

音響的特徴量を用いた
エレキギターの演奏評価手法

**Performance Assessment Methods for
Electric Guitars Using Acoustic Features**

37-166462 下尾 波輝

指導教員 矢谷 浩司 准教授

東京大学大学院
工学系研究科 電気系工学専攻

2018 年 2 月 1 日 提出

Acknowledgements

数多くの人たちの支えがあり、無事に研究を終えて修士論文を書くことができました。その中でも特に、学部時代を合わせて3年間、研究の指導をしてくださった矢谷浩司准教授に大変感謝しております。主に毎週1回の1on1ミーティングを通して、研究に対する考え方そのものから実際の研究手法まで、様々な面で指導していただきました。また、研究をする意欲がなくなってどうしようもなくなったときも、精神的な面で支えてくださりました。他の研究室のような他に指導を任せられる人がいない中、学生一人一人にこのような非常に手厚い支援をしてくださり、非常に感謝しております。

研究室の同期である吉谷拓真君と橋爪崇弘君には数多くのアドバイスをいただきました。吉谷君には特にプログラム実装の面で貴重なアドバイスをいただき、橋爪君には環境構築や論文の体裁などの面で様々なことを教えていただきました。優秀な同期に囲まれたおかげで、効率良く研究を進めることができました。大変感謝しております。

研究室の後輩たちにもお世話になりました。M1の柴藤大介君には、日常生活の中の会話での気分転換に付き合ってもらったことが多かったです。B4にもお世話になりましたが、その中でも特に正木博明君、松井秀憲君、三島潤平君の3人には予備実験のデータ収集に協力していただき、非常に助かりました。研究室生活を支えてくださった後輩たちに感謝します。

研究室の秘書である元岡みさ子さんには、事務的な作業の面で大変お世話になりました。事務的なことに関していつも迅速かつ適切な対応をしてくださり、研究に専念することができました。縁の下の力持ちとして研究を支えてくださり、ありがとうございました。

分析に用いるデータ収集のために、所属するバンドサークルの7人の知人に協力していただきました。彼らが演奏音源収集や演奏評価などに協力してくれたおかげで、無事に研究を進めることができました。大変感謝しております。

最後になりますが、6年もの間、私が東京に出て学ぶ環境を整えてくださった両親に本当に感謝しています。私のために懸命に働いてくださった両親のおかげで、6年間の勉学および研究をやり切ることができました。本当にありがとうございました。

Abstract

Playing electric guitars requires rhythmical and dexterous finger movements. Novices may have to spend an enormous amount of time and effort for practice to master such movements. However, practicing electric guitars can be boring and easy to lose motivations. Prior research developed various systems that support and motivate electric guitar practices. Although these systems make practice enjoyable, a limitation exists that they evaluate users' performances only by timing and pitch accuracy. In contrast, people evaluate electric guitar performances by more various factors, such as tone and technique. Thus, there is a gulf of evaluation methods of electric guitar performance between existing systems and human. To solve this problem, I develop several acoustic features for evaluating electric guitar playing from the perspectives of tone and fluency: *Lag*, *Tone Similarity*, *Tone Instability*, and *Harmonic-Percussive Ratio*. I then built a machine learning model for assessing electric guitar performance using performance data of five novices with subjective ratings by three experienced players. This model can predict the subjective evaluation score ($2 \sim 14$, $Mean = 8.26$, $SD = 2.64$) with $MSE = 1.08$, achieving sufficient accuracy. Furthermore, I built a method for judging whether a player successfully holds strings down. This model can classify electric guitar sounds into three categories of "Good hold," "Not enough hold," and "hold on fret" at the accuracy of 88.4%. These results show that my acoustic features have a potential to assess electric guitar performance. As an application example of this research, I also present an electric guitar practice system that evaluates users' performance from the viewpoint of tone and fluency.

概要

エレキギター演奏においてはリズム良くかつ細かく手を動かすことが要求される。そのような動作を習得するためには、多くの時間と労力を練習に費やす必要がある。しかし、その練習は地味でときには孤独なものであるため、初心者は途中で挫折してしまうことも多い。そのため、エレキギター練習を支援するための数々のシステムがこれまでに開発されてきたが、これらのシステムには「正しいタイミングで正しい音程で弾けたか」のみしか演奏評価基準が存在しない。実際に人間が演奏評価を行う際には、リズムや音程の正確さ以外にも音色の良さなどの要素も考慮する。このように、既存のエレキギター練習システムと人間の演奏評価基準には乖離が生じている。本研究ではこの問題を解決するために、リズムや音程の正確さ以外の観点からもエレキギター演奏を評価するための音響的特徴量 (*Tone Similarity*, *Tone Instability*, *Harmonic-Percussive Ratio*, *Lag*) を提案した。その後、5人のエレキギター初心者から収集した演奏音源と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、定義した音響的特徴量を用いてエレキギター演奏を評価するモデルを構築した。750個の演奏音源データを用いて 10-fold cross validation でそのモデルの性能検証を行ったところ、経験者による評価 (2 ~ 14, *Mean* = 8.26, *SD* = 2.64) を平均二乗誤差 1.08 の精度で予測できた。さらに同じ音響的特徴量の一部と既存の音響的特徴量によって、エレキギターの押弦が適切に行われているかを判定する手法を考案した。実験の結果、音を「正しく押弦できている音/十分に弦を押さえられていない音/フレット上を押さええている音」に 88.4% の精度で分類することができた。これらの結果により、本論文で提案した音響的特徴量がエレキギター演奏評価に利用できる可能性を示した。さらに本研究成果の応用例として、リズムや音程の正確さにとどまらない演奏評価が可能なエレキギター練習用のシステムを作成した。

Table of contents

List of figures	vii
List of tables	viii
1 はじめに	1
1.1 背景	1
1.2 貢献	2
2 関連研究	4
2.1 音楽演奏の主観的評価手法	4
2.2 信号処理による音楽演奏分析	4
2.3 ギター演奏の運指決定	6
2.4 ギター練習提供システム	6
2.5 楽器練習支援システム	7
2.6 まとめ	8
3 エレキギター演奏評価に利用可能な音響的特徴量の提案	9
3.1 ギター演奏における評価項目	9
3.2 Technique に関する音響的特徴量: <i>Lag</i>	10
3.3 Tone に関する音響的特徴量	12
3.3.1 <i>Tone Similarity, Tone Instability</i> : 音色の安定度に着目した特徴量	12
3.3.2 <i>Harmonic-Percussive Ratio</i> : 撥弦音と調波音に着目した特徴量	14
4 エレキギター演奏に対する主観的評価の予測	17
4.1 データセット	17
4.1.1 演奏音源データの収集	17
4.1.2 演奏音源データの前処理および特徴量の計算	18
4.1.3 エレキギター経験者による評価付け	20
4.2 定義した特徴量と評価値との関連性の分析	22

4.2.1	SVR による予測性能比較	22
4.3	考察	24
4.3.1	10-fold と LONO および LOGO の結果の比較	24
4.3.2	Lag が Tone の評価に与える影響	24
4.3.3	Tone 評価における TS_{np} の影響力	27
4.3.4	Tone 評価の予測誤差が大きかった音源データ	27
5	押弦ミスの自動分類	29
5.1	左手の押弦ミスの種類	30
5.2	押弦におけるミスの自動検知手法	31
5.2.1	データセット	31
5.2.2	使用する音響的特徴量	31
5.2.3	SVM による分類	32
5.3	考察	32
5.3.1	HPR の分類精度への影響力	32
5.3.2	エレキギターの個体差が TS_{np} に与える影響	33
5.3.3	弦の違いが分類精度に与える影響	34
6	考察	36
6.1	特徴量に関する考察	36
6.1.1	TS_{np}	36
6.1.2	HPR	36
6.2	本研究で得られた知見の汎用性向上に向けての課題	37
6.2.1	より多様なフレーズの演奏音源データを用いた分析	37
6.2.2	より多くの演奏者の演奏音源データを用いた分析	37
6.2.3	コード演奏評価の分析	37
7	アプリケーション	39
7.1	システム詳細	39
7.2	本システムの問題点	41
7.3	その他の想定されるアプリケーション	42
8	おわりに	43
	Publications	44
	References	45

List of figures

3.1	検出した onset と release および <i>Lag</i> の例	11
3.2	エレキギターの波形のスペクトログラム表示	11
3.3	エレキギター音の対数振幅スペクトルをメルフィルタバンクに通した例	12
3.4	<i>Tone Similarity</i> の計算に用いる \mathbf{M}_p , \mathbf{M}_{np} , \mathbf{M}_r の位置	13
3.5	<i>Tone Instability</i> の計算に用いる \mathbf{M}_i ($i = 1 \sim N$) の位置	14
3.6	エレキギター音の波形に HPSS を実行した例	16
4.1	演奏音源収集の様子	18
4.2	演奏音源収集に用いたフレーズ	19
4.3	定義した音響的特徴量の計算	20
4.4	経験者による評価用に作成した Web ページの画面	21
4.5	3 人のエレキギター経験者がつけた <i>Score</i> の平均の分布	21
4.6	P2 と P4 の演奏音波形の比較	26
4.7	各演奏者 (P1~P5) の <i>Lag</i> の分布と <i>Lag</i> の有無による <i>Tone</i> 予測誤差の分布差	26
4.8	各音源データの <i>Tone</i> の評価と 3 音の平均 TS_{np} の散布図	28
4.9	<i>Tone</i> 評価予測誤差が大きかった演奏波形の例	28
5.1	不適切な押弦の例	30
5.2	音の種類による <i>HPR</i> の分布の差	33
5.3	各エレキギターの正しく押弦できている音の弦別の TS_{np} の分布	34
5.4	弦別の <i>HPR</i> の分布の差	35
7.1	作成したシステムの画面	40

List of tables

2.1	先行研究で得られた Guitar Performance Rating Scale	5
4.1	演奏音源収集の手順	18
4.2	録音環境の詳細	19
4.3	説明変数として使用する音響的特徴量の一覧	22
4.4	SVR による <i>Score</i> 予測性能評価	23
4.5	leave-one-novice-out cross validation と leave-one-phrase-out cross validation の結果	23
4.6	SVR による <i>Tone</i> および <i>Fluency</i> の評価予測における平均二乗誤差	25
4.7	<i>Tone</i> と <i>Fluency</i> の評価予測における leave-one-novice-out cross validation と leave-one-phrase-out cross validation の結果	25
5.1	SVM による押弦の分類性能評価	32
5.2	MFCC, Δ MFCC, HPR を説明変数として leave-one-guitar-out cross validation を 行ったときの分類結果	33
5.3	MFCC, Δ MFCC, <i>HPR</i> を説明変数として leave-one-guitar-out cross validation を 行ったときの弦別の分類精度	34
5.4	MFCC, Δ MFCC, <i>HPR</i> を説明変数として leave-one-guitar-out cross validation を 行ったときの6弦の音の分類結果	35

Chapter 1

はじめに

1.1 背景

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 [23]. 楽器演奏においては手足や指を細かくリズム良く動かす必要があるが、日常生活でそのような動きを要求される機会は少ない. 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. しかしながら、楽器の練習は地味で孤独な努力が求められるため、初心者が途中で挫折してしまうことも多い.

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、練習を支援する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に Yousician¹や Rocksmith²がある. Yousician は流れてくるギターの譜面に合わせて実際にエレキ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.

Yousician や Rocksmith に代表されるシステムは、モチベーションや楽しさといった面においては確かにエレキギター練習の効率を向上させるが、重大な欠点を抱えている. それは、ユーザの演奏に対する評価基準が基本的には「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ことである. 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り、「綺麗な音色で弾けているか」「なめらかに弾けているか」など、リズムと音程の正確さ以外にも考慮すべき点は多い. しかし現状ではその評価は人間が行う必要があるため、上記のようなシステムを利用して練習をする場合には、音色などに関するフィードバックを得る機会がない. そのため、特に初心者にとっては本当に自分が上手く弾けているのかの判断

¹<https://yousician.com/>

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

が難しい。また、適切なフィードバックが得られないまま練習を続けてしまうと、誤った動きなどを身につけてしまう恐れがある。

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.”や“Played fluently.”など、評価者の感覚に委ねられるものが多い [27]。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、まず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。

そこで本研究では、エレキギター演奏の評価に利用できる音響的特徴量について検討し、それを用いて以下の2つを実現することを目的とする。

- 音色と演奏のなめらかさの観点から、単音のエレキギター演奏を自動的に評価する。
- 単音のエレキギター演奏において押弦が適切に行われているかを自動的に判定する。

本研究では最初に、人間によるギター演奏評価に関する文献などを参考にし、エレキギター演奏評価に利用できそうな4つの音響的特徴量 (*Lag*, *Tone Similarity*, *Tone Instability*, *Harmonic-Percussive Ratio*) を定義した。その後、5人のエレキギター初心者による演奏音源に対して3人のエレキギター経験者が主観的評価を付与した。そして、定義した音響的特徴量から主観的評価を予測するモデルを構築したところ、定義した4つの音響的特徴量がエレキギター演奏評価予測に利用できることがわかった。次に、エレキギター初心者が最も難しく感じるものの1つである左手の指による押弦の良さを、定義した音響的特徴量と既存の音響的特徴量を用いて評価できるかを検証した。具体的には、音を「正しく押弦できている音」、「十分に弦を押さえていない音」、「フレット上を押さえている音」の3つに8割以上の精度で分類できることを確認した。最後にこれらの結果をもとにして、ユーザのエレキギター演奏に対して点数および押弦に関するフィードバックを与えるシステムをアプリケーション例として作成した。

1.2 貢献

本研究の貢献は以下の通りである。

エレキギター演奏の評価に利用できる音響的特徴量の提案

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)、*Tone Similarity* (メル周波数ケプストラム係数を用いた音色の安定度を表す指標)、*Tone Instability* (メル周波数ケプストラム係数を用いた音色のばらつきを表す指標)、*Harmonic-Percussive Ratio* (音の Harmonic 成分と Percussive 成分の最大振幅の比の対数) を提案した。

音色と演奏のなめらかさに着目したエレキギター演奏評価手法の検証

定義した上記の音響的特徴量を説明変数として、音色と演奏のなめらかさに関して経験者が付与したスコアを予測する Support Vector Regression (SVR) のモデルを構築した。その結果、2 から 14 ($Mean = 8.26$, $SD = 2.64$, $N = 750$) の範囲のスコアを平均二乗誤差が 1.08 の精度で予測するモデルを構築することができた。

エレキギター演奏における押弦ミスの検知方法の検証

定義した上記の音響的特徴量を用いて、エレキギターの音を「正しく押弦できている音/十分に弦を押さえられていない音/フレット上を押さええている音」の 3 つに分類する手法を考案した。その結果、88.4% の精度でこれら 3 つの音を分類することに成功した。

上記の結果を用いたエレキギター練習システムの構築

上記の 2 つの実験結果をもとにして、BPM100 の 8 分音符 4 つから成る単音のフレーズを演奏したときに自動で点数を与え、さらに各音について押弦が適切に行われていたかを判定するシステムを構築した。

Chapter 2

関連研究

2.1 音楽演奏の主観的評価手法

音楽演奏評価は評価者の主観に任されることが多く、その結果として評価者ごとに評価の差が発生してしまうことがある。そのため、音楽演奏の評価をどのように行うべきか、あるいは音楽演奏のどのような要素が評価に影響を与えるのかに関する研究が昔から行なわれている [18]。その中でも、因子分析などを用いて演奏評価項目を定める研究が多く見られる。例えば、ヴァイオリンなどのストリングス楽器 [32]、合唱 [4]、ギター [27] の演奏評価における主要因子とその項目を見つけ出した研究が存在する。本研究においてはこの中でも特に、ギターの演奏評価（Guitar Performance Rating Scale）の選定を行った Brian らの研究 [27] で得られた結果が参考になる。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、ギター演奏評価における 5 つの主要な因子（Interpretation/Musical Effect, Intonation, Rhythm/Tempo, Tone, Technique）と、それら进行评估するための 31 個の評価項目が得られている（表 2.1）。しかしコンピュータによって演奏に評価を与えるためには、演奏の音響信号を処理して適切な特徴量を抽出し、これらの演奏評価項目と結びつける必要がある。

2.2 信号処理による音楽演奏分析

ギター

ギターから得られる音響信号を用いた分析も数多く行われてきた。その中でも、ギターの特性に着目した研究を紹介する。Reboursière らはギター特有の奏法を音楽信号から検出する

表 2.1: Brian らによる研究で得られた Guitar Performance Rating Scale (GPRS) の 5 つの主要因子と、それらに対応する演奏評価項目 [27].

