

エレキギター演奏自動評価のための音響的特徴量の調査

下尾 波輝^{1,a)} 矢谷 浩司^{1,b)}

概要：独学でも効率良くエレキギターを練習できるシステムはこれまでに開発されてきた。しかしこれらの大半は、指定したタイミングで正しい高さの音を鳴らせたかのみを演奏評価の基準としている。実際の演奏においては音色やテクニックなど、他の観点からも演奏を評価する必要がある。そこで本研究では、より多角的な観点から自動的にエレキギター演奏評価を行うシステムを実現することを目指し、エレキギター演奏の評価に活用できる音響的特徴量を探る。5人のエレキギター初心者による演奏音源と3人のエレキギター経験者による主観的評価をデータセットとし、様々な音響的特徴量との関係性を検証した。本稿では、音響的特徴量とエレキギター経験者による評価との関係性の分析結果を報告する。

Investigating Acoustic Features for Automatic Assessment of Electric Guitar Performance

NAMIKI SHIMOO^{1,a)} KOJI YATANI^{1,b)}

1. はじめに

楽器演奏は広く一般的に行われているが、“*the most complex forms of skilled serial actions*”の1つであると言われるほど難解な動作を要するものである [1]。そのため自分の思い通りに楽器を演奏できるようになるには、多くの時間と労力を練習に費やす必要がある。広く親しまれている楽器の1つであるエレキギターにおいてもこのことは例外ではなく、エレキギター練習支援のためのシステムがこれまでに数多く開発されてきた。

例えば、エレキギター練習のためのシステムやゲームに Yousician^{*1}や Rocksmith^{*2}がある。Yousician は流れてくるギターの譜面に合わせて実際にギターを弾くシステムであり、基本的な演奏スキルの習得や楽曲演奏の練習を行うことができる。Rocksmith には基礎的な練習のためのミニゲームがあり、ゲーム感覚で演奏スキルを高められ

る。しかしこれらのようなシステムやゲームは、十分に練習を支援できているとは言えない。なぜなら、演奏に対する評価が基本的には「正しいタイミングで正しい音程で弾けたか」のみしか存在しないためである。実際には楽器演奏の評価項目は音色や表現の質など多岐に渡る。しかしその評価は人間が行う必要があるため、個人で練習をしているときはフィードバックを得る機会がなく、特に初心者にとっては本当にうまく弾けているのかの判断が難しい。

この問題を解決する方法として、人間による演奏評価をコンピュータで代替することが考えられる。しかしながら、人間による演奏評価項目は抽象的なものが多い。例えば、適切な演奏評価項目を定めるための研究がこれまでにいくつか行なわれてきたが、導き出された項目は“Tone quality is beautiful.”など評価者の感覚に委ねられるものが数多く見られる [2]。そのため、演奏の音響信号と人間による評価を結びつけるためには、聴き手の評価に影響を与える音響的特徴量を見つけ出す必要がある。

本研究では、エレキギターの音響信号から演奏を自動評価するための音響的特徴量に関する考察を行った。具体的には、5人のエレキギター初心者による演奏音源と3人のエレキギター経験者による主観的評価をデータセットとし、私たちの定義した音響的特徴量との関係性を検証した。

¹ 東京大学大学院 工学系研究科
Interactive Intelligent Systems Laboratory,
Graduate School of Engineering, The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan

a) shimoo@iis-lab.org

b) koji@iis-lab.org

*1 <https://yousician.com/>

*2 <https://rocksmith.ubisoft.com/rocksmith/en-us/home/>

その結果、エレキギター演奏の自動評価に利用できる音響的特徴量として、*Lag*（音の鳴り止み始めから次の音が鳴るまでの時間）と、*Tone Stability* および *Tone Variance*（メル周波数ケプストラム係数を用いた音色の不安定さを表す指標）が発見された。本稿ではエレキギターの演奏データの収集とラベル付け、および得られたデータを用いた分析の結果について報告する。

