

Music Signal Analysis Methods and Their Applications to Bowed String Instruments Based on Expressiveness Modeling

KOIZUMI, Yuma / 小泉, 悠馬

(出版者 / Publisher)

法政大学大学院情報科学研究科

(雑誌名 / Journal or Publication Title)

法政大学大学院紀要. 情報科学研究科編 / 法政大学大学院紀要. 情報科学研究科編

(巻 / Volume)

9

(開始ページ / Start Page)

101

(終了ページ / End Page)

106

(発行年 / Year)

2014-03

(URL)

<https://doi.org/10.15002/00010529>

Music Signal Analysis Methods and Their Applications to Bowed String Instruments Based on Expressiveness Modeling

Yuma Koizumi

Graduate School of Computer and Information Sciences, Hosei University

E-mail: 12t0005@cis.k.hosei.ac.jp

Abstract

The essence of music is the “expression” of each performer, namely, the deviations in amplitude, pitch, timbre and tempo/rhythm that they add to their performance. Hence, in computational applications of music, deviation analysis is important. However, because the musical tone of excitation-continuous musical instruments changes complexly in accordance with the level of controllability, it is difficult to analyze the deviations. This paper proposes five deviation analysis methods, focused on statistical consistency and repetition, for bowed string instruments. Results of analysis using each of the methods are presented. By using CMKLD, which is an acoustic feature based on aural characteristics, the error rate of musical score alignment was decreased to 63.2 percentage points. By modeling the sound control indeterminacy due to performance expression, the error rate of intra-note segmentation was decreased to 89.4 (A-to-S) and 48.8 (S-to-R) percentage points, respectively. By using a generative model of amplitude contour, focused on statistical consistency, a performer’s phrasing and variation of articulation could be analyzed. By using a physical model of a violin in the frequency domain, high-quality sound could be synthesized via quantitative expressiveness parameters. By removing deviations that have no statistical consistency, misplayed sounds could be adjusted. These results show that the proposed methods can be used to analyze the expressive deviation of bowed string instruments, and can be applied to musical information processing systems such as automatic performance and automatic skill evaluation.

1 序論

音楽は、音の心理要素である音量、音色、音高とその時間的秩序(リズム、テンポ)を巧みに制御し情報伝達する音メディアである。奏者は、演奏意図や楽譜の解釈などに基づき、楽譜制約される相対的な音量や音高に対し逸脱を加え、個性や芸術性を伝達する[1]。よって、楽譜が同一でも、奏者ごとに演奏が異なる。これら逸脱は、楽器[2]や作曲家[3]でも特徴が異なる。そのため、逸脱のない“機械的な演奏”^{*1}は、“音楽的意味”を欠く。よって楽音合成などの応用では、逸脱を解析し、音楽的かつ定量的に扱う必要がある。しかし、連続励起振動楽器(擦弦楽器など)では、励起振動源の制御の自由度が高く、逸脱が複雑に変化するため、逸脱解析が困難であり、応用技術の対象は、逸脱解析が容易な撥/打弦楽器(ピアノなど)が中心だった。

そこで本稿では、擦弦楽器のための音楽演奏中の演奏表現に起因する逸脱量の解析手法の確立をめざす。また、解析した逸脱量に基づく楽音合成、制御、転写手法を提案する。逸脱量解析を、音楽および物理的な観点からの制約条件を設けた逆推定問題として定式化する。また従来の問題となっていた楽音制御の柔軟性を、統計モデルの自由度として扱うことで解決する。

本稿では参考文献表を省略するが番号は本論と同一である。

2 演奏表現解析の先行研究

フレーズ中の逸脱成分は、音色、音高、音量、テンポの4要素に表れる。これら逸脱が複雑に作用しあうことにより、演奏

表現が生成される。しかし、連続励起振動楽器の逸脱は複雑に変動するため、各要素の解析法も整備されていない。そのため多くは、1つのパラメータに焦点を絞って議論を行っている。

また楽音合成、制御、転写などの応用を考えた場合逸脱は、“定量的”かつ“音楽/物理的に意味があり”、“音声波形に可逆”な情報として解析しなくてはならない。本章では、連続励起振動楽器の特性と楽譜情報を説明した後、従来の逸脱解析法の多くが上記の観点から見た場合不十分であることを示す。

2.1 連続励起振動楽器

連続励起振動楽器は、演奏中に励起源を常に制御する楽器である。制御が発音時のみの撥/打弦楽器と比べ、任意の時刻で楽音を制御でき、多彩な演奏表現が可能である。一方でこの楽音の不確定性が、連続励起振動楽器の逸脱解析を困難にする。

連続励起振動楽器の楽音解析で特に問題となるのが、音響信号と楽譜の対応付け(スコアアライメント)と音符内状態推定である。スコアアライメントでは音符の開始時刻(発音時刻)を検出する。撥/打弦楽器であれば発音時刻で振幅が急激に増大するため、検出が容易だが、連続励起振動楽器の *legato* 奏法など、振幅に変化が表れにくい演奏音の検出は難しい。

また、1つの音符の楽音には、発音区間、定常区間、減衰区間の3つの音符内区間が存在する。連続励起振動楽器の楽音知覚では、区間ごとに作用する音響特徴が異なるため、分析・認識[4]・生成・制御[5][6]では、音符内区間の考慮が必要である。従来法[7][8]はパワーの変化に基づき音符内状態推定を行うため、演奏表現により音量が複雑に変すると推定精度が低下する。

よって、連続励起振動楽器の逸脱解析では、楽音制御の不確定性をいかにモデル化するかが重要となる。

2.2 楽譜に記載されている定量的な情報

楽譜とは、作曲者が意図した音響信号をシンボルとして記載したものである。シンボル化の過程で、定量的な情報のほとんどが損出する。本節では、できるだけ一般性を失わないように、定量化できる情報について述べる。

音色は、楽器の種類と奏法記号、発想記号で指定される。前者の2種類は励振機構の物理特性を指定し、後者は励振機構のニュートラルな状態からの逸脱の仕方を感性語で指定したものである。後者は、奏者の解釈の幅が大きいため、音色の指定は、励振機構がニュートラルな状態で振動したものとす。

音高は、チューニング音高からの相対的な音高差であり、半音の音高差が100centとなるように演奏される。したがって、音高の指定は、100centごとに遷移する基本周波数とする。

音量は、フォルテやピアノなどの音量記号で指定される大局的な変化であるダイナミクスと、レガートやスタッカートなどの発想記号で指定される局所的な変化であるアーティキュレーションの2種類が指定される。音量の定量的な表現は対数パワー(dB)やラウドネスが一般的であるが、楽譜で指定される情報は相対的な音量差のみであり、また基準となる指標も存在しない。よって、音量には定量的な情報は存在しない。