<i>Factors</i>	<i>GPRS items</i>	<i>Factors</i>	<i>GPRS items</i>
Interpretation/ Musical Effect	Melodic expression.	Technique	String crossing is controlled and smooth.
	No contrasts in performance.		Played fluently.
Tone	The interpretaion was musical.	Rhythm/ Tempo	Poor synchronization of pick and frethand fingers.
	Spiritless playing.		Lack of clarity in picked passages.
	Performance not expressive.		Attacks were clean.
	Performance reflected sensitivity.		The tempo was steady.
	Melodic phrasing.		Correct rhythms.
Tone	Ton is strong.	Intonation	Offbeats played properly.
	Tone is full.		Hurried repeated notes.
	Thin tone quality.		Lack of a steady pulse.
	Tone is rich.		Tempo not controlled.
	Sound is clear and resonant.		The tempo was in good taste.
Tone quality is beautiful.	Played out of tune.	Intonation	Intonation is good.
There was lack of tonal color.	Intonation is inconsistent.		Intonation is inconsistent.
	Ignored key signature.		Ignored key signature.

手法について検討した [26]. ギター演奏においては、単に右手で弦を弾くだけではなく、ハンマリング、プリング、チョーキング、スライドなどの左手の動きによって音を出す奏法がある。Reboursière らの提案した手法では、エレキギターの音響信号からこれらの奏法を 90% 以上の精度で検出することに成功している。Hrybyk は、音響信号とカメラ映像を組み合わせた和音認識手法を提案した [11]. その結果、音響信号で和音認識を行ったときよりも認識精度が高くなった。本研究においてはカメラ映像は用いなかったが、指が余計に動くなどの音に現れない不適切な動きを検出してフィードバックするためには、最終的にカメラによる映像が必要となってくると考えられる。

その他の楽器および歌

本研究で行った左手の押弦ミス検知に類似する研究としては、既存の楽器練習システムが音程とリズムの正確さしか考慮していないことに着目し、ヴァイオリン初心者に頻繁に見られるミスを検知しようとした Luo らの研究 [31] が挙げられる。Luo らはメル周波数ケプストラム係数 (MFCC) などの代表的な音響的特徴量を用いて、Support Vector Machine (SVM) によってこれらのミスを検知しようとした。しかし検出精度は決して高くはなく、検出が最も上手くいった部分でも F 値が 0.6 程度であったため、このままでは実際の練習に用いるのは難しい。また Nonogaki らは、およそ 8 小節分の 16 分音符からなる特定のフレーズ (Czerny's Etude No.40) のピアノ演奏音から 4 つの音楽的特徴量 (onset time/velocity/duration/tempo) を抽出し、それらの平均や標準偏差などの統計値を用いてピアノの習熟度を推定しようとした [21]. Nonogaki らの手法で求めた習熟度とプロのピアニストが付与した習熟度を比べたところ、予測性能を表す決定係数は BPM75 と BPM150 の場合においてそれぞれ .694 と .581 であり、特に BPM150 の場合においては予測性能が高くはなかった。これらの研究で高精度な

予測モデルを構築できなかった原因としては、これまでの音楽情報処理で定義されてきた一般的な音響的特徴量しか用いなかったことが考えられる。

楽器ではなく歌による演奏分析に関しては、音程の正確さとヴィブラートに関する特徴量を定義して歌のうまさを2値 (good または poor) で判定する Nakano らの研究 [20] がある。Nakano らの手法による判定と人間による判定を比較したところ、8割以上が一致していた。Nakano らの研究においては、歌において重要な要素である音程とヴィブラートに焦点を当てた特徴量を用いたことによって高精度な予測ができたと考えられる。本研究においても、エレキギターの特徴などを入念に考慮しながら特徴量を定義することで評価予測の精度が上がり期待される。

2.3 ギター演奏の運指決定

ギターの練習を支援するために、ギター演奏に関する様々な研究が行われてきた。その中でも特に多く見られるのが、弦を押さえる左手の運指 (fingering) に関わる研究 [24] である。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。そのため、動的計画法や遺伝的アルゴリズムなどを用いて最適な運指を決定する研究が行われてきた [25, 28]。また、Hejjink 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 [10]。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。Yazawa らは、ユーザのギター習熟度に合わせて演奏難易度を調節可能な TAB 譜生成手法について研究を行った [30]。彼らの手法を用いることで「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」などが可能である。しかしギターの習熟度に関してはユーザが自ら入力する必要があり、初心者にとっては自らの習熟度を判断することが難しいため、自動で演奏評価をするシステムが求められている。

2.4 ギター練習提供システム

練習に利用できる楽曲は非常に多い。このため、ユーザのレベルに応じた、あるいはユーザの興味の持ちやすい練習教材を提供するシステムの研究が行われている。Barthet らは、音楽推薦システムを応用したギター練習提供システムである Hottabs [3] を開発した。Hottabs は、YouTube から取得した最新の人気曲のギター演奏動画と、Web 上からクロールして取得した曲の TAB 譜をユーザに提示するシステムである。Hottabs を使って最新の人気曲の練習をすることで、ユーザのモチベーションが維持されることが期待される。また Ariga らは、コードの難易度と出現頻度などのデータにもとづいて練習を提供するシステムである Strummer [2] を開発した。Strummer を使うことで簡単かつ重要なコードや曲から練習を行う

ことができる。習得しやすい簡単なコードを練習しながら弾ける曲を増やしていくことは、ユーザの良いモチベーションとなる。

2.5 楽器練習支援システム

楽器練習を支援するシステムの研究は多岐に渡る。本節では、ギターの練習支援システムを中心にこれらの研究について紹介する。

Augmented Reality を用いたシステム

ギターの演奏に慣れないうちは自分の手元を見ながら弦を押さえたり弾いたりする必要があるが、譜面を読みながら自分の手元を見るのは難しい。そこで、Augmented Reality を用いてギターの練習を視覚的に支援するシステムが存在する。Motokawa らはカメラと PC を用いて、PC に映されたギターの指板にコードの押さえ方を提示するシステムを開発した [19]。また Löchtfeld らは、モバイルプロジェクションを用いて押さえる弦をギターの指板に投影するシステムのプロトタイプを作成した [15]。特に初心者にとっては、譜面に書かれた内容と実際の指の動きを素早く結びつけることが難しいので、このような直感的に情報を提示できるシステムは初心者の練習を効率的に支援できる。

ウェアラブルデバイスを用いたシステム

ウェアラブルデバイスによる練習支援システムは機材を用意するコストが大きいものの、ソフトウェアによる処理のみでは実現できない触覚などのフィードバックを与えることができる。そのため、ウェアラブルデバイスによる楽器練習支援システムがこれまでも数多く開発されている。Huang らの開発した PianoTouch [12] は、振動によってピアノ練習を支援する手袋型デバイスである。PianoTouch を用いることで、ピアノの曲を聴くとそれに合わせて演奏に使う指が振動し、ユーザはピアノがない場所でもピアノの練習をすることができる。Johnson らはヴァイオリンの練習の際に振動を用いて運弓の指示をするデバイス [14] を開発した。演奏練習においてウェアラブルデバイスによって触覚的なフィードバックを与えることは、タイミングや音の強弱に関する練習に効果的である [8]。特に楽器初心者にとっては、楽器演奏に必死で視覚的なフィードバックを確認する余裕がないため、振動などの触覚的なフィードバックは有効である。

ギターの練習に関するウェアラブルデバイスとしては、Matsushita らが開発したギターのストローク検知デバイスがある [17]。このデバイスを使用することで、右手のストロークのタイミングや強弱を測定することができる。また Matsushita らの研究においては、ストロークのタイミングや強弱のぶれにより演奏習熟度を推定できる可能性も示されている。本研究においては特別なデバイスは用いずにソフトウェアによる音楽信号処理しか行わなかった

が, Matsushita らの研究のように指の動きなどを直接的に取得することにより, 音に現れない部分のフィードバックを与えることができると考えられる.

音楽ゲーム形式のシステム

Guitar Hero¹はエレキギターを模した専用のコントローラを使って遊ぶ音楽ゲームである. しかし, 実際のエレキギターではなく専用のコントローラを使用するため, 練習ではなく娯楽目的の側面が大きい [5]. 一般向けに公開・販売されている代表的なギター練習用のシステムとしては, Yousician², Rocksmith³などが挙げられる. これらのシステムは実際にエレキギターを接続して演奏することができる. Yousician はテクニックごとの練習や曲の演奏を行うことができ, Rocksmith では曲の演奏に加えてミニゲーム形式の基礎練習モードなどがある. このような機能は練習のモチベーション向上には役立つものであるが, 前章でも述べた通り, 「正しいタイミングで正しい音程で弾けたか」のみしか演奏評価基準が存在しないという問題点がある.

2.6 まとめ

人間によるギター演奏の評価分析はこれまでも行われている. しかし音響信号処理によるギター演奏の分析は, ギターの奏法の検出や和音認識などにとどまっており, 演奏を評価するところまで至った研究はない. そのため, 多様なギター練習システムが提案されてきたものの, どのシステムもユーザの演奏音の評価は行わない, あるいはタイミングと音程の正確さのみしか演奏判断基準を持っていない. 本研究ではこの問題を解決するために, エレキギター演奏評価に利用可能な音響的特徴量を見つけ出し, それを用いてエレキギター演奏評価を行う手法を考案することを目的としている. この目的が達成できた場合, 例えば演奏者の演奏習熟度が推定可能になると考えられる. 演奏習熟度が推定可能になることで, 本章で述べたような練習提供・支援システムがより個人に適応したものになることが期待される.

¹<https://www.guitarhero.com/>

²<https://yousician.com/>

³<https://rocksmith.ubisoft.com/>

Chapter 3

エレキギター演奏評価に利用可能な 音響的特徴量の提案

本研究では、リズムと音程の正確さ以外の観点からエレキギター演奏を評価するための音響的特徴量を提案する。そのためにまず人間が行う演奏評価に関する研究を参考にし、エレキギター演奏評価における評価項目を調査する。そしてそれらの演奏評価項目に対応する音響的特徴量を定義する。

3.1 ギター演奏における評価項目

前章で述べた通り、Brian らが行った Guitar Performance Rating Scale (GPRS) の研究 [27] により GPRS には 5 つの主要な因子、Interpretation/Musical Effect, Intonation, Rhythm/Tempo, Technique, Tone があることがわかっている。この知見は、エレキギター演奏評価に利用できる音響的特徴量を考案する際に参考になると考えられる。

上記の 5 つの因子のうち 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個性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ほとんど考慮されない。また、Intonation はピッチの正確性に関するものであるが、ギターにおいては演奏に関係のないチューニングなどの要素の影響が大きい。Rhythm/Tempo はあらゆる種類の音楽演奏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ことを想定しているため、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 Technique と Tone の 2 つに着目し、これらに対応する音響的特徴量について検討する。

3.2 Technique に関する音響的特徴量: Lag

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>”¹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、左手人差し指で2弦5フレットを押さえて右手で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事が重要であり、左手で弦を押さえてから（あるいは押さえている弦を離してから）できるだけ早く右手で弦を弾く必要がある。

左手と右手の同期（hand synchronization）のうまさは、Heijink らの研究 [10] や Brian らの作成した GPRS [27] 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っていないと、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薬指を早く離した場合、<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（release）から次の音の立ち上がり（onset）までの時間を Lag（図3.1）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 の検出には Spectral Flux を利用した手法 [13] を用いる。release の検出は、Spectral Flux を応用した以下の SF' を用いて行う。

$$SF'(n) = \sum_{k=1}^N \max \{0, a_k \cdot (|X(n-1, F_k)| - |X(n, F_k)|)\} \quad (3.1)$$

SF' は、図3.2に見られるように弦の振動中は基本周波数とその倍音成分の振幅が大きいことに着目し、これらの振幅の減少分を考慮した指標である。 $|X(n, F_k)|$ 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。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 samples とし各時刻 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 release の時刻とする。今回は $N = 2$, $a_1 = 1$, $a_2 = 2$ とし計算を行った。この手法を用いて検出した release および Lag の例を図3.1に示す。図3.1右の3つ目の波形は押弦がうまく行われなかった音の波形であるが、このような歪な形の波形に対しても release が適切に検出されている。これは、基本周波数と倍音成分の振幅のみに着目したことで、不適切な押弦によって発生するノイズなどの影響を排除できたからであると考えられる。

¹<s,f> で s 弦 f フレットの音を表すこととする。

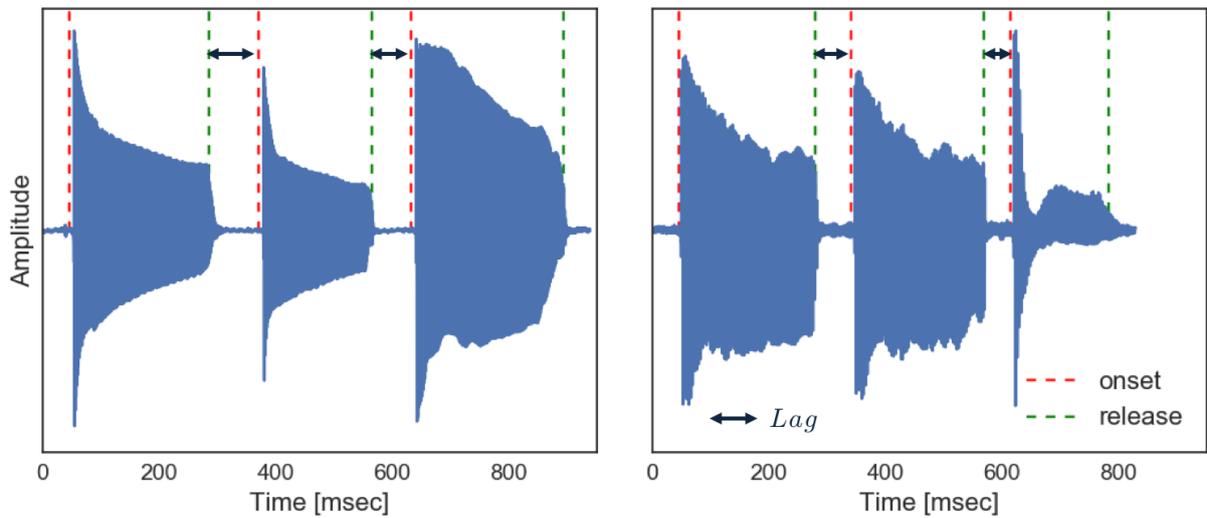


図 3.1: 検出した onset と release および *Lag* の例. 右図の 3 音目のような歪な形の波形に対しても適切に release が検出できている. release から次の音の onset までの時間が *Lag* である.

