

音楽表現の生成モデリングの検討 ～ 熟練度に依存しない演奏表現の解析技術を目指して～

小泉 悠馬^{1,a)} 伊藤 克巨^{2,b)}

概要: 人間が演奏する音楽は、楽譜通りの“機械的”な演奏ではなく、リズム・音高・音量・音程に、奏者の意図した演奏表現による変動と、低熟練度起因する演奏誤差による奏者の意図していない変動が含まれる。前者の演奏表現による変動や逸脱は、計算機上で感性を扱う情報処理の実例として、近年の音楽情報科学の中心的話題の一つである。しかし後者の、演奏誤差による逸脱は深く議論されていない。本稿では、演奏誤差を含んだ演奏からの奏者の演奏表現情報の解析について議論を行うために、従来の演奏表現解析技術を整理し、著者らが今までに提案した音量・リズムに関する変動のモデル化と解析手法を紹介する。

キーワード: 演奏表現, 演奏誤差, 演奏表現の生成モデル, 楽音合成・修正

Discussion for establishing an analysis method of musical expression from low-proficiency performance

YUMA KOIZUMI^{1,a)} KATUNOBU ITOU^{2,b)}

Abstract: Musicians do not play exactly what is written in the score. Deviations and/or fluctuations are included in the tempo, the timbre, the pitch and the amplitude. These fluctuations consist of two types; intended fluctuations and not intended fluctuations due to low-proficiency. Analysis and modeling of intended fluctuations from music recording is a central topic in the field of music information science. However, not intended fluctuations in the performance are not discussed well. In this paper, in order to discuss modeling of the two types of fluctuations, we introduce to our previous studies about modeling of not intended fluctuation.

Keywords: Musical expression, mis-play component, generative model of musical expression, sound synthesis/adjustment.

1. まえがき

音楽の本質は、演奏に込められた演奏者それぞれの「芸術性」や「音楽表現」である。音楽は音声と比べ、制約が強いメディアであるため、それらは、音楽の要素であるリズムやテンポ、音量、音高、音色に、繊細な揺らぎや逸脱

を付与することで実現する。

しかし、残念ながら、その繊細さゆえに、多くのアマチュア楽器奏者は、思い通りに演奏できない。そのような人々が CGM (Consumer Generated Media) を制作しようと思ったら、楽器の練習をするか、多大な労力で修正するしかなく、潜在的に多くの制作者が挫折してしまっているだろう。

他方で、「芸術性」や「音楽表現」は楽器を演奏できない人にも伝わる。実際、熟達した奏者は、熟練度の低い奏者の演奏から彼らの所望の演奏を推定し、的確な指導を行うことで教育している。

¹ 法政大学 大学院 情報科学研究科
Graduate School of Computer and Information Sciences, Hosei University

² 法政大学 情報科学部
Faculty of Computer and Information Sciences, Hosei University

^{a)} 12t0005@cis.k.hosei.ac.jp

^{b)} it@fw.ipsj.or.jp

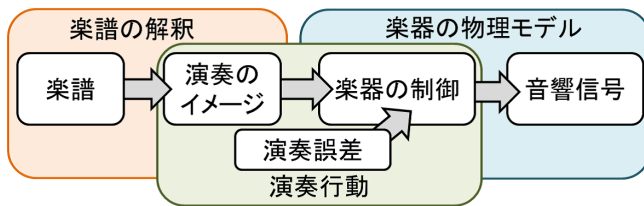


図 1 音楽演奏の生成過程

Fig. 1 The generating process of a musical sound.