テンポは、速度記号とリズムで指定される。前者は1分間に拍子上の1拍を何音演奏するか絶対的な指標である。後者は、拍子上の1拍を基準に、何倍の長さで演奏するかを指定したものであり数値的な情報(秒)に変換可能である。したがって、テンポおよびリズムの指定は、各音符の持続時間とする。

2.3 音色の解析

音色解析の研究は、人間の楽音知能力の解明を目的とし、分析合成を用いて行われてきた。古くから取り組まれている周波数領域の分析合成では、擦弦楽器音の知覚/合成には発音区間の非調波性や、定常区間ではビブラートと連動したスペクトル包絡の変動[9]が重要なことが解明された。しかしこの手法は、楽音制御は容易だが、演奏表現をパラメータ化できないため、

Supervisor: Prof. Katunobu Itou

^{*1} 楽譜に書かれている相対音高や、音価通りの演奏

音色の逸脱解析には至っていない。

そこで近年、力学的センサ付きの楽器から奏法情報を取得し、楽音合成する物理モデル方式 [10] が提案された。この手法は、演奏表現に由来する奏法の物理パラメータを用いて楽音制御が出来るため、音楽的に意味のある逸脱の解析/付与が容易である。しかし解析には専用の機材や演奏技術が要求されるため、アプリケーションへの実装が困難である。

よって音色の解析技術では、楽音の物理特性を保持しつつ、周波数領域で逸脱を制御する手法が必要である。

2.4 音高の解析

音高逸脱は、ピブラートの深さや速さなど、基本周波数 (F_0) の動特性である [11]。 F_0 変動の解析は、歌唱解析の分野で広く取り組まれており、最も解析技術が発達している。解析では、発声機構の物理モデルを線形 2 次系で近似し、隠れマルコフモデル (HMM) を利用して逸脱成分を解析する手法 [12]、応用では F_0 軌跡の短時間スペクトルを判別特徴とする歌唱力自動評価 [13] などがある。

2.5 音量の解析

音量は、楽譜に絶対的な指標が存在しないため、演奏解釈の自由度が高い。また連続励起振動楽器は定常部でも励起振動源を制御可能であるため、本来別個の情報であるダイナミクスとアーティキュレーションを分離して扱うことが困難である。

よって、音量の解析では、ダイナミクスを持続部の平均音量 [14] や、持続部を直線で結んだもの [7] とし、アーティキュレーションは、発音時刻の間隔と音符の継続時間の比率とする [15] などの簡易な手法がとられてきた。また生成では、ADSR のようにアーティキュレーションを固定する手法や、2 成分を明示的に区別しない手法 [16] がとられてきた。

これらの手法は、アーティキュレーションの一貫性を考慮しない、局所的に含まれる演奏表現情報を解析できない、などの問題がある。よって音量軌跡の解析技術では、軌跡をダイナミクスとアーティキュレーションに分離する自由度の高い不良設定問題を、音楽的な制約を守りながら解く必要がある。

2.6 テンポ変動の解析

テンポ変動は、楽器による逸脱の差異が小さい特徴量であり、応用は、自動採譜 [17] や、自動演奏 [18][19] がある。解析法は、独奏音では発音時刻検出、混合音では tempogram [20] がある。

熟練した奏者のテンポ変動は、フレーズに沿ってなめらかな曲線を描く (テンポ曲線)。一方、テンポ曲線だけでは説明できない微細なテンポ変動が存在する [21] という主張もあり、近年では、意図的なテンポの微細変動を表現するために、マイクロテンポとマクロテンポ [22] を解析すべきという主張もある。このように、熟練した奏者を対象のテンポ変動に対しては様々な議論が行われている。しかし、解析結果の応用技術を幅広いユーザーが使用するには、演奏を習熟していない奏者の演奏も解析しなくてはならない。よって、演奏ミスによるテンポ変動を含んだ演奏からテンポ曲線を推定する手法が必要である。

2.7 本論文の構成

これらの問題点より本稿では、音色、音量、テンポ変動の逸脱解析法を検討する。まず、3 章で連続励起振動楽器の独奏音を対象としたスコアアライメント法、4 章で音符内区間推定法を述べる。次いで 5 章では、連続励起振動楽器音解析で深く検討されていない、音量軌跡のダイナミクスとアーティキュレーションへの分離法を述べる。また、楽音合成や修正を行う際には、楽音の自然性を保ったまま、演奏意図を反映させて楽音制御を行う必要がある。よって 6 章では、2.3 節で挙げた問題点を解決する擦弦楽器音合成法を述べる。最後に 7 章で、テンポ変動における奏法誤差の推定および除去法について検討する。

3 複素メル KL 情報量によるスコアアライメント

演奏音と楽譜を対応付けるスコアアライメントとして発音時刻検出を行う。ここで楽譜とは音高^{*2}と音価^{*3}を指す。発音時刻とは励振機構の発振の開始時刻であり、音符を知覚できる最

も早い時刻である。発音時付近では、楽器の種類や奏法に対応した音響特徴が急激に変化する。従来法 [26][27] は、混合音やピアノ、ギターなどの発音時刻を検出を目標とし、振幅変化を利用するため、連続励起振動楽器の検出精度は悪い。

音符の変化は、どのような演奏表現でも、聴衆が知覚できるように演奏されるため、音響特徴量には、人の聴覚が受ける刺激の変化を尺度とする。そこで発音時刻検出の特徴量として、複素メルスペクトルの KL 情報量 (CMKLD: Complex Mel-spectrum Kullback-Leibler Divergence) を提案する。CMKLD は、人が知覚する、予測できない急激な複素スペクトルの変化の尺度である。CMKLD は、時刻 t で観測された複素メルスペクトル $S_{\mu,t}$ と、微小時間 τ -ms 前から予測される時刻 t の複素メルスペクトル $\hat{S}_{\mu,t}$ の KL 情報量として計算される。CMKLD の問題点として、位相の周期性があるため、以下の近似式で求める。

$$\mathcal{D}[t] \approx \sum_{\mu} |S_{\mu,t}| \sqrt{\left(\log \left(|S_{\mu,t}| / |\hat{S}_{\mu,t}| \right) \right)^2 + \Phi_{\mu,t}^2} \quad (1)$$

次に $\mathcal{D}$ から局所的なピーク値を検出し、発音時刻の候補集合 $\mathcal{O}$ を生成する。各局所ピークは、大きさのばらつきやノイズの影響で様ではないため、動的閾値を以下のように求める。

$$\mathcal{D}_{th}[t] = \lambda \text{Median}(\mathbf{d}_t) + \text{Median}(\mathcal{D}) / 2 \quad (2)$$