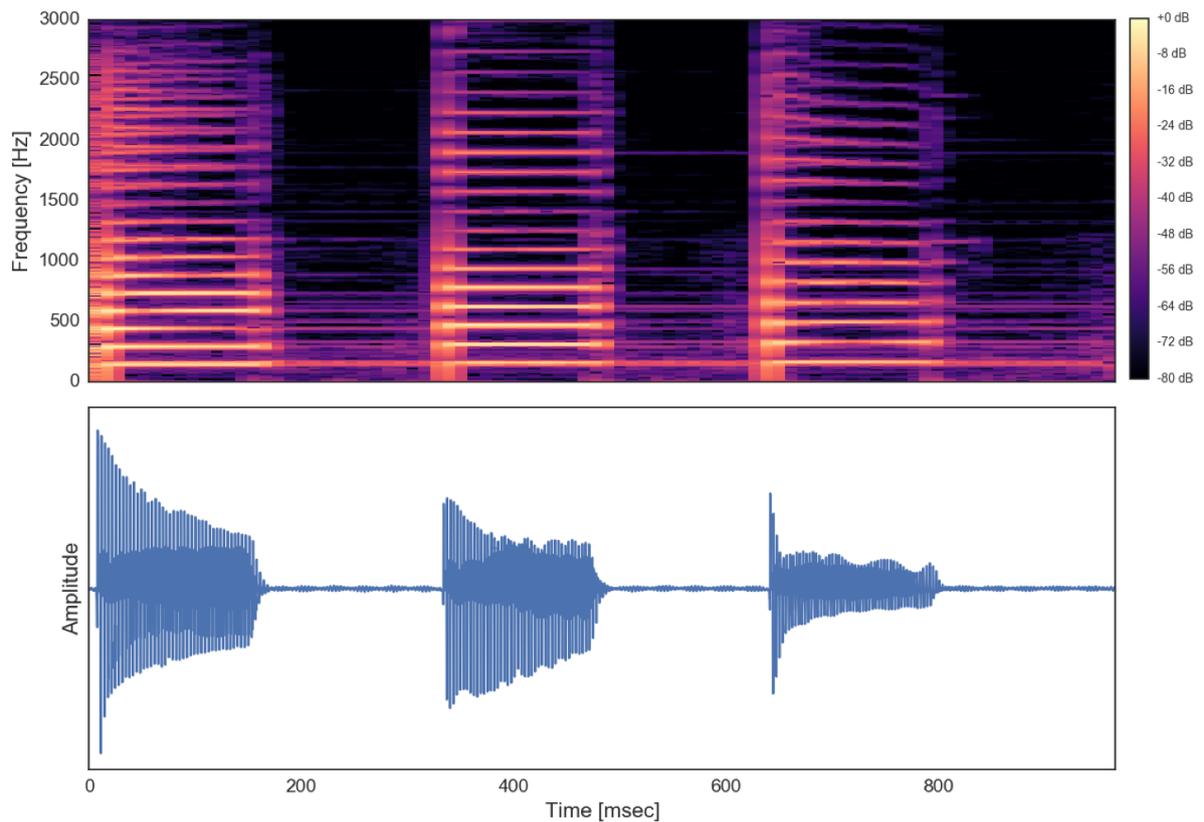


図 3.2: エレキギター波形のスペクトログラム表示. 音の立ち上がり (onset) 直後においては様々な周波数成分が入り混じっているが, それ以降は基本周波数成分とその倍音成分が大きい.

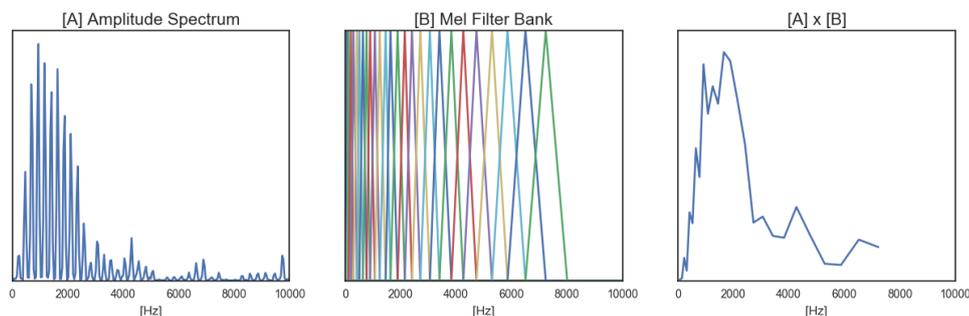


図 3.3: エレキギター音の対数振幅スペクトルをメルフィルタバンクに通した例.

3.3 Tone に関する音響的特徴量

3.3.1 *Tone Similarity, Tone Instability*: 音色の安定度に着目した特徴量

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。Brian らによる GPRS [27] の Tone 評価項目も “Tone is rich.”, “Sound is clear and resonant.”, “Tone quality is beautiful.” などの聴き手の主観による抽象的なものが多いと見られる (表 2.1)。

しかし、演奏の際の不適切な指の動きによって明らかに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ていなかったときは、音が鳴ってから鳴り終わるまでの間で音色が均一でなく不安定になる。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。

音色に関する代表的な特徴量としては、メル周波数ケプストラム係数 (MFCC) があげられる。MFCC は、メルフィルタバンク (図 3.3) を用いることで人の聴覚特性を考慮しながらスペクトル包絡の概形を表現する特徴量である。音声認識の分野で頻繁に用いられてきたが音楽信号の分析にも利用されており [16], 例えば音楽ジャンル識別 [29] や楽器音分類 [6] などを行う際の特徴量として用いられる。

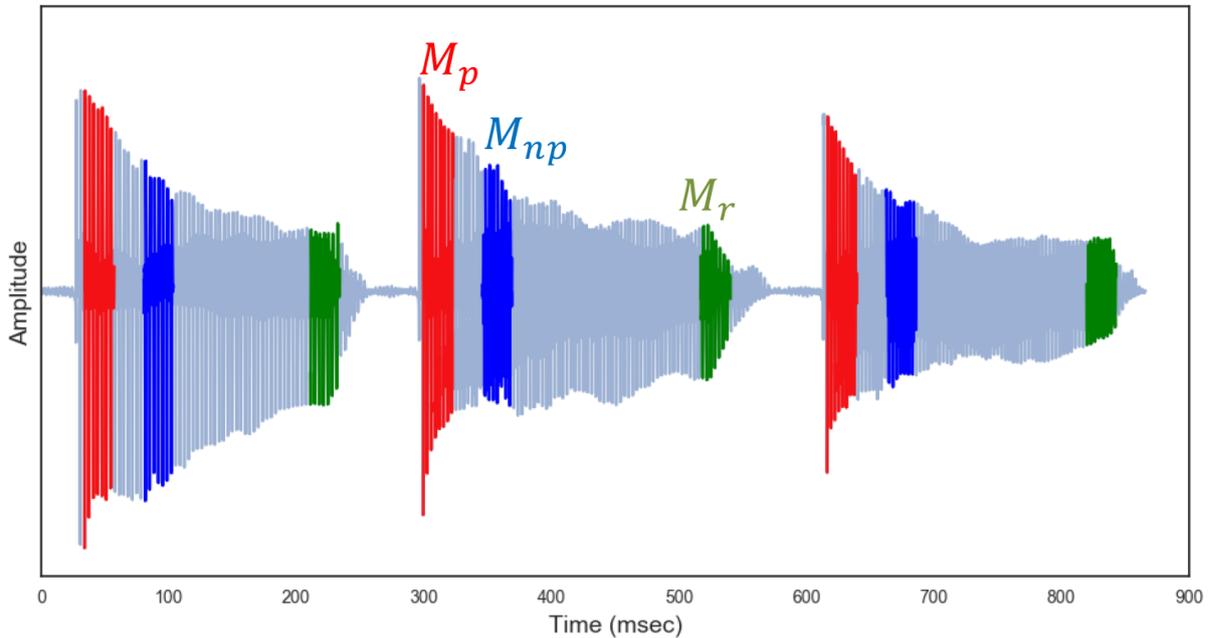


図 3.4: *Tone Similarity* の計算に用いる M_p , M_{np} , M_r の位置. M_p は peak (振幅最大時点), M_{np} は peak から約 46 ms 後, M_r は release における 12 次元の MFCC である.

得られた MFCC の係数の大半は直接的に解釈することが難しいため, 分析に用いるには何らかの計算が必要である [22]. そこで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 Similarity* (TS) を MFCC を用いて以下のように定義する.

$$TS_{np} = \frac{\mathbf{M}_p \cdot \mathbf{M}_{np}}{\|\mathbf{M}_p\| \|\mathbf{M}_{np}\|} \quad (3.2)$$

$$TS_r = \frac{\mathbf{M}_p \cdot \mathbf{M}_r}{\|\mathbf{M}_p\| \|\mathbf{M}_r\|} \quad (3.3)$$

\mathbf{M}_p , \mathbf{M}_{np} , \mathbf{M}_r はそれぞれ peak (振幅が最大となる時点), near_peak (peak から約 46 ms 後), release (音の鳴り止め始め) の 3 時点における, 周辺約 23 ms (1024 samples)² の 12 次元の MFCC である (図 3.4). 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.

²本研究ではサンプリングレートは 44100 Hz とする.

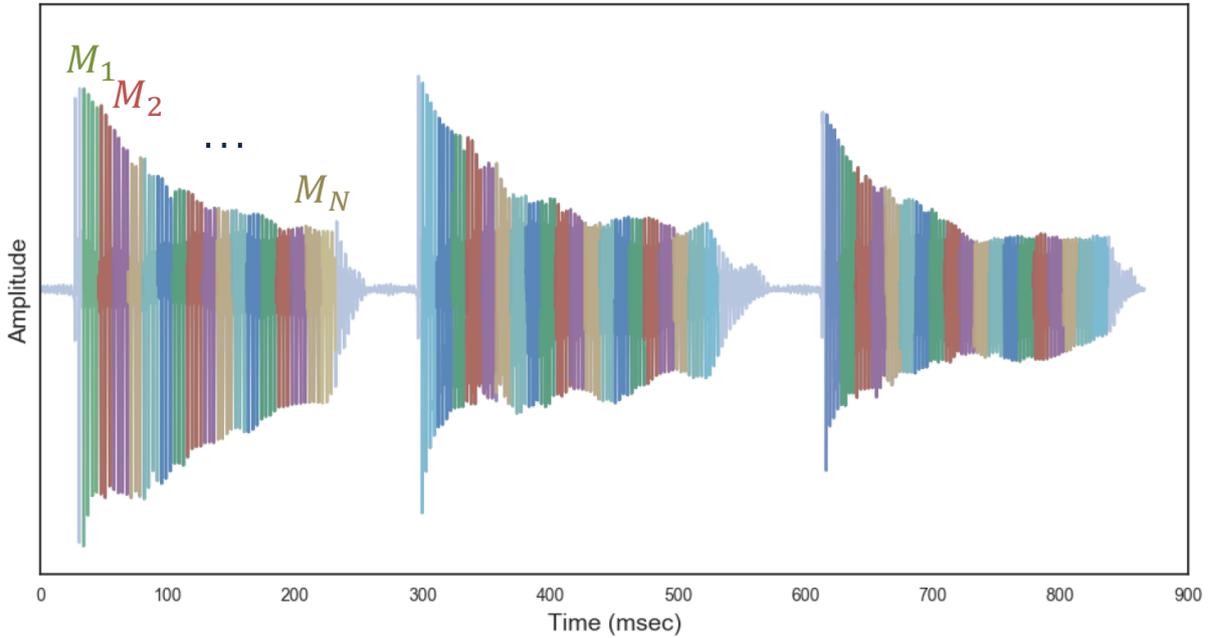


図 3.5: *Tone Instability* の計算に用いる \mathbf{M}_i ($i = 1 \sim N$) の位置. onset から release までの各フレーム (1024 samples/frame, overlap=50%) における MFCC を求める.

Tone Similarity は 2 時点間の音色を考慮したものであるが、それに対して onset から release までの全時間の音色を考慮した指標として以下の *Tone Instability* (*TIS*) を定義する.

$$TIS = 1 - \|\overline{\mathbf{M}}\| \quad (3.4)$$

$$\overline{\mathbf{M}} = \frac{1}{N} \sum_{i=1}^N \hat{\mathbf{M}}_i$$

\mathbf{M}_i は onset から release までの各フレーム (1024 samples/frame, overlap=50%) における MFCC であり (図3.5), $\hat{\mathbf{M}}_i$ は \mathbf{M}_i を正規化したものである. *TIS* は \mathbf{M}_i ($i = 1 \sim N$) をベクトルと見なしたときの向きのばらつきを表す指標であり [1], 0 から 1 までの値をとる. 音色が不安定であるほど *TIS* が大きくなると想定される.

3.3.2 Harmonic-Percussive Ratio: 撥弦音と調波音に着目した特徴量

エレキギター演奏は、左手で押さえた弦を弾くことによって弦が振動して音が鳴る. このとき、左手で弦が適切に押さえられていなければ弦が十分に振動せず、弦の振動音より撥弦時の打楽器のような音が目立った悪い音色になってしまう. このことを考慮し、撥弦音とその後の弦振動による調波音を比較した特徴量を考える.