2. 関連研究

音楽演奏に関する研究は数多く行なわれている。本章では音楽演奏に関する研究の中から、1) ギター演奏分析、2) 演奏評価、3) 演奏評価を利用したシステム、の3つに焦点を当てた研究例を紹介する。

ギター演奏の分析に関する研究

ギターの演奏に関する研究はこれまでも数多く行われてきた。その中でも、弦を押さえる左手の運指（*fingering*）に関する研究 [3] が数多く見られる。ギターは同じフレーズでも運指が何通りも存在するため、ギターの譜面（TAB 譜）を生成する際には運指をいかにして決定するかという問題が必然的に発生する。その運指の決定には、人間の身体的特徴などを考慮した演奏可能性や演奏難易度が制約として課される。そこで Heijink らはプロのギタリストの演奏を分析することで、ギター演奏における左手の運指と演奏難易度の関係について調査した [4]。その結果、左手の指の開く幅が広いフレーズほど難しいことなどが判明している。また有賀らは、ポピュラー音楽におけるコードの出現率とコードの難易度を考慮した曲の重要度を定義し、その重要度に基づいてギターの練習を行うことができる *Strummer* を作成した [5]。Strummer のような大量の既存曲の分析結果に基づいて練習を提供するシステムは、初心者が演奏可能曲を効率的に増やすことを支援できる。

ギター演奏特有の音楽信号処理に関する研究も存在する。例えばギターは右手で弦の振動をわずかに押さえながら弾くブリッジ・ミュートや、左手の指で弦を叩くように押さえ音を出すハンマリングなどの奏法があるが、Reboursière らはそのようなギター特有の奏法を音楽信号から検出する手法を考案し、90%以上の精度で奏法を検出することに成功した [6]。本研究においても Reboursière らの研究のように、ギター特有の奏法や指の動きなどに着目して分析を進めていく必要がある。

演奏評価手法に関する研究

楽器演奏から習熟度を推定する研究の1つに、Nonogaki らの行ったピアノ習熟度推定に関する研究 [7] がある。この研究で Nonogaki らは、ピアノ演奏における各音から4つの音楽的特徴量（onset time/velocity/duration/tempo）を用いてピアノの習熟度を推定しようとしたが、これらの

特徴量はすべて時間に関するものであり、音の強弱などについては触れていない。また、楽器ではなく歌のうまさの評価に関する研究も多い。Nakano らは音程の正確さだけでなく、歌にとって重要な要素の1つであるビブラートを用いた歌声評価手法を考案した [8]。Nakano らの研究のように、本研究においてもギターに特徴的な要素を考慮することは重要である。

演奏者の楽器習熟度を利用したシステム

演奏者の楽器習熟度を利用したシステムに関する研究もいくつか行われている。Yazawa らはギターのTAB譜の生成に関して、藤田らはピアノ譜の生成に関して、演奏者の楽器習熟度に応じた楽譜を生成する手法 [9], [10] を提案した。例えば Yazawa らの手法においては、「指を大きく開く運指があるフレーズは難しいため初心者の譜面にはなるべく現れないようにする」ことなどが可能である。しかしこれらの研究は、習熟度を演奏者自らが判断して入力する必要がある。本研究で行う自動演奏評価によって、このような習熟度を入力として用いるシステムがさらに発展することが期待される。

3. 演奏評価項目と対応する音響的特徴量の検討

本研究では、エレキギターの演奏自動評価に利用できる音響的特徴量を見つけるために、エレキギターの演奏音源から得られる音響的特徴量と人手による演奏評価の関係性を調べる。そのためにまず、エレキギター演奏における評価項目と、その評価項目に対応すると考えられる音響的特徴量について検討する必要がある。