ただし係数 λ は、初期値 ξ から始め、 $|\mathcal{O}| \geq N$ とならなければ $\Delta \xi$ 減少させ再度ピーク検出を行う。ここで、 N はアライメントしたい楽譜に記載されている音符の数である。また $\mathbf{d}_t$ は $\mathcal{D}$ を区間 $[t - T/2, t + T/2]$ で切り出したベクトルである。そして $\mathcal{D}[t]$ から、動的閾値 $\mathcal{D}_{th}[t]$ よりも大きなピーク値を選択し、その時刻を候補集合 $\mathcal{O}$ とする。

最後に候補集合 $\mathcal{O}$ から、 N 個の音符の発音時刻を選択する。まず、推定した F_0 軌跡と楽譜情報の DTW (Dynamic Time Warping) によるスコアアライメントでスコア F_0 軌跡を生成する。そして、スコア F_0 軌跡が変化する時刻を、発音時刻の初期値 $z[n]$ とする。但し、隣接する音符のノート高が変化しない場合は、隣接するノートの音価の比率を用いて $z[n]$ を決定する。最後に、候補集合 $\mathcal{O}$ の中から、以下の CMKLD 重み付距離が最小の候補 $\mathcal{O}[i]$ を選択し、 n 音目の発音時刻 $o[n]$ とする。

$$o[n] = \arg \min_{\mathcal{O}[i] \in \mathcal{O}} (|\mathcal{O}[i] - z[n]| / \mathcal{D}[\mathcal{O}[i]]) \quad (3)$$

3.1 精度評価実験

提案法の精度評価を行う。実験には、吹奏楽器からフルート、クラリネット、サキソフォン、トランペット、バイオリンの *legato*, *marcato* などの、様々な奏法を含む 10 フレーズを用いた。実験データの総音符数は 349 であり、演奏時間は 7-30 秒である。正解データは、時間波形、基本周波数、スペクトログラム、音量の変化を元に、3 人がアノテーションを行い、その結果の平均値を正解時刻とした。

提案法と従来法 [26] の発音時刻検出の精度を、正解時刻を中心とする時間窓を許容誤差とする適合率で比較した結果、10-ms から 100-ms の全ての窓幅で、提案法の適合率が上回った。また、発音時刻検出の許容誤差とされる 50-ms の窓幅 [23] では、エラー率が 63.2% 減少した。これは従来法がスペクトルの乖離度を、線形周波数領域で均一な重みで評価したのに対し、CMKLD は聴覚特性を考慮したメル対数周波数領域で、ピッチを有する楽音の知覚に重要となる調波周波数の乖離に対して重みを置いて評価したためと考えられる。このことから、提案法は連続励起振動楽器の独奏音の発音時刻検出に有効である。

3.2 まとめ

本章では、連続励起振動楽器音のためのスコアアライメント法として、聴覚特性を考慮した音響特徴量である複素メル KL 情報量基準に基づく発音時刻検出法を提案した。提案法と従来法の適合率の比較では、全ての時間窓幅で提案法の適合率が上回り、またエラー率が 63.2% 減少した。このことから、提案法は連続励起振動楽器の独奏音の発音時刻検出に有効である。

4 HMM を入れ子にする無限混合正規分布を用いた音符内状態推定

楽音の不確定性を生成モデルに内包する音符内状態推定法として、ディリクレ過程を出力する Nest 型 HMM を応用した、

^{*2} 各音高について割り振られた値である。本稿では MIDI と同様に、“Middle C” (261.6 Hz) を 60、A3 (440 Hz) を 69 とする。

^{*3} “音価”は楽譜上の音符の長さである。本稿では、4 分音符を 1、2 分音符を 2、8 分音符を 0.5 のように定義する。

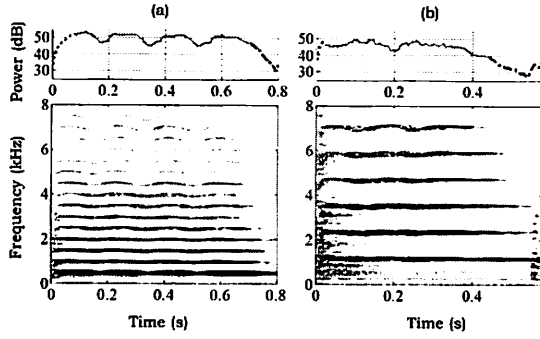


図 1. 音符内区間による音響特性の変化例 (上: 音量軌跡, 下: スペクトログラム). バイオリンの通常の音量で演奏した音 (a) と強く演奏した音 (b). 点線が発音区間, 実線が定常区間, 破線が減衰区間を示す.

音符内状態推定法を提案する. 発音区間, 定常区間, 減衰区間は, 励起機構の振動の変化で区別される. 音符内状態推定では, 観測音の音響特徴が変化する時刻を検出し分割する.

発音区間は, 発音時刻から励振機構が安定するまでの区間を指す. 音響特徴は, ほとんどの種類の楽器や奏法で音量が上昇する [8]. また一部の奏法では, 励振機構の非周期振動に起因して, ノイズのような音色となる (図 1 (b)). 定常区間は, 励振機構が安定振動する区間であり, 音量がほぼ一定の区間を指す [8]. また, 演奏表現によってはビブラートが存在し, その影響で振幅や音色も変化する. 減衰区間は, 奏者による励振機構の直接的な制御が終了してから, 楽音が知覚できなくなるまで減衰する区間を指す. 音響特徴は, 音量が急激に減衰をはじめ, 高次倍音から強度が徐々に減少する (図 1 (a)).

4.1 音符内区間を考慮した楽音の生成過程

楽音中の音符内状態の生成モデルを考える. 以降では, 時刻 t での音量 x_t を対数パワー (dB) で考える. 音符内区間によって変化する特徴量は主に音量と音色である. よって本研究では, セグメンテーションのための音響特徴量として, 音量と音色に関する音響特徴量を用いる.

音量は, その動的特性で特徴づけられる. よって本研究では音量の特徴として, 音量軌跡の一階差分値 $\Delta x_t = (x_t - x_{t-1})/\Delta t$ を用いる. 音色は, 非周期性や調波倍音比で特徴づけられる. 非周期性を表す音響特徴量には, 調波雑音比や線形予測残差などが考えられるが, 本稿では, 時刻 t の振幅スペクトルの調波成分がどの程度支配的なのかを推定したい. そこで本稿では, スペクトルの白色性の指標であるスペクトルエントロピー [28] を用いる. また楽音スペクトル包絡に関する特徴量は, スペクトルセントロイドやスペクトルカートシスが有効と言われている [29]. そこで本稿では, スペクトル包絡を確率密度関数とみなし, 正規化周波数に対する 1 次から 4 次のモーメントを計算する. さらに各モーメントの相関を除去するため, 得られたスペクトルエントロピーと 4 つのモーメントを主成分分析する. そして寄与率の高い順から 3 次元 ($c_t = (c_t^1, c_t^2, c_t^3)^T$) を特徴量とする. 以上より, $y_t = (\Delta x_t, c_t^1, c_t^2, c_t^3)^T$ を音響特徴量とする. ただし $\dagger$ は転置を意味する.