ここで、演奏音源を調波音と打楽器音に分離する Fitzgerald の Harmonic/Percussive Source Separation (HPSS) [7] に着目する. これは 1 つの音源をギターやピアノなどの調波音とドラ

ムなどの打楽器音に分離する手法である。演奏音源をスペクトログラム表示した際に、調波音は時間軸方向（横）に滑らかで打楽器音は周波数軸方向（縦）に滑らかであることを利用して、その両者を分離するという原理に基づいている。

エレキギター演奏音のスペクトログラム（図3.2）を確認すると、撥弦による onset 付近では周波数軸方向に滑らかで、その後の弦の振動音は時間軸方向に滑らかである。そこで Fitzgerald の HPSS をエレキギターの音に適用したところ、撥弦による音と弦振動による調波音に分離することができた（図3.6）。この分離した調波音と撥弦音の最大振幅をそれぞれ A_h および A_p とし、*Harmonic-Percussive Ratio* (HPR) を以下のように定める。

$$HPR = \log_{10} \left(\frac{A_h}{A_p} \right) \quad (3.5)$$

HPR は撥弦音と調波音の最大振幅の比の対数であり、上手く押弦できている音ほど HPR が高くなると想定される。

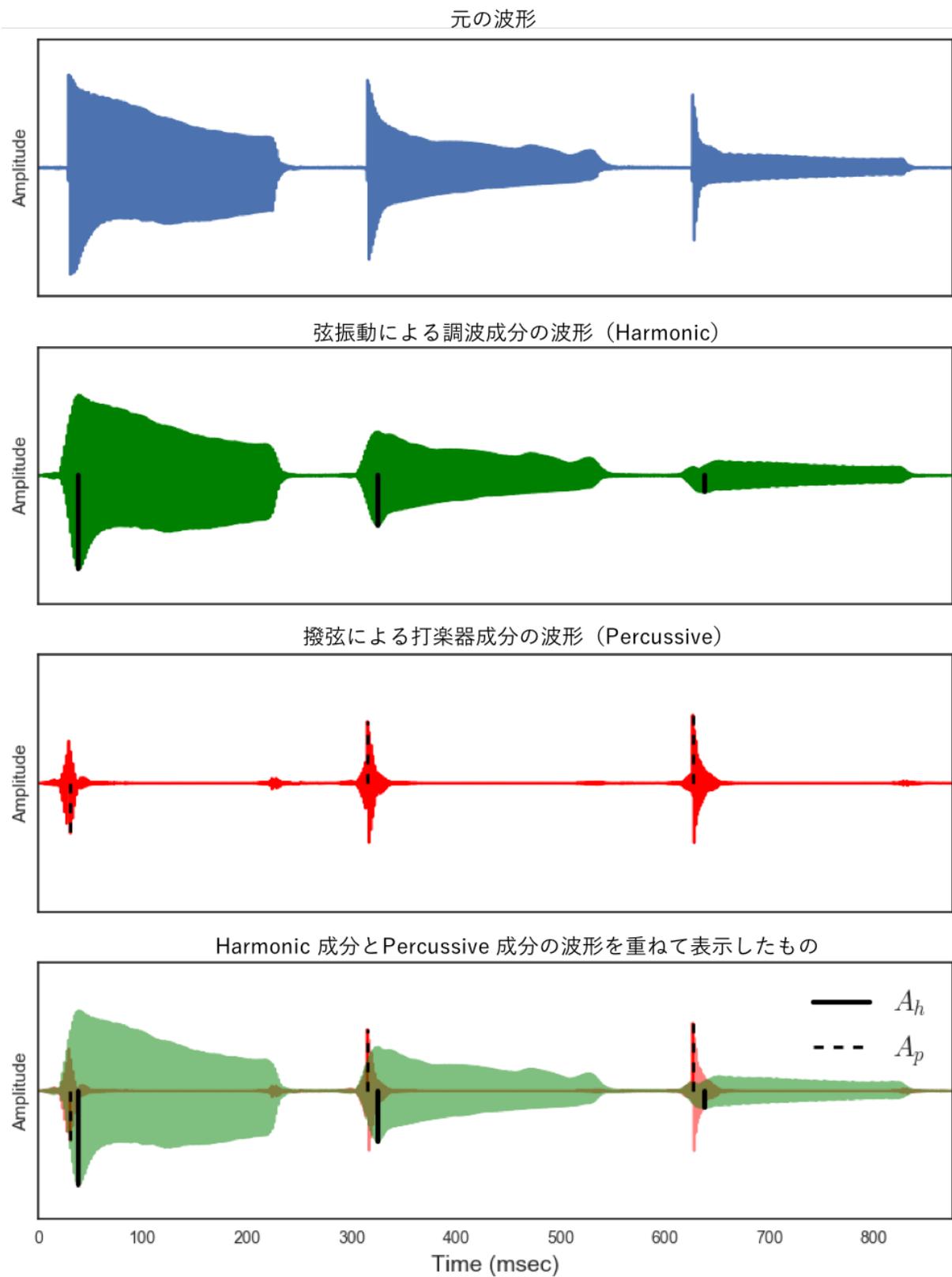


図 3.6: エレキギター の 3 つの音の波形に Harmonic/Percussive Source Separation を実行した例. この 3 つの音の HPR はそれぞれ左から順に, 0.27, 0.09, -0.61 である.

Chapter 4

エレキギター演奏に対する主観的評価の予測

前章で定義した音響的特徴量を用いて、エレキギター演奏を自動的に評価するモデルを構築する。データセットとして5人のエレキギター初心者からエレキギター演奏音源を収集し、定義した音響的特徴量を計算した。そして筆者を含む3人のエレキギター経験者が上記の演奏音源を評価し、定義した音響的特徴量から経験者による評価を予測するモデルを構築した。本章では、これらの内容の詳細について述べる。

4.1 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および音源データの処理について述べる。

4.1.1 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（図4.1）。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が、バンド演奏における他の楽器やボーカルなどの経験があり、音楽演奏に関する基本的な知識は有していた。演奏するフレーズは図4.2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4.2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クロマチック・スケールの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。



図 4.1: 演奏音源収集の様子.

実験手順は表4.1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，その演奏を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．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
フェクトなどをかけていない音を録ったが，演奏者を飽きさせないためにエフェクトをかけた音を
PCから鳴らすようにした．録音環境の詳細については表4.2に示す．

4.1.2 演奏音源データの前処理および特徴量の計算

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4.2の青枠線で囲った連続する4音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，つまり既存システムでは正しく弾けていると見なされる部分
を抽出することとした．その手順は以下の通りである．

1. 演奏音源からすべての onset を検出する [13].

表 4.1: 演奏音源収集の手順

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	〃
その他	約60分	今回は分析に用いなかったパワーコードフレーズの練習と録音
インタビュー	約15分	感想などについての open-ended 形式のインタビュー

Figure 4.2 displays musical notation for guitar exercises and simplified song phrases. (A) shows a chromatic scale in 2/4 time with fret numbers 7-8-9-10. (B) shows a 4/4 phrase with fret numbers 8-6, 7-5, 5-5, 5-7, 6-8, 10. (C) shows a 4/4 phrase with fret numbers 15-13, 15-13, 13-13, 13-15, 17. A tempo marking of 100 BPM is shown.

図 4.2: 演奏音源収集に用いたフレーズ. (A) はエレキギターの基礎練習でよく用いられるクロマチック・スケールのフレーズで, (B) と (C) は有名なエレキギター楽曲の一部を簡略化したフレーズである. 今回の分析には 5 つの青枠の部分 (8 分音符間隔の 4 音) を用いた.

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 200 ms から 450 ms の間 (BPM100 の 8 分音符の長さは 300 ms) に収まっている場合, その 4 音を抽出してデータセットに入れる.

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) のそれぞれの音源データの中から, 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. その後, 選択した 750 個の音源データを用いて, 前章で定義した音響的特徴量を計算した. 1 つの音に対して, 図 4.3 に示すような手順で処理を行った.

表 4.2: 録音環境の詳細

	使用機材	注釈
エレキギター	テレキャスター	ピックアップ (シングルコイル) はリア側.
オーディオインタフェース	LINE6 POD HD500	24 bit/44100 Hz, エフェクト等は無し.
録音ソフト	QuickTime Player (Mac)	

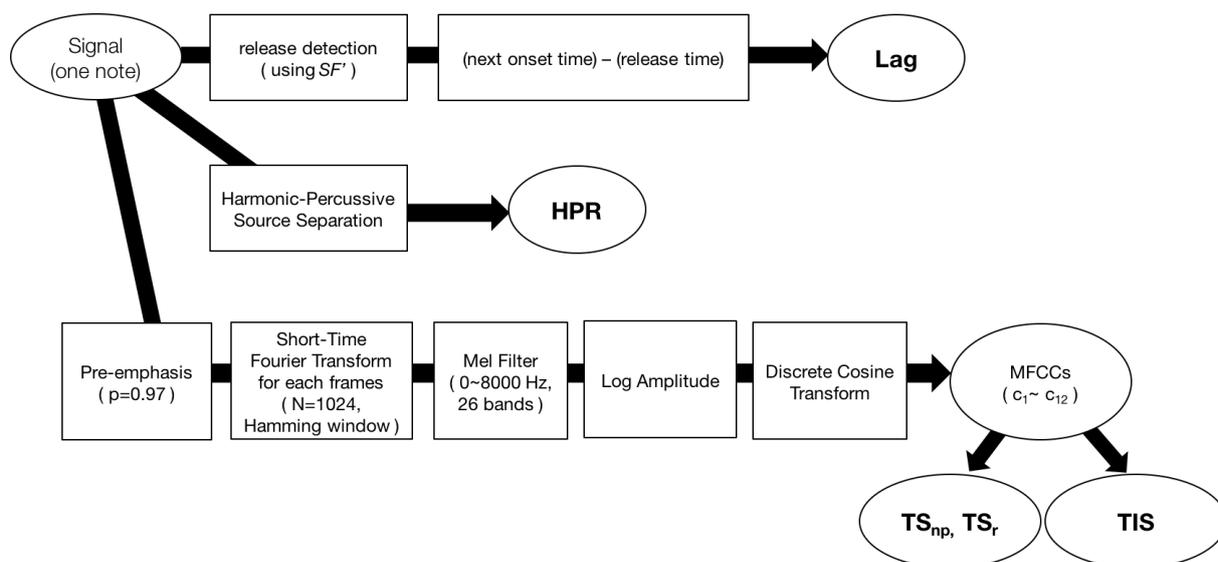


図 4.3: 定義した音響的特徴量の計算.

4.1.3 エレキギター経験者による評価付け

750 個の音源データに対して、著者を含む 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 は音色の綺麗さ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、それぞれの音源データの 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. 今回の分析で用いる 750 個の音源データ以外の 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 個の音源データに対して評価を行う。
3. 3 人の評価者の回答を比較しながら、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、著者が用意した Web ページ (図 4.4) を用いて、750 個の音源データに対して Tone と Fluency の点数をつける。
5. Tone と Fluency の点数の和を、それぞれの音源データの 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 4.5 のようになった ($Mean = 8.26$, $SD = 2.64$)。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 [9] は 0.73 であったため、程よく一致していると言える。以降の分析では、Tone と Fluency の評価および Score は 3 人の経験者が付与した点数の平均を用いる。

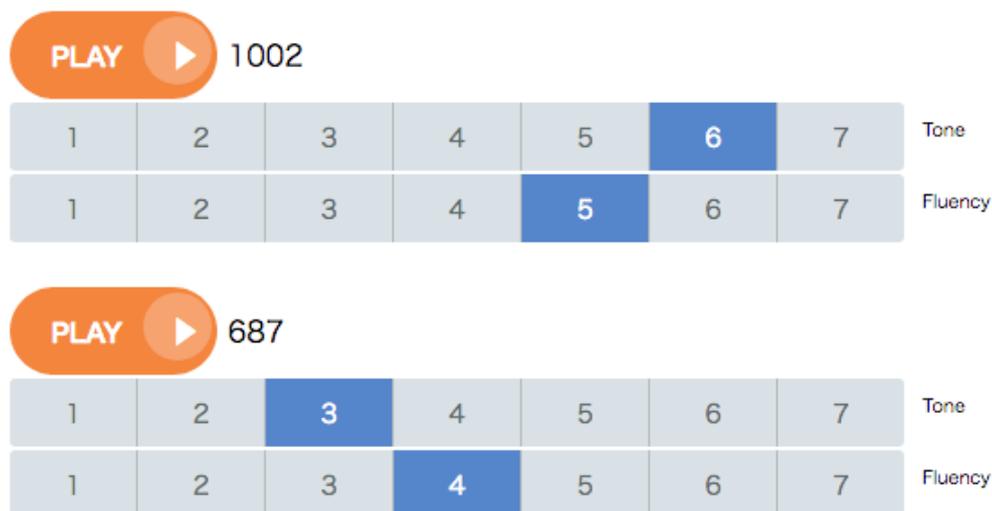


図 4.4: 経験者による評価用に作成した Web ページの画面.

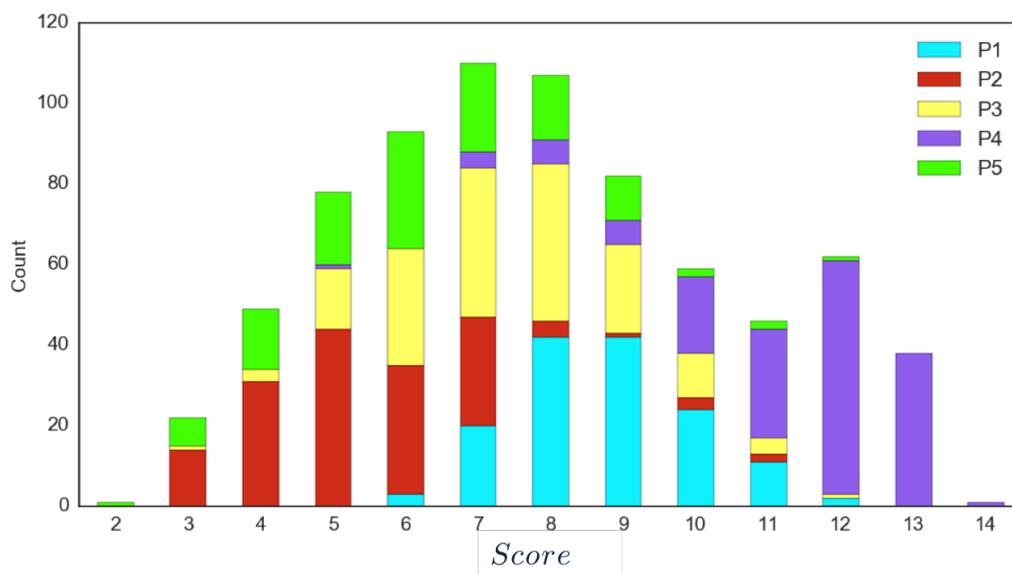


図 4.5: 3 人のエレキギター経験者がつけた *Score* の平均 (小数点以下四捨五入) の分布. $Mean = 8.26$, $SD = 2.64$ である. 5 人の演奏者ごとのデータで色を分けている.

4.2 定義した特徴量と評価値との関連性の分析

定義した音響的特徴量とエレキギター経験者による 750 個の音源データの *Score* の関係性を調べるために、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 *Score* の予測性能を比較する。使用する特徴量の一覧は表4.3の通りである。それぞれの音源データは 4 音から成るため、*Lag* は各音間の 3 個である。また、前章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ため、 TS_{np} 、 TS_r 、 TIS はそれぞれ 1~3 音目の 3 個である。これら 5 種類 × 3 個 (合計 15 個) 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。

4.2.1 SVR による予測性能比較

以下に示す 3 つの方法で cross validation を行い、それぞれの場合におけるテストデータに対する平均二乗誤差 (MSE) を求める。

10-fold cross validation (10-fold)

データをランダムに 10 分割にして、そのうち 9 個を学習データ、残り 1 個をテストデータとする。

leave-one-novice-out cross validation (LONO)

4 人の演奏者によるデータを学習データ、残り 1 人の演奏者によるデータをテストデータとする。

leave-one-phrase-out cross validation (LOPO)

4 つのフレーズのデータを学習データ、残り 1 つのフレーズのデータをテストデータとする。

これら 3 つ cross validation で SVR (RBF カーネル, $C = 10.0$, $\gamma = 0.01$) による予測性能評価を行った結果を表4.4に示す。 Lag 、 TS_{np} 、 TIS 、 HPR の 4 種類の特徴量を用いたときの予測性能が最も高かった。また、これら 4 種類の特徴量を用いたときの leave-one-novice-out cross validation と leave-one-phrase-out cross validation の結果の詳細を表4.5に示す。

表 4.3: 定義した音響的特徴量。これら 5 種類の特徴量 (計 15 個) を説明変数とする。

名称	個数	説明
Lag	3	release から次の onset までの時間
TS_{np}	3	peak と peak から 46 ms 後の 2 時点における MFCC の cos 類似度
TS_r	3	peak と release の 2 時点における MFCC の cos 類似度
TIS	3	peak から release までの各フレームにおける MFCC のばらつき
HPR	3	Harmonic 成分と Percussive 成分の最大振幅比の対数

表 4.4: SVR による *Score* 予測性能評価. 5 種類の特徴量の組み合わせを変えながら, 10-fold cross validation, leave-one-novice-out cross validation (LONO), leave-one-phrase-out cross validation (LOPO) の 3 種類の cross validation でテストデータに対する MSE を求めたところ, *Lag*, *TS_{np}*, *TIS*, *HPR* の 4 種類の特徴量を用いたときの予測性能が最も高かった.