音楽の演奏評価に関する研究は数多く行われてきた。その中でも、Brian らが行った Guitar Performance Rating Scale (GPRS) に関する研究 [2] は、本研究において特に参考になる。Brian らは過去の文献やギター講師たちの助言を参考に、129個のクラシックギター演奏評価項目を列挙した。さらに、ギター初心者からプロまで幅広いレベルの実験参加者からクラシックギター演奏の音源を集め、それを他の実験参加者に評価してもらったデータを用いて因子分析を行った。その結果、GPRS には5つの主要な因子 (a) Interpretation/Musical Effect, (b) Intonation, (c) Rhythm/Tempo, (d) Tone, (e) Technique があることがわかっている。

上記の5つの因子のうち (a) Interpretation/Musical Effect は音楽表現に関するものであり、演奏者や聴き手による評価の差が大きいと考えられるうえ、与えられた譜面通りに演奏するときは考慮されない要素が多い。また、(b) Intonation はピッチの正確性に関するものであるが、ギターのチューニングなどの演奏に関係ない要素の影響が大きい。(c) Rhythm/Tempo はあらゆる種類の音楽演奏

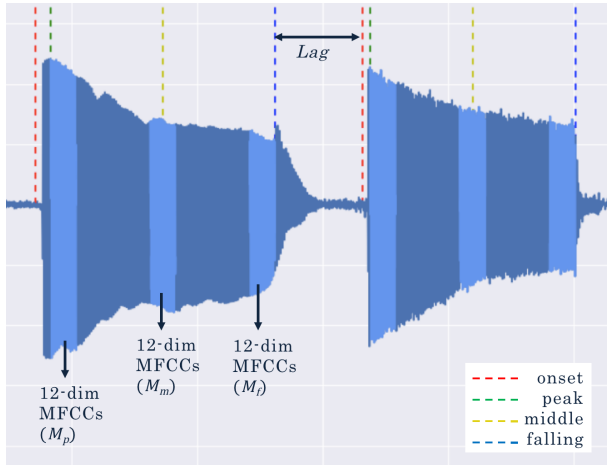


図 1: エレキギターの声の波形例と今回の分析で用いる特徴量

Fig. 1 An example of guitar sound waveform and features used in this work.

において重要な要素であるが、今回は既存の練習システムのようにある程度正しいタイミングで音を鳴らしている部分のみを抽出して分析を行うため、評価の差が小さいと考えられる。そこで本研究では残りの (d) Technique と (e) Tone の 2 つを演奏評価の項目とする。次に、これらに対応する音響的特徴量について検討する。

Technique

ギターでは基本的に、左手の指で弦を押さえて右手の指かピックで弾いて音を出すという動作の繰り返しによってメロディーを奏でる。例えば “< 3, 7 >, < 2, 5 >”^{*3} という連続する 2 音を弾くとき, “(1) 左手薬指で 3 弦 7 フレットを押さえて 3 弦を右手で弾く. (2) 左手薬指を 3 弦から離して左手人差し指で 2 弦 5 フレットを押さえ, 右手で 2 弦を弾く” という一連の動作が必要となる。このとき左手と右手の動きをうまく合わせることが重要である。

左手と右手の同期のうまさは、Heijink らの研究 [4] や Brian らの作成した GPRS [2] でも評価項目の 1 つとして用いられている。左手と右手の動きがあていないと、音が十分に伸びないなどの問題が発生する。上記の例を用いると、< 2, 5 > を弾くために左手の指を早く動かすすぎた場合、< 3, 7 > の音が早めに途切れてしまって 2 つの音がなめらかに繋がらない。そこで “現在鳴っている音の鳴り止め始め (falling) から次の音の立ち上がり (onset) までの時間” を *Lag* と定義し、なめらかに弾けているかに関する音響的特徴量とする。

onset の検出は、Spectral Flux を利用した手法 [11] を用いて行う。falling の検出には、Spectral Flux を応用した以下の指標を用いる。

$$SF'(n) = \sum_{k=1}^N \max \{0, a_k \cdot (|X(n-1, F_k)| - |X(n, F_k)|)\}$$

*3 <s,f> で s 弦 f フレットの音を表すこととする。

$|X(n, F_k)|$ は時刻 n における周波数 F_k の振幅であり、 F_k は基本周波数に対する第 k 倍音の周波数である。 $X(n, F_k)$ を求めるのには Constant-Q Transform [12] を用いる。また、 a_k はそれぞれの周波数に対する重み付け係数である。hop size を 256 sample として各時刻 n における $SF'(n)$ を求め、各 onset 区間において $SF'(n)$ が最大となる n を falling の時刻とする。今回は $N = 2$, $a_1 = 1$, $a_2 = 2$ として計算を行った。