4.2 音響特徴量の生成過程

楽音の音量と音色は, ビブラートや奏法などの要因で変動する. またその変動の様子は, 数個の単純な関数で近似可能なものから複雑なものまで様々である. よってあらゆる演奏表現による音響特徴の変化を表現するためには, 自由度を固定することは妥当ではない. 自由度は, 音響信号の変化の複雑さに合わせて変化するべきである.

一方で本稿で考える発音, 定常, 減衰の音符内状態は, 楽音中の励起振動の特性を 3 つに分類したものである. つまり奏法によって特定の区間が出現しないことはあっても, 楽音の複雑さに応じて区間数が 4 以上になることはない. よって音符内状態は, 演奏表現による音響信号の複雑さの変化より上位の概念として考えることが妥当である. 以上の要件を満たすために, 音符内状態 z_t と音響特徴量 y_t に階層的な生成過程を考える.

まず音符内区間の切り替わりを隠れ状態の遷移とみなし, 状態のスキップを含む $K = 3$ 状態間のマルコフ遷移 $z_t \sim \pi_{z_t-1}$

で表現する. また図 1 上図の状態遷移からもわかるように, 音符内状態の遷移には自己遷移が多い. よって, 音符内状態がゆっくりと遷移するよう制約を掛けるために, スティック HMM [30] を用いる.

次に, 各状態に対応した音響特徴量の生成を考える. 各区間には特徴的な音響特徴が存在するものの, 実際に出力される特徴量を事前にパターン化することは困難である. これは, 音響特徴が奏者や奏者の演奏表現によって決定するものであり, 無限個のパターンを有するためである. しかし, 実際に観測される音響パターンが有限であることを考えると, 演奏行動とは, 無限の演奏パターンの中から, 自身の演奏表現に対応した音響特徴を出力する演奏パターンを, 選択的に組み合わせるものと捉えることが出来る. このことを統計モデルとして表現するために, 音響特徴量 y_t を, 分布パラメータ $\Theta_{k,j} = \{\mu_{k,j}, \Lambda_{k,j}\}$ の基底測度 $\mathcal{H}_k$ が音符内区間ごとに異なるネスト型のディリクレ過程 [32] により生成されたとみなし, $y_t \sim \mathcal{N}(\hat{\mu}_{z_t, s_t}, \hat{\Lambda}_{z_t, s_t}^{-1})$, $s_t \sim \psi_{z_t}$ と記述する. ここで ψ_k は, 無限混合正規分布の混合比に対応し, $\varsigma > 0$ をパラメータとする Stick-breaking 過程 [31] により生成される. また分布パラメータの事前分布は, 共役事前分布である, 正規-ウィシャート分布とする.

4.3 状態推定アルゴリズム

パラメータ推論はマルコフ連鎖モンテカルロ法的一种である Gibbs Sampler で推論を行う. モデル中の各パラメータは各潜在変数の条件付き事後分布からサンプルする.

音符内状態の遷移は, 奏法による状態のスキップを含む Left-to-Right のオートマトンだが, スティック HMM は ergodic な HMM であるため, “onset $\rightarrow$ steady $\rightarrow$ onset” などの “逆戻り” が推定されることもある. よって, サンプリング後に後処理により z_t を修正する. 逆戻りを含む時間区間 $\tau \in \{t_1, \dots, t_2\}$ で修正可能なパターン z_τ^p が P 種類考えられるとする. HMM のパラメータ $\Upsilon = \{\pi_k, \psi_k, \Theta_k\}$ が与えられた際, 各パターンに対する尤度は以下の式で求められる.

$$p(z_\tau^p, y_\tau | \Upsilon) = \prod_{\tau=t_1}^{t_2+1} \pi_{z_\tau^p-1, z_\tau^p} \sum_{j=1}^{J_{z_\tau^p}} \psi_{z_\tau^p, j} \mathcal{N}(y_\tau | \mu_{z_\tau^p, j}, \Lambda_{z_\tau^p, j}^{-1}) \quad (4)$$

本手法では, 式 (4) を最大とする z_τ^p を用いて, z_t を修正する.

4.4 評価実験

提案法のノート内セグメンテーションの精度を従来法 [7] とエラー率で比較した. 検出すべき時刻は発音区間から定常区間 (A-to-S) と, 定常区間から減衰区間の切り替わり時刻 (S-to-R) の 2 つである. 検出結果が正解時刻を中心とする 50 ms の時間窓に含まれた場合を正解とした.

提案法のエラー率は, A-to-S が 89.4%, S-to-R が 48.8 % 減少した. 従来法は, 奏者認識 [4] や, 音色モデリング [5] に応用されており, 提案法は, 複雑な音量変動をする演奏に対しても高性能であることから, 連続励起振動楽器の音符内状態推定法として有効である.

S-to-R の推定精度が大幅に向上した理由は, 従来法が音量変化のみに着目したのに対し, 提案法は音色変化も考慮したためである. 擦弦楽器の定常区間における弦を擦りながら音量を減衰させる奏法では, 自然減衰とは音色変化の特徴が異なる.

一方, A-to-S の推定精度向上が小さい理由は, スティック HMM の遷移行列には自己遷移確率を高めるバイアスがかかるため, ある状態に留まる時間が短いとその状態が消滅する方向に向かう. よって発音区間が短い音符は, 発音区間が存在しないと誤判別する. このような音符は, 音量の上昇が他と比べ非常に急峻である特性を利用する改良方法を検討する.

4.5 まとめ

ディリクレ過程を出力する Nest 型隠れマルコフモデルを応用した音符内区間推定法を提案した. 評価実験では, セグメンテーションのエラー率が, 従来法と比べ, A-to-S が 89.4%, S-to-R が 48.8 % 減少した. 従来法は, 奏者認識 [4] や, 音色モデリング [5] に応用されており, 提案法は, 複雑な音量変動をする演奏に対しても高性能であることから, 連続励起振動楽器の音符内状態推定法として有効である.

