放弦から第 5 フレットまでの各ポジションで、ピック弾きと指弾きのをそれぞれ 5 回ずつ録音。演奏の強さは両奏法で弱、中、強で対応付けを行い、隣接弦をミュートした単音演奏のみを対象とし、和音やコード演奏は含まず、合計 720 個のデータを録音した。

加えて、写像結果を確認するためにピック弾き弱、中、強それぞれ 48 データずつ、合計 144 個のデータを録音した。

4.3 GMM のパラメータ設定

PCAにより低次元化された振幅スペクトル特徴ベクトルに対して、ピック弾きと指弾きの対応関係を統計的に学習するため、Gaussian Mixture Model を適用した。GMM の学習には、録音データ 720 個から構成される結合ベクトル

$\mathbf{z} = [\mathbf{y}; \mathbf{x}]$ を用い、ピック弾き特徴 \mathbf{y} が与えられたときの指弾き特徴 \mathbf{x} を条件付き期待値として推定可能な枠組みを構築した。

GMM の主要なパラメータは以下の通り設定した。

- 成分数 $K = 5$: Bayesian Information Criterion (BIC) に基づき最適化。
- 共分散行列のタイプ: `full`。
- 正則化項: `RegularizationValue = 1e-6` (共分散行列の非正定値化を回避)。
- 初期化方法: k-means による 10 回の再初期化。
- EM アルゴリズム: 最大反復回数 500, 収束判定閾値 `TolFun = 10-6`。

本研究では、主成分数を 3 に制限して GMM を学習した。この設定により、EM アルゴリズムの収束安定性および数値的安定性が向上することを確認している。高次元の主成分を用いた場合には、クラスタリングの一致性が低下し、共分散行列の条件数が大きくなる傾向が見られた。

また、BIC により選択された成分数 $K = 5$ に基づいてクラスタリングを行った結果、Fig. 7 に示すように、特徴空間上で各クラスが適切に分離されていることを確認した。このことから、 $K = 5$ は本研究で扱う奏法特徴を過不足なく表現可能な成分数であると判断した。

加えて、主成分寄与率の上位 3 次元のみを用いた特徴空間においても、類似探索の結果は Fig. 8 に示すように、高次元特徴空間における分布の大きな外形を良好に保持していることが確認された。

以上の理由から、本研究では主成分数を 3, GMM の成分数を $K = 5$ とする設定を最終的に採用した。この設定により、同一演奏強度内におけるニュアンスのばらつきを保持しつつ、奏法差の統計的写像を安定して学習できる。

5 変換後音の特徴

ピック弾き音の特徴量を入力とし、GMM により学習した条件付き期待値を用いて、対応する指弾き音の特徴量へ変換を行った。この変換により、特徴量空間上においてピック弾き分布から指弾き分布へ写像された特徴量が得られた。

変換後の特徴量を用いて再構成した音を確認した結果、ピック弾き音を指弾きに近い音色へ変換できていることが確認された。

Fig9 に示すように、変換前後のスペクトルを比較すると、変換後の赤色のスペクトルは 100Hz 以降の周波数帯域が全体的に低下していることがわかる。これは、ピック弾きに比べて指弾き音は高次倍音成分が控えめで、音の硬さが抑えられることに対応しており、変換後の音がより柔らかく滑らかな印象を持つことと一致している。このスペクトルの変化は、変換によって指弾きの音色特徴が再現されていることを示している。加えて、Fig10 に示すように PCA の上位 3 次元で構成した信号と、探索によって得られたスペクトルと比較すると、大まかな包絡を捉えていることが確認できる。

5.1 結果

提案手法の性能を、クラス単位で算出した正規化距離 D_z およびその減少率 R_z に基づいて示す。各クラスは演奏強度の違いを反映しており、クラス番号が大きいほど強い演奏に対応する。