Tone

同じ種類の楽器であっても、使用する機材や演奏者によって音色は異なる。エレキギターにおいてもそれは例外でなく、たとえ同じ機材を使っていたとしても演奏者によって音の柔らかさなどに違いが生じる。音色は聴き手の好みの差もあり、一概にどの音色が良いと決めることは難しい。しかし、演奏の際の不適切な指の動きによって悪い音色が引き起こされる場合がある。例えば左手の指でうまく弦を押さえられていなかったときなどは、音に「ビビリ」が生じて音色が不安定になる。そこでこのような音色の不安定さを定量化することにより、悪い音色を検知することができると考えられる。この音色の安定さを求めるための指標として、1 つの音に対する *Tone Stability* (TS) を以下のように定義する。

$$TS_{pm} = \frac{M_p \cdot M_m}{\|M_p\| \|M_m\|}$$

$$TS_{pf} = \frac{M_p \cdot M_f}{\|M_p\| \|M_f\|}$$

M_p , M_m , M_f はそれぞれ peak (振幅が最大となる時点), middle (peak と falling の中間), falling (音の鳴り止め始め) の 3 時点における、周辺 1024 sample の 12 次元のメル周波数ケプストラム係数 (MFCC) である。2 時点間の MFCC の cos 類似度を計算したものが TS であり、音色が不安定であるほど TS が小さくなると想定される。また M_p , M_m , M_f の 3 つを同時に考慮した指標として以下の *Tone Variance* (TV) を定義する。

$$TV = 1 - \|\overline{M}\|$$

$$\overline{M} = \frac{1}{3}(\hat{M}_p + \hat{M}_m + \hat{M}_f)$$

ここで $\hat{M}_p$, $\hat{M}_m$, $\hat{M}_f$ はそれぞれ、 M_p , M_m , M_f を正規化したものであり、音色が不安定であるほど TV が大きくなると想定される。

4. 使用するデータセット

分析に使用する演奏音源データを収集するために、エレキギター初心者 5 人に 3 つの簡単なフレーズを繰り返し弾いてもらい、その音を録音した。その後、録音した演奏に対してエレキギター経験者 3 人が評価を行った。本節ではこれらのデータの収集方法について述べる。

表 1: 演奏音源収集の手順
Table 1 Performance Data Collection Procedure.

内容	時間	注釈
実験説明	10 分	
エレキギター指導	20 分	ピックの持ち方や弦の押さえ方など (以下を 4 回繰り返す。適当に休憩を挟む。)
(A) の練習	約 8 分	
(A) の録音	約 1 分	
(B) の練習	約 10 分	P2・P5 はフレーズを前後半にわけた。
(B) の録音	約 1 分	"
(C) の練習	約 10 分	P2・P5 はフレーズを前後半にわけた。
(C) の録音	約 1 分	"