本稿では、低熟練度奏者の演奏からの音楽表現の推定について取り上げる。その実現のためには、先行研究で得られている音楽表現解析の知見に加えて、奏者の楽器の制御ミスを検討しなければならない。

これらの問題について議論を行うために、先行研究の流れを整理し、楽器の制御ミスを検討した音楽表現生成全体のモデリングの検討を行う。

まず、従来研究で行われている演奏表現の解析技術を整理する。次に、演奏誤差成分を含んだ演奏の、音楽表現の生成過程をモデリングする検討を行う。5章では、演奏誤差成分を含んだ演奏からの演奏表現解析で問題となる、評価手法やデータセットの形式、また効果的な統計モデルについての、議論を要する今後の課題を述べる。

2. 音楽表現の解析およびモデリングの先行研究

図 1 に、音楽音響信号の生成過程を示す。まず奏者が楽譜を知覚することにより、演奏表現、すなわち音楽的な揺らぎや逸脱をイメージする。次に奏者はそのイメージに基づき、楽器の励振機構を制御する。楽器の制御の際に、熟練度の低い奏者は意図通りに楽器をコントロールすることが出来ず、励起振動に演奏誤差成分が加算される。最後に、励振機構の振動が楽器の構造により変換され、空气中に放射される。

本稿では、奏者の演奏イメージと観測された音楽表現を区別するために、奏者の意図した音楽的な揺らぎや逸脱を“意図特徴量”，観測された音響信号に含まれる音楽的な揺らぎや逸脱を“演奏特徴量”と呼ぶことにする。

音楽表現の解析やモデル化は、意図特徴量と、楽譜や楽器の制御パラメータを対応付けることを目的とする。しかし、奏者の意図は非常に抽象的な情報であり、直接観測を行うのは困難である。そこで多くの研究では、意図特徴量と物理的に観測が可能な演奏特徴量が等しいと仮定を置き、演奏特徴量と、楽譜や楽器の物理的な制御パラメータを対応付けている。

2.1 楽譜と演奏特徴量を対応付ける手法

楽譜情報*1と演奏特徴量を対応付ける手法は、合成音へ

*1 MIDI や MusicXML など

の演奏情報の転写や奏者検索、自動演奏への表情付けなどへの応用を目的として行われている。これは、楽譜が入力された際、音響信号にどのような演奏特徴量が付与されるかの解析およびモデリングである。また、演奏特徴量と意図特徴量が等しい場合、スコアから生成される演奏意図の解析および生成モデリングとなる。

2.1.1 信号処理による演奏特徴量の解析と応用

音響信号から直接演奏特徴量を解析する手法は、発音時刻検出 [1] やビートトラッキング [2]、また基本周波数推定 [3] や多重音解析 [4] などの信号処理技術で解析された演奏情報と楽譜情報を対応付けることで実現される。ゆえにその精度は信号処理技術の精度に依存するが、ユーザーが自由に音響信号を入力できるため、演奏の特徴の自動解析や合成音への演奏特徴量の転写、奏者認識などに用いられる。

リズムやテンポの変動は音楽表現の知覚の中でも特に重要な要素であり [5]、MIDI などの合成音でも容易に制御できるパラメータであるため、多くの先行研究で音響信号からの解析が行われおり、ポピュラー音楽からのビートトラッキングでは、8割以上の精度でビートが認識されている [6, 7]。またテンポはゆるやかに時間変化することが知られており、その時間変化はテンポ曲線 (tempo curve) と呼ばれる滑らかな曲線として解析されている [8, 9]。

一方、解析結果の楽音合成への応用では、演奏特徴量の転写やモーフィングなどがある。VocaListener では、入力歌唱データから音量、音高の変化を解析し、それを合成歌唱へ転写する [10]。v.morish では、声質 (音色) と歌いまわし (リズム、音量、音高) を独立にモーフィングすることにより二歌唱者の中間的な歌唱の合成を実現している [11]。これらの技術による合成音は動画共有サイトなどで多くの反響を生んでおり、VocaListener は Vocaloid のプラグインとして製品化され、一般ユーザーが使用可能な技術となっている。

また奏者認識では、演奏特徴量を解析し、ケルトバイオリンやジャズサキソフォンの奏者認識を行っている [12, 13]。ジャズサキソフォンの奏者認識では、音量変化などの演奏特徴量を用いることにより、Support Vector Machines と Neural Networks で 7割強の識別率を得ている。