Table1 に、各クラスにおけるサンプル数 N , 変換前後の正規化距離の平均値と標準偏差, および減少率を示す。

クラス 1 については、評価対象となるピック弾きおよび変換後のサンプルが存在せず、サンプル数 $N = 0$ であっ

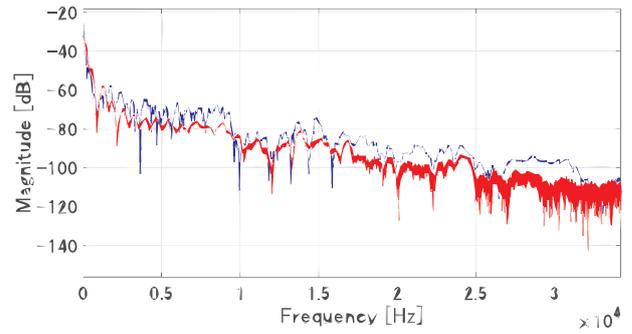


Fig.9: 変換前後のスペクトル比較

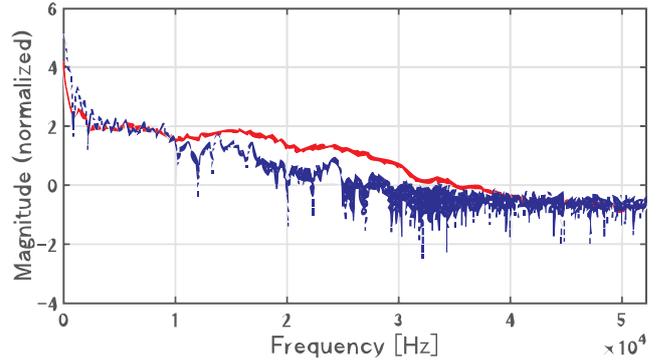


Fig.10: 指弾きスペクトルとの比較

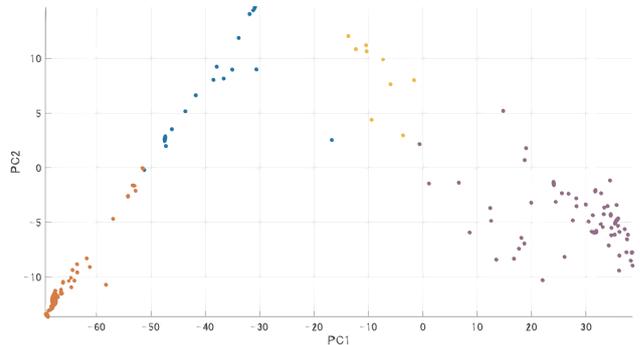


Fig.11: 変換後の特徴量分布

ため、正規化距離および減少率は算出されていない。したがって、本研究ではクラス 1 を定量評価の対象外とし、以下ではクラス 2 からクラス 5 までの結果について述べる。

まず、全ての評価対象クラスにおいて、変換後の正規化距離 D_z^{after} は、変換前の D_z^{before} と比較して明確に低下していることが確認された。この結果は、提案手法によって特徴量が指弾き演奏に由来する特徴量分布へ一貫して近づいたことを示している。

弱い演奏に対応するクラス 2 では、サンプル数は $N = 62$ と最も多く、変換前の正規化距離は $D_z^{\text{before}} = 2.16 \pm 0.57$ であった。これに対し、変換後の正規化距離は $D_z^{\text{after}} = 0.83 \pm 0.54$ まで減少しており、減少率は $R_z = 0.61$ であった。この結果は、多数のサンプルに対して安定した改善効果を得られていることを示している。

中程度の演奏強度に対応するクラス 3 では、サンプル数は $N = 9$ と比較的少ないものの、変換前の正規化距離は 3.60 ± 0.62 , 変換後は 1.89 ± 0.98 となり、正規化距離の低下が確認された。減少率は 0.48 と他のクラスに比べて低