5 音量軌跡のアーティキュレーションとダイナミクスへの分解に基づく演奏表現分析

音量軌跡の生成過程を階層ディリクレ過程遷移線形動的システム [33] (HDP-SLDS) を用いて表現し、アーティキュレーションとダイナミクスに分離する。音量軌跡の解析結果 [(C-3)] より、アーティキュレーション f_t とダイナミクス g_t に対数音量領域 (dB) での加法性を仮定する (i.e. 時刻 t の音量 y_t は $y_t = f_t + g_t$ を満たす)。また奏者は、同一フレーズ、もしくは楽曲の小区間内では、いくつかのアーティキュレーションや奏法を選択し、再利用して演奏すると仮定する。

5.1 音量軌跡の数理表現と推論

各音符の演奏動作を細かく見たとき、それは“弓を加速する”や“息を減衰させる”などの、いくつかのプリミティブな動作 (以降、奏法プリミティブと呼ぶ) の組み合わせである。そして、各奏法プリミティブが励起振動の物理特性を変化させ、音量軌跡が変化する。よって、本稿ではアーティキュレーションの変化を、奏法プリミティブごとに係数を持つ自己回帰 (AR) モデルで表現する。

$$f_t = \sum_{i=1}^r a_i^t f_{t-i} + e_i^t(z_t), \quad e_i^t(z_t) \sim \mathcal{N}(0, \sigma_f^2(z_t)) \quad (5)$$

ただし $z_t \in \{1, 2, \dots, K\}$ は奏法プリミティブのインジケータである。つまり、AR 係数が再利用されることにより、アーティキュレーションの繰り返し性を表現している。

奏法プリミティブの組み合わせ方を考えたとき、各音符ごとに、“弓の加速”→“音量の維持”→“弓の減速”などの規則的な遷移が存在すると考えられる。よって本稿では、奏法プリミティブの遷移をマルコフ過程で表現し、 z_t の生成過程にスティッキー HDP-HMM [30] を適用する。

ダイナミクスの変動はフレーズ感などに関係し、*sub.p* (急に弱く) などの指定がある場合を除いて緩やかに変化する。また、*sub.p* などの指示があった場合でも、音量が急激に変化したあとは、また緩やかに変化する。これは時系列解析における“トレンド”とみなすことが出来る。時系列解析ではトレンドになんらかの特性が仮定できる場合、直線近似や季節調整法などのトレンド関数を導入する。しかし、本稿でのダイナミクス解析は、奏者の演奏表現に依存した楽譜に記載されないダイナミクスの変動をも解析することを狙っており、事前に関数を当てはめることは困難である。よって本稿ではダイナミクスを、関数形を仮定しない一階の和分プロセスで表現する。

$$g_t = g_{t-1} + e_t^g, \quad e_t^g \sim \mathcal{N}(0, \sigma_g^2) \quad (6)$$

よって、 t での音量 y_t は、パラメータ $\Theta_k = \{A^k, \sigma_f^2(k), \sigma_g^2\}$ によって制御される HDP-SLDS として記述できる。

$$x_t = \begin{bmatrix} A^k & 1 \end{bmatrix} x_{t-1} + e_t^t, \quad y_t = U x_t \quad (7)$$

ただし、 $x_t = (f_t, f_{t-1}, \dots, f_{t-r+1}, g_t)^T$, $e_t^t = (e_t^f(z_t), 0, \dots, 0, e_t^g)^T$ であり、 A^k は k 番目の奏法に対応する VAR(r) 行列、 $U = (1, 0, \dots, 0, 1)$ である。

また、奏法が変化すると、音量以外に音色も変化する。よって本稿では、奏法プリミティブのインジケータ系列 $z_{1:T}$ を効率的に求めるために、4 章と同様の音色の音響特徴量も用いる。時刻 t で観測される音色特徴量 c_t は、パラメータ $\Upsilon_k = \{\mu_k^c, \Sigma_k^c\}$ を持つ、無限混合正規分布 [34] から出力されたものとみなし、 $c_t^k \sim \mathcal{N}(\mu_k^c, \Sigma_k^c)$ とする。

モデル中の各パラメータは、Gibbs Sampler により各変数の条件付き事後分布からサンプルする。

5.2 評価実験

まず、アーティキュレーションとダイナミクスの推定精度を調べるために、MIDI データを利用して作成したデータを用いて分離精度を評価した。評価に用いる楽器は、クラリネット、トランペット、バイオリンとした。楽曲は、レガートやスタカートなどのアーティキュレーションを含む 3 フレーズずつとした。精度を、正解データと推定結果の標準絶対誤差 (MAE: mean absolute error) で評価した結果、MAE は 0.75dB であった。MIDI の Velocity で音量を制御する場合、音量記号の 1 段階変化 (e.g. *mp* から *mf*) の差約 4dB と比べ誤差は小さい。よって、人工的な音量軌跡を用いた場合、提案法の分離アルゴ

リズムは、局所解問題を抑制できていることが確認できた。

誤差が大きくなった楽曲には、フレーズ中に 2 分休符以上の休符が含まれていた。休符中の音量変動はアーティキュレーションにもダイナミクスにも依存しない。演奏行動の観点から考えると、奏者はまず“音を出すか出さないか”を決定するが、提案モデルは奏者が“音符を演奏すること”を前提としたモデルである。今後、楽譜情報などを参照し、休符を提案モデルより 1 段上のレベルで扱えるよう、生成モデルの改善を考える。

次に提案法の実演奏音分析への有効性を示すために、プロ奏者とアマチュア奏者によるバイオリン演奏音の分離実験を行った。アマチュア奏者は、プロ奏者の聴き、30 分間の練習を行ったもとの、プロ奏者の演奏を模倣するように演奏した。図 2 に分離結果例を示す。プロ奏者のダイナミクスは、4 拍ごとに 5 から 10dB 程度の起伏がおきている。これは、奏者が旋律の構造を理解し、それを音楽的に表現するフレージングを行った結果と解釈できる。一方、アマチュア奏者のダイナミクスは、フォルテからピアノの変動幅が 60dB から 50dB と、プロ奏者と一致しているが、4 拍ごとの起伏は確認できない。聴感的には、アマチュア奏者の演奏は“フレーズ感”が感じられない。これはアマチュア奏者が、奏者が旋律の構造を理解していない、もしくは理解したものを演奏として出力する技術を身に付けていないためと考えられる。

これらの結果から提案法は、奏者のフレーズの解釈や演奏表現の変化、演奏技術によるアーティキュレーションのバリエーションなどの演奏解析を行えることが示唆される。

5.3 まとめ

連続励起振動楽器の音量軌跡を、ダイナミクスとアーティキュレーションに分解する手法を提案した。ダイナミクス変動を和分プロセス、アーティキュレーション変動を AR モデルでモデル化し、遷移型線形動的システムを用いて音量軌跡を分離した。MIDI を用いた人口データの分解実験では、平均絶対誤差が 0.75dB で分解可能であることから、局所解を抑制しつつ音量軌跡を分解できることが分かった。実演奏音の分離実験では、奏者のフレーズの解釈やそれに基づく演奏表現の変化、演奏技術によるアーティキュレーションのバリエーションなどの演奏解析を行えること示した。これらの結果から、提案法は連続励起振動楽器の音量解析に有効である。

