2013年度修士論文

擦弦楽器のための
演奏表現のモデル化に基づく楽音分析と応用

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

小泉 悠馬
Yuma Koizumi
学籍番号 12T0005
法政大学大学院 情報科学研究科 情報科学専攻
E-mail: 12t0005@cis.k.hosei.ac.jp

指導教官 伊藤克亘 教授

目 次

第 1 章　序論 **4**

第 2 章　演奏表現解析の先行研究 **5**
2.1　連続励起振動楽器 5
2.2　楽譜に記載されている定量的な情報 6
2.3　音色の解析 7
2.4　音高の解析 7
2.5　音量の解析 8
2.6　テンポ変動の解析 8
2.7　本論文の構成 8

第 3 章　複素メル KL 情報量によるスコアアライメント **9**
3.1　提案法 9
3.1.1　複素メルスペクトルの計算手順 10
3.1.2　複素メル KL 情報量による発音時刻検出 11
3.2　精度評価実験 12
3.2.1　発音時刻検出の精度評価 12
3.2.2　雑音・残響耐性実験 13
3.3　まとめ 14
3.4　関連研究 14

第 4 章　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 **16**
4.1　音符内区間ごとの音響特性 16
4.2　音符内区間を考慮した楽音の生成過程 16
4.2.1　音符内区間推定のための音響特徴量 16
4.2.2　音響特徴量の生成過程 17
4.3　状態推定アルゴリズム 18
4.3.1　パラメータのギブスサンプリング 19
4.3.2　z_t の後処理 20
4.4　評価実験 21
4.4.1　実験条件 21
4.4.2　精度評価実験 21
4.5　まとめ 22
4.6　関連研究 22

第 5 章　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 **24**
5.1　連続励起振動楽器の音量軌跡 24
5.2　音量軌跡の生成モデル 25
5.2.1　音量軌跡の線形動的システム表現 25

5.2.2 奏法プリミティブによる音色変化 26
5.3 推論アルゴリズムの実装 26
5.4 評価実験 28
5.4.1 MIDI データを用いた分離実験 28
5.4.2 実演奏音を用いた分離実験 29
5.5 おわりに 31
5.6 関連研究 31

第 6 章 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 32
6.1 擦弦楽器音の生成過程 32
6.1.1 擦弦振動 32
6.1.2 楽器の共鳴 33
6.2 奏法モデルの構築 33
6.2.1 調波モデル 34
6.2.2 発音区間の非調波モデル 35
6.2.3 定常区間の非調波モデル 37
6.3 奏法モデルを用いた楽音合成実験 38
6.3.1 楽音の合成 38
6.3.2 評価実験 38
6.4 まとめ 41
6.5 関連研究 41

第 7 章 真のテンポ曲線の推定に基づく演奏音の伸縮修正 42
7.1 真のテンポ曲線の推定と音響信号の修正 42
7.1.1 真のテンポ曲線の推定 42
7.1.2 音響信号の伸縮修正 44
7.2 評価実験 44
7.3 まとめ 46
7.4 関連研究 46

第 8 章 結論 47

謝辞 48

付 録 A アライメントデータセット 49

参考文献 50

研究業績 56

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 decreased by 63.2 percentage points. By modeling the sound control indeterminacy due to performance expression, the error rate of intra-note segmentation decreased by 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.

第1章　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）を巧妙に制御することで情報を伝達する音メディアである．奏者は，演奏意図（performance intention）や楽譜の解釈，習熟度などに基づき，楽譜によって制約される相対的な音量や音高に対し逸脱（deviance）を加えることで，個性や芸術性を伝達する [1]．よって，同一の楽譜を用いても，奏者によって演奏が異なる．これら逸脱は，音楽構造，楽器 [2] や作曲者 [3] によっても特徴が異なる．そのため，逸脱のない “機械的な演奏[1]” は，“ 音楽的意味 ”を欠く演奏となる．よって楽音の認識や合成など，ほとんどの音楽メディアの工学的の応用では，これら逸脱を解析し，音楽的かつ定量的に扱う必要がある．

逸脱解析の研究の古くは，音楽知覚の解明を目的とし，音響心理学や人間科学の分野で行われてきた．それら研究の中で，音楽の知覚には音の動特性 (時間変化) や予測可能性 (刺激の同期・周期性) が重要であると解明されてきた．また演奏の生成は，奏者は楽譜に記載される音符の並びから音楽的な意味を創出し，それを逸脱に変換して音メディアに付与していることがわかった．この逸脱の元となる抽象度の高い情報を “演奏意図 (performance intention)” と呼び，演奏意図を逸脱へ変換していく際の “ノイズ”（奏法誤差）の除去能力を “技巧” や “習熟度 (technique)” と呼んでいる．しかし，連続励起振動楽器（擦弦楽器や吹奏楽器）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．そのため，逸脱解析が困難であり，聴衆がどのようにして逸脱を知覚しているかも完全には解明されていない．さらに，低熟練度の奏者の演奏意図に対し，どのように奏法誤差が加わるのかも分かっていない．よって自動演奏や熟達度評価などの工学的な応用技術は，逸脱解析が容易な撥/打弦楽器（ピアノやギター）を対象としたものが中心であった．

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（expressiveness）に起因する逸脱量の解析/推定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．ただし，本研究では，音響信号からの逸脱量の解析に注力し，逸脱量の楽譜との対応付けた学習や，演奏意図などの抽象的な情報の推定は行わない．

[1]楽譜に書かれている相対音高や，音価通りの演奏

第2章　演奏表現解析の先行研究

音楽フレーズ中の楽譜からの逸脱成分は，発話における韻律変化と同様に，音色，音高，音量，テンポ/リズムの 4 つの要素に表れる．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音楽演奏中の逸脱は非常に複雑な変動量であり，後述するように個々の要素の逸脱解析法も十分に整備されていない．そのため，多くの逸脱解析の研究が，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，転写などの応用を見据えた逸脱量解析では，以下の点を考慮しなくてはならない．

1. 逸脱を定量的に表現する必要がある．

   人間の音楽指導と違い，計算機では感性語などの抽象的な情報を扱うことが困難である．よって，逸脱量は数値的に取り扱わなくてはならない．人間による合成や制御に，抽象的な情報を用いたい場合は，逸脱量と感性語などの抽象的な情報を対応付ける必要がある．

2. 音楽/物理的に意味のある逸脱量を解析する必要がある．

   音楽/物理的に意味のない逸脱は，観測ノイズや奏法誤差の可能性がある．よって，逸脱量を解析する際は，音楽/物理的に適切な制約条件を設けて解析する必要がある．

3. 音声波形に可逆な情報を解析する必要がある．

   音合成/制御では，解析した逸脱をもとに時間波形を制御する．よって逸脱を計測する音響特徴量は，時間波形に可逆なものを用いる必要がある[1]．これが，認識と合成の特徴抽出で大きく異なる点である．

本章では，連続励起振動楽器の特性および本論文での楽譜情報の定義について説明した後，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 (excitation-continuous musical instruments) は，各音符の演奏中に励起源（擦弦楽器あれば弦振動，リード楽器であればリードの振動）を常に制御する楽器のことを指す．連続励起振動楽器の内部分類には諸説あるものの，本稿での分類は擦弦楽器，吹奏楽器，その他歌声などの 3 分類とする（図 2.1）[2]．

擦弦楽器は弓に張られた毛，もしくは棒などで励起源である弦を擦り，その振動を楽器の胴体で増幅させることによって音を出す楽器である．内部分類として，バイオリンなどが属すバイオリン属，コントラバスなどが属すヴィオール属と，その他に二胡などがある．本論文では，オーケストラで用いられるバイオリン属とヴィオール属を対象とする．

吹奏楽器は呼吸などの空気の流れにより管を振動させ，音を出す楽器である．内部分類として，板の振動を励起源とするクラリネットやオーボエ，またハーモニカなどのリード楽器，管内部の

[1] よって，本稿での周波数解析は，逆変換が容易なフーリエ変換によって行う．
[2] 詳細な分類法については，楽器分類学の文献を参照されたい．

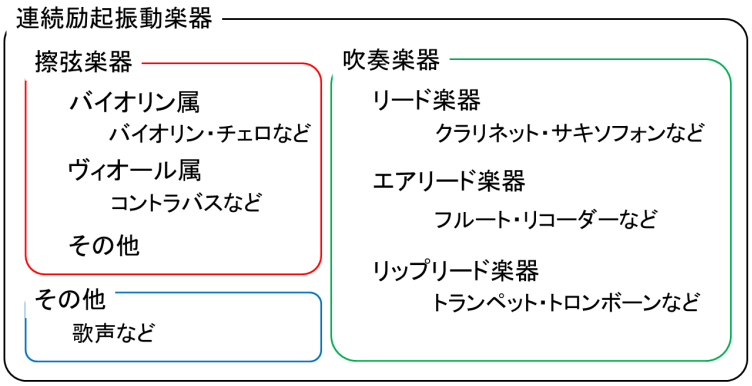

図 2.1: 本研究での連続励起振動楽器の分類.

気流を利用して音を出すフルートなどのエアリード楽器，唇の振動を励起源とするトランペットなどのリップリード楽器がある．

連続励起振動楽器は，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．一方でこの奏者がいつまでも励起源にロードをかける不確定性が，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（Attack），定常区間（Steady/Sustain），減衰区間（Release）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 [5][50]・制御 [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，音響信号のように定量的に表現できる情報のほとんどが損出する．よって，シンボルに含まれる数値情報を定義するのは困難であるが，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（e.g. *sul tast*: 指板の上で弦を擦って, *pizzicato*: 弦をはじいて），および発想記号（e.g. *feroce*: 荒々しく，*dolce*: 甘く）で指定される．特に前者の 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，奏者の解釈による揺らぎが大きく，また定量的に表現することが困難であるため，楽譜上での音色の指定は，楽器の励振機構がニュートラルな状態で振動したものとする．

音高は，任意のチューニング音高からの相対的な音高差が記載されている．本稿で対象とする西洋音楽では，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，

任意のチューニング音高から，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，まだ基準となる指標も存在しない．したがって，音量に関しては定量的に表現できる情報は存在しない．

テンポは，速度記号/標語（e.g. **Allegro**: 快速に）で指定される．前者は 1 分間に拍子上の一拍を何音演奏するかを指定した絶対的な指標である．また後者は感性語ではあるが，速度記号で置き換えられることが多い．テンポの他に，音楽の時間秩序を表す記号として音価（リズム）が存在する．これは，拍子上の一拍を基準に，何倍の長さで演奏するかを指定したものである．これも速度記号が指定されると，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，速度記号より算出される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分析合成は，周波数領域で処理を行うものである．初期の分析では，音色の知覚は，スペクトルの時間平均によって決まる [51] とされていた．しかし，音質が劣化したレコードからも音色が知覚可能 [52] であることや，楽音を逆再生すると自然性が大きく低下する [53] ことから，音色の知覚には，音の時間的な変化が重要であることが分かった．また，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．吹奏楽器の楽音合成でも，音符内の音色変化が重要であることが知られている [54]．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，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，ビ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．