2.1.2 演奏特徴量のモデリングと自動演奏への表情付け

楽譜情報からの演奏特徴量の生成モデリングは、自動演奏への表情付けなどに応用される。自動演奏への表情付けでは、入力楽譜を自動ないし手動で音楽構造解析*2し、それに対応する演奏特徴量を合成音に自動で付与する。処理手法はルール型、事例参照型、学習型に 3つに大別される。

ルール型では、スタッカートでは音符の継続長を短くするなどの、演奏記号やフレーズに対する経験則的な逸脱のルールを設定する [14]。そのルールに基づき、楽譜の構造

*2 フレーズ、和声、拍節など

解析結果を用いて合成音に演奏特徴量を付与する。この手法は音楽信号からの演奏特徴量解析を必要とせず、処理の中身が分かりやすいのが特徴である。

事例参照型では、入力楽譜と音楽構造が類似する演奏事例をデータベースから検索し、演奏特徴量を転写する [15]。よって、生成される演奏特徴量はデータベースに依存するため、ジャズの曲にウィンナワルツ風の演奏特徴量を付与することが可能である。ピアノなどの一人の奏者が同時に複数の音符を演奏したのでは、前述の信号処理ベースの演奏特徴量解析では高精度な演奏特徴抽出が難しいため、ハンドラベリングによる演奏特徴量解析結果のデータベースを用いることが多い [16]。

学習型は、事前に用意された実演奏の演奏特徴量データベースから演奏ルールを統計的手法を用いて学習し、合成音に演奏特徴量を付与する。学習にはベイジアンネットワーク [17] やガウシアンプロセス [18]、Linear-Chain CRF [19] を用いる手法などが存在する。生成される演奏特徴量は学習データに依存するため、ある一人の名音楽家の演奏のみを用いて学習することで、その音楽家の演奏特徴量を反映した楽音が合成できることが示されている。

2.2 楽器の物理的な制御パラメータと演奏特徴量を対応付ける手法

近年、モーションキャプチャや力学的センサ [20] を用いて人間の奏法行動や楽器制御の物理パラメータを取得し、それを演奏特徴量と対応付ける研究がおこなわれている。楽器制御の物理パラメータとは、擦弦楽器ならば擦弦位置や擦弦速度、吹奏楽器であれば息の吹き込む量などである。これらのパラメータは、人間の奏法行動に直結するため、楽器や奏者ごとの表現の特徴および楽器の制御方法の解析が可能である。制御パラメータの変化が音響信号にどのような影響を与えるかを解析し、楽音合成に応用したり、奏者や演奏形態ごとの演奏の特徴を解析したりする。また、演奏特徴量と意図特徴量が等しい場合、演奏意図と楽器制御の対応付けとなる。

2.2.1 奏法パラメータと音響信号の統計的な対応付けによる楽音合成

楽音合成の代表的な手法として、事前に用意した楽音のデータベースを参照し、素片を接続しながら合成する方式 [21] と力学的センサから楽器の制御パラメータを取得し、楽器の物理モデルを用いて合成する方式 [22] が挙げられる。しかし、特に擦弦楽器や吹奏楽器などの、演奏中に励起機構に常にロードがかかる楽器 (excitation-continuous musical instruments) では、前者は高品質な音色で合成可能だが音楽表現 (演奏特徴量) の作りこみが難しく、反対に後者はユーザーの意図を容易に合成音に反映できるが合成音の音質が悪いという問題があった。

そこで近年、両モデルの長所を生かしたハイブリッド

合成方式が検討されている [23]。この手法は、素片接続合成方式の素片選択を楽器の制御パラメータを用いて行う。ユーザーが力学的センサ付き楽器を用いて楽器を演奏すると、その制御パラメータに合った素片がデータベースから選択され、楽音が合成される。制御パラメータと素片の対応付けは、マルチモーダル楽器により奏法パラメータと素片を同時に記録したデータを学習データに用いた、機械学習により行われる [24, 25]。これらの手法は、物理モデル合成方式の柔軟性を保ったまま、高品質な合成音を得られることが示されており、その合成音は著者らの web サイトで公開されている [25]。