Features					MSE		
<i>Lag</i>	<i>TS_{np}</i>	<i>TS_r</i>	<i>TIS</i>	<i>HPR</i>	10-fold	LONO	LOPO
✓					3.53	7.00	4.09
	✓				3.98	6.69	4.11
		✓			6.86	9.57	7.72
			✓		6.10	8.59	6.48
				✓	5.23	8.98	5.35
✓	✓				1.56	2.21	1.66
✓		✓			2.83	5.52	3.55
✓			✓		1.82	2.81	2.00
✓				✓	1.77	3.13	2.48
	✓	✓			3.69	6.69	3.91
	✓		✓		3.77	6.60	3.97
	✓			✓	3.74	7.25	3.99
		✓	✓		5.78	8.63	6.53
		✓		✓	5.10	9.20	6.19
			✓	✓	5.00	9.10	5.71
✓	✓	✓			1.50	2.17	1.68
✓	✓		✓		1.41	1.99	1.61
✓	✓			✓	1.27	1.81	1.68
✓		✓	✓		1.69	2.68	2.29
✓		✓		✓	1.21	1.65	1.49
✓			✓	✓	1.16	1.58	1.31
	✓	✓	✓		3.68	6.81	4.00
	✓	✓		✓	3.41	6.88	3.71
	✓		✓	✓	3.44	6.88	3.75
		✓	✓	✓	4.93	9.69	5.71
✓	✓	✓	✓		1.38	1.98	1.67
✓	✓	✓		✓	1.11	1.45	1.40
✓	✓		✓	✓	1.08	1.43	1.30
✓		✓	✓	✓	1.13	1.49	1.39
	✓	✓	✓	✓	3.38	7.11	3.69
✓	✓	✓	✓	✓	1.08	1.46	1.38

表 4.5: *Lag*, *TS_{np}*, *TIS*, *HPR* を説明変数としたときの予測性能評価.

(1) leave-one-novice-out CV			(2) leave-one-phrase-out CV		
学習データ	テストデータ	MSE	学習データ	テストデータ	MSE
P2, P3, P4, P5	P1	1.33	(B1), (B2), (C1), (C2)	(A)	1.43
P1, P3, P4, P5	P2	1.58	(A), (B2), (C1), (C2)	(B1)	1.32
P1, P2, P4, P5	P3	1.44	(A), (B1), (C1), (C2)	(B2)	1.35
P1, P2, P3, P5	P4	1.91	(A), (B1), (B2), (C2)	(C1)	1.19
P1, P2, P3, P4	P5	0.91	(A), (B1), (B2), (C1)	(C2)	1.21

4.3 考察

4.3.1 10-fold と LONO および LOGO の結果の比較

leave-one-novice-out cross validation においては、特に P4 をテストデータとしたときの MSE が、10-fold cross validation における MSE に比べて大きくなっている。これは図4.5からわかるように、Score の高い部分の大半を P4 が占めており、この部分の学習が十分にできていなかったためであると考えられる。より多くの演奏者に適合したモデルにするためには、さらに多くの演奏者からデータを集める必要がある。

leave-one-phrase-out cross validation においては、特に (A) をテストデータとしたときの MSE が大きい。これは、他の4つのフレーズは異なる弦を含む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るなどの違いによるものであると考えられる。このようなフレーズの特徴の違いが、定義した音響的特徴量では捉えられていない部分に影響を与えている可能性がある。今回は実験に参加するエレキギター初心者への考慮もあり、用意したフレーズの弦移動のパターンや使用する弦の種類を少数にしたが、今後はより多様なフレーズの音源データを用いた分析が必要である。

4.3.2 Lag が Tone の評価に与える影響

表4.6より、Tone 単体における点数予測においても Lag, TS_{np} , TIS, HPR の4種類の特徴量を用いたときの精度が最も良いことがわかる。ここで、これらの特徴量から Lag を除いた3種類を用いたときの MSE (表4.6赤背景部分)を確認すると、これら4種類の特徴量を用いたときに比べて MSE が大きくなっている。前章で述べた通り Lag は演奏のなめらかさを評価するための指標として定義したが、この結果から Lag は Tone の評価にも影響を与えているとわかる。

Lag が Tone の評価に影響を与える原因としては、今回の実験で収集した演奏音源の中には、押弦した指を離すのが早すぎて音がすぐに途切れてしまう (つまり Lag が非常に大きい) ものが多かったことが考えられる。図4.6の左のように音の長さが短すぎた場合は、その音に関する音色そのものの良さを判断することは難しい。そこで、演奏者 (P1~P5) ごとの3つの Lag の平均の分布 (図4.7 (左)) と、特徴量に Lag を含めた場合と含めなかった場合における leave-one-novice-out cross validation の Tone の点数予測誤差の分布 (図4.7 (右))を確認する。その結果、例えば Lag が他の演奏者よりも長い P2 においては、Lag を含めずに Tone の点数予測を行ったときは予測値が実際の評価値よりも高くなる傾向があることがわかる。これは P2 の演奏は音が短いため音色の評価が難しいことに加え、演奏がぎこちなく聞こえるがゆえに音色も悪いであろうというバイアスがデータ収集における Tone の評価時にかかっており、Tone が低めに評価されていたためであると推測される。

表 4.6: SVR による Tone および Fluency の評価予測における平均二乗誤差 (MSE). 5 種類の特徴量の組み合わせを変えながら, 3 種類の cross validation でテストデータに対する MSE を求めた. Tone の予測性能に着目すると, TS_r 以外の 4 種類の特徴量を用いたとき (青部分) が最も高い. そこから Lag を除いたとき (赤部分) Tone の予測性能が落ちており, 本来は演奏のなめらかさの指標である Lag が Tone の評価にも影響を与えていることがわかる.

Lag	Features				Tone MSE			Fluency MSE		
	TS_{np}	TS_r	TIS	HPR	10-fold	LONO	LOPO	10-fold	LONO	LOPO
✓					1.35	2.71	1.54	0.72	1.20	0.83
	✓				0.90	1.11	0.90	1.44	2.73	1.54
		✓			1.90	2.93	2.26	1.92	2.73	2.07
			✓		1.48	1.86	1.59	1.86	2.94	1.96
				✓	1.21	1.92	1.26	1.69	3.09	1.74
✓	✓				0.55	0.71	0.58	0.45	0.66	0.50
✓		✓			1.14	2.14	1.44	0.58	0.97	0.70
✓			✓		0.69	1.01	0.78	0.47	0.65	0.52
✓				✓	0.62	1.03	0.80	0.49	0.82	0.66
	✓	✓			0.84	1.10	0.88	1.32	2.65	1.49
	✓		✓		0.89	1.18	0.97	1.30	2.51	1.43
	✓			✓	0.78	1.10	0.82	1.38	2.86	1.57
		✓	✓		1.36	1.78	1.57	1.82	3.13	2.13
		✓		✓	1.15	1.81	1.46	1.67	3.26	2.00
			✓	✓	1.09	1.67	1.28	1.65	3.30	1.89
✓	✓	✓			0.53	0.71	0.59	0.43	0.61	0.50
✓	✓		✓		0.51	0.66	0.62	0.42	0.55	0.48
✓	✓			✓	0.42	0.54	0.50	0.42	0.66	0.56
✓		✓	✓		0.61	0.89	0.83	0.45	0.69	0.56
✓		✓		✓	0.44	0.56	0.53	0.40	0.53	0.51
✓			✓	✓	0.39	0.50	0.45	0.39	0.54	0.47
	✓	✓	✓		0.83	1.17	0.90	1.27	2.61	1.47
	✓	✓		✓	0.72	1.10	0.80	1.26	2.60	1.44
	✓		✓	✓	0.73	1.13	0.79	1.23	2.62	1.38
		✓	✓	✓	1.07	1.76	1.25	1.63	3.57	1.98
✓	✓	✓	✓		0.48	0.60	0.55	0.42	0.58	0.50
✓	✓	✓		✓	0.37	0.49	0.47	0.39	0.55	0.50
✓	✓		✓	✓	0.35	0.43	0.40	0.39	0.55	0.49
✓		✓	✓	✓	0.39	0.49	0.47	0.38	0.55	0.51
	✓	✓	✓	✓	0.72	1.16	0.78	1.22	2.73	1.41
✓	✓	✓	✓	✓	0.36	0.44	0.46	0.39	0.57	0.52

表 4.7: Lag , TS_{np} , TIS , HPR を説明変数としたときの Tone と Fluency の予測性能.

(1) leave-one-novice-out CV				(2) leave-one-phrase-out CV			
学習データ	テストデータ	MSE		学習データ	テストデータ	MSE	
		Tone	Fluency			Tone	Fluency
P2, P3, P4, P5	P1	0.44	0.55	(B1), (B2), (C1), (C2)	(A)	0.40	0.62
P1, P3, P4, P5	P2	0.50	0.59	(A), (B2), (C1), (C2)	(B1)	0.38	0.41
P1, P2, P4, P5	P3	0.44	0.47	(A), (B1), (C1), (C2)	(B2)	0.38	0.51
P1, P2, P3, P5	P4	0.39	0.81	(A), (B1), (B2), (C2)	(C1)	0.43	0.38
P1, P2, P3, P4	P5	0.41	0.38	(A), (B1), (B2), (C1)	(C2)	0.42	0.56

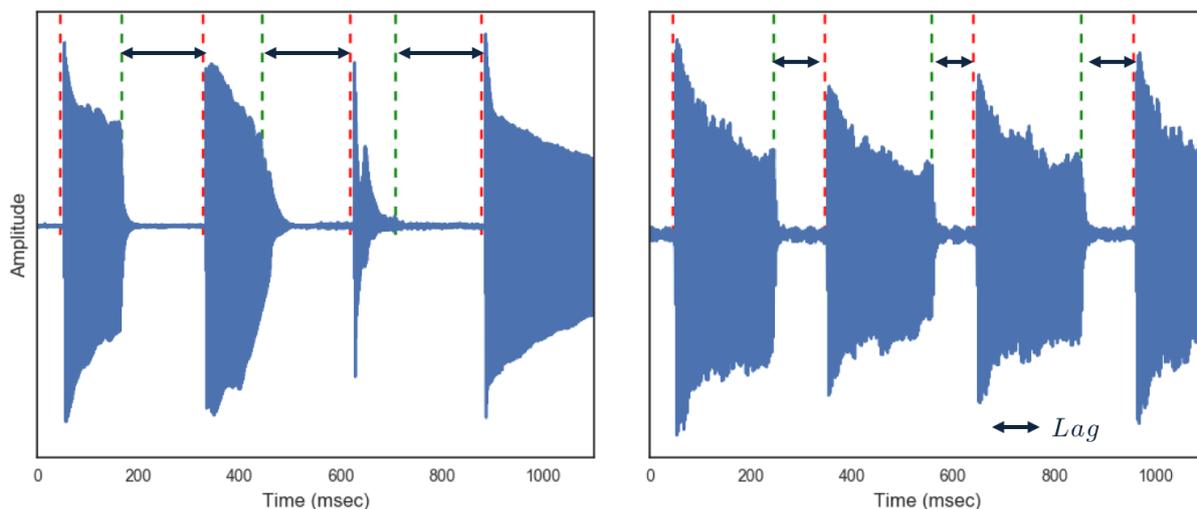


図 4.6: *Lag* が大きい傾向がある P2 の演奏音波形 (左) と, 逆に *Lag* が小さい傾向がある演奏音波形 (右). これらの波形の *Lag* の平均はそれぞれ 168 ms と 95 ms である. *Lag* が大きいと音が鳴っている時間が短く, Tone の評価が行いにくい.

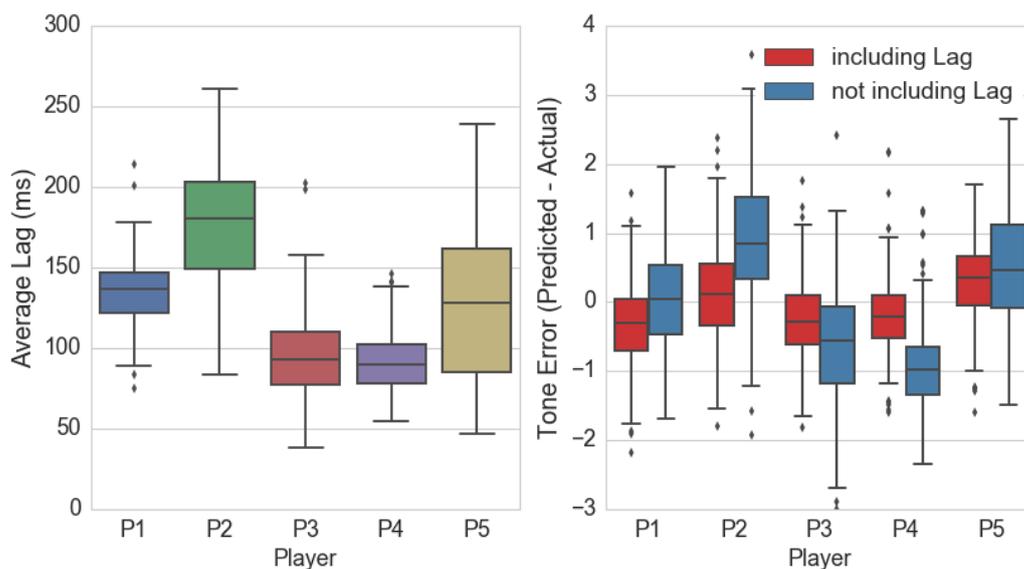


図 4.7: 各演奏者 (P1~P5) の *Lag* の分布 (左) と, 説明変数に *Lag* を含めたときと含めなかったときの Tone の予測誤差の分布 (右). 特に *Lag* が大きい P2 の Tone 評価予測においては, 説明変数に *Lag* を含めなかったときに実測値よりも予測値のほうが高くなる傾向がある.