歌唱 F0 の逸脱には，発声器官の物理制約に起因する逸脱成分（オーバーシュートや微細構造など）と，演奏意図に起因する逸脱成分（ビブラートやポルタメント）の 2 種類が存在するといわれている．大石らは，人間の発声機構の物理モデルを線形 2 次系で近似した藤崎モデルを応用し，隠れマルコフモデル（HMM）を利用して 2 種類の逸脱成分を解析する手法を提案した [12]．また中野らは，歌唱力の自動評価の手法として，F_0 軌跡の短時間スペクトルを判別特徴として利用する手法を提案している [13]．

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，別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信号分離問題（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る．解析法として，独奏音では発音時刻検出，混合音では tempogram[20] などが提案されている．解析結果の応用には，自動採譜 [17] や，自動演奏 [18][19] などがある．

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も必要である．

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．楽譜からの逸脱を考えるためには，楽譜と演奏音のアライメントがとられている必要がある．また，連続励起振動楽器音の解析には，各音符の音符内区間推定が必須である．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．

第3章　複素メルKL情報量によるスコアアライメント

演奏音を録音しディジタル信号として扱う場合，演奏は1次元の数値列として記録される．よって，演奏音の楽譜からの逸脱を求めるためには，まず，数値列（音声波形）と楽譜を対応付ける前処理（スコアアライメント）を行う必要がある．ここで楽譜とは音高[1]と音価[2]を指す．

スコアアライメントは，独奏音であれば発音時刻検出（各音符の開始時刻の検出）[23] や基本周波数（F_0）推定 [55][56]，多重音であればビートトラッキング [24] や多重音基本周波数推定で行われる．本研究の解析対象は連続励起振動楽器の独奏音のため，発音時刻検出か F_0 推定を用いることになるが，F_0 は *marcato* 奏法などの発音時に非調波成分が含まれる楽音では正確に求まらないことが多い．そのため，本研究で対象とする，様々な演奏表現や奏法のための楽音解析には不向きである．よって本論文では，スコアアライメントとして発音時刻検出を採用する．

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 [23]．そのため発音時刻検出では，まず楽音の変化を特徴量として抽出し，次に特徴量のピーク点を発音時刻として選択する．特徴量は楽器や奏法の種類で有効なものが異なり，先行研究では特徴量として，位相の変化 [25]，複素スペクトルのユークリッド距離 [26]，振幅スペクトルの KL 情報量 [27] などが提案されている．従来法は，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 *legato* 奏法など，振幅変化の小さい演奏の検出精度は悪い．そこで本章では，連続励起振動楽器に対応できる発音時刻検出法を考える．

3.1　提案法

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．よって発音時刻検出のための音響特徴量には，人の聴覚が受けとる刺激の変化を尺度とするのが妥当である．

人間の聴覚特性については文献 [61] が詳しい．聴覚機構での音分析は，蝸牛基底膜の共振による，対数線形周波数軸上での周波数分析である．この手続きは，対数周波数上でのフーリエ解析とみなすこともできるため，音響特徴量には対数周波数スペクトルを用いるのが妥当である．そこで音響特徴量に，複素スペクトルを音高の知覚的尺度のメル対数周波数から見た，複素メルスペクトルを用いる．

次に変化尺度について考える．音の変化の知覚の手掛かりとなる音色とピッチ[3]の知覚は，基本周波数とその整数倍にパワーを持つ倍音成分の間隔と強度比（調波構造）に起因する．よって変化尺度は，人間が楽音変化を知覚する手がかりとなる，調波構造の変化に鋭感な物が望ましい．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，時刻 t で観測された複素メルスペクトル $\boldsymbol{S}_{\mu,t} = |\boldsymbol{S}_{\mu,t}| \exp(j\phi^{\mathrm{mel}}_{\mu,t}) \in \mathbb{C}^{M\times T}$ と，微小時間 τ-ms 前から予測される時刻 t

[1]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．

[2]“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．

[3]本稿では連続励起振動楽器を扱うため，楽譜に記載できる明確な音高を “ピッチ” と呼ぶ．

の複素メルスペクトル $\hat{\boldsymbol{S}}_{\mu,t} = |\hat{\boldsymbol{S}}_{\mu,t}| \exp(j\hat{\phi}^{\mathrm{mel}}_{\mu,t}) \in \mathbb{C}^{M\times T}$ の KL 情報量として，以下のように定義される．

$$\mathrm{CMKLD}[t] = \sum_{\mu} \left| \boldsymbol{S}_{\mu,t} \log \frac{\boldsymbol{S}_{\mu,t}}{\hat{\boldsymbol{S}}_{\mu,t}} \right| = \sum_{\mu} |\boldsymbol{S}_{\mu,t}| \sqrt{\left(\log \frac{|\boldsymbol{S}_{\mu,t}|}{|\hat{\boldsymbol{S}}_{\mu,t}|} \right)^2 + (\phi^{\mathrm{mel}}_{\mu,t} - \hat{\phi}^{\mathrm{mel}}_{\mu,t})^2} \tag{3.1}$$

ここで μ はメル対数周波数軸を均等に分割した際の周波数ビンである．式 (3.1) より CMKLD は，観測した調波構造に対して大きな重みを与える係数 $|\boldsymbol{S}_{\mu,t}|$ を乗じて，振幅と位相の，予測との乖離を同時に考慮する特徴量である．

CMKLD では，人が知覚する複素スペクトルの変化による聴覚的な “驚き” のモデル化を狙う．KL 情報量は，真の分布と予測した分布の対数尤度の差の期待値である．この演算を振幅スペクトルで考えたとき，正規化振幅スペクトルの KL 情報量は，観測対数振幅スペクトルと，予測対数振幅スペクトルの差を，観測振幅スペクトルの値で重みづけて和を取ったものと考えることが出来る．これを複素拡張すると，式 (3.1) のように，振幅と位相の差を同時に考慮することが可能になる．よって，CMKLD でモデル化される聴覚的な “驚き” は，音色や音高，音量などの様々な音響特徴の予測できない急激な変化である．

本節ではまず複素メルスペクトル $\boldsymbol{S}_{\mu,t}$ と，予測複素メルスペクトル $\hat{\boldsymbol{S}}_{\mu,t}$ の計算手順を示し，次に CMKLD を用いた発音時刻推定法を述べる．

3.1.1 複素メルスペクトルの計算手順

観測スペクトル $\boldsymbol{X}_{\omega,t} = |\boldsymbol{X}_{\omega,t}| \exp(j\psi_{\omega,t}) \in \mathbb{C}^{\Omega\times T}$ から複素メルスペクトルを計算する手順を述べる．まず，線形周波数領域の観測振幅スペクトル $|\boldsymbol{X}_{\omega,t}|$ をメル対数周波数領域に変換し，時刻 t の振幅メルスペクトル $|\boldsymbol{S}_{\mu,t}|$ を求める．

$$\left| \boldsymbol{X}^{\mathrm{mel}}_{1,\ldots,M,t} \right| = \mathrm{mel}\left[|\boldsymbol{X}_{1,\ldots,\Omega,t}| \right] \tag{3.2}$$

$$|\boldsymbol{S}_{\mu,k}| = \frac{\left| \boldsymbol{X}^{\mathrm{mel}}_{\mu,t} \right| + C}{\sum_{\mu} \left| \boldsymbol{X}^{\mathrm{mel}}_{\mu,t} \right| + C} \tag{3.3}$$

ここで $\mathrm{mel}[\cdot]$ は，線形周波数領域のスペクトルをメル周波数軸上で均等になるように各周波数ビンをリサンプリングして，メル対数周波数領域に変換する処理，C は短時間フーリエ変換（STFT）による白色雑音の振幅スペクトルの不確定性を抑える正の定数である．

次に，$|\boldsymbol{S}_{\mu,t}|$ と対応する位相スペクトルを求める．まず線形周波数領域の予測位相スペクトル $\hat{\psi}_{\omega,t}$ を先行研究 [25] の手法で，観測位相スペクトル $\psi_{\omega,t}$ から求める．そして，各位相スペクトルをメル周波数領域に変換する．

$$\phi^{\mathrm{mel}}_{1,\ldots,M,t} = \mathrm{princarg}\left[\mathrm{mel}[\psi_{1,\ldots,\Omega,t}] \right] \tag{3.4}$$

$$\hat{\phi}^{\mathrm{mel}}_{1,\ldots,M,t} = \mathrm{princarg}\left[\mathrm{mel}[\hat{\psi}_{1,\ldots,\Omega,t}] \right] \tag{3.5}$$

すると，各複素メルスペクトルは以下の式で求められる．

$$\boldsymbol{S}_{\mu,t} = |\boldsymbol{S}_{\mu,t}| \exp(j\phi^{\mathrm{mel}}_{\mu,t}) \tag{3.6}$$

$$\hat{\boldsymbol{S}}_{\mu,t} = |\boldsymbol{S}_{\mu,t-\tau}| \exp(j\hat{\phi}^{\mathrm{mel}}_{\mu,t}) \tag{3.7}$$

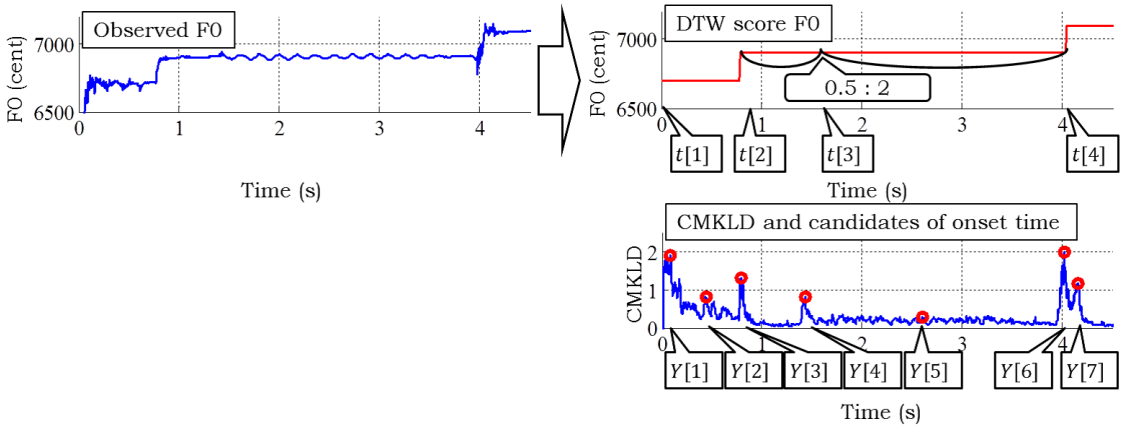

図 3.1: 音価 $= (0.5, 0.5, 2, 0.5, ...)$，音高 $= (64, 69, 69, 71, ...)$ の発音時刻選択の例．ここで x 軸は時間 (秒) を示す．