6 擦弦楽器の音色分析合成のためのハイブリッドソースフィルターモデル

楽音合成では、演奏表現制御の容易さと合成自体の容易さが求められる。代表的な楽音合成方式である、物理モデル方式 [35] や素片接続方式 [36] では、前者は専用の機材や演奏技術が要求され合成が困難、後者は演奏表現がコーパス依存のため演奏表現制御が困難である。よって両者を達成には、物理的な特性などを踏まえたモデル化と、音色を容易に制御できるパラメータが必要となる。この達成のために、擦弦楽器の合成音の柔軟な音色制御のための、物理モデルとスペクトルモデルのハイブリッドな分析合成系である、奏法モデルを提案する。奏法モデルでは、奏法による調波構造の変化、発音区間の非調波成分の変化、定常区間における非調波成分を制御する。

6.1 擦弦楽器音の生成過程

擦弦楽器音は、励起源である擦弦振動が駒を通して楽器本体で共鳴した放射音である [37]。奏法により、弦の各モードの振動の強さや非周期成分が決定し、特徴的な音色が生成される。

擦弦中の基本的な弦の運動はヘルムホルツ振動と呼ばれ、Stick-Slip 運動により生成される。調波成分にのみパワーを持つ振動である。しかし、実際の擦弦振動は、奏法による各モードの振動の強度の比率の変化 [38] や、弦のスティフネスの効果などによる不規則振動 [39] が含まれおり、三角波ではない。この不規則振動には、擦弦運動の非線形性や、非周期成分の間欠性、slip 現象の周期からのランダムなずれと関係するカオス理論 [40] などが関係し、未解明の部分も存在する。

また発音区間では、発音時の弓圧と加速度が釣り合わない場合、数十 ms 程の不安定な slip 現象が発生し、毎振動ごとの周期が安定しない擦弦振動が発生する [41]。この不規則な slip 現象には、一周中に複数回の slip が起きる multiple-flyback や、

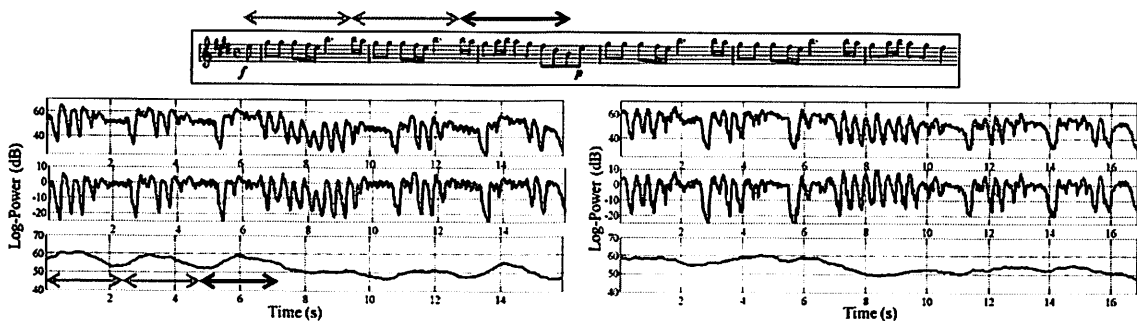


図 2. プロ奏者 (左) とアマチュア奏者 (右) のバイオリン演奏音への推定結果. 3 つ並んだ音量グラフは上から, 実測音量軌跡, アーティキュレーションの推定結果, ダイナミクスの推定結果を示す.

本来 slip が起きる位置よりも遅れて slip が起きる prolonged periods が存在する. 特に後者は, 音に迫力を付与する意図表現の演奏の際に用いられる [42].

楽器の共鳴は線形時不変系と仮定され, インパルス応答により計測される [43][44].

6.2 奏法モデルの構築

擦弦楽器の演奏で, 人間が直接制御を行う部分は擦弦振動である. よって演奏表現による音色変化は, ヘルムホルツ振動からの乖離と考える. このことを時間周波数領域で表現するために, 奏法による各モードの振動の強度の比率の変化を, 時変の線形伝達系として表現する. 一方で非調波成分は非線形な要因で生成されるため, 線形な伝達系では記述出来ない. さらに, 発音区間と定常区間の非調波成分は発生原因が異なる. よって非調波成分はヘルムホルツ振動とは分け, 発音区間と定常区間の非調波モデルは個別の確率モデルで構築する.

6.2.1 調波モデル

ヘルムホルツ振動からの乖離を, 奏法によって変化する線形時変伝達系で記述し, 観測信号の調波成分の振幅スペクトログラム $Y_{\text{harm}}(\omega, t)$ を $Y_{\text{harm}}(\omega, t) = S_{\text{harm}}(\omega, t)E(\omega, t)I(\omega)$ と書く. すなわち, ヘルムホルツ振動のスペクトル $S_{\text{harm}}(\omega, t)$ に, 奏法線形時変伝達系の周波数特性 $E(\omega, t)$ が乗じられ, さらに楽器共鳴の周波数特性 $I(\omega)$ がかったものと解釈する. 本節では, $E(\omega, t)$ を調波成分の奏法モデル (調波モデル) とする.

調波モデルは音色制御を取り扱うものであるため, $E(\omega, t)$ は信号のパワーを変化させないものとする. するとヘルムホルツ振動の調波パワー比より $S_{\text{harm}}(\omega, t)$ が求まる. よって, n 番目の調波成分における調波モデルの伝達特性は, $E(nF_0, t) = Y_{\text{harm}}(nF_0(t), t) / \{S_{\text{harm}}(nF_0(t), t)I(nF_0(t))\}$ で求まる. しかし一般に伝達特性は周波数方向に滑らかに変化するため, 周波数方向に補間を行い, 滑らかな調波モデル $E(\omega, t)$ を推定する.

6.2.2 発音区間の非調波モデル

発音時の非調波成分は, 楽譜によって定義される情報がないため, 発音区間の非調波モデル (非調波発音モデル) は, スペクトルを直接生成する確率的モデルとして表現する. 発音時の非調波成分は, 基本周期 F_0 と異なる周期の slip 現象に起因するスパースなスペクトルである. そこで周波数軸を F 個の区間 $f = \{1, 2, \dots, F\}$ 分割し, スパースなスペクトル $S_{\text{attack}}(\omega, t)$ を, $N_f(t)$ 個の周波数成分が周波数位置 $\omega_f(n_f, t)$ に, 対数パワー $G_{\text{attack}}^f(n_f, t)$ で立ち上ったものとする. よって, $S_{\text{attack}}(\omega, t)$ を生成するためのパラメータは確率変数 $\omega_f(k, t), G_{\text{attack}}^f(k, t), N_f(t)$ であり, 非調波発音モデルの構築問題はパラメータの分布推定問題となる.