4.3.3 Tone 評価における TS_{np} の影響力

表4.6から、 TS_{np} は単体でも Tone の評価予測における MSE が 0.90 になるなど、Tone の評価予測に与える影響が他の特徴量に比べて高いことが推測される。そこで 750 個すべての音源データの Tone の評価と 3 音の平均 TS_{np} の散布図 (図4.8左) を表示したところ、両者に正の相関が見られた ($r = 0.58$)。また前小節で述べたように Lag が大きい音の音色は正しく評価しにくいことを考慮し、3 音の Lag の最大値がその中央値以上であるものを除いた 375 個の音源データに関して、同様の散布図 (図4.8右) を表示した。その結果、全音源データを用いたときと比べて Tone と 3 音の平均 TS_{np} に強い正の相関が見られた ($r = 0.79$)。このことから、特に Lag が小さい音の Tone 評価においては TS_{np} の影響力が大きいことが確認できた。

4.3.4 Tone 評価の予測誤差が大きかった音源データ

10-fold cross validation において Tone 評価の予測誤差が大きかったものとして、リズムの悪さによって Lag が大きくなってしまった演奏音源データがあげられる。例えば図4.9の演奏音の波形は、2 音目の 3 音目の間の Lag が比較的大きいが、これは 2 音目の音が短いからではなく 3 音目を鳴らすタイミングが遅かったからである。この演奏の経験者による Tone の評価は 5.33 であったが、このリズムの悪さによって起因する Lag の大きさによって Tone 評価の予測が 3.78 となってしまった。今回の分析では正しいタイミングと音程で弾けている演奏音源を抽出して使用したが、タイミングが正しいとみなす閾値を低めに設定したためこのような問題が起こったと考えられる。

また、演奏中に他の弦を鳴らしてしまった演奏音源データの Tone 評価予測誤差も大きかった。エレキギター演奏においては、押弦している指を離す際にその指で弦を弾いてしまったり、ピックがミュートしていない弦に触れてしまったりすることで、演奏に不要な弦が鳴ってしまうことがある。経験者は不要な弦が鳴っている演奏音源の Tone を低く評価したが、本研究で定義した音響的特徴量では不要な弦が鳴っていることを検知できなかったため、不要な弦が鳴っている演奏音源の Tone が実際の評価値よりも高く予測されたと考えられる。

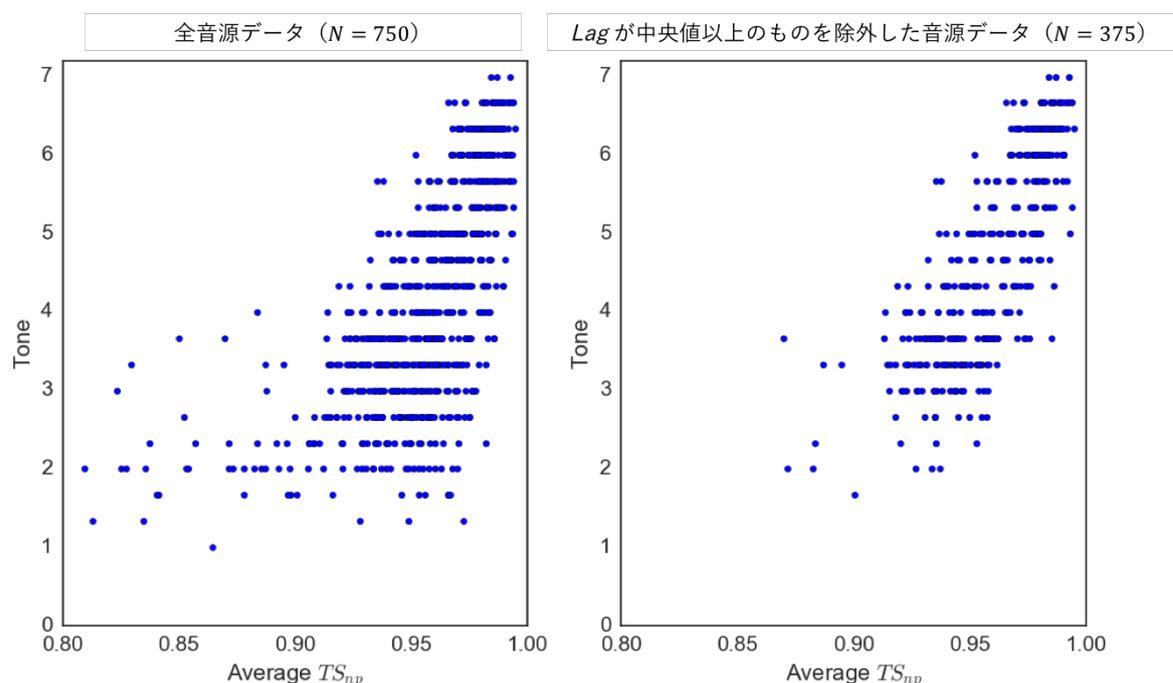


図 4.8: 各音源データの Tone の評価と 3 音の平均 TS_{np} の散布図. 左は 750 個全ての音源データを用いたとき, 右は 3 音の Lag の最大値がその中央値以上であるものを除いた 375 個の音源データを用いたときの結果であり, 相関係数はそれぞれ 0.58 と 0.79 であった. Tone の評価が適切に行える音, つまり Lag ある程度は小さく十分に長く鳴っている音に対しては特に, TS_{np} を用いた Tone 評価予測が有効であることがわかる.

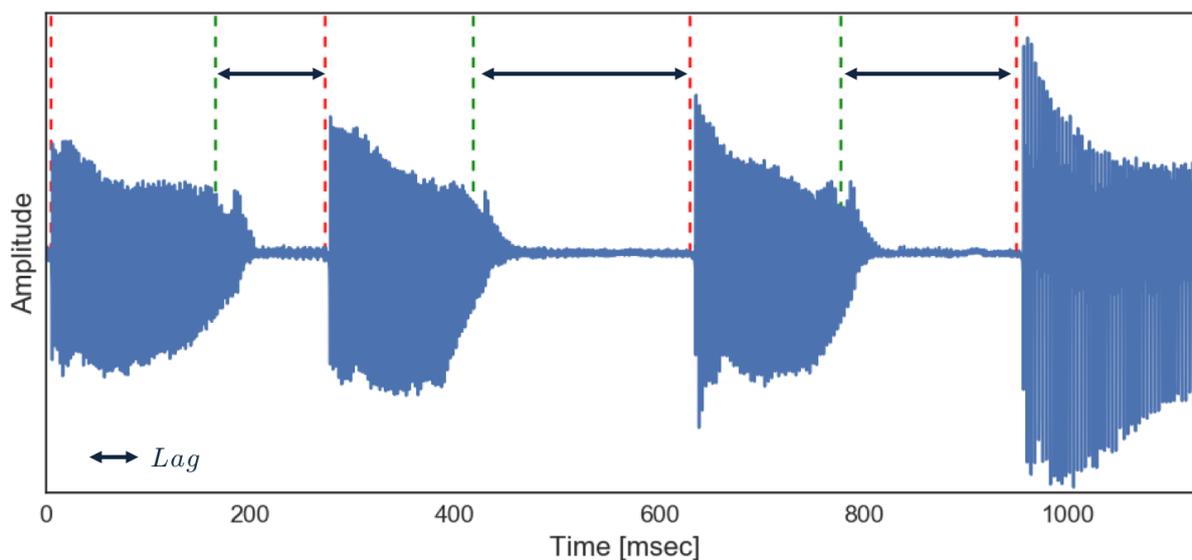


図 4.9: Tone 評価予測誤差が大きかった演奏波形の例. 3 音目を鳴らすタイミングが遅れてしまったため, 2 音目の 3 音目の間の Lag が比較的大きく (212 ms) になっている. この演奏音源の経験者による Tone 評価は 5.33 と高かったが, この Lag の大きさによって Tone 評価予測が 3.78 と低くなってしまった.

Chapter 5

押弦ミスの自動分類

前章では単音のエレキギター演奏を自動評価するために、定義した音響的特徴量から経験者による主観的評価を予測するモデルを構築した。しかし、このモデルを用いてユーザのエレキギター演奏に対してスコアを提示するだけでは、練習を十分に支援することはできない。低いスコアが出た場合、特に初心者にとっては自分の演奏のどの部分が悪かったのかわからないためである。演奏音源データ収集に参加した5人の初心者に、エレキギター演奏後に簡単なインタビューをした結果、練習システムに求める機能について実際に以下のような声を得られている。

「なんかちゃんと、得点として、何回ミスりましたよみたいなもの、数字として現れるとモチベとして繋がるかもしれない。テストの点数みたいな。…うまく弾けてなかったら、こういうところがダメだからみたいな言ってくれるといいよね。」(P2)

「自分が鳴らしている音がどれくらいいい感じなのかをほしいかもしれない。弾いているのに必死だと（自分の音を）聴くのが大変。…カラオケの採点みたいな。」(P4)

「その場で、どう違ったのかが出ると嬉しいな。…何が良くないのか。…あとは弾き終わった後に自分の音を再生してくれる機能とか。（自分の音を）聴く余裕がない。アプリを見る余裕はないから音で知らせてほしいな。」(P5)

インタビュー中の発言からもわかるように特に初心者は演奏に必死になるため、自分の演奏音を聴いて良し悪しを判断することやどこが悪かったのかを見つけることが難しい。そのため、演奏中の不適切な音を検知してユーザに改善案を提示するシステムがあれば、効率良くエレキギター練習ができると考えられる。

そこでまずは、エレキギター演奏において初心者が特に難しいと感じる左手の押弦に焦点を当てる。インタビューにおいても、左手の難しさに関する発言が多く得られている。



図 5.1: 不適切な押弦の例。左の写真のようにフレットのすぐ近くで押弦するのが正しいが、真ん中の写真はフレットから指が遠く離れており、十分に押弦できない可能性がある。右の写真はフレット上で押弦しており、不適切な音が発せられる。

「左手の指を動かすことに慣れてなさすぎる。」(P2)

「単音は（左手の指で）どこを押さえるのかに神経を使う。」(P4)

「指が思うように動かない。ピロピロするときの左手のこの、細かい動作が動かない。早すぎる。細かい。」(P5)

インタビューの結果も考慮し、左手の押弦が適切に行なわれているかを判定する手法について考える。そのためにまず、左手の押弦のミスの種類について検討する。

5.1 左手の押弦ミスの種類

エレキギター経験が6年以上である3人から意見を収集し、エレキギターで単音のフレーズを弾く際の押弦に関する主なミスを列挙した。

十分に弦を押さえられていない。

フレットから離れた部分を押弦する（図5.1中央）、あるいは指の腹などで押弦するなどした場合、弦が十分に押さえられずに音がうまく鳴らない。

フレット上で弦を押さええている。

押弦する際にはフレット（金属製の棒）のすぐ近くを押さえるのが正しいが、フレット上で弦を押さえってしまうと不適切な音が出る（図5.1右）。

押弦している指を離すのが早すぎる。

押弦している指を離すと鳴っている音が止まるが、指を離すのが早すぎると音が滑らかに繋がらない。

以上のミスを検知してフィードバックするためには、以下を実現する手法について考える必要がある。

A. 音を「正しく押弦できている音/十分に弦を押しえられていない音/フレット上を押しえている音」の3つに分類する.

B. 音が止まったタイミングを検知する.

B に関しては, *Lag* を求める際に用いた音の鳴り止め始めを検知する手法が使用できるため, A を実現する手法について考える.

5.2 押弦におけるミスの自動検知手法

音から押弦に関係しそうな音響的特徴量を抽出し, その特徴量を用いて音を分類する. 本節では, まず分析に用いるデータセットについて説明し, 使用する音響的特徴量とそれを用いた分類性能について述べる.

5.2.1 データセット

音を「正しく押弦できている音 (Good) /フレット上を押しえている音 (On Fret) /十分に弦を押しえられていない音 (Not Enough)」の3つに分類する手法を考えるために, 実際にこれらの音を鳴らしたデータセットが必要となる. そこでエレキギターの各弦 (1~6 弦) の 1~12 フレットで上記の3種類の音をそれぞれ3回ずつ鳴らしたデータセットを, 筆者が実際に演奏することで用意した. また3種類のエレキギターを用いて, それぞれのエレキギターで上記の録音を行った. そのため, 録音した音は合計で 1944 個である.

5.2.2 使用する音響的特徴量

MFCC (M), Δ MFCC (Δ M)

音の振幅が最大になる時点周辺の MFCC と, その時点から 5 フレーム (1024 samples/frame, overlap=50%) における Δ MFCC を使用する. それぞれ 1~12 次の係数を用いる.

Tone Similarity (TS_{np})

各音の振幅が最大になる時点とそこから約 46 ms 後の時点における MFCC の \cos 類似度を *Tone Similarity* (TS_{np}) とし, 音の安定度の指標とする. 前章で用いた TS_{np} と同一である.

Harmonic-Percussive Ratio (HPR)

音の Harmonic 成分と Percussive 成分の最大振幅の比の対数を *Harmonic-Percussive Ratio* (HPR) とする. 前章で用いた HPR と同一である.

表 5.1: 4 種類の特徴量の組み合わせを変化させたときの分類精度比較. 10-fold cross validation においては 4 種類の特徴量すべてを用いたときの精度が最も高いが, leave-one-guitar-out cross validation (LOGO) においては, MFCC, Δ MFCC, HPR の 3 つを説明変数として用いたときの精度が最も高い.

Features				Mean ACC		Features				Mean ACC	
MFCC	Δ MFCC	TS_{np}	HPR	10-fold	LOGO	MFCC	Δ MFCC	TS_{np}	HPR	10-fold	LOGO
✓				.662	.562		✓		✓	.844	.820
	✓			.722	.709			✓	✓	.730	.696
		✓		.389	.384	✓	✓	✓		.801	.702
			✓	.744	.709	✓	✓		✓	.879	.853
✓	✓			.794	.689	✓		✓	✓	.812	.739
✓		✓		.714	.620		✓	✓	✓	.846	.828
✓			✓	.792	.723	✓	✓	✓	✓	.884	.844
	✓	✓		.722	.709						

5.2.3 SVM による分類

上述した 4 種類の特徴量の組み合わせを変えながら, Support Vector Machine (SVM) による分類精度を比較する. 評価は以下の 2 種類の cross validation における Accuracy の平均とする.

10-fold cross validation (10-fold)

データをランダムに 10 分割にして, そのうち 9 個を学習データ, 残り 1 個をテストデータとする.

leave-one-guitar-out cross validation (LOGO)

2 つのエレキギターによるデータを学習データ, 残り 1 つのエレキギターによるデータをテストデータとする.

これら 2 種類の cross validation による性能評価の結果を表 5.1 に示す. 10-fold においては 4 種類の特徴量全てを用いたときの予測精度が最も高いが, LOGO においては TS_{np} を除いたときのほうが予測精度が高くなっている. この原因に関しては次節で考察する. また, MFCC を除いた 3 種類の特徴量を用いて LOGO を行ったときの結果を混同行列で表示したものが表 5.2 である. この表から, 十分に弦を押しえられていない音の検知が最も精度よく行えていることがわかる.