2.2.2 奏者や演奏形態による奏法パラメータと演奏特徴量の変化の解析

合奏では、奏者はテンポや音量などを他の奏者と合わせる必要があるため、演奏表現の自由度は独奏と比べて低い。また、メロディや伴奏といったパートごとの演奏上の役割によっても、演奏表現の自由度は変化する。これら演奏形態の違いによる演奏行動の変化の理解は、計算機による演奏意図の理解の上でも重要な課題である。Maestre らのグループでは、マルチモーダル楽器を用いて、演奏表現の自由度の変化や奏者同士の楽器制御の同期を解析している [26, 27]。彼らの研究では、バイオリンの演奏は、弦楽四重奏はソロ演奏に比べてテンポの揺らぎが制限されることが確認されており、さらに合奏練習を重ねることにより、奏者間の演奏の同期や相関が強くなることが確認されている。

3. 演奏誤差を含んだ演奏からの意図特徴量の推定

演奏特徴量と楽譜や楽器の物理的な制御パラメータの対応付けは、音響信号および楽器の制御パラメータに演奏誤差が含まれていない場合、意図特徴量と、楽譜や楽器の制御パラメータの対応付けとなる。よって、対象が熟練した奏者の演奏の場合は、意図特徴量の解析・モデル化が可能である。

しかし、多くのアマチュア奏者は楽器を熟練しておらず、演奏には演奏誤差が含まれる。よって、前述の手法では意図特徴量の解析やモデル化ではなく、音響信号や演奏特徴量の解析やモデル化となる。

そのため、楽器を熟練していない奏者からの意図特徴量解析では、演奏誤差を考慮した解析技術が必要となる。演奏誤差成分は楽器の制御ミスにより生成されるため、力学的センサから得られた奏法パラメータから、楽器の制御ミスらしき成分を除去することで実現可能である。しかし、一般ユーザの使用を考えた際、高価な力学的センサ付の楽器での演奏は非現実的であるため、音響信号から直接、意図特徴量を推定する手法が必要である。

その実現のためには、音響信号からの正確な演奏特徴量の解析技術と、音響信号の生成過程全体のモデリングが必

要となる。これはつまり、演奏特微量を意図特微量と演奏誤差成分へ分解することを意味し、観測演奏特徴からの演奏イメージの逆推定問題としてとらえることが出来る。

しかしこの逆問題の推定手法は研究として、正解データが未知、という重大な問題をはらんでいる。推定された演奏イメージが、奏者の意図を精度よく推定できているかの評価が非常に難しい。そこで我々は、アマチュア奏者が、熟練した奏者（プロ奏者）の演奏を模倣しようとして演奏したデータを用いて実験を行っている。つまり、アマチュア奏者の意図特微量はプロ奏者の演奏特微量と等しくなり、推定結果がプロ奏者の演奏特微量と一致すれば、推定法が精度よく演奏イメージを推定できているとみなすことが出来る。

次節では、我々が提案した、演奏誤差成分を含んだ音響信号からのテンポ変動と音量変化に関する意図特微量の解析技術を紹介する。

4. 演奏誤差を含んだ演奏からの演奏イメージの推定法

我々の研究では、演奏特微量には、意図特微量と演奏誤差成分が加法的に含まれていると仮定する。つまり、意図特微量の推定問題は、演奏特微量の信号分解問題となる。信号分解問題では、所望の信号に対して何らかの統計的な制約条件が必要である。我々のアプローチでは、この制約条件に対して、2章で紹介した先行研究で得られている知見や、音楽理論や演奏理論的制約を用いている。

4.1 観測テンポ変動からの奏者の意図したテンポ変動の推定 [28, 29]

テンポ変動解析の先行研究では、テンポはフェルマータや速度記号の変化などの楽譜上で指示されたテンポ変動を除けば、滑らかかつゆったりと変化する (tempo curve) という知見が得られている [8, 9]。さらに三浦らの、奏者の意図したテンポ変動をスプライン曲線で表現し、曲線の逸脱量を用いて習熟度を評価する手法は、専門家の評価とほぼ同等の評価性能を示している [30]。これらのことから、意図特微量を観測テンポ変動から曲線フィッティングを用いて推定し、その曲線からの逸脱を演奏誤差とすることで、演奏誤差を含んだテンポ変動から奏者の意図したテンポ変動を推定できそうである。