3.1.2 複素メル KL 情報量による発音時刻検出

CMKLD の問題点として，式 (3.1) の平方根内の第二項 $(\phi^{\mathrm{mel}}_{\mu,t} - \hat{\phi}^{\mathrm{mel}}_{\mu,t})$ が，位相の周期性により，原点の選択に依存する点がある．すなわち，$\phi^{\mathrm{mel}}_{\mu,t} = \pi - \epsilon$，$\hat{\phi}^{\mathrm{mel}}_{\mu,t} = -\pi + \epsilon$，$0 < \epsilon \ll \pi$ であるとき，極座標系での偏角の距離は 2ϵ であるが，式 (3.1) では $2\pi - 2\epsilon$ である．そこで実際の計算時には，絶対値が π 以上の位相を，2π の補数を用いて範囲 $[-\pi, \pi]$ に変換する関数 princarg$[\cdot]$ を用いて，$\Phi_{\mu,t} = \mathrm{princarg}[\phi^{\mathrm{mel}}_{\mu,k} - \hat{\phi}^{\mathrm{mel}}_{\mu,t}]$ とし，CMKLD を以下の近似式で求める．

$$\mathcal{D}[t] \approx \sum_{\mu} |\boldsymbol{S}_{\mu,t}| \sqrt{\left(\log \frac{|\boldsymbol{S}_{\mu,t}|}{|\hat{\boldsymbol{S}}_{\mu,t}|}\right)^2 + \Phi^2_{\mu,t}} \tag{3.8}$$

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一様ではないため，閾値を動的に変化させる必要がある．動的閾値は先行研究を拡張し以下のように求める．

$$\mathcal{D}_{th}[t] = \lambda \mathrm{Median}(\boldsymbol{d}_t) + \frac{\mathrm{Median}(\mathcal{D})}{2} \tag{3.9}$$

ただし係数 λ は，初期値 ξ から始め，$|\mathcal{Y}| \geq N$ 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 $\boldsymbol{d}_t$ は以下のように求める．

$$\boldsymbol{d}_t = \left(\mathcal{D}\left[t - \frac{\mathcal{T}}{2}\right], \mathcal{D}\left[t - \frac{\mathcal{T}}{2} + t_s\right], \cdots, \mathcal{D}\left[t + \frac{\mathcal{T}}{2}\right]\right) \tag{3.10}$$

ここで t_s は離散時刻 t の刻み幅である．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，F_0 軌跡を推定する．次に，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．例として図 3.1 では，$z[3]$ はノート高が変化しないため検出されない．そこで $z[2]$ と

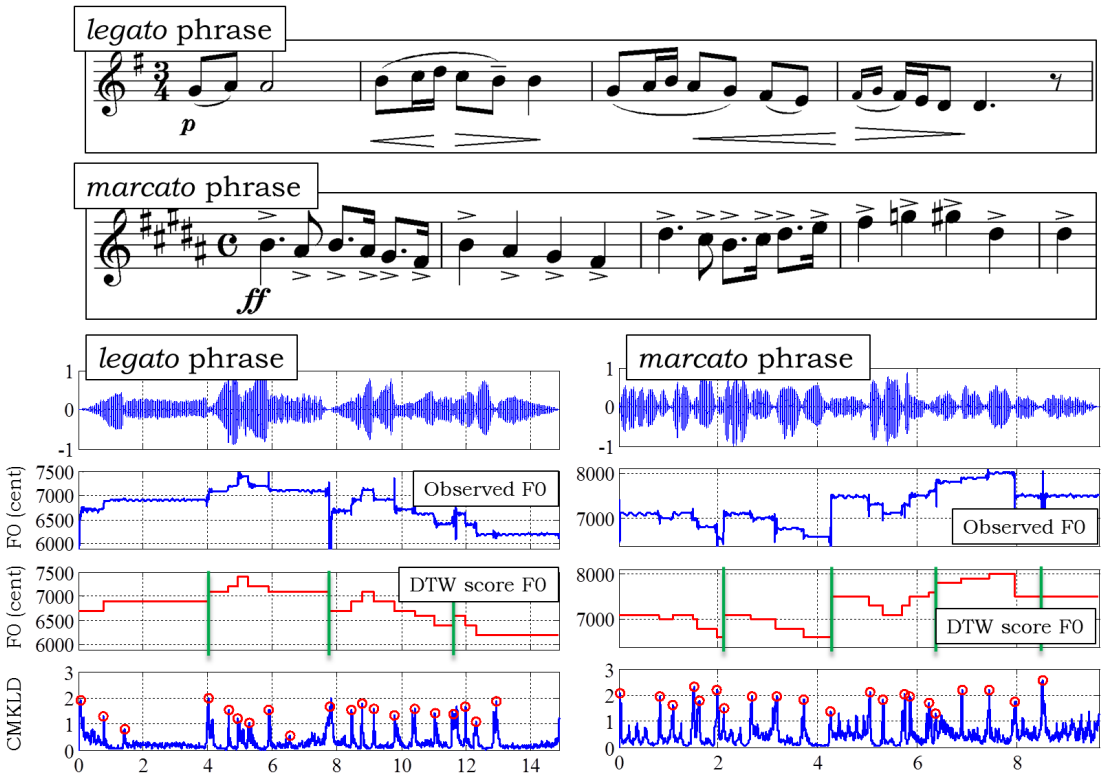

図 3.2: *legato* 奏法（左）と *marcato* 奏法（右）からの発音時刻検出結果例．図は上から，時間波形，観測 F_0 軌跡，アライメントされたスコア F_0 軌跡，CMKLD と発音時刻検出結果（赤丸）を示す．また，アライメントされたスコア F_0 軌跡中の緑の縦線は，小節線を表す．

$z[4]$ を音価を用いて 0.5 : 2 で分割し $z[3]$ を決定する．最後に，候補集合 $\mathcal{O}$ の中から，以下の式で定義される CMKLD 重み付距離が最小の候補 $O[i]$ を選択し，n 音目の発音時刻 $o[n]$ とする．

$$o[n] = \arg\min_{O[i]\in\mathcal{O}} \frac{|O[i] - z[n]|}{\mathcal{D}\left[O[i]\right]} \tag{3.11}$$

図 3.2 に *legato* と *marcato* 奏法によるバイオリン演奏音からの発音時刻検出結果例を示す．従来，検出が困難とされていた *legato* 奏法の演奏からも，精度よく発音時刻が検出出来ていることが確認できる．また，*marcato* 奏法からも精度よく検出が出来ていることも確認できる．

3.2 精度評価実験

提案法の精度評価，および雑音・残響への耐性実験を行う．実験には，付録 A のデータを用いた．提案法の各パラメータは，STFT の窓長を 10ms，シフト幅を $t_s = 5$ms をとした．発音・消音時刻検出の各パラメータは，τ =10 ms，$\mathcal{T}$ =100 ms，$C = 0.2$，式 (3.9) の変数 λ の初期値は $\xi = 1$ とし，刻み幅は $\Delta\xi = 0.1$ とした．

3.2.1 発音時刻検出の精度評価

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，検出結果の適合率で比較した．提案法は規定の音符数の発音時刻を検出するため，再現率・調和

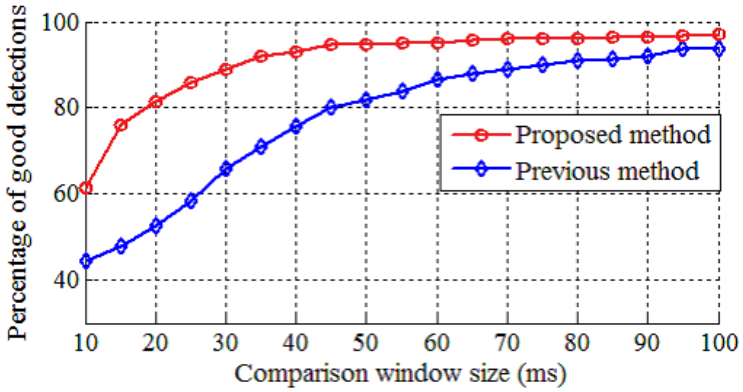

図 3.3: 発音時刻検出の適合率.

平均の評価は行わない．従来法の検出閾値と各パラメータは，評価実験と同様である．従来法は，アライメントを用いないが，比較のために，候補集合からの選択で発音時刻を検出した．

図 3.3 に，提案法と従来法の適合率を示す．全ての窓幅で，提案法の適合率が上回った．また，発音時刻検出の精度比較で多く用いられる 50 ms の時間窓 [23] [4]では，エラー率が 63.2%減少した．さらに，評価実験で用いたフレーズ数は 10 であるため，明確に結論付けることは困難であるものの，提案法は従来法と比べ，*legato* などの音符の移り変わり時の振幅変化が滑らかなフレーズでも，精度の低下が小さい傾向が見られた．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 は人間の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．

テンポ検出や楽音合成のための発音時刻検出では，検出誤差などがテンポ推定に悪影響を及ぼすため，音響イベント検出のための発音時刻検出で求められる “再現率” よりも，短い時間窓での “適合率” が重要となる．図 3.3 から提案法は，短い時間窓での適合率が大幅に向上する．一方で提案法は，スペクトルのメルスケール変換を行うため，従来法より実行速度が遅く，1 分を超える音響信号を大量に処理する音響イベント検出には不向きである．そのため提案法は，従来法のような汎用的な発音時刻検出ではなく，スコアアライメントに特化した発音時刻検出法である．このことから提案法は，演奏表現解析を対象にしたスコアアライメントのため，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．

3.2.2 雑音・残響耐性実験

提案法を適用可能な環境を示すために，残響および背景雑音を付与したデータで評価した．残響は，評価データにインパルス応答を畳み込み付与した．インパルス応答には，室内の残響として “Aachen Impulse Response database[57]” より音楽スタジオ（RT=0.11sec）と講義室（RT=0.7sec），コンサートホールの残響として “The open acoustic impulse response library [5] ” より “St Andrew’s Church”（RT=1.45sec）を用いた．雑音は，SN 比がそれぞれ 20，30，40dB となるようにホワイトノイズを付与した．評価尺度は，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 [23]．この時間窓は，正解データアノテーションのヒューマンエラーを吸収する意味を持つ．