$\omega_f(k, t)$ は弦振動が押弦位置により決まる基本周期から確率的に揺らぐことにより発生する. よって $\omega_f(k, t)$ は, 平均を中心に前後に等確率で揺らぎ, かつ解析的に扱いやすい分布である正規分布 $\mathcal{N}(\mu_f, \sigma_f^2)$ でモデル化する. $G_{\text{attack}}^f(k, t)$ はヘルムホルツ振動の対数ゲイン $\log(\sin(f\pi\psi)/f^2)$ が奏法により揺らぐことにより発生する. よって $\omega_f(k, t)$ と同様に正規分布 $\mathcal{N}(\mu_G^f, \sigma_G^2)$ でモデル化する. $N_f(t)$ は区間 f に何本のピンが立ち上がるかを示す変数であるため, ポアソン分布 $\text{Poisson}(\lambda_f)$ でモデル化する. そして, 事後分布および予測分布を解析的に解ける形で導出するために, 分布パラメータに共役事前分布を

設定し, 観測スペクトルから予測分布を求める.

6.2.3 定常区間の非調波モデル

定常区間の擦弦ノイズは slip 現象時に弦のスティフネスの効果により発生する雑音である. 物理モデル [45] では生成された擦弦振動に確率的にノイズを付与している. パラメータは時間変化するノイズの強度 $Q(t)$ である. 本稿では, $Q(t)$ を定常区間の非調波モデル (非調波定常モデル) として推定する.

観測された擦弦振動スペクトルの非調波成分 $S_{\text{noise}}(\omega, t)$ を, 観測擦弦振動のスペクトログラムから, 調波擦弦振動スペクトルを窓関数で周波数軸方向に補間したスペクトルを減算すること求め, $Q(t)$ を求める.

6.3 奏法モデルを用いた楽音合成実験

モデルの有効性の評価のために, 分析合成音と制御合成音の品質を, 主観評価によって評価する. 合成には, 3 種類の発想記号 (*feroce*, *marcato*, *dolce*) で演奏したバイオリンの 4 種類の音高の単音を用いた. 比較音声は, 実演奏音, 提案法, 提案法で音高を短 3 度上昇させた合成音と下降させた合成音, 奏法モデルを用いない合成音と, 非調波発音モデルを用いない合成音 (ただし *dolce* は除く) の計 68 種類とした. 被験者は, 擦弦楽器経験 5 年以上の 5 名とし, 自身の発想記号のイメージに対する音色の合致度および音色の自然さを, 5 段階の MOS (Mean Opinion Score) で評価した.

実験結果に Dunnett の多重比較検定を行ったところ, 提案法による合成音は, 全ての発想記号で実演奏音と危険率 5% で有意差が認められなかった. また音高操作を行った合成音も, 全ての発想記号で実演奏音と危険率 5% で有意差が認められなかった. 調波モデルを用いない合成音は, 全ての発想記号で実演奏音と比較して危険率 5% で有意差が認められ, 発音非調波モデルを用いない合成音は, *marcato* で実演奏音と危険率 5% で有意差が認められた. この結果から奏法モデルを用いた楽音合成は, 実演奏音の発想記号による音色のイメージを保ったまま分析合成および音高制御をすることが可能である. 認められたことから, 意図表現の知覚において音色の変化が重要な要素であることが示された.

6.4 まとめ

本章では, 擦弦楽器の合成音の柔軟な音色制御のための, 物理モデルとスペクトルモデルのハイブリッドな分析合成系である, 擦弦楽器の奏法モデルとその構築法を提案した. 奏法モデルを用いた分析合成音の発想記号に基づく演奏表現のイメージに対する音色の合致度および音色の自然さの評価では, バイオリンの単音では元音声と有意差のない楽音を分析合成できることを示した. また音高制御では, 発想記号のイメージを保ったまま, 元音声と同等の品質で楽音制御を行えることを示した. よって提案法は, 聴衆の耳になじみ深い発想記号や, 特徴的な音色を持つ発想記号の分析合成に有効である.

7 真のテンポ曲線の推定に基づく演奏音の伸縮修正

熟練した奏者^{*4}は意図表現に基づき, フレーズ中にテンポを滑らかに変動させる (図 3 左). が, 熟練度の低い奏者^{*5}のテンポは滑らかに変化せず, ばらつく (図 3 右) 本稿では, アマ

^{*4} 以降, 本来の定義とは異なるが, “プロ奏者” と呼ぶ.

^{*5} 以降, 本来の定義とは異なるが, “アマチュア奏者” と呼ぶ.

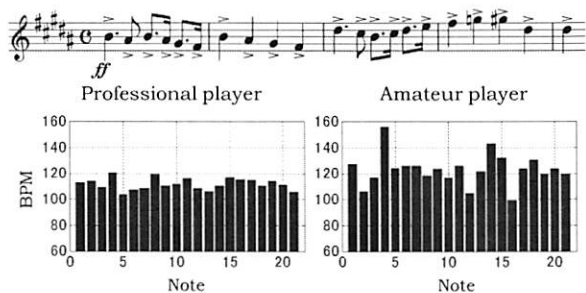


図 3. プロ奏者とアマチュア奏者のテンポ変動例

チュア奏者も滑らかなテンポ変動を意図して演奏するが、楽器の制御ミスによりテンポがばらつくことと仮定し、奏者の意図した滑らかなテンポ変動である“真のテンポ曲線”を推定し、奏法誤差に由来するテンポ変動を除去する。

7.1 真のテンポ曲線の推定と音響信号の修正

音符の持続時間は対象とする音符の発音時刻から次の音符の発音時刻までとする。すると、奏法誤差成分を含まない n 音目の発音時刻 $y[n]$ は、1 音目から $(n-1)$ 音目までの持続時間の和となる。しかし実際の発音時刻には奏法誤差成分が含まれる。ここで奏法誤差成分は、真のテンポ曲線によって決まる発音時刻に対し加法的に作用すると仮定し、 $y[n] = \sum_{m=1}^{n-1} (60/b[m])h[m] + e[n]$ と書く。ここで $h[m]$ は m 音目の音価、 $b[m]$ は m 音目の真のテンポ曲線の値、 $e[n]$ は n 音目の奏法誤差成分の値 (秒) である。

さらに、テンポ変動を曲線として推定するために、武田らの曲線フィッティング [17] を参考に、真のテンポ曲線の逆数を多項式カーネルを用いて $b[n]^{-1} = \sum_{p=0}^P w_p g[n]^p$ と定義する。ここで P は多項式の次数である。また $g[n]$ は累積された拍の相対位置を示し、 $g[n] = \sum_{m=1}^{n-1} h[m]$ である。よって n 音目の持続時間 $\Delta y[n] = y[n+1] - y[n]$ は $\Delta y[n] = 60 \sum_{p=0}^P w_p g[n]^p h[n] + e[n+1] - e[n]$ となる。ただし、音響信号中に存在しない $(N+1)$ 音目の発音時刻は $y[N+1] = L_x/f_s$ とする。ここで L_x は音響信号のデータ点数であり、 f_s はサンプリングレートを表す。