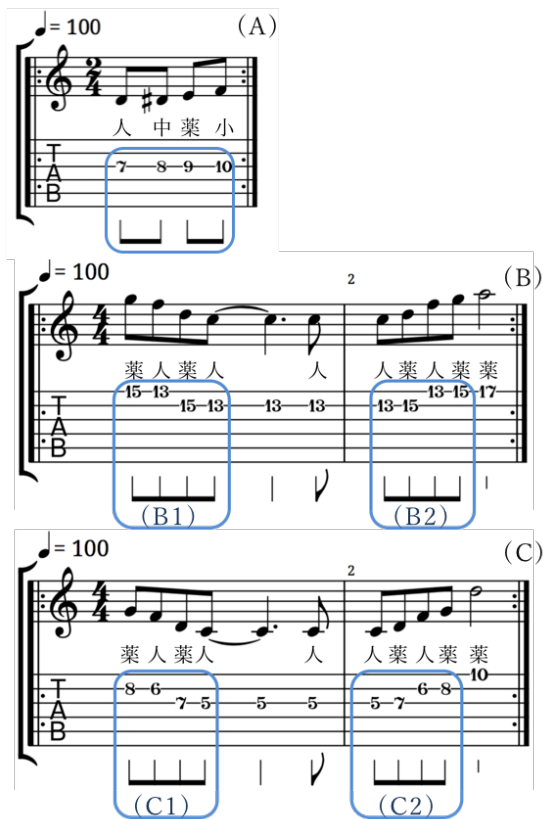


図 2: 演奏音源収集に用いたフレーズ。青枠の部分は今回の分析に用いた箇所を示している。

Fig. 2 The phrases used in our data collection. We extracted the five parts highlighted in blue for analysis.

4.1 演奏音源データの収集

5 人のエレキギター初心者 (P1~P5) に演奏音源データ収集に協力してもらった。5 人はエレキギター未経験あるいは数週間以内でやめてしまった者である。演奏するフレーズは図 2 に示す (A), (B), (C) の 3 つを用意した。実際に分析に用いた箇所は図 2 の青枠で囲った連続する 4 音である。(A) はエレキギターの基礎練習によく用いられるフレーズであるため採用した。(B) と (C) は有名なエレキギター楽曲の一部を簡略化したもので、実験参加者のモチベーションを維持する目的もあり採用した。なお、今回はすべてダウン・ピッキングでの演奏とした。

実験手順は表 1 のとおりである。3 つのフレーズそれぞれについて、ある程度練習した後に 1~2 分間ほど繰り返して弾いてもらい録音するという手順を 4 回繰り返した。演奏するときは常にメトロノームを鳴らしており、そのリズムに合うように弾いてもらった。なお録音の際には、オーディオインターフェイスを用いてエレキギターと PC を接続し、エフェクトなどはかけなかった。また、サンプリング周波数と量子化ビット数は 44100 Hz/24bit で行った。

4.2 演奏音源データの前処理

録音した演奏音源から、分析に使用する部分を適切に抽出する必要がある。今回の分析では、図 2 の青枠線で囲った連続する 4 音のすべてが、指定された音の高さかつある程度正しいタイミングで鳴っている部分 (*accurate_notes*) を抽出することとした。*accurate_notes* の抽出は以下の手順で行った。

- (1) 演奏音源からすべての onset を検出する [11]。
- (2) 各 onset から約 23 ms における基本周波数 F_0 をもとに、音名 (音の高さ) を求める。
- (3) 連続する 4 つの onset について、音の高さが正しくかつ各 onset 区間が 450 ms から 750 ms の間に収まっている場合、その 4 音を *accurate_notes* とする。

これらの処理の後、分析を行う 5 つのフレーズ (A), (B1), (B2), (C1), (C2) のそれぞれの *accurate_notes* の中から、各実験参加者 (P1~P5) の演奏データ数がおおよそ均等になるという条件のもとで無作為に 150 個ずつ、合計 750 個を選択した。

4.3 エレキギター経験者による主観的評価

前小節で述べた 750 個の *accurate_notes* に対して、3 人のエレキギター経験者が評価を行った。評価項目は Tone と Fluency の 2 つであり、7 段階のリッカート尺度 (1:最も悪い, 7:最も良い) を用いた。Tone は音色に対する評価項目であり、Fluency は Technique の「なめらかに弾けているか」に該当する項目である。なお、Tone の評価の際、それぞれの *accurate_notes* の 4 つ目の音に関しては考慮しないこととした。1~3 つ目の音はそれぞれ長さが BPM100 の 8 分音符であるが、4 つ目の音は長さが一定ではなく、それによる特徴量の差が生じる恐れがあったためである。具体的な評価手順を以下に示す。