[4] この時間窓は，正解データアノテーションのヒューマンエラーを吸収する意味を持つ．
[5] http://www.openairlib.net/

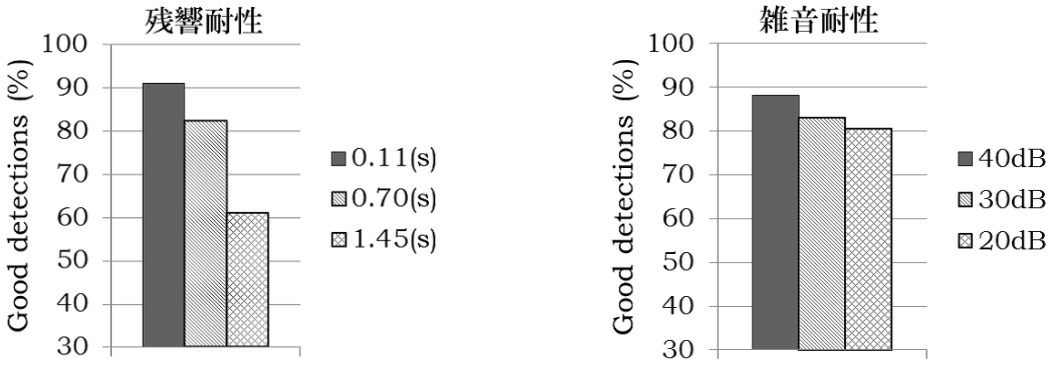

図 3.4: 雑音・残響耐性実験結果

図 3.4 に実験結果を示す．残響時間に反比例して精度が急激に低下している．CMKLD はスペクトルの時間変化量の尺度であり，残響によりスペクトルが時間方向に平滑化されると精度が低下する．一方，雑音下での検出精度は SNR に比例するが，残響に比べ影響が小さい．これは，CMKLD が時間定常な雑音に対しては頑健であるためである．さらに，音楽演奏をレコーディングする環境は高 SNR かつ定常雑音であることが多い．よって今後，残響除去を組み込むことにより，提案法の適用可能な環境の拡大を考えなくてはならない．

3.3 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（CMKLD）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．雑音・残響耐性実験では，残響時間と比例して検出精度が低下することを示した．このことから提案法は，クリーンな環境で録音された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．一方で演奏録音に残響が含まれている場合，発音時刻の検出精度が大きく低下するため，今後，残響除去を組み込むことにより，提案法の適用可能な環境の拡大を考えなくてはならない．

3.4 関連研究

発音時刻検出は，音楽情報処理のトップカンファレンス ISMIR（The International Society for Music Information Retrieval）で行われる，音楽情報検索の要素技術コンテスト MIREX（Music Information Retrieval Evaluation eXchange）の主要タスクの一つである．MIREX で発音時刻検出は 2005 年から今年まで，2008 年を除き毎年行われている．MIREX での発音時刻検出の目的はスコアアライメントではなく，音響信号中に含まれる音響イベントの検出である．そのため，多重音，撥/打弦楽器，連続励起振動楽器の全てで有効な検出手法を目指している．また，情報検索への応用も考え，実行速度にも規定が定められている．

連続励起振動楽器の発音時刻検出には位相の変化が有効である [23] ことが知られているが，多重音からの発音時刻検出では，あらゆる楽器の位相情報が混在するため，位相が予測できず有効に機能しない．よって MIREX で提案される手法の多くは，振幅スペクトログラムの変化に着目している．これらの手法では，前後フレームの振幅スペクトルの変化を，何らかの距離尺度で評価する．2013 年 11 月に行われた MIREX で最も高精度であった “畳み込みニューラルネット

[58]” は，変化検出に画像処理のテクニックを応用した手法である．スペクトログラムを画像とみなしてエッジ検出し，発音時刻をニューラルネットワークで検出する．

本研究とほぼ同時に独立して行われた擦弦楽器に焦点を絞った手法に，位相変化を群遅延でとらえるものがある [59]．また，位相を特徴量に用いる場合，ビブラートやトレモロの影響で検出精度が悪化するが，この手法では同時にこれらを抑圧する処理を行う．

また発音時刻検出は，時系列の変化点を検出するという点で，統計的時系列解析の “Change point detection（CPD）[60]” と関連が深い．CPD では時系列をフレーム分割し，そのフレームの時系列を生成した確率密度関数を推定する．そして，フレームごとの確率密度関数を，KL 情報量や f-ダイバージェンスで比較する．

提案法は，スペクトルを複素領域の確率密度関数ととらえ[6]，位相情報を考慮しつつ前後フレームの変化を KL 情報量で比較するため，これらの手法のハイブリッドとみなすこともできる．

[6]ただし，複素メルスペクトルは正規化されていない（周波数積分の結果が 1 にならない）ため確率密度関数ではない．

第4章　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

奏者認識 [4] などで用いられる音符内区間推定法 [7] は，まず音量軌跡を事前に決定した数の直線で近似し，各直線の傾きを元に区間推定する．これは音量軌跡を，自由度を固定したモデルでフィッティングすることに相当するため，軌跡がビブラートなどに起因して複雑に変化[1]した場合，推定精度が低下する問題があった．

そこで本章では，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．音響信号からの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．提案法では，区間ごとの音響特徴量の変化を，音響特徴量出力分布の基底測度の遷移によって表現する．

4.1　音符内区間ごとの音響特性

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．音響信号からの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処理である．バイオリンの音符内区間ごと音響特性例を図 4.1 に示す．

発音区間は，発音時刻から励振機構が安定振動状態とな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 4.1 (b))．定常区間は，励振機構が安定して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，その影響によって振幅や音色も変化する．減衰区間は，連続励起振動楽器では，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 4.1 (a))．

4.2　音符内区間を考慮した楽音の生成過程

本章では，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．また，$\mathcal{N}$ を正規分布，$\mathcal{W}$ をウィシャート分布，$\mathcal{D}$ をディリクレ分布，Ber をベルヌーイ分布，Bin を二項分布とする．

4.2.1　音符内区間推定のための音響特徴量

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．

[1] 振幅の変化点が多く，直線数本では近似が困難なもの．

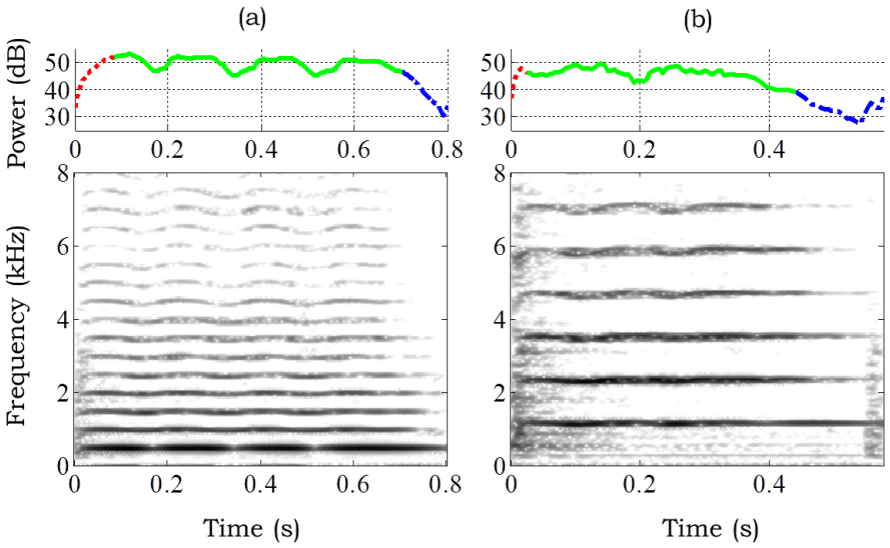

図 4.1: 音符内区間による音響特性の変化例 (上: 音量軌跡，下: スペクトログラム)．バイオリンの通常の音量で演奏した音 (a) と強く演奏した音 (b)．点線が発音区間，実線が定常区間，破線が減衰区間を示す．

音量は，増加や減衰など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 ($\boldsymbol{c}_t = (c_t^1, c_t^2, c_t^3)^\dagger$) を特徴量として用いる．

以上より，本研究では $\boldsymbol{y}_t = (\Delta x_t, c_t^1, c_t^2, c_t^3)^\dagger$ を音響特徴量として用いる．ただし $\dagger$ は転置を意味する．

4.2.2 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．この変動は，奏者の演奏解釈などに基づき生成される．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させるべきである．

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 $\boldsymbol{y}_t$ に階層的な生成過程を考える．

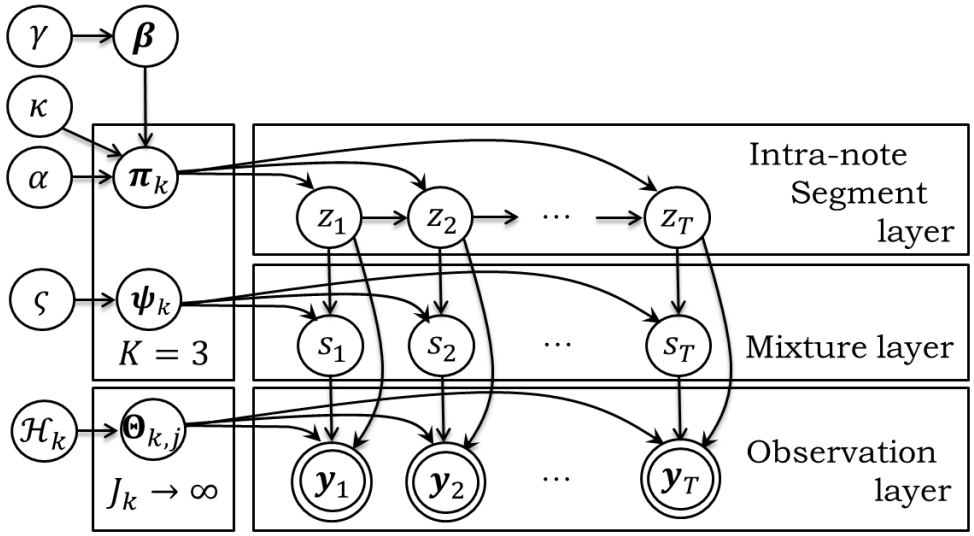

図 4.2: 提案法のグラフィカル表現

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 $z_t \sim \boldsymbol{\pi}_{z_{t-1}}$ で表現する．また図 4.1 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，スティッキー HMM[30] を用いる．

$$\boldsymbol{\pi}_k \sim \mathcal{D}\left(\alpha\boldsymbol{\beta} + \kappa\boldsymbol{\delta}_k\right), \quad \boldsymbol{\beta} \sim \mathcal{D}\left(\gamma/K, \gamma/K, \gamma/K\right) \tag{4.1}$$