我々の手法では、発音時刻と音価を用いてテンポ変動を多項式回帰分析することにより、奏者の意図したテンポ変動である“真のテンポ曲線”を推定している [28, 29]。図2にプロ奏者の演奏のテンポ変動および真のテンポ曲線（左）と、アマチュア奏者がプロ奏者の演奏を模倣して演奏した録音のテンポ変動および真のテンポ曲線（右）を示す。プロ奏者の真のテンポ曲線は観測テンポ変動と非常に似通った値を示している。これはプロ奏者が滑らかなテンポ変動

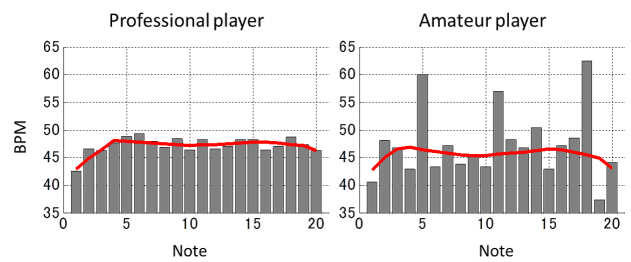


図2 観測テンポ変動（バープロット）と推定された真のテンポ曲線（赤線）の例。左図はプロ奏者が演奏した者であり、右図はアマチュア奏者がプロ奏者のテンポ変動を模倣して演奏したもの。

Fig. 2 Observed tempos (bar plot) and estimated true tempo curves (red line). The tempo of the left figure was estimated from a professional recording and that of the right figure was estimated from an amateur recording that mimicked the professional one.

を意図し、非常に少ない演奏誤差成分のみで楽器を制御できることを示している。一方、アマチュア奏者の観測テンポ変動は真のテンポ曲線からばらついた値を示している。しかし、真のテンポ曲線の形状は、模倣したプロ奏者のテンポ曲線の形状と非常に似通っており、アマチュア奏者がプロ奏者のテンポ変動を意図して演奏したことを読み取ることが出来る。

さらに我々は、推定した意図特微量を用いて音響信号を自動伸縮修正することにより、演奏特微量から奏法誤差成分を除去する手法を提案している。アマチュア奏者がプロ奏者の演奏を模倣した録音を、提案法を用いて修正することにより、修正前の録音に比べ修正後の録音がプロの演奏に有意に近づいたことが、聴取実験により示された。

4.2 観測音量変動からの奏者の意図したアーティキュレーションとダイナミクスの推定 [31]

音楽演奏の音量変化は、対数音量スケールでフォルテやクレシェンドなどの音量のベースラインとなる“ダイナミクス”と、マルカートやレガートなどの個々の音符の音量変化である“アーティキュレーション”の加算により生成されているといわれている。Maestreらは観測音量系列から、音量系列を数本の直線で近似するパラメータ (ADSR) の推定法を提案した [32]。この手法は、演奏のダイナミクスとアーティキュレーションの変化を高々数個のパラメータで表現することができ、奏者推定・同定に用いる特微量として有効だといわれている [12, 13]。このことから、ダイナミクスとアーティキュレーションの変化は、音量変化における奏者の特徴、つまり意図特微量や演奏特微量である可能性がある。

しかし、ADSRは打弦や撥弦などの、振幅が指数減衰する楽器の音量変化の表現には有効な手法であるが、擦弦楽器や吹奏楽器などの振幅変化の自由度が高い楽器の正確な

表現は難しい。よって推定結果を様々な楽器の楽音合成や楽音修正に用いる場合、より柔軟に音量変化を記述する手法が必要である。

我々の手法では、音量変化は、意図したダイナミクス変化、意図したアーティキュレーション変化と、演奏誤差成分の三種類の変動の重畳により生成されていると仮定する[31]。さらにアーティキュレーションの変化は、奏者はフレーズの統一感のために、高々数種類のアーティキュレーション素片を組み合わせでイメージしていると仮定する。よって、音量系列からダイナミクス変動と数種類のアーティキュレーション成分を推定することにより、奏者の意図特徴量を推定する。