Table 1: クラスタ別の評価結果

Cluster	N	D_z^{before}	D_z^{after}	R_z
1	0	—	—	—
2	62	2.16 ± 0.57	0.83 ± 0.54	0.61
3	9	3.60 ± 0.62	1.89 ± 0.98	0.48
4	18	4.00 ± 0.27	0.79 ± 0.32	0.80
5	55	2.99 ± 0.90	0.33 ± 0.24	0.89

Table 2: 学習データにおけるクラスタ別 CovDet

Cluster	CovDet
1	2.67×10^{-3}
2	5.77×10^{-4}
3	6.64×10^{-5}
4	8.55×10^{-3}
5	2.21×10^{-2}

い値を示したが、変換後の距離が一貫して小さくなっている点から、提案手法の有効性は維持されているといえる。

強い演奏に対応するクラスタ 4 では、サンプル数は $N = 18$ であり、変換前の正規化距離は 4.00 ± 0.27 と全クラスタ中で最も大きな値を示した。しかし、変換後には 0.79 ± 0.32 まで大きく減少し、減少率は 0.80 に達した。これは、提案手法が強い演奏条件において顕著な分布接近効果を持つことを示している。

さらに、最も強い演奏に対応するクラスタ 5 では、サンプル数は $N = 55$ と十分に確保されており、変換前の正規化距離は 2.99 ± 0.90 、変換後は 0.33 ± 0.24 となった。このときの減少率 R_z は 0.89 と全クラスタ中で最大であり、提案手法が高強度演奏において特に高い性能を発揮することが確認された。

加えて、定性的な検証として、Fig. 11 に示す変換後特徴量の散布図を確認した。この図では、変換後の特徴量が指弾き演奏の分布方向へ移動している一方で、各クラスタに対応する分布の外形や相対的な配置関係は大きく崩れていないことが確認できる。すなわち、提案手法は基準分布への接近を達成しつつ、演奏強度に起因するクラスタ構造を保持したまま特徴量を変換できていることが示唆される。

以上の結果から、提案手法はサンプル数の多少にかかわらず、全ての評価対象クラスタにおいて正規化距離を低減させることに成功しており、特に演奏強度が高いクラスタにおいて顕著な改善効果を示したことが明らかとなった。

6 考察

前章で示したクラスタ別評価結果に加え、本章では学習データにおける特徴量分布の統計的性質を踏まえ、演奏強度ごとの距離減少率の違いについて考察する。特に、共分散行列式 (CovDet) による分布の広がり、学習段階における Pick 弾きと Finger 弾きの分布間距離が、奏法変換性能に与える影響に着目する。

Table 2 に、z-score 正規化した PC1-PC3 空間における学習データのクラスタ別 CovDet を示す。CovDet は特徴量分布の体積的な広がりを表す指標であり、値が大きいほど分布が広く、小さいほど特徴量が狭い領域に集中していることを意味する。

さらに、学習データにおけるピック弾きと指弾きの分布間距離を Table 3 に示す。各クラスタにおける分布間距離は、奏法差に起因する特徴構造の明確さを示す指標として解釈

Table 3: 学習データにおけるクラスタ別 Pick-Finger 分布間距離

Cluster	Mean Distance (Pick → Finger mean)
1	67.22
2	28.12
3	43.87
4	52.94
5	53.04

できる。

まず、クラスタ 4 およびクラスタ 5 では、CovDet がそれぞれ 8.55×10^{-3} 、 2.21×10^{-2} と他のクラスタに比べて大きく、分布間距離も高い値を示している。これは、強い演奏条件に対応する特徴量分布が、特徴空間内で明確かつ十分な広がりを持ち、奏法差が顕著であることを示している。このようなクラスタでは、GMM による分布構造の捉えやすさと条件付き期待値変換の方向性が安定しているため、評価結果において高い距離減少率が得られたと考えられる。

一方、クラスタ 3 では CovDet が 6.64×10^{-5} と極めて小さく、分布間距離も他のクラスタと比較して低い。このクラスタは中強度の演奏に対応し、弱から強への連続的な変化を示す遷移的な領域に位置している。PCA 空間ではその変化が低次元に圧縮され、特徴差が顕在化しにくいいため、条件付き期待値による変換において明確な方向が定まりにくく、距離減少率が相対的に小さくなったと考えられる。