ここで $\kappa \geq 0$ は自己遷移確率を高めるバイアスパラメータであり，$\alpha, \gamma > 0$ はハイパーパラメータである．

次に，各状態に対応した音響特徴量の生成を考える．前述の通り，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 $\boldsymbol{y}_t$ を，分布パラメータ $\Theta_{k,j} = \{\boldsymbol{\mu}_{k,j}, \boldsymbol{\Lambda}_{k,j}\}$ 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] により生成されたものとみなす．

$$\boldsymbol{y}_t \sim \mathcal{N}\left(\hat{\boldsymbol{\mu}}_{z_t,s_t}, \hat{\boldsymbol{\Lambda}}^{-1}_{z_t,s_t}\right), \quad s_t \sim \boldsymbol{\psi}_{z_t} \tag{4.2}$$

ここで $\boldsymbol{\psi}_k$ 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，パラメータ $\mathcal{H}_k = \{\boldsymbol{\lambda}_k, R_k, \boldsymbol{W}_k, \nu_k\}$ の正規–ウィシャート分布とする．

すなわち提案法は，音符内状態の遷移を，音響特徴量出力分布の基底測度の遷移によって表現する．提案法のグラフィカル表現を図 4.2 に示す．

4.3 状態推定アルゴリズム

本章では状態系列 $z_{1,\ldots,T}, s_{1,\ldots,T}$ の推定法を説明する．まず，観測信号を音符ごとに分割し，その後各音符ごとに推論を行う．本稿ではマルコフ連鎖モンテカルロ法の一種である Gibbs Sampler

で推論を行う．基本的な推論アルゴリズムについては文献 [30] と同様であるため，詳細な議論および導出は省略し，アルゴリズムおよび各更新式のみを説明する．

4.3.1 パラメータのギブスサンプリング

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．サンプリングは，$z_t, s_t, \boldsymbol{\beta}, \alpha, \kappa, \varsigma, \mathcal{H}_k$ の順に行う．

Step 1: z_t と s_t のサンプリング

$$z_t \sim \sum_{k=1}^{K} f_k(\boldsymbol{y}_t)\delta(z_t, k) \tag{4.3}$$

$$s_t \sim \sum_{j=1}^{J} f'_{z_t,j}(\boldsymbol{y}_t)\delta(s_t, j) + f'_{z_t,J_k+1}(\boldsymbol{y}_t)\delta(s_t, J_k+1) \tag{4.4}$$

ここで

$$f_k(\boldsymbol{y}_t) = \left(\alpha\beta_k + n^-_{z_{t-1},k}\right)\left(\frac{\alpha\beta_{z+1} + n^-_{k,z_{t+1}} + \kappa\delta(k, z_{t+1})}{\alpha + n^-_{k,\cdot} + \kappa}\right)\sum_{j=1}^{J_k}\mathcal{N}(\boldsymbol{y}_t|\hat{\boldsymbol{\mu}}_{k,j}, \hat{\boldsymbol{\Lambda}}^{-1}_{k,j}) \tag{4.5}$$

$$f'_{z_t,j}(\boldsymbol{y}_t) = \left(\frac{m^-_{z_t,j}}{\varsigma + m^-_{z_t,\cdot}}\mathcal{N}(\boldsymbol{y}_t|\hat{\boldsymbol{\mu}}_{z_t,j}, \hat{\boldsymbol{\Lambda}}^{-1}_{z_t,j})\right) \tag{4.6}$$

$$f'_{z_t,J_{z_t}+1}(\boldsymbol{y}_t) = \left(\frac{\varsigma}{\varsigma + m^-_{z_t,\cdot}}\mathcal{N}(\boldsymbol{y}_t|\hat{\boldsymbol{\mu}}_{z_t,J_{z_t}+1}, \hat{\boldsymbol{\Lambda}}^{-1}_{z_t,J_{z_t}+1})\right) \tag{4.7}$$

であり，$n_{k,k'}$ は状態 k から k' への遷移回数，$m_{k,j}$ は状態 k で j 番目の正規分布がアクティブになった回数，上付き文字の $-$ は $\boldsymbol{y}_t$ に関する情報を取り除くことを意味する．また，$\cdot$ はその変数に関する総和を意味し，$\delta(i,j)$ はクロネッカーのデルタである．また，$\hat{\boldsymbol{\mu}}_{z_t,j}$ と $\hat{\boldsymbol{\Lambda}}_{z_t,j}$ は以下の式に従いサンプリングする．

$$\hat{\boldsymbol{\mu}}_{z_t,j} \sim \mathcal{N}\left(\left(m^-_{z_t,j}\hat{\boldsymbol{\Lambda}}_{z_t,j} + R_{z_t}\right)^{-1}\left(\hat{\boldsymbol{\Lambda}}_{z_t,j}\bar{\boldsymbol{y}}^-_{z_t,j} + R_{z_t}\boldsymbol{\lambda}_{z_t}\right), \left(m^-_{z_t,j}\hat{\boldsymbol{\Lambda}}_{z_t,j} + R_{z_t}\right)^{-1}\right) \tag{4.8}$$

$$\hat{\boldsymbol{\Lambda}}_{z_t,j} \sim \mathcal{W}\left(\left(\nu_{z_t}\boldsymbol{W}_{z_t} + \Phi^-_{z_t,j}\right)^{-1}, \nu_{z_t} + m^-_{z_t,j}\right) \tag{4.9}$$

$$\bar{\boldsymbol{y}}_{k,j} = \sum_{t' \in (z_t=k, s_t=j)} \boldsymbol{y}_{t'} \tag{4.10}$$

$$\Phi_{k,j} = \sum_{t' \in (z_t=k, s_t=j)} (\boldsymbol{y}_{t'} - \hat{\boldsymbol{\mu}}_{z_t,j})(\boldsymbol{y}_{t'} - \hat{\boldsymbol{\mu}}_{z_t,j})^\dagger \tag{4.11}$$

全ての $t \in 1, ..., T$ についてサンプリングが終了した後，$m_{z_t,j} = 0$ となるような j が存在する場合，その j を消去する．

Step 2: $\boldsymbol{\beta}$ のサンプリング

スティッキー HDP-HMM の自己遷移バイアスを表現するために，$\boldsymbol{\beta}$ は，補助乱数 $\boldsymbol{q}, \boldsymbol{r}, \bar{\boldsymbol{q}}$ を用いてサンプリングする．

$$q_{k,k'} = \sum_{i=1}^{n_{k,k'}} u_i, \;\; u_i \sim \mathrm{Ber}\left(\frac{\alpha\beta_{k'} + \kappa\delta(k,k')}{i + \alpha\beta_{k'} + \kappa\delta(k,k')}\right) \tag{4.12}$$

$$r_k \sim \mathrm{Bin}\left(q_{k,k}, \frac{\rho}{\rho + \beta_k(1-\rho)}\right) \tag{4.13}$$

$$\bar{q}_{k,k'} = \begin{cases} q_{k,k'} & (k \neq k') \\ q_{k,k'} - r_k & (k = k') \end{cases} \tag{4.14}$$

$$\boldsymbol{\beta} \sim \mathcal{D}\left(\bar{q}_{\cdot,1}, \bar{q}_{\cdot,2}, ..., \bar{q}_{\cdot,K}\right) \tag{4.15}$$

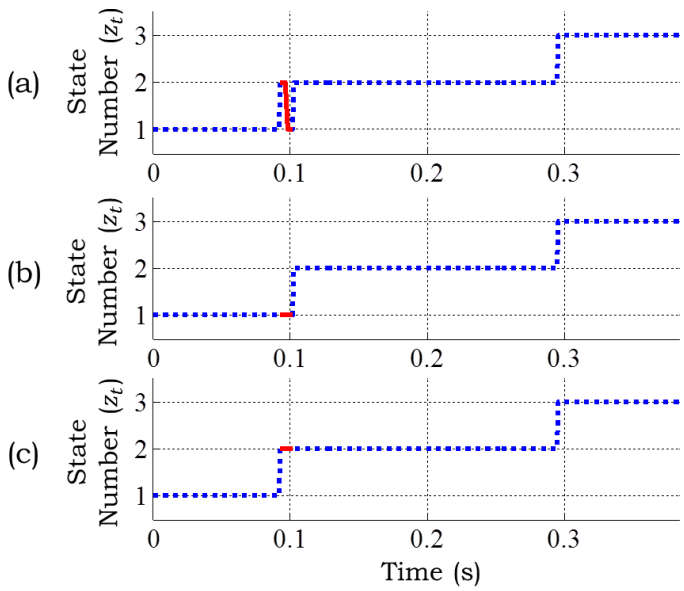

図 4.3: z_t の後処理例．ギブスサンプリングにより推定された音符内区間の遷移 (a) と修正パターン 1(b) および修正パターン 2(c).