- (1) 前小節の操作で選ばなかった *accurate_notes* の中から 30 個を無作為に抽出する。
- (2) 3 人の評価者がそれぞれ、それら 30 個の *accurate_notes* に対して評価を行う。
- (3) 3 人の評価者の回答を比較し、3 人ともある程度共通した評価基準を持つように各々の評価基準を調整する。
- (4) 3 人の評価者がそれぞれ、750 個の *accurate_notes* に対して Tone と Fluency の点数をつける。

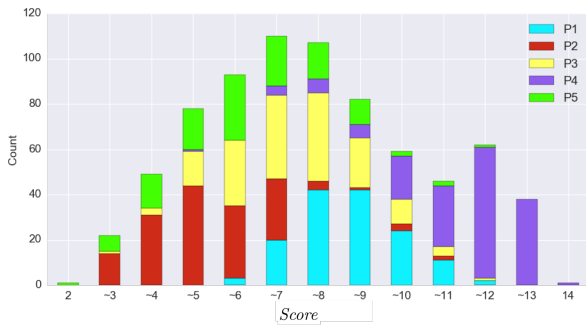


図 3: 3 人のエレキギター経験者がつけた *Score* の平均の分布.

Fig. 3 The distribution of average *Scores* rated by the three experienced players.

表 2: 使用する特徴量. これら 4 種類の特徴量 (合計 12 個) を説明変数として用いる.

Table 2 Four types of features used in our evaluation. The length of each feature vector is three.

特徴量	個数	説明
<i>Lag</i>	3	falling から次の onset までの時間
TS_{pm}	3	peak と middle における MFCC の cos 類似度
TS_{pf}	3	peak と falling における MFCC の cos 類似度
<i>TV</i>	3	peak, middle, falling における MFCC のばらつき

(5) *Tone* と *Fluency* の和を, それぞれの *accurate_notes* の *Score* とする.

この操作を通して得られた 3 人の評価者の *Score* の平均の分布は図 3 のようになった. また, 3 人の評価者による *Score* の一致度を調べたところ, Krippendorff の α 係数 [13] は 0.73 であったため, 程よく一致していると言える. 次節の分析では, 本節で述べた 750 個の *accurate_notes* から得た音響的特徴量と, それに対して付与されたエレキギター経験者による評価を用いる.

5. 音響的特徴量と主観的評価の関連性の分析

第 3 節で定義した音響的特徴量とエレキギター経験者による *accurate_notes* の評価の関係を調べるために, 評価予測のための回帰モデルの構築と性能評価を行う. 回帰モデルには Support Vector Regression (SVR) を使用し, 説明変数として用いる特徴量の組み合わせを変えながら予測性能を比較する. 使用する特徴量の一覧は表 2 の通りである. *accurate_notes* は 4 音から成るため, *Lag* は各音間の 3 つである. また, 前節で述べた通り今回の分析では 4 音目の音色に関しては考慮しないので, TS_{pm} , TS_{pf} , *TV* はそれぞれ 1~3 音目の 3 つである. これら 4 種類 $\times$ 3 つの特徴量を, それぞれの種類の中で値順に並び替えてから説明変数として用いる.

5.1 Cross Validation による性能評価

使用する特徴量の組み合わせを変化させながら, RBF カーネルを用いた SVR ($C = 10.0, \gamma = 0.1$) による回帰

表 3: SVR による *Score* および *Tone* と *Fluency* の予測性能評価. 4 種類の特徴量の組み合わせを変えながら, テストデータに対する決定係数 R^2 の変化を求めたところ, 4 種類すべての特徴量を用いたときが *Score* 予測の決定係数 R^2 が最も高かった.

Table 3 The accuracy result of SVR for estimating *Score*, *Tone* and *Fluency*. R^2 was the highest when SVR uses all the four features.