まず、観測音量変化をカルマンフィルタのアルゴリズムによりダイナミクスの変化と個々の音符ごとの音量変化に分離する。ここで、意図的なダイナミクス変化が推定される。次に、音符ごとの音量変化を混合ガウス分布(GMM)によりクラスタリングする。混合数はAICにより決定する。各GMMの平均値ベクトルをアーティキュレーションの意図特徴量とする。

図3に、プロ奏者の演奏録音と、アマチュア奏者がプロ奏者の演奏を模倣して演奏した録音から推定した各成分を示す。ダイナミクス変化は似通った変化を推定できているが、アーティキュレーション素片からは、モデルの自由度が高すぎるため、演奏誤差成分の除去できていない。今後、音符内の状態(発音・持続・減衰)と、楽器の制御機構の物理制約を考慮することで、推定精度の向上を目指す。

さらに我々は、演奏録音から推定したダイナミクス変化とアーティキュレーション素片、および真のテンポ曲線を用いて、合成音に意図特徴量を転写する実験を行った。この手法では、まずMIDIを用いて演奏特徴量を制御していない合成音を生成する^{*3}。次に、前述の二つの手法を用いて奏者の意図特徴量を推定し、合成音に転写する^{*4}。聴取実験などの評価実験は行っていないが、合成音に表現力が転写されていることが確認できた。これらの合成音は、著者らのホームページで公開しているので、興味のある読者の方に是非聴いていただきたい。

5. 議論を要する今後の課題

本研究の対象は音楽演奏の演奏表現の解析であるため、「音楽」を前提とした評価を行わなくてはならない。一般に、音楽の評価は聴衆が行うものであるため、演奏特徴量の雑音除去などを行った演奏の聴取実験が効果的であると考えられる。しかし、聴衆が音楽を聴くポイントは人によって様々である。そのため、聴取実験を行うべきか、推定演奏特徴の数値的な誤差の評価を行うべきかが議論の対象となる。

^{*3} <http://slp.cis.k.hosei.ac.jp/student/2013/wav/rawMidi.wav>

^{*4} <http://slp.cis.k.hosei.ac.jp/student/2013/wav/transcription.wav>

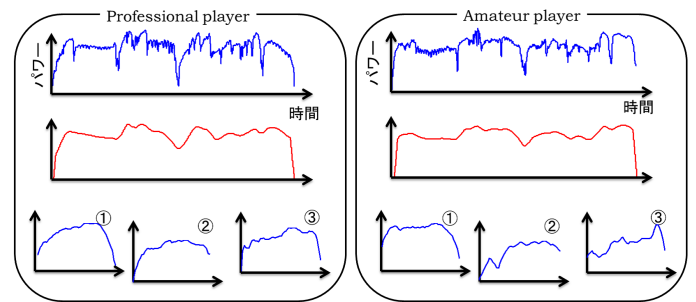


図3 観測音量変動(上), 推定されたダイナミクス変化(中), 推定されたアーティキュレーション素片(下). 左図はプロ奏者が演奏した者であり, 右図はアマチュア奏者がプロ奏者の演奏を模倣したもの。

Fig. 3 The observed amplitude series (top), estimated dynamics component (middle), estimated articulation components (bottom). The amplitude series of the left figure was estimated from a professional recording and that of the right figure was estimated from an amateur recording that mimicked the professional one.

また、評価実験を行う際、奏者の意図した真の意図特徴量が既知のデータセットが必要となる。さらに演奏の生成過程全体を統計モデル化するためには、演奏誤差がどのような分布に従って発生するかを知る必要があるため、楽譜情報、意図特徴量、楽器の制御パラメータ、演奏特徴量、音響信号がすべてそろったデータセットが必要となる。音楽の、強い制約条件の下での音表現は、声優や俳優の感情表現と似た構造があり、演技音声の大規模コーパスには「千の声コーパス[33]」などが存在する。しかし、このようなデータセットを用意するためには高いコストや機材が必要となる。また、どの程度の量のデータを取集める必要があるかも議論の対象である。