ただし $\rho = \kappa/(\alpha + \kappa)$ である．

Step 3: ハイパーパラメータ $\alpha, \kappa, \varsigma, \mathcal{H}_k$ のサンプリング

$\alpha, \kappa, \varsigma$ については冗長となるのでここでは割愛するが，文献 [30] と同様に補助乱数を用いてサンプリングする．$\mathcal{H}_k$ は $\boldsymbol{y}_{t\in(z_t=k)}$ のデータを用いて事後分布を求め，サンプリングする [34]．指定した反復回数を満たせば更新を終了し，満たさなければ **Step 1** に戻る．

4.3.2 z_t の後処理

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 のオートマトンである．しかし，スティッキー HMM は ergodic な HMM であるため，“発音 → 定常 → 発音” などの “逆戻り” が推定されることもある (図 4.3 (a))．このような場合，本稿では後処理により z_t を修正する．

逆戻りを含む時間区間 $\tau \in \{t_1, ..., t_2\}$ で修正可能なパターン $\hat{z}^p_\tau$ が P 種類考えられるとする (e.g. 図 4.3 では, $P = 2$ であり，$\hat{z}^1_\tau$ =(b), $\hat{z}^2_\tau$ =(c) である)．HMM のパラメータ $\Upsilon = \{\boldsymbol{\pi}_k, \boldsymbol{\psi}_k, \Theta_k\}$ 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．

$$p(\hat{z}^p_\tau, \boldsymbol{y}_\tau|\Upsilon) = \prod_{\tau=t_1}^{t_2+1} \pi_{z^p_{\tau-1}, z^p_\tau} \sum_{j=1}^{J_{z^p_\tau}} \psi_{z^p_\tau, j} \mathcal{N}(\boldsymbol{y}_\tau|\boldsymbol{\mu}_{z^p_\tau, j}, \boldsymbol{\Lambda}^{-1}_{z^p_\tau, j}) \tag{4.16}$$

本稿では，式 (4.16) を最大とする $\hat{z}^p_\tau$ を用いて，z_t を修正する．

提案法の動作例（バイオリン音: 468Hz) を図 4.4 に示す．提案法による推定結果は，正解データと若干のずれはあるものの，20msec 程度の誤差で推定できている．さらに，各区間での演奏表現による振幅や音色の変化を ($\sum_K J_k$ =)11 個の状態で表現している．

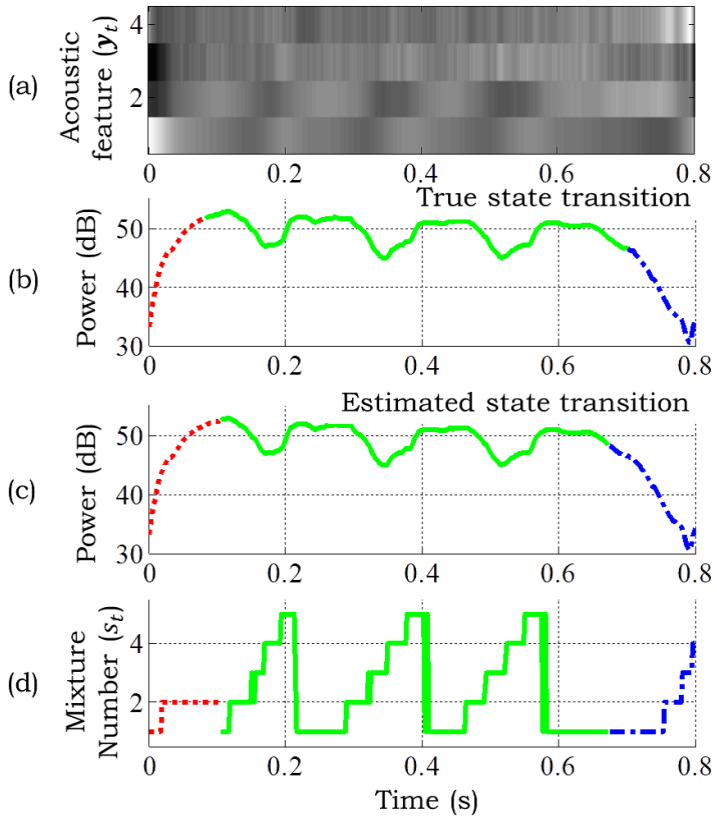

図 4.4: 音符内状態推定の結果例．音響特徴系列 $\boldsymbol{y}_{1,\ldots,T}$(a)，正解データ (b)，推定結果 $z_{1,\ldots,T}$(c)，正規分布のインジケータ $s_{1,\ldots,T}$(d)．点線が発音区間，実線が定常区間，破線が減衰区間を示す．

4.4 評価実験

4.4.1 実験条件

実験には，付録 A のデータを用いた．ただし後述する正解を判別する時間窓の関係から，持続時間が 200-ms 以下の音符は評価の対象外とした．

STFT の切り出し長は 20-msec，シフト幅は 1-msec，STFT 長は 1024 点とした．文献 [30] で用いられている $\alpha, \kappa, \varsigma$ のハイパーパラメータは $a, b, c, d = 1$ とした．正規分布のインジケータ s_t の初期値は，$J_k = 30$ として乱数を用いて決定した．状態 z_t の初期値は，収束性を高めるため，$z_{1,\ldots,T/4} = 1$，$z_{T/4+1,\ldots,3T/4} = 2$，$z_{3T/4+1,\ldots,T} = 3$ とした．推論の繰り返し回数は，各音符につき 1000 回とした．

4.4.2 精度評価実験

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．音符内状態推定は，1 つの音符を 3 つの区間へ分割する処理のため，検出すべき時刻は発音区間から定常区間への切り替わり時刻 (A–to–S) と，定常区間から減衰区間への切り替わり時刻 (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．この時間窓は，正解データアノテーションのヒューマンエラーを吸収する意味を持つ．発音時刻と消音時刻はハンドラベリングしたものを用いた．

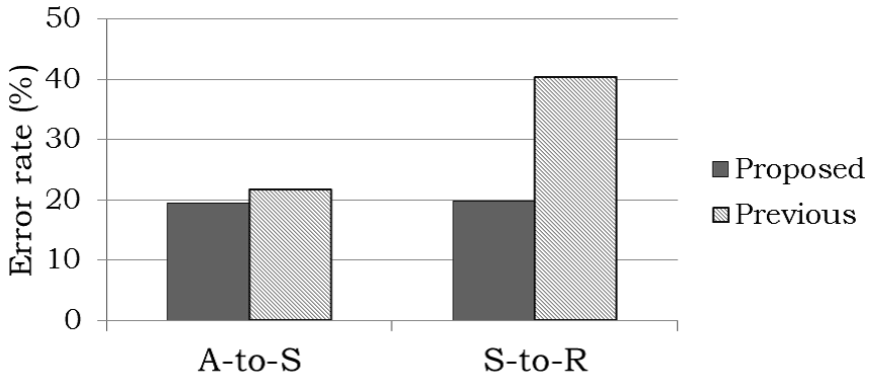

図 4.5: 実験結果

図 4.5 にセグメンテーションのエラー率を示す．提案法のエラー率は，従来法より，A–to–S が 10.6%，S–to–R が 51.2 %減少した．また，二群の比率の差の検定を行ったところ，S–to–R に有意水準 $\alpha = 0.01$ で有意差が認められた．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として，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と考えられる．擦弦楽器では定常区間で，弦を擦りながら音量を減衰させる奏法があるが，この音量減衰は，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，スティッキー HMM の遷移行列に自己遷移確率を高めるバイアスがかかるためである．スティッキー HMM は，ある状態に留まる時間が短い場合，その状態が消滅するように推論が働く．よって *sttacato* などの発音区間が非常に短い音符は，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，という特性を利用する改良方法を検討する．

4.5 まとめ

本稿では，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，従来法より，A–to–S が 10.6%，S–to–R が 51.2 %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．

本稿では，HDP-HMM のエルゴード性による状態の逆戻りを後処理によって解決したが，音符内状態遷移行列 $\boldsymbol{\pi}_k$ を上三角行列に制限することで，この後処理は不要となる．さらにこの制約は，各状態が持つ音響特徴量の出力分布の推定精度も向上することができると考えられる．よって今後，音符内状態遷移行列を上三角行列に制限した場合の更新式の導出を行う．

4.6 関連研究

従来の音符内状態推定は，事前に設定した自由度固定の楽音生成モデルを観測にフィッティングさせ，音符内区間を推定する．よって生成モデルが，連続励起振動楽器の楽音変動や定常区間の伸縮を考慮できず推定精度が低下する．提案法は音符内状態を，振幅スペクトルを生成する確

率分布の基底測度とみなすモデルを立て，観測をフィッティングさせた．この際，生成モデルの自由度を観測データの複雑さに応じて決定させるため，モデルがアンダー/オーバーフィッティングしづらくなる．

事前に自由度を設定しない楽音生成モデルという観点で提案法は，中野らの無限状態スペクトルモデル [63] と関連が深い．無限状態スペクトルモデルは，音源分離や自動採譜を目的とした楽音生成モデルであり，各楽器の１つの音符のスペクトルが，数種類の振幅スペクトルテンプレートが時間的に遷移することにより生成されたとみなすモデルである．この時間遷移のスペクトルインジケータを HDP-HMM で表現する．

第5章　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

音符内の音量の時間的変化は，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 種類の変動に起因する．前者は旋律のフレーズ感，後者はキャラクター性などに関連する．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して解析を行う必要があるが，これは非常に自由度の高い不良設定問題であり，未解決の問題であった．

本章では，音量の動特性に含まれる演奏表現や演奏技術の情報を抽出/解析するために，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離する手法を提案する．2 成分を別個に扱うことにより，例えばアーティキュレーションを別の奏者と入れ替る，ダイナミクスレンジを広げるなどの個別の操作や，ダイナミクスを手描きで修正するなどの，MIDI のような直観的な楽音操作も可能になる．熟達度評価では，音楽構造に沿ったフレージング（ダイナミクス変動）が出来ているかや，“音の粒（アーティキュレーション）がそろっているか” などの観点で評価を行うこともできるようになる．

次節で述べる，ダイナミクスのゆるやかに変化する特性と，アーティキュレーションの類似した局所変動が繰り返す点に着目し，音量軌跡の生成過程を階層ディリクレ過程遷移型線形動的システム [33] (HDP-SLDS: Hierarchical Dirichlet processes switching Linear dynamical system) を用いて表現し，軌跡を分離する．ただし，本手法は音量軌跡を分離することにのみ注力し，楽譜音符列との対応付けは行わない．

5.1　連続励起振動楽器の音量軌跡

本稿では，音量のベースラインの変化をダイナミクス，音符ごとの音量の上下をアーティキュレーションと呼ぶ．

図 5.1 は，2 人のバイオリンプロ奏者による A. Vivaldi の “四季” より “春” の 1 楽章冒頭のフレーズの演奏の音量軌跡である．この楽曲は前半 3 小節はフォルテ，後半 3 小節はピアノで，音高と音価がほぼ同一のフレーズを演奏する．

図 5.1 からは，以下の 3 点が読み取れる．パターン 1 は，音価や音高が同一で音量記号が異なる箇所である．音量軌跡は，平均的な音量は前半は 40dB，後半は 30dB と異なっているが，局所的な上下の起伏は類似している．パターン 2 の，奏者および音量記号が同一で，音価と音高もほぼ一致している箇所では，音量軌跡も類似した変動を示している．パターン 3 では，同一の楽譜を用いても，奏者が異なる場合，音量軌跡も異なることを示している．

以上より，連続励起振動楽器の音量軌跡に，以下の 3 点の仮定を置く．まずパターン 1 より，ダイナミクスとアーティキュレーションに対数音量領域 (dB) での，加法性と独立性を仮定する．次にパターン 2 より，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．そしてパターン 3 より，奏者