5.3 考察

5.3.1 HPR の分類精度への影響力

HPR は単体で用いた場合でも 10-fold における accuracy が .744 になる. このことから, HPR は音の分類に大きな影響を与えていることがわかる. そこで分類する 3 種類の音それぞれの

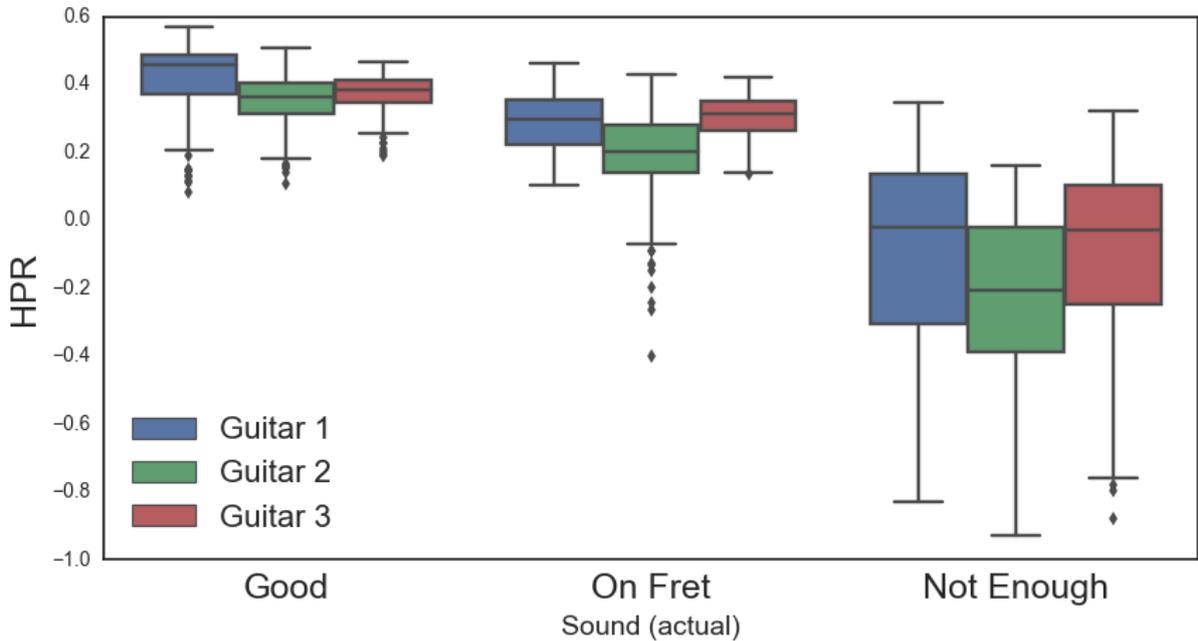


図 5.2: 音の種類による HPR の分布の差. どのエレキギターにおいても, 十分に押弦できていない音 (Not Enough) は HPR が低い傾向がある.

HPR の分布を確かめるために, 箱ひげ図によるプロットを行った結果を図5.2に示す. この図からわかるように, どのエレキギターにおいても「十分に押弦できていない音」に関しては HPR が低い傾向がある. この傾向が, 表5.2で見られるような「十分に押弦できていない音」の検知精度の高さにつながっていると考えられる.

5.3.2 エレキギターの個体差が TS_{np} に与える影響

LOGO においては TS_{np} を用いたときに精度が落ちていることが表5.1からわかる. これは, TS_{np} がエレキギターの個体差に影響を受けやすい特徴量であることを示唆している. そこで「正しく押弦できている音」に関して, 各エレキギターの弦別の TS_{np} の分布を確認した (図5.3). その結果, エレキギターによっては特定の弦の TS_{np} の値が大きく異なっているこ

表 5.2: 3 種類の特徴量 (MFCC, $\Delta MFCC$, HPR) を説明変数として LOGO を行ったときの分類結果. 十分に弦を押さえていない音の検知精度が最も高い. (○... 正しく押弦できている音. △... フレット上を押さえている音. ×... 十分に弦を押さえていない音.)

		predicted					
		○	△	×			
actual	○	556	65	27	○	.869	.858
	△	70	533	45	△	.805	.823
	×	14	64	570	×	.888	.880

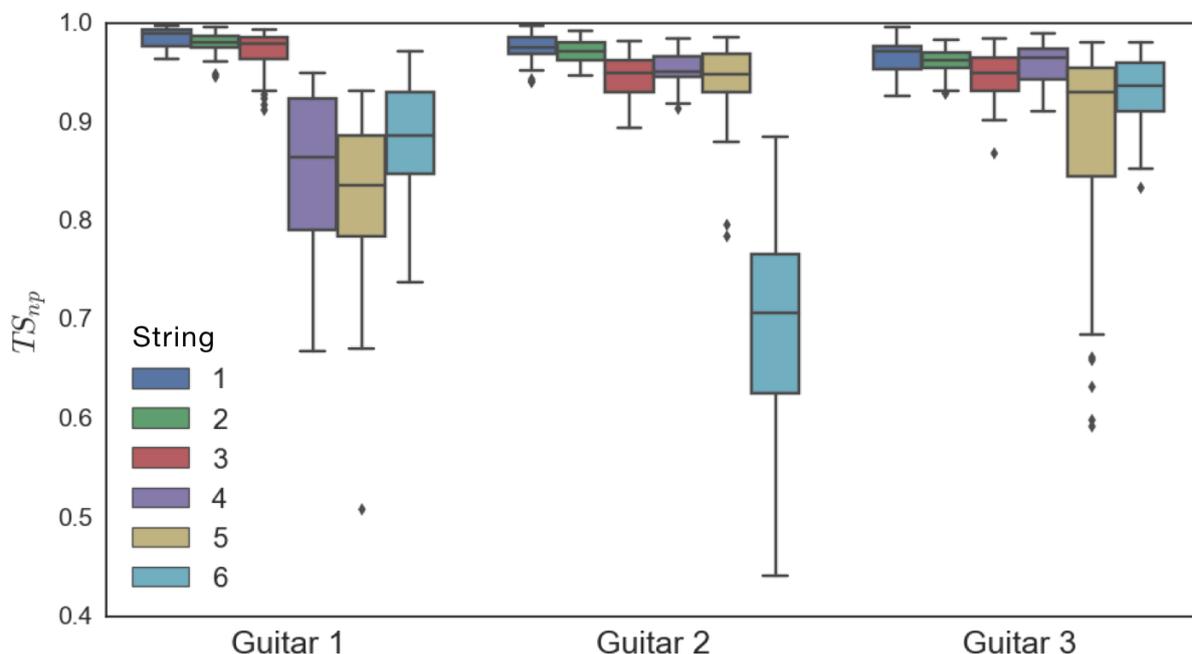


図 5.3: 各エレキギターの正しく押弦できている音の弦別の TS_{np} の分布. 低音弦における TS_{np} が低い傾向があり, 特に Guitar 1 においてはそれが顕著である.

とがわかった. 例えば Guitar 2 においては, 6 弦の TS_{np} が他の弦に比べて低くなっている. この原因として, 弦高をはじめとしたエレキギターのコンディションによって, 正しく押弦できていても特定の弦の音に「ビビリ」が生じて TS_{np} が下がっていることが考えられる. エレキギターにおいては特に低音弦 (6 弦が最も低い) に「ビビリ」が発生しやすく, 図 5.3 からも低音弦の TS_{np} が低い傾向があることから, TS_{np} にエレキギターのコンディションによる弦の「ビビリ」が影響を与えていると推測される.

5.3.3 弦の違いが分類精度に与える影響

エレキギターの各弦はそれぞれ微妙に音色が異なる. 例えば, 6 弦 7 フレットの音と 5 弦 2 フレットの音は音の高さは同じであるが音色は異なる. 直感的には, 低音弦のほうが丸みを帯びた音, 高音弦のほうが固い音が鳴る. このような音色の違いが, 分類精度に影響を与える可能性がある. そこで, MFCC, Δ MFCC, HPR の 3 種類の特徴量で LOGO を行ったときの結果を弦ごとに表示したものが表 5.3 である. この表からわかるように, 6 弦の音の分類

表 5.3: 3 種類の特徴量 (MFCC, Δ MFCC, HPR) を説明変数として LOGO を行ったときの分類精度を弦ごとに表示したもの. 6 弦は他の弦に比べて分類精度が低くなっている.

	1 弦	2 弦	3 弦	4 弦	5 弦	6 弦
ACC	.873	.842	.920	.944	.935	.605

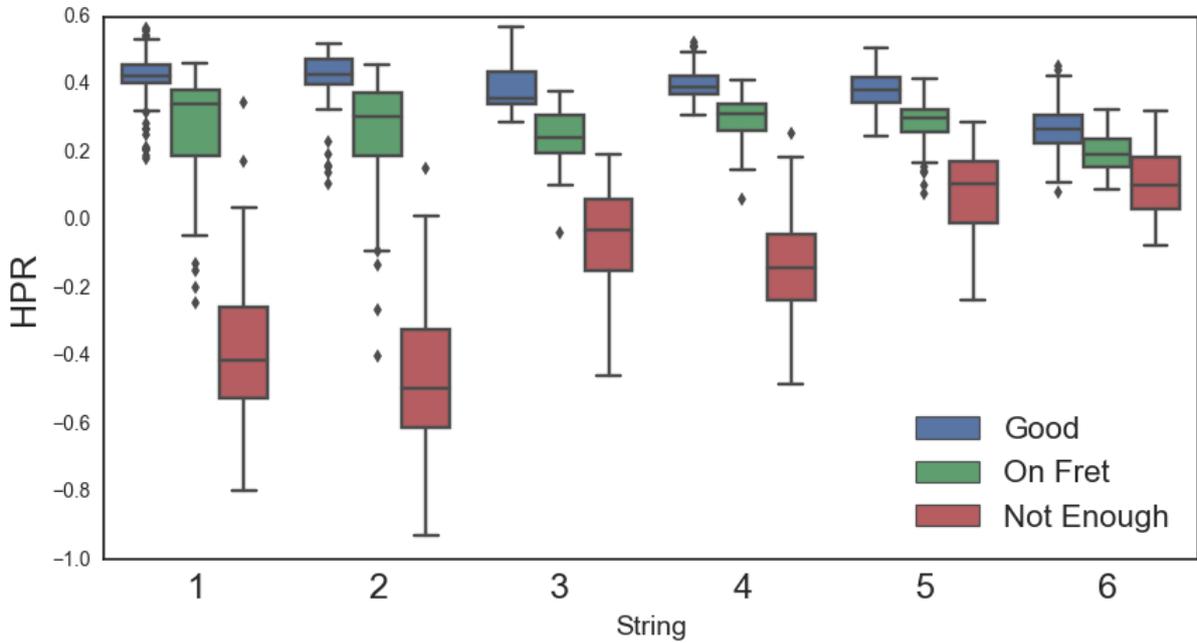


図 5.4: 弦別の HPR の分布の差. 低音弦ほど音の種類による HPR の差が小さい傾向があり, これが 6 弦の音の分類精度の低さの一因であると考えられる.

精度は他の弦に比べて非常に低くなっている. この 6 弦の音の分類結果を混同行列で表した表 5.4 を確認すると, 「正しく押弦できている音」と「十分に弦を押さえられていない音」の多くが「フレット上を押さええている音」に誤分類されていることがわかる. これは, 丸みを帯びたような 6 弦の音色が「フレット上を押さええている音」と似ていることが原因の 1 つであると考えられる. また, HPR も弦の違いの影響を受ける可能性がある. 弦ごとの HPR の分布を音の種類別に表示した図 5.4 を確認したところ, 6 弦においては音の種類による HPR の差が小さくなっていることがわかる. このことも, 6 弦の音の分類精度が低い原因の 1 つとなっていると考えられる.

表 5.4: 3 種類の特徴量 (MFCC, Δ MFCC, HPR) を説明変数として LOGO を行ったときの 6 弦の音の分類結果.

		predicted					
		○	△	×			
actual	○	58	38	12	○	.829	.537
	△	9	88	11	△	.486	.815
	×	3	55	50	×	.685	.463

Chapter 6

考察

第3章で定義した音響的特徴量を用いて、第4章ではエレキギターの演奏評価予測モデルの構築、第5章では押弦が適切に行われているかの分類モデルの構築を行った。本章ではこれらの章で得られた知見に関して横断的な考察を行った後、本研究で残された課題について述べる。

6.1 特徴量に関する考察

6.1.1 TS_{np}

第3章で述べた通り、Toneの評価予測性能には TS_{np} が大きな影響を与えている。図4.8からは、 TS_{np} が高いほどToneの評価が高いことがわかる。しかしながら、弦高などのエレキギターのコンディションによって、特に低音側の弦では TS_{np} が低くなってしまふことが第4章の分析で明らかになった(図5.3)。第4章の演奏音源収集で用いたエレキギターは第5章のGuitar 1と同一であり、図5.3からはGuitar 1の4~6弦の TS_{np} が低くなっていることがわかる。演奏音源収集で弾いてもらったフレーズは1~3弦のみを使用するものであったため、第3章の分析においては4~6弦の TS_{np} の低さの影響を受けなかったと考えられる。

エレキギター演奏評価に TS_{np} を利用するためには、弦の「ビビリ」などが生じにくいエレキギターを使用しなければならない。 TS_{np} は音の長さに依存しない特徴量であり、他の特徴量に比べて汎用性が高い指標であると考えられるため、 TS_{np} を適切に扱うための条件などについてより深く考えていく必要がある。

6.1.2 HPR

音のHarmonic成分とPercussive成分の最大振幅比の対数である HPR は、演奏評価予測と押弦判定の両方で有効な特徴量であることがわかった。しかし、高音弦においては音による HPR の差が大きい、低音弦においては音による HPR の差が小さくなるのが第5章の分

析で明らかになった (図5.4). そのため, 例えば5弦や6弦 (低音弦) を弾いたときに適切に弦を押さえられずに不適切な音色が発せられた場合, *HPR* による演奏評価が上手く機能しないと考えられる. この問題を解決するために, 弦ごとの *HPR* の分布差をなくすための正規化の方法を見つけるなどする必要がある.

6.2 本研究で得られた知見の汎用性向上に向けての課題

本研究では単音のエレキギター演奏評価に向けた分析を行ったが, 分析に用いるフレーズの種類を少数に絞るなどの様々な制約を課した. そのため, 本研究で得られた知見をあらゆる単音のエレキギター演奏に適用可能であるかは未確認である. 本節では本研究における制約を踏まえ, 本研究の知見が適用できる範囲を広げるための今後の課題について述べる.

6.2.1 より多様なフレーズの演奏音源データを用いた分析

第4章で行った分析に用いた5つのフレーズは, BPM100の8分音符間隔の4つの単音から成るフレーズであった. 最終的に演奏評価予測に有効であるとわかった4つの音響的特徴量 (*Lag*, TS_{np} , *TIS*, *HPR*) のうち, *Lag* はフレーズの音の長さに大きく依存する特徴量であるため, 異なるリズムやテンポのフレーズの演奏評価予測を行う際にそのまま用いることは難しい. *TIS* も音が鳴り始めてから鳴り止むまでの全時間の MFCC を用いる特徴量であるため, 音の長さの影響を受ける可能性がある. そのため, 異なるテンポやリズムのフレーズにおいても本研究で定義した音響的特徴量が有効であることを検証する必要がある.

6.2.2 より多くの演奏者の演奏音源データを用いた分析

第4章では5人のエレキギター初心者から演奏音源を収集した. 今回は具体的な分析は行っていないものの, ピッキングが他の演奏者よりも強いなどの特徴が見られる者もいた. より様々な演奏者に対応するための一般的な評価予測モデルを構築するためには, より多くの演奏者から演奏音源を収集する必要がある.

また, 第5章の押弦判定に関する分析で用いた演奏音源は筆者1人によるものであった. そのため, その演奏音源には筆者の個性が表れている可能性があり, そこから構築したモデルで他の演奏者による演奏音の押弦判定が精度よく行えるかは今の段階ではわからない. 本研究で構築したモデルを用いて他の演奏者の演奏に対して押弦判定を行う実験などを行い, モデルの汎用性を検証しなければならない.