ここで、 $N \times (P+1)$ の説明変数行列を $G_{n,p} = \{g[n]^{(p-1)}h[n]\}$ と置くことにより、音符の持続時間ベクトル $\Delta y = (\Delta y[1], \dots, \Delta y[N])^T$ は $\Delta y = 60Gw + \Delta e$ と書ける。ここで w は回帰係数を並べたベクトル $w = (w_0, \dots, w_P)^T$ であり、 Δe は奏法誤差成分のデルタベクトル $\Delta e = (e[2] - e[1], e[3] - e[2], \dots, -e[N])^T$ である。

ここで先行研究 [46] を参考に、 $e[n] \sim \mathcal{N}(0, \sigma^2)$ と仮定すると、正規分布の再生性より、 Δe の各要素も正規分布に従う。よって、最小二乗法により回帰係数ベクトル w を求めることで、テンポ曲線が求まる。多項式カーネル回帰の問題として、最適な多項式の次数 P の決定が挙げられるが、本稿では赤池情報量基準 (AIC) [47] の最小化で次数 P を決定する。

音響信号の修正は、“各音符の持続時間から奏法誤差による変動を除去すること”と定義できる。奏者の意図した音符の持続時間 $\hat{z}[n]$ は、真のテンポ曲線を用いて $\hat{z}[n] = (60/b[n])h[n]$ となる。また、観測された n 音目の持続時間は $y[n+1] - y[n]$ であるため、音響信号の修正は、 n 音目の持続時間を伸縮係数 $\alpha[n] = \hat{z}[n]/\Delta y[n]$ 倍することとなる。

音響信号の修正伸縮には、パワースペクトログラムの逆短時間フーリエ変換 (IDFT) のシフト幅の伸縮による速度変換手法 [48] を用いる。本稿では、各音符ごとの IDFT のシフト幅を $\alpha[n]$ 倍して、音響信号を伸縮する。シフト幅の変化による位相の不整合は、Griffin らの位相再構成法 [49] で除去する。

7.2 評価実験

提案法により、奏者の意図したテンポ変動に修正されるかを、聴取実験で評価した。対象楽器は、バイオリン、チェロ、エレキギターとした。推定する真のテンポ曲線は、奏者の意図したテンポ変動であり、正解データが存在しない。そこで、目標テンポ変動に、プロ奏者の演奏を用いた。楽器歴 3 年以上のアマチュア奏者が、プロ奏者の演奏を聴き 30 分間練習し、そのテンポ変動を模倣するよう演奏した。よって、正解データはプロ

奏者の演奏のテンポ変動であり、修正が正しく行われているならば、修正後のテンポ変動はプロ奏者のものに近づく。

アマチュア奏者は各楽器 2 名ずつ、楽曲は各楽器 3 曲ずつとした。聴取実験では 5 年以上の音楽経験を持つ、演奏者と別の 5 名が、実演奏音と修正音のテンポ変動の、プロ奏者の演奏との近さを評価した。評価には 5 段階の MOS を用いた。各評定は 1 が非常に遠い、5 が非常に近いを表す。

修正音の評定は、全ての楽器で実演奏の評定よりも上昇した。 t -検定では、バイオリンとギターは危険率 1% で、チェロは危険率 5% で有意差が認められた。アマチュア奏者はプロ奏者の演奏のテンポ変動を意図して演奏しており、提案法を用いた修正により、修正音がプロ奏者の演奏に有意に近づいたことから、提案法は、奏者の意図したテンポ変動を推定し、その変動に合わせて音響信号を伸縮修正できる。

7.3 まとめ

本稿では、奏法誤差成分を含んだ独奏音から、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する手法を提案した。また、真のテンポ曲線を用いて、音響信号のテンポ変動を奏者の意図したものに自動修正する手法を提案した。聴取実験では、修正音と目標演奏のテンポ変動の類似性が修正前と比べ有意に向上した。よって提案法は、奏者の意図したテンポ変動を推定し楽音修正を行える。

8 結論

本研究では、擦弦楽器のための演奏表現に起因する逸脱の解析法を提案した。3 章では、聴覚を意識した特徴量である複素メル KL 情報量をスコアアライメントに用い、従来法よりエラー率が 63.2% 減少することを示した。4 章では、楽音制御の不確定性を内包する生成モデルを立てることにより、音符内状態推定のエラー率が従来法と比べ、A-to-S が 89.4%、S-to-R が 48.8% 減少した。5 章では、演奏表現による音響変動の、時間方向への統計的一貫性に着目した音量生成モデルにより、ダイナミクスとアーティキュレーションを平均絶対誤差 0.75dB で分解した。また、分離結果を用いた楽音解析では奏者のフレーズの解釈や演奏表現の変化、演奏技術による音量軌跡のバリエーションなどを解析できること示した。6 章では、擦弦楽器の物理モデルを周波数領域で信号処理/統計的に扱うことにより、音色の逸脱を定量的に扱いつつ、高品質に楽音合成できることを示した。7 章では、統計的一貫性を持たない逸脱を奏法誤差とみなし除去することで、楽音修正できることを示した。

本研究の意義は、従来、データだけを頼りに統計的なモデリングに終始した数理統計的アプローチで行われてきた音楽音響信号解析に、音響心理学や物理学の知見を融合させた点である。音楽の特徴である繰り返しや音響特徴量変化の時間秩序を効果的に利用することで、従来法より解析精度が向上した。本研究で解析可能となった、ダイナミクスとアーティキュレーションや音色の逸脱量、奏法誤差を含んだ演奏のテンポ変動は、音楽情報処理に発展に重要な知見をもたらすと考える。

本研究の結果に基づいた新たな研究対象として、2 章冒頭で述べた、各逸脱の演奏表現知覚への関係性や相關解析が挙げられる。例えば擦弦楽器や歌唱の音高変動は、音量や音色に影響する。様々な演奏表現で演奏されたデータから逸脱を解析し、音響特徴量自体の生成モデルを立てることにより、連続励起振動楽器の自動演奏や、習熟度自動評価も可能になる。将来的には人間の演奏と区別のつかない、自然かつ演奏表現豊かな楽音合成システムや、演奏技術取得支援システムを開発したい。

主要発表文献

- (1) “連続励起振動楽器を対象としたノート内セグメンテーション” 信学論, 2014, (in press).
- (2) “Intra-note Segmentation via Sticky HMM with DP Emission” ICASSP-14, 2014, (accepted).
- (3) “連続励起振動楽器を対象とした音量軌跡のダイナミクスとアーティキュレーションへの分解法” 情報研報-MUS-102, 2014.
- (4) “擦弦楽器の意図表現合成のための奏法モデル” 情報処, 2013.
- (5) “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伸縮修正法” FIT-2013, 2013.
- (6) “Expressive Oriented Time-Scale Adjustment for Mis-played Musical Signals based on Tempo Curve Estimation” DAFx-16, 2013.