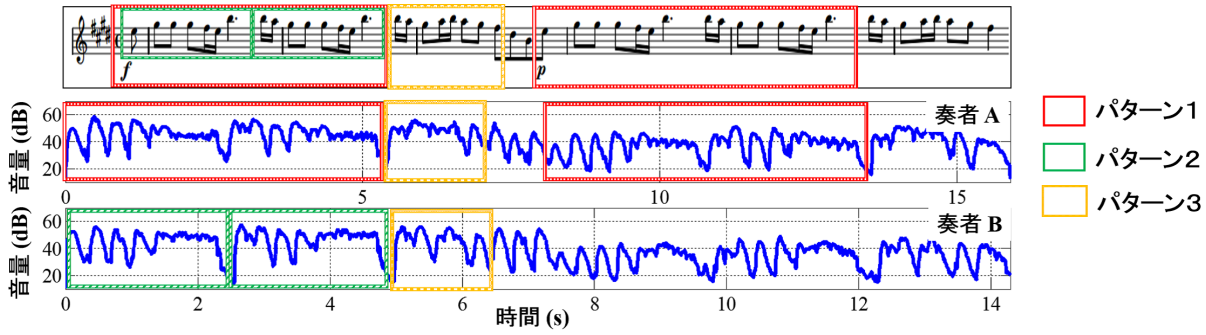

図 5.1: バイオリンの音量軌跡例．楽曲は A. Vivaldi の “四季” より “春” の 1 楽章冒頭．

の演奏解釈や奏法の違いによって，音量軌跡が変化することを仮定する．以上のことを踏まえ次章では，音量軌跡の動特性を数理的に表現する方法を考える．

5.2 音量軌跡の生成モデル

5.2.1 音量軌跡の線形動的システム表現

観測演奏音を短時間フーリエ変換（STFT）して得られるスペクトログラムを $\boldsymbol{X}_{\omega,t} \in \mathbb{C}^{\Omega \times T}$，時刻 t での音量を y_t，アーティキュレーションを f_t，ダイナミクスを g_t とし，それぞれに以下の関係が成り立つと仮定する．

$$y_t = 20 \log_{10} \sum_{\Omega} |\boldsymbol{X}_{\omega,t}| = f_t + g_t \tag{5.1}$$

ここで t, ω はそれぞれ時間と周波数のインデックスであり，y_t，f_t，g_t の単位は dB である．

アーティキュレーションの変化は擦弦楽器であれば弦をこする強さや速さ，吹奏楽器であれば息を吹き込む強さなどによって制御される．さらに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．

$$f_t = \sum_{i=1}^{r} a_i^{z_t} f_{t-i} + e_t^f(z_t), \quad e_t^f(z_t) \sim \mathcal{N}(0, \sigma_f^2(z_t)) \tag{5.2}$$

ただし $z_t \in \{1, 2, ..., K\}$ 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．

ここで奏法プリミティブの総数 K について考える．奏法プリミティブは，楽器制御の物理パラメータに対応するため，その実際のパラメータは実数であり，K は非可算無限である．よって，式 (5.2) は厳密には成立しない．ここで計算の簡単のために，極めて類似した奏法の変化を一つの奏法として扱い，奏法プリミティブの可算無限個へのクラスタリングを考える．この近似により，z_t をカテゴリー変数としてみなすことができる．

また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階層ディリクレ過程隠れマルコフモデル（HDP-HMM）[30] を適用する．

$$z_t \sim \boldsymbol{\pi}_{z_{t-1}}, \quad \boldsymbol{\pi}_k \sim \mathrm{DP}\left(\alpha+\kappa, \frac{\alpha\boldsymbol{\beta}+\kappa\delta_j}{\alpha+\kappa}\right) \tag{5.3}$$

$$\beta_k = \nu_k \prod_{l=1}^{k-1}(1-\nu_l), \quad \nu_k \sim \mathrm{Beta}(1,\gamma) \tag{5.4}$$

ダイナミクスの変動はフレーズ感などに関係し，$sub.\boldsymbol{p}$（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，$sub.\boldsymbol{p}$ 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．

$$g_t = g_{t-1} + e_t^g, \ \ e_t^g \sim \mathcal{N}(0, \sigma_g^2) \tag{5.5}$$

よって，式 (5.1)(5.2)(5.5) より時刻 t での音量 y_t は，パラメータ $\Theta_k = \{\boldsymbol{A}^k, \sigma_f^2(k), \sigma_g^2\}$ によって制御される HDP-SLDS として記述できる．

$$\boldsymbol{x}_t = \begin{bmatrix} \boldsymbol{A}^{z_t} & \\ & 1 \end{bmatrix} \boldsymbol{x}_{t-1} + \boldsymbol{e}_t^{z_t}, \ \ y_t = \boldsymbol{U}\boldsymbol{x}_t \tag{5.6}$$

ただし，$\boldsymbol{x}_t = (f_t, f_{t-1}, ..., f_{t-r+1}, g_t)^\dagger$，$\boldsymbol{e}_t^{z_t} = (e_t^f(z_t), 0, ..., 0, e_t^g)^\dagger \in \mathbb{R}^{(r+1)\times 1}$ であり，$\boldsymbol{A}^k$ は k 番目の奏法に対応する VAR(r) 行列，$\boldsymbol{U} = (1, 0, ..., 0, 1) \in \mathbb{N}^{1\times(r+1)}$ である．ここで $\dagger$ は転置を表す．したがって，音量軌跡の分解問題は，状態ベクトル系列 $\boldsymbol{x}_{1,...,T}$ の推定問題となる．

各パラメータの事前分布はそれぞれ，AR 係数 a_i^k は正規分布 $\mathcal{N}(0, \sigma_A^2)$，AR モデルの分散 $\sigma_f^2(k)$ は逆ガンマ分布 $\mathcal{IG}(\nu, \psi)$ とし，和分プロセスの分散 σ_g^2 は推論の安定のため固定とする．

5.2.2 奏法プリミティブによる音色変化

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 章で用いた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 $\boldsymbol{c}_t$ は，パラメータ $\Upsilon_k = \{\boldsymbol{\mu}_k^c, \Sigma_k^c\}$ を持つ，無限混合正規分布 [34] から出力されたものとみなす．

$$\boldsymbol{c}_t^k \sim \mathcal{N}(\boldsymbol{\mu}_k^c, \Sigma_k^c) \tag{5.7}$$

Υ_k 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．提案法のグラフィカルモデルを図 5.2 に示す．

5.3 推論アルゴリズムの実装

マルコフ連鎖モンテカルロ法の一種である 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．サンプリングは，$z_t, \boldsymbol{\pi}, \boldsymbol{\beta}, \Upsilon, \Theta, \alpha, \kappa, \varsigma, \mathcal{H}, \boldsymbol{x}_t$ の順に行う．基本的なアルゴリズムについては文献 [33][34][64] と同様であるため，導出は省略し，アルゴリズムおよび各更新式のみを説明する．

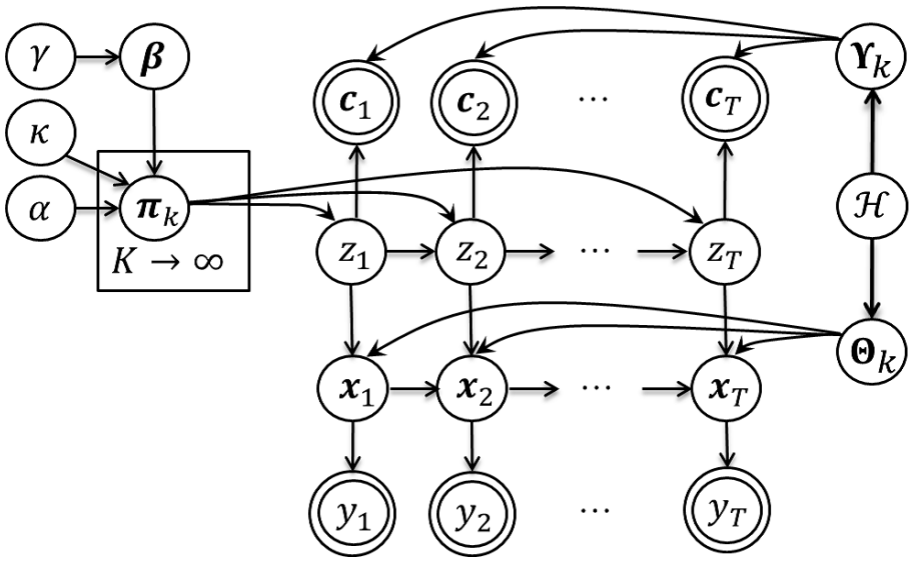

図 5.2: 提案法のグラフィカル表現．二重丸が観測データを表す.

Step 1: z_t のサンプリング

推論の高速化のために，z_t のサンプリングには Blocked sampler を用いる．ここで式 (5.6) 中で z_t に依存する項がアーティキュレーションの項のみなことに注意すると z_t の条件付き事後分布は,

$$p(z_t|z_{t-1}, \boldsymbol{\pi}, f_{1-r:T}, \boldsymbol{c}_t, \Theta, \Upsilon) \propto p(z_t|\boldsymbol{\pi}_{z_{t-1}})p(f_t|\boldsymbol{f}_{t-1}, \Theta_{z_t})p(\boldsymbol{c}_t|\Upsilon_{z_t})m_{t+1,t}(z_t) \tag{5.8}$$

となる．ここで $\boldsymbol{f}_t = (f_t, f_{t-1}, ..., f_{t-r+1})^\dagger$ であり，$m_{t+1,t}(z_t)$ は，遷移 $z_{t+1} \to z_t$ のバックワードメッセージである．よって z_t のサンプリングは

$$z_t \sim \sum_{k=1}^{K} \pi_{z_{t-1},k} P_k(\boldsymbol{f}_t, \boldsymbol{c}_t)\delta(z_t, k) \tag{5.9}$$

となる．ただし $P_k(\boldsymbol{f}_t, \boldsymbol{c}_t)$ と $m_{t,t-1}(k)$ は以下となる．

$$P_k(\boldsymbol{f}_t, \boldsymbol{c}_t) = \mathcal{N}\left(f_t \left| \sum_{i=1}^{r} a_i^k f_{t-i}, \sigma_f^2(k)\right.\right)\mathcal{N}(\boldsymbol{c}_t|\boldsymbol{\mu}_k^c, \Sigma_k^c)m_{t+1,t}(k) \tag{5.10}$$

$$m_{t,t-1}(k) = \sum_{j=1}^{K} \pi_{k,j}\mathcal{N}\left(f_t \left| \sum_{i=1}^{r} a_i^j f_{t-i}, \sigma_f^2(j)\right.\right)\mathcal{N}\left(\boldsymbol{c}_t|\boldsymbol{\mu}_j^c, \Sigma_j^c\right)m_{t+1,t}(j) \tag{5.11}$$

Step 2: $\boldsymbol{\pi}, \boldsymbol{\beta}, \Upsilon, \Theta$ のサンプリング

まず，補助乱数を用いて $\boldsymbol{\pi}, \boldsymbol{\beta}$ をサンプリングする [30]．次に，$\boldsymbol{c}_{t|z_t=k}$ のデータを用いて Υ_k の事後分布を求め，サンプリングする [34].