6.2.3 コード演奏評価の分析

本研究では単音の演奏にのみ焦点を当てた分析を行った. しかし実際のエレキギター演奏では, 単音ではなく和音 (コード) による演奏も行われる. 特にエレキギターにおいては, 2

本あるいは3本の弦を用いたパワーコードと呼ばれるコードの演奏を行うことが多い。パワーコード演奏はエレキギターの基本の1つであるため、第4章の演奏音源収集の際にパワーコードを用いたフレーズの音源収集も同時に行ったが、その分析までは至らなかった。演奏音源収集後のインタビューにおいてパワーコード演奏の難しさについて述べる声も多数得られたため、今後はこの収集したパワーコードの演奏音源データを活かして、パワーコード演奏評価に関する分析を行っていく必要がある。

Chapter 7

アプリケーション

第4章と第5章にて、音色と演奏のなめらかさに着目してエレキギター演奏を評価する手法と押弦が適切に行われているかを判定する手法を考案した。これらは既存のギター練習支援システムでは考慮されていない部分であったため、本研究で得られた知見を活かして更に効率的なエレキギター練習支援システムを作成することができると考えられる。そこで本研究では最後に、本研究の応用例としてエレキギター練習を支援するシステムを構築した。その画面を図7.1に示す。このシステムは、エレキギターをPCに接続して演奏することで点数と各音の押弦に関するフィードバックが得られるシステムである。これは、第4章で述べたインタビュー中の発言にある「評価が数字として表されるとモチベーションにつながる」「自分の演奏のどの部分が悪かったのかを教えてほしい」などの要望に適ったものとなっている。

このシステムは Web ブラウザ上で動作するシステムとして作られており、バックエンドとフロントエンドの作成にはそれぞれ Python (Flask) と JavaScript (D3.js) を、演奏音の読み取りには Python のライブラリである PyAudio を使用した。Python の実行環境と Web ブラウザおよびエレキギターを接続するための機材があれば、他のコンピュータでも実行可能である。本章では以下、このシステムの説明および問題点と他に考えられるアプリケーションについて述べる。

7.1 システム詳細

演奏するフレーズと繰り返す回数を選択し、そのフレーズをエレキギターで演奏することで点数と押弦の判定が表示される。以下にその詳細を述べる。



図 7.1: 作成したシステムの画面。黒背景の部分には演奏するフレーズが表示されている。そのフレーズを何回か弾いた後、それぞれの演奏に対する点数と押弦の判定が表示される。この図は、フレーズ (B2) を 5 回繰り返して弾いたときの結果を表示している場面である。

演奏するフレーズの選択

第 3 章にて用いた 5 つのフレーズ (A), (B1), (B2), (C1), (C2) から 1 つ選んで演奏する。選んだフレーズは Yousician などのようなギターの指板を模した形で画面上部に表示される。また、今回はこの 5 つのフレーズに制限したが、他にも BPM100 で 4 つの 8 分音符から成る単音のフレーズであれば対応可能である。

演奏

実際に PC に接続したエレキギターを用いて演奏する。指定した回数分だけ選んだフレーズを繰り返し弾く。演奏の際にはメトロノームが常に鳴っており、タイミングを合わせるためのマーカーがフレーズを表示している部分を動く。

点数の表示

選んだフレーズを演奏したときの点数を 0~100 点で表示する。繰り返し演奏したときは、その各々の演奏に対して 1 つずつ点数が与えられる。図7.1は、フレーズ (B2) を 5 回繰り返し演奏したときの結果を表している。

押弦の判定

選んだフレーズを演奏したとき、各音の押弦が適切に行われていたかを判定する。具体的には第 4 章で述べた通り、(○) 正しく押弦できている音、(△) フレット上を押さえている音、(×) 十分に弦を押さえられていない音」の 3 つに各音を分類する。

7.2 本システムの問題点

このシステムを使用することで、第 3 章と第 4 章で示した精度で点数と押弦に関するフィードバックが得られる。しかしながら、このシステムには以下のような問題点がある。

演奏するフレーズのリズムとテンポが一定である。

第 4 章では BPM100 の 4 つの 8 分音符からなるフレーズを用いて分析を行った。そのときに用いた特徴量の中でも *Lag* と *TIS* は音の長さに影響を受ける可能性がある。そのため、第 4 章で構築した点数予測モデルを異なるリズムとテンポのフレーズに適用できるかは現段階では不明である。しかし、押弦の判定に用いる特徴量は時間に依存しないものであるため、押弦の判定は異なるリズムとテンポのフレーズであっても行うことができる。

正しいリズムと音程で弾けなかった場合は点数が付与されない。

第 4 章で行った分析では、正しいリズムと音程で弾けている部分（つまり既存システムでは正しく弾けていると見なされる部分）を抽出し、それをもとにして点数予測モデルを構築した。この点数予測モデルを使用している以上は、フレーズの演奏中にリズムか音程が間違っている部分があった場合は、その演奏に対する点数を付与することができない。しかし、BPM100 の 4 つの 8 分音符からなるフレーズは、実際にそれらの演奏データを初心者から収集できたことを考えると、リズムと音程を正しく弾くだけであれば初心者にとってもそれほど難しいことではない。そのため、簡単なフレーズをより綺麗な音色で滑らかに弾くことを支援するシステムとしては十分に機能すると考えられる。

上記の問題点以外にも、実際にこのシステムを使用してもらうことで新たな問題点が発見される可能性がある。そのため、特にエレキギター初心者を中心にユーザスタディを行っていくことが今後の課題である。

7.3 その他の想定されるアプリケーション

本研究で作成したシステムの他にもいくつかのアプリケーションが想定される。例えば、既存のエレキギター練習システムの裏で実行できる押弦判定システムが考えられる。Yousician や Rocksmith などエレキギターを練習しているときに同時に押弦判定システムを起動しておけば、既存システムの優れたインタフェースを本研究の押弦判定手法が補完する形で上手く機能することが期待される。

Chapter 8

おわりに

既存のエレキギター練習システムは「正しいタイミングで正しい音程で弾けたか」のみしか演奏評価基準がない。そこで本研究では、リズムと音程の正確さ以外の観点からもエレキギター演奏を評価するための音響的特徴量 (*Tone Similarity*, *Tone Instability*, *Harmonic-Percussive Ratio*, *Lag*) を提案した。そしてそれらの音響的特徴量が、エレキギター演奏の自動評価および押弦が適切に行われているかの判定に利用できることを示した。さらにそれらの知見の応用例として、音色と演奏のなめらかさを考慮したエレキギター練習システムを作成した。

本研究の今後の方向性としては、より多くの演奏者あるいはエレキギターによる演奏音源データセットを収集し、今回の分析で得られた結果の汎用性を確かめることがあげられる。それを通して、演奏者の個性やエレキギターの個体差に影響を受けにくい特徴量、あるいは特徴量の補正方法などについて考えなければならない。また最終的には、ユーザが効率良くエレキギターを練習できるシステムについて考えていく必要がある。本研究で得られた知見をもとにして、エレキギター練習システムがさらに発展していくことを期待している。

Publications

本研究に関する発表

- 下尾波輝, 矢谷浩司, 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, 第 117 回情報処理学会音楽情報科学研究会 (SIGMUS), Vol.2017-MUS-117, No.3, 2017, 学生奨励賞受賞.
- 下尾波輝, 矢谷浩司, エレキギター演奏におけるミスの自動検出, 情報処理学会第 80 回全国大会 (発表予定).

その他の発表

- 下尾波輝, 矢谷浩司, KnowledgeDeck: ビジネス資料作成向け情報収集・整理支援システム, 情報処理学会第 80 回全国大会 (発表予定).

References

- [1] F. H. Allen and O. Johnson. 1991. Automated conformational analysis from crystallographic data. IV. Statistical descriptors for a distribution of torsion angles. *Acta Crystallographica Section B* 47 (1991), 62–67. <https://doi.org/10.1107/S0108768190010382>
- [2] S. Ariga, M. Goto, and K. Yatani. 2017. Strummer: An interactive guitar chord practice system. In *2017 IEEE International Conference on Multimedia and Expo (ICME)*. 1057–1062. <https://doi.org/10.1109/ICME.2017.8019338>
- [3] Mathieu Barthet, A Anglade, György Fazekas, Şefki Kolozali, and R Macrae. 2011. Music recommendation for music learning: Hotttabs, a multimedia guitar tutor. In *2nd Workshop on Music Recommendation and Discovery*, Vol. 793. 7–13.
- [4] John M. Cooksey. 1977. A Facet-Factorial Approach to Rating High School Choral Music Performance. *Journal of Research in Music Education* 25, 2 (1977), 100–114. <https://doi.org/10.2307/3345190>
- [5] Christian Dittmar, Estefanía Cano, Jakob Abeßer, and Sascha Grollmisch. 2012. Music Information Retrieval Meets Music Education. In *Multimodal Music Processing*, Meinard Müller, Masataka Goto, and Markus Schedl (Eds.). Dagstuhl Follow-Ups, Vol. 3. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Dagstuhl, Germany, 95–120. <https://doi.org/10.4230/DFU.Vol3.11041.95>
- [6] A. Eronen. 2001. Comparison of features for musical instrument recognition. In *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics*. 19–22. <https://doi.org/10.1109/ASPAA.2001.969532>
- [7] Derry Fitzgerald. 2010. Harmonic/percussive separation using median filtering. In *13th International Conference on Digital Audio Effects (DAFX10)*. Dublin Institute of Technology.
- [8] G. Grindlay. 2008. Haptic Guidance Benefits Musical Motor Learning. In *2008 Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*. 397–404. <https://doi.org/10.1109/HAPTICS.2008.4479984>
- [9] A. F. Hayes and K. Krippendorff. 2007. Answering the Call for a Standard Reliability Measure for Coding Data. *Communication Methods and Measures* 1, 1 (2007), 77–89. <https://doi.org/10.1080/19312450709336664>
- [10] Hank Heijink and Ruud GJ Meulenbroek. 2002. On the complexity of classical guitar playing: functional adaptations to task constraints. *Journal of motor behavior* 34, 4 (2002), 339–351.
- [11] Alex Hrybyk and Youngmoo Kim. 2010. Combined audio and video analysis for guitar chord identification. In *In Proceedings of the 11th International Society on Music Information Retrieval (ISMIR)*. 159–164.

- [12] K. Huang, E. Y. L. Do, and T. Starner. 2008. PianoTouch: A wearable haptic piano instruction system for passive learning of piano skills. In *2008 12th IEEE International Symposium on Wearable Computers*. 41–44. <https://doi.org/10.1109/ISWC.2008.4911582>
- [13] S. Abdallah C. Duxbury M. Davies J. P. Bello, L. Daudet and M. B. Sandler. 2005. A Tutorial on Onset Detection in Music Signals. *IEEE Transactions on Speech and Audio Processing* 13, 5 (Sept 2005), 1035–1047. <https://doi.org/10.1109/TSA.2005.851998>
- [14] Rose Johnson, Janet van der Linden, and Yvonne Rogers. 2011. Real-time Feedback for Learning the Violin. In *Proceedings of the 25th BCS Conference on Human-Computer Interaction (BCS-HCI '11)*. British Computer Society, Swinton, UK, 535–538. <http://dl.acm.org/citation.cfm?id=2305316.2305414>
- [15] Markus Löchtefeld, Sven Gehring, Ralf Jung, and Antonio Krüger. 2011. guitarAR: Supporting Guitar Learning Through Mobile Projection. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems (CHI EA '11)*. ACM, New York, USA, 1447–1452. <https://doi.org/10.1145/1979742.1979789>
- [16] Beth Logan. 2000. Mel Frequency Cepstral Coefficients for Music Modeling. In *In International Society on Music Information Retrieval*.
- [17] Soichiro Matsushita and Daisuke Iwase. 2013. Detecting Strumming Action While Playing Guitar. In *Proceedings of the 2013 International Symposium on Wearable Computers (ISWC '13)*. ACM, New York, NY, USA, 145–146. <https://doi.org/10.1145/2493988.2494345>
- [18] Gary E. McPherson and William F. Thompson. 1998. Assessing Music Performance: Issues and Influences. *Research Studies in Music Education* 10, 1 (1998), 12–24. <https://doi.org/10.1177/1321103X9801000102>
- [19] Y. Motokawa and H. Saito. 2006. Support system for guitar playing using augmented reality display. In *2006 IEEE/ACM International Symposium on Mixed and Augmented Reality*. 243–244. <https://doi.org/10.1109/ISMAR.2006.297825>
- [20] Tomoyasu Nakano, Masataka Goto, and Yuzuru Hiraga. 2006. An automatic singing skill evaluation method for unknown melodies using pitch interval accuracy and vibrato features. In *Ninth International Conference on Spoken Language Processing*.
- [21] Asami Nonogaki, Shohei Shimazu, Norio Emura, Masanobu Miura, Seiko Akinaga, and Masuzo Yanagida. 2011. Use of spline curve to evaluate performance proficiency of a Czerny piano piece. In *Proceedings of International Symposium on Performance Science*. 69–74.
- [22] Douglas O' Shaughnessy. 2008. Invited paper: Automatic speech recognition: History, methods and challenges. *Pattern Recognition* 41, 10 (2008), 2965 – 2979. <https://doi.org/10.1016/j.patcog.2008.05.008>
- [23] Caroline Palmer. 1997. Music performance. *Annual review of psychology* 48, 1 (1997), 115–138.
- [24] Daniele Radicioni and Vincenzo Lombardo. 2005. Guitar fingering for music performance. In *Proceedings of the International Computer Music Conference*, Vol. 40. 527–530.
- [25] Aleksander Radisavljevic and Peter F. Driessen. 2004. Path Difference Learning for Guitar Fingering Problem. In *Proceedings of the International Computer Music Conference*.
- [26] Loïc Reboursière, Otso Lähdeoja, Thomas Drugman, Stéphane Dupont, Cécile Picard-Limpens, and Nicolas Riche. 2012. Left and right-hand guitar playing techniques detection. In *Proceedings of the International Conference on New Interfaces for Musical Expression*.

-
- [27] Brian E. Russell. 2010. The Development of a Guitar Performance Rating Scale using a Facet-Factorial Approach. *Bulletin of the Council for Research in Music Education* 184 (2010), 21–34.
- [28] Daniel R. Tuohy and Walter D. Potter. 2005. A Genetic Algorithm for the Automatic Generation of Playable Guitar Tablature. In *Proceedings of the International Computer Music Conference*.
- [29] G. Tzanetakis and P. Cook. 2002. Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing* 10, 5 (2002), 293–302. <https://doi.org/10.1109/TSA.2002.800560>
- [30] K. Yazawa, K. Itoyama, and H. G. Okuno. 2014. Automatic transcription of guitar tablature from audio signals in accordance with player’s proficiency. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 3122–3126. <https://doi.org/10.1109/ICASSP.2014.6854175>
- [31] Yi-Hsuan Yang Yin-Jyun Luo, Li Su and Tai-Shih Chi. 2015. Detection of Common Mistakes in Novice Violin Playing. In *In International Society on Music Information Retrieval*.
- [32] Stephen F. Zdzinski and Gail V. Barnes. 2002. Development and Validation of a String Performance Rating Scale. *Journal of Research in Music Education* 50, 3 (2002), 245–255. <https://doi.org/10.2307/3345801>