現在提案しているモデルは、楽器の制御機構を考慮したものではない。演奏誤差成分は、楽器の制御パラメータにより生成されるものなので、制御パラメータの直接的なモデリングは行わないにしても、それに準ずるモデル化が必要である。潜在変数として制御パラメータを考慮した演奏特徴量の分解に、どのような統計モデルや確率過程を用いるのが妥当かを議論したい。

6. あとがき

本稿では、演奏誤差成分を含んだ演奏からの演奏表現推定のために、音楽表現の生成過程をモデリングする検討を行った。従来の演奏表現推定・モデリングについての整理を行い、演奏誤差を含んだ音響信号からの演奏表現推定法に必要とされる事項について整理した。

今後、情報処理技術の発展に伴い、演奏を熟練していない奏者が自身の演奏を公開する場が増加するであろう。よって、そのような演奏から楽音合成や楽音修正、音楽検索を行う需要が高まっていくと考えられる。そのような技

術には、演奏誤差を含んだ録音からの、奏者の意図した演奏表現推定が必須である。熟練度に依存しない演奏表現の解析技術の確立のために、音楽情報科学に精通した多くの研究者の方々から、本研究に対して助言いただきたい。

謝辞 本研究に対し、有益なご助言を頂いたポンペウ・ファブラ大学の Xavier Serra 准教授, Esteban Maestre 博士, Music Technology Group の皆様に感謝いたします。また本研究の一部は、科研費基盤研究 (C) の支援を受けた。