Features				R^2		
<i>Lag</i>	TS_{pm}	TS_{pf}	<i>TV</i>	<i>Score</i>	<i>Tone</i>	<i>Fluency</i>
✓				.475	.280	.608
	✓			.222	.382	.047
		✓		.154	.309	.018
			✓	.230	.379	.060
✓	✓			.717	.669	.695
✓		✓		.793	.722	.770
✓			✓	.796	.740	.772
	✓	✓		.204	.388	.003
	✓		✓	.217	.410	.006
		✓	✓	.242	.404	.062
✓	✓	✓		.807	.753	.760
✓	✓		✓	.805	.756	.761
✓		✓	✓	.800	.738	.767
	✓	✓	✓	.251	.422	.051
✓	✓	✓	✓	.809	.761	.762

表 4: 演奏者 5 人のうち 4 人のデータを訓練用データ, 残り 1 人のデータを評価用データとしたときの *Score* 予測の性能評価.

Table 4 The accuracy result of SVR for *Score* estimation under the leave-one-player-out conditions.

訓練用データ	評価用データ	R^2	MSE
P2, P3, P4, P5	P1	.168	1.27
P1, P3, P4, P5	P2	.190	1.95
P1, P2, P4, P5	P3	.402	1.52
P1, P2, P3, P5	P4	-.573	3.74
P1, P2, P3, P4	P5	.638	1.26

モデルの性能評価を行った. 10-fold Cross Validation で無作為にデータを分割し, テストデータに対する決定係数 R^2 を求めたところ, 表 3 のような結果が得られた. *Score* の予測において, 4 種類の特徴量をすべて使用したときの決定係数 R^2 が .809 で最も高い. また, このときのテストデータに対する平均二乗誤差は 1.25 であった.

表 3 から, 定義した 4 種類の特徴量は *Score* と相関があることがわかる. 次に, これら 4 種類の特徴量を用いたとき, 無作為ではなく演奏者やフレーズ単位でデータを分割した場合でも予測性能が保てるのかの検証を行う.

5.2 演奏者ごとに分割したデータを用いた性能評価

演奏者 (P1~P5) ごとにデータを分割し, 4 人の演奏者のデータから学習したモデルで残りの 1 人の演奏者の *Score* を予測した際の性能評価結果を表 4 に示す. 無作為にデータを分割したときに比べて, 決定係数が大きく下がって

表 5: 5つのフレーズのうち4つのデータを訓練用データ, 残り1つのデータを評価用データとしたときの Score 予測の性能評価.

Table 5 The accuracy result of SVR for Score estimation under leave-one-phrase-out conditions.

訓練用データ	評価用データ	R^2	MSE
(B1), (B2), (C1), (C2)	(A)	.589	2.05
(A), (B2), (C1), (C2)	(B1)	.823	1.46
(A), (B1), (C1), (C2)	(B2)	.784	1.50
(A), (B1), (B2), (C2)	(C1)	.820	1.29
(A), (B1), (B2), (C1)	(C2)	.851	1.09

ることがわかる. この原因として, 演奏者ごとにスコアの分布に偏りがあることが考えられる. 特に P4 の Score の予測に関しては, 図 3 から見て取れる通り Score の高い部分の大半を P4 が占めており, 残りの 4 人のデータによる学習では不十分であったため決定係数が小さくなったと考えられる. この結果から, 最終的にあらゆる演奏者に対して Score の予測ができるようになるには, さらに多くの演奏者からデータを集める必要があると言える.

5.3 フレーズごとに分割したデータを用いた性能評価

次に, フレーズ ((A), (B1), (B2), (C1), (C2)) ごとにデータを分割し, 4つのフレーズのデータから学習したモデルで残り1つのフレーズの Score を予測した. その結果を表5に示す. (A) を評価用データとして用いたとき以外は, 無作為にデータを分割して Cross Validation を行ったときと同程度の決定係数が得られている. (A) に対する予測性能が他と比べて低くなった理由として, 他の 4 つは 2 音ごとに異なる弦を弾くフレーズであるのに対して (A) は同一の弦を弾き続けるフレーズであり, このことが定義した特徴量以外に影響を与えていることが考えられる. そのため, より高精度なモデルを構築するためには新たな特徴量を探していく必要があると言える.