また、クラスタ 2 では CovDet が 5.77×10^{-4} とクラスタ 3 よりも大きく、分布間距離も一定程度確保されている。このクラスタでは、変換後に一貫した距離低減が観測され、中程度の距離減少率が得られたと解釈できる。

以上より、本研究における奏法変換性能は、単純なサンプル数の多寡ではなく、学習データにおける分布の広がり、ピック弾きと指弾きの分布間距離の二つの要因に強く依存していることが明らかとなった。特に、CovDet が大きく、分布構造が明確な弱強クラスタでは、GMM によるクラスタリングおよび条件付き変換が効果的に機能し、高い距離減少率につながったと考えられる。一方で、中強度クラスタのように分布が狭く遷移的な領域では、変換方向の不確実性が高まり、性能向上に限界が生じることが示唆された。

7 おわりに

本研究では、物理モデルに基づく枠組みのもと、エレキベース演奏におけるピック弾き音を指弾き音へと変換するための音響信号処理手法を提案した。奏法差を共鳴特性とは独立した励起信号の違いとして捉え、発音直後の励起信号を抽出・正規化・整列した上で、振幅スペクトルに対する主成分分析により低次元特徴空間へ写像した。

得られた特徴空間において、指弾き音の特徴分布をガウス混合モデルによりクラスタリングし、演奏強度や音色特性の違いを反映したクラスタ構造を構築した。さらに、ピック弾き音と指弾き音の特徴を結合した GMM を学習し、条件付き期待値に基づく特徴変換を行うことで、ピック弾き音を統計的に最も尤もらしい指弾き音特徴へ写像する手法を実現した。

評価実験では、PCA 空間における z-score 正規化距離を用いて、変換前後の特徴分布の近接度を定量的に評価した。その結果、すべてのクラスタにおいて変換後特徴が指弾き音の分布に近づくことが確認され、特に分布の広がりが大きく構造が明確なクラスタでは高い距離減少率が得られた。こ

これらの結果から、提案手法が奏法差に起因する励起信号の特徴を捉え、クラスタ適応的に奏法変換を行えることが示された。

今後の課題としては、学習データ数を増加させることで、主成分分析において使用可能な主成分数を増やすことが挙げられる。現状では、上位3次元の主成分を用いた解析に基づいて奏法変換を行っているが、データ数の制約により、高次の主成分に含まれる微細な奏法差や個体差を十分に活用できていない可能性がある。学習データを拡充することで、高次元特徴空間においても安定した統計モデリングが可能となり、より多様な励起信号の構造を反映した奏法変換が実現できると期待される。

以上より、励起信号の物理的解釈と統計的特徴写像を組み合わせることで、録音済みエレキベース音源に対する奏法変換を可能とする枠組みを示した。提案手法は、音楽制作や演奏支援における奏法制御の新たな可能性を示すものであり、今後の発展が期待される。

参考文献

- [1] Paavo Alku. Glottal wave analysis with pitch synchronous iterative adaptive inverse filtering. In *Proceedings of the 2nd European Conference on Speech Communication and Technology (EUROSPEECH '91)*, pp. 1081–1084, Genova, Italy, September 1991. ISCA. Helsinki University of Technology, Acoustics Laboratory.
- [2] Matti Karjalainen, Vesa Välimäki, and Zoltán Janosy. Towards high-quality sound synthesis of the guitar and string instruments. In *Proceedings of the International Computer Music Conference (ICMC)*, pp. 56–63, Tokyo, Japan, 1993. International Computer Music Association.
- [3] Raymond V. Migneco and Youngmoo E. Kim. A component-based approach for modeling plucked-guitar excitation signals. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, pp. 63–66, Ann Arbor, Michigan, USA, 2012. University of Michigan.