参考文献

- [1] J. P. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies and M. Sandler: "A tutorial on onset detection in music signals", *IEEE Transactions on Audio, Speech, and Language Processing*, **13**, 5, pp. 1035–1047 (2005).
- [2] D. P. W. Ellis: "Beat tracking by dynamic programming", *J. New Music Res.*, **36**, 1, pp. 51–60 (2007).
- [3] A. de Cheveigne and H. Kawahara: "Yin, a fundamental frequency estimator for speech and music", *J. Acoust. Soc. Am.*, **111**, 4, pp. 1917–1930 (2002).
- [4] H. Kameoka and et.al.: "Complex nmf: A new sparse representation for acoustic signals", *Proc. ICASSP 2009*, pp. 3437–3440 (Apr. 2009).
- [5] R. Parncutt: "A perceptual model of pulse salience and metrical accent in musical rhythms", *Music Perception*, pp. 409–464 (1994).
- [6] M. Goto and Y. Muraoka: "A beat tracking system for acoustic signals of music", the Second ACM Intl. Conf. on Multimedia, pp. 365–372 (1994).
- [7] P. Grosche and M. Muller: "Extracting predominant local pulse information from music recordings", *IEEE Transactions on Audio, Speech, and Language Processing*, **19**, 6, pp. 1688–1701 (2011).
- [8] D. A. Jaffe: "Ensemble timing in computer music", *Computer Music Journal*, **9**, 4, pp. 38–48 (1985).
- [9] P. Desain and H. Honing: "Tempo curves considered harmful", *Music Review*, **7**, 2, pp. 123–138 (1994).
- [10] T. Nakano and M. Goto: "Vocalistener: A singing-singing synthesis system based on iterative parameter estimation", the 6th Sound and Music Computing Conference (SMC 2009), pp. 343–348 (2009).
- [11] 森勢, 河原, 片寄: "Straight によるリアルタイム歌唱モーフィングシステムの実装", 情報処理学会研究報告 [音楽情報科学] (2008).
- [12] R. Ramirez, E. Maestre and X. Serra: "Automatic performer identification in commercial monophonic jazz performances", *Pattern Recognition of Non-Speech Audio*, **31**, 12, pp. 1514–1523 (2010).
- [13] R. Ramirez, E. Maestre, A. Perez and E. Maestre: "Automatic performer identification in celtic violin audio recordings", *Music and Machine Learning*, **40**, 2, pp. 1514–1523 (2011).
- [14] K. Hirata and R. Hiraga: "Ha-hi-hun: Performance rendering system of high controllability", *ICAD 2002 Rencon Workshop*, pp. 40–46 (2002).
- [15] K. Hirata and R. Hiraga: "Ha-hi-hun: Performance rendering system of high controllability", *ICAD 2002 Rencon Workshop*, pp. 40–46 (2013).
- [16] 橋田, 松井, 北原, 片寄: "ピアノ名演奏の演奏表現情報と音楽構造情報を対象とした音楽演奏表情データベース crestmusedb の構築", 情報処理学会論文誌, **50**, 3, pp. 1090–1099 (2009).
- [17] G. Widmer, S. Flossmann and M. Grachten: "Yqx plays chopin", *AI Magazine*, **30**, 3, pp. 35–48 (2009).
- [18] K. Teramura, H. Okuma, Y. Taniguchi, S. Makimoto and S. Maeda: "Gaussian process regression for rendering music performance", *International Conference on Music Perception and Cognition (ICMPC)*, pp. 167–172 (2008).
- [19] T. H. Kim, S. Fukayama, T. Nishimoto and S. Sagayama: "Polyhymnia: An automatic piano performance system with statistical modeling of polyphonic expression and musical symbol interpretation", *NIME*, pp. 96–99 (2011).
- [20] T. Machover: "Hyperinstruments - a progress report 1987 - 1991", MIT Media Laboratory (1992).
- [21] J. Bonada and A. Loscos: "Sample-based singing voice synthesizer by spectral concatenation", the Stockholm Music Acoustics Conference (SMAC) (2003).
- [22] M. Demoucron: "On the control of virtual violins: Physical modelling and control of bowed string instruments", PhD thesis, Universite Pierre et Marie Curie (UPMC), Paris & Royal Institute of Technology (KTH), Stockholm (2008).
- [23] E. Maestre: "Modeling instrumental gestures: an analysis/synthesis framework for violin bowing", PhD Thesis, Universitat Pompeu Fabra (2009).
- [24] E. Maestre, M. Blaauw, J. B. nad E. Guaus and A. Perez: "Statistical modeling of bowing control applied to violin sound synthesis", *IEEE Transactions on Audio, Speech, and Language Processing*, **18**, 4 (2010).
- [25] A. Perez, J. Bonada, E. Maestre, M. Blaauw and E. Guaus: "Performance control driven violin timbre model based on neural networks", *IEEE Transactions on Audio, Speech, and Language Processing*, **20**, 3 (2012).
- [26] M. Marchini, P. Papiotis and E. Maestre: "Timing synchronization in string quartet performance: a preliminary study", *International Workshop on Computer Music Modeling and Retrieval (CMMR12)*, pp. 117–185 (2012).
- [27] P. Papiotis, M. Marchini and E. Maestre: "Computational analysis of solo versus ensemble performance in string quartets: Dynamics and intonation", *12th International Conference on Music Perception and Cognition (ICMPC) (2012)*.
- [28] 小泉, 伊藤: "演奏意図関数に基づく表現力を反映させた音響信号の伸縮修正", 情報処理学会研究報告 [音楽情報科学] (97) (2012).
- [29] 小泉, 伊藤: "奏者の意図したテンポ変動の推定に基づく演奏録音の自動修正", FIT2013 第 12 回情報科学技術フォーラム (submitted) (2013).
- [30] 三浦, 江村, 秋永, 柳田: "ピアノによる 1 オクターブの上下行長音階演奏に対する熟達度の自動評価", *日本音響学会誌*, **66**, 5, pp. 203–212 (2010).
- [31] 小泉, 伊藤: "演奏意図関数に基づく表現力を反映させた音響信号の伸縮修正", 情報処理学会第 75 回全国大会講演論文集, 第 2 巻, pp. 257–258 (2013).
- [32] E. Maestre and E. Gomez: "Automatic characterization of dynamics and articulation of expressive monophonic recordings", *AES 118th Convention* (2005).
- [33] 菊池, 宮島, 前川: "表現豊かな音声の収録における多様性の追求", *日本音響学会周期研究発表会講演論文集*, pp. 263–264 (2012).