6. おわりに

本研究では, エレキギターの音響信号から演奏を自動評価するための音響的特徴量に関する考察を行った. Technique (Fluency) と Tone に関する音響的特徴量として *Lag*, *Tone Stability*, *Tone Variance* を定義したところ, エレキギター経験者による主観的評価との間に関連性が見られた. これらの音響的特徴量から経験者による評価を予測する回帰モデルを構築したところ, テストデータに対して決定係数 $R^2 = .809$ で予測することができた.

今後の課題としては, 学習データ数を増やすことと, 本稿で定義した以外の有用な特徴量を見つけ出すことがあげられる. また, 今回は等間隔の 4 音という単純なフレーズのみを用いた分析を行ったが, 異なるリズムのフレーズなどに対しても応用する方法を考えていく必要がある. 以上の課題の解決を通してより汎用的なモデルを構築したのち,

演奏の自動評価結果をどのように演奏者にフィードバックすれば, 演奏者にとって有益であるかを探っていく.

参考文献

- [1] Palmer, C.: Music performance, *Annual review of psychology*, Vol. 48, No. 1, pp. 115–138 (1997).
- [2] Russell, B. E.: The Development of a Guitar Performance Rating Scale using a Facet-Factorial Approach, *Bulletin of the Council for Research in Music Education*, No. 184, pp. 21–34 (2010).
- [3] Radicioni, D. and Lombardo, V.: Guitar fingering for music performance, *In Proceedings of the International Computer Music Conference*, Vol. 40, No. 45, pp. 527–530 (2005).
- [4] Heijink, H. and Meulenbroek, R. G.: On the complexity of classical guitar playing: functional adaptations to task constraints, *Journal of motor behavior*, Vol. 34, No. 4, pp. 339–351 (2002).
- [5] 有賀竣哉, 後藤真孝 and 矢谷浩司: Strummer: インタラクティブなギターコード練習システム, 研究報告音楽情報科学 (SIGMUS), Vol. 2017, No. 24, pp. 1–6 (2017).
- [6] Reboursière, L., Lähdeoja, O., Drugman, T., Dupont, S., Picard-Limpens, C. and Riche, N.: Left and right-hand guitar playing techniques detection., *NIME* (2012).
- [7] Nonogaki, A., Shimazu, S., Emura, N., Miura, M., Akinaga, S. and Yanagida, M.: Use of spline curve to evaluate performance proficiency of a Czerny piano piece, *Proceedings of International Symposium on Performance Science*, pp. 69–74 (2011).
- [8] Nakano, T., Goto, M. and Hiraga, Y.: An automatic singing skill evaluation method for unknown melodies using pitch interval accuracy and vibrato features, *Ninth International Conference on Spoken Language Processing* (2006).
- [9] Yazawa, K., Itoyama, K. and Okuno, H. G.: Automatic transcription of guitar tablature from audio signals in accordance with player's proficiency, *International Conference on Acoustics, Speech and Signal Processing*, IEEE, pp. 3122–3126 (2014).
- [10] 藤田顕次, 大野博之 and 稲積宏誠: 習熟度を考慮した複数楽譜からのピアノ譜生成手法の提案, 研究報告音楽情報科学 (SIGMUS), Vol. 2008, No. 89, pp. 47–52 (2008).
- [11] Bello, J. P., Daudet, L., Abdallah, S., Duxbury, C., Davies, M. and Sandler, M. B.: A tutorial on onset detection in music signals, *IEEE Transactions on speech and audio processing*, Vol. 13, No. 5, pp. 1035–1047 (2005).
- [12] Brown, J. C.: Calculation of a constant Q spectral transform, *The Journal of the Acoustical Society of America*, Vol. 89, No. 1, pp. 425–434 (1991).
- [13] Hayes, A. F. and Krippendorff, K.: Answering the call for a standard reliability measure for coding data, *Communication methods and measures*, Vol. 1, No. 1, pp. 77–89 (2007).