最後に SLDS のパラメータ Θ をサンプリングする．ここで σ_g^2 が固定であることに着目すると，Θ の推論は AR モデルのパラメータ $\boldsymbol{A}^k, \sigma_f^2(k)$ の推論であることがわかる．まず $\sigma_f^2(k)$ の条件付き事後分布は，$\tau \in \{t|z_t = k\}$ のデータを用いることで，ベイズ推論の標準的な結果から以下のようになる．

$$\sigma_f^2(k) \sim \mathcal{IG}\left(\nu + \frac{N_k}{2}, \psi + \frac{S}{2}\right) \tag{5.12}$$

ただし，$N_k = |\tau|$，$S = \sum_{t|z_t=k}(f_t - \sum_{i=1}^{r} a_i^j f_{t-i})^2$ である．次に $\boldsymbol{A}^k$ の条件付き事後分布は，$\bar{\boldsymbol{F}}_k$ を $\boldsymbol{f}_{\tau-1}$ を行方向に並べた行列，F_k を f_τ を並べた列ベクトル，$\boldsymbol{a}_k = (a_1^k, ..., a_r^k)^\dagger$ (i.e. $\mathrm{vec}(\boldsymbol{A}^k)$)，$\Sigma_A = \sigma_A \boldsymbol{I}_r$ と置くと，

$$\boldsymbol{a}_k \sim \mathcal{N}\left(\boldsymbol{S}\boldsymbol{m}, \sigma_f^2(k)\boldsymbol{S}\right) \tag{5.13}$$

となる [33]．ただし $\boldsymbol{m}$ と $\boldsymbol{S}$ は以下となる．

$$\boldsymbol{m} = \bar{\boldsymbol{F}}_k^\dagger F_k, \quad \boldsymbol{S} = \left(\Sigma_A^{-1} + \bar{\boldsymbol{F}}_k \bar{\boldsymbol{F}}_k^\dagger\right)^{-1} \tag{5.14}$$

Step 3: $\alpha, \kappa, \varsigma, \mathcal{H}$ のサンプリング

まず $\alpha, \kappa, \varsigma$ を，文献 [30] と同様に補助乱数を用いてサンプリングする．次に，$\mathcal{H}$ の iGMM に関するパラメータを，$\boldsymbol{c}_{t|z_t=k}$ から求めた事後分布よりサンプリングする [34]．$\mathcal{H}$ の Θ に関するパラメータは，局所解を避けるために固定する．

Step 4: $\boldsymbol{x}_t$ のサンプリング

線形動的システムのパラメータと，全ての時刻の観測ベクトルが既知の下での状態ベクトル $\boldsymbol{x}_{1,...,T}$ の推論は，カルマンスムーザで行うことが出来る．サンプリング法を用いたカルマンスムーザは，シミュレーションスムーザ [65] とも呼ばれる．本稿では推論の高速化のために，Fox らの Block Samling による手法 [33] で推論を行う．

5.4 評価実験

提案法を用いた音量軌跡の分解実験を行う．実験に用いる演奏データはあらかじめ全て標本化周波数 48kHz にリサンプリングした．音量 $y_{1,...,T}$ は式 (5.1) から求め，STFT のパラメータは，シフト幅を 5-msec，STFT 長を 2048 点とした．

パラメータ推論の設定値を以下に示す．AR 次数は $r = 3$ とした．AR 係数の事前分布は $\sigma_A^2 = 1$ とし，AR モデルの事前分布のパラメータは $\nu = \psi = 500$ とした．音色特徴量の出力分布の超パラメータおよび超々パラメータは文献 [34] と同様に，観測データから設定した．z_t の初期値は，$K = 20$ として乱数を用いて決定した．g_t の初期値は観測音量系列 y_t の移動平均（窓幅 1.5sec）とし，f_t の初期値は $f_t = y_t - g_t$ とした．和分プロセスの分散 σ_g^2 は，g_t の初期値の一階差分をとり，その分散の最尤推定量を 2 倍したもので固定した．Gibbs Sampler は，焼き入れを 50 回とし，繰り返し回数は 1000 回とした．

5.4.1 MIDI データを用いた分離実験

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人口データを用いて分離精度を評価した．

正解データの作成方法を説明する．まず，“Volume”，“Velocity” および “Expression” の値を固定した MIDI データを作成する．この MIDI データの音量変動は，ダイナミクスが固定であるため，MIDI 音源のプリセットアーティキュレーションのみに依存する．その MIDI データを wav ファイルに変換し，アーティキュレーションの正解データを得る．次に，ダイナミクス記号や音高の上下に基づき，人手でダイナミクスの概形を指定する．それをスプライン関数でなめらかに補完し，ダイナミクスの正解データを作成する．最後に，作成したアーティキュレーションとダイナミクスを加算し，観測音量系列を作成する．

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，レガートやスタッ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（表 5.1）．本実験で

表 5.1: 使用楽曲

クラリネット

作曲者	楽曲名	小節番号
J. Brahms	Clarinet Quintet - I	5–13
	Clarinet Quintet - II	1–7
	Clarinet Quintet - III	44–53

トランペット

作曲者	楽曲名	小節番号
F.J.Haydn	Trumpet Concerto - I	101–105
L. Anderson	Bugler's Holiday	9–23
G. Verdi	Aida - Triumphal March -	1–6

バイオリン

作曲者	楽曲名	小節番号
A. Vivaldi	The Four Seasons - Spring - I	1–7
F. Schubert	Death and the Maiden - III	1–23
E. Grieg	Holberg Suite - II	1–4

は，サウンドフォントは “TimGM6mb.sf2[1]” を用いた．音色特徴量は，MIDI データのものを使用した．

提案法は式 (5.1) を満たすように分離を行うため，誤差を求めるのはダイナミクスかアーティキュレーションの片方となる．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. $\boldsymbol{mp}$ から $\boldsymbol{mf}$）が，Velocity の 15 段階変化に該当し，その差が約 4dB であることから，小さな誤差であるといえる．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．本稿では，AR モデルが無音区間の音量変動を吸収することを期待し，明示的に休符の音量変動を取り扱わなかった．しかしいくつかの無音区間では，ダイナミクスが休符による音量低下の一部を吸収するように推論が働き，結果として推定精度が低下した．演奏行動の観点から考えると，奏者はまず最初に “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．

5.4.2 実演奏音を用いた分離実験

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．プロ奏者の演奏は空調を切った防音室で，標本化周波数 192kHz で

[1] `http://ocmnet.com/saxguru/Timidity.htm#sf2`（2014/01/24 アクセス）

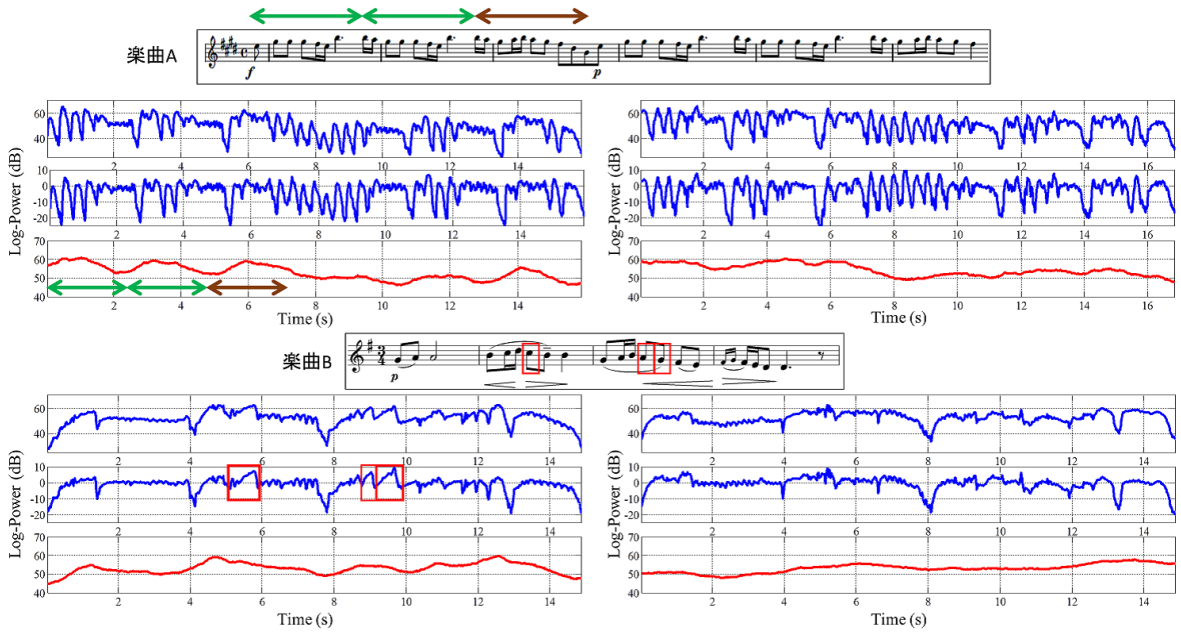

図 5.3: プロ奏者（左）とアマチュア奏者（右）のバイオリン演奏音への推定結果．3 つ並んだ音量グラフは上から，実測音量軌跡，アーティキュレーションの推定結果，ダイナミクスの推定結果を示す．

録音した．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で，プロ奏者の演奏を模倣するように演奏した．図 5.3 に分離結果を示す．

楽曲 A は，最初の 4 拍の音列をモデルとして 2 回繰り返し，その後 4 拍の補充を入れるという旋律パターンを，フォルテとピアノで繰り返す．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が出来る．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ものの，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” が感じられず，“平たい” 印象を受ける．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着けていないためと考えられる．

楽曲 B は，音量記号がピアノで，クレシェンドおよびデクレシェンドが記載されている．ダイナミクスには，プロ・アマチュア共にクレシェンドなどに起因する起伏が見て取れるが，プロ奏者の方がダイナミクスレンジが広く，また変化が急峻である．またプロ奏者のアーティキュレーションには，5.5，9.0，9.5 秒付近に，音量を上昇させながら 1 つの音符を演奏する，“似た形” のアーティキュレーションがある．これは，音符中で弓を加速することで実現するが，これは弓速の細やかなコントロールを必要とする難易度の高い奏法である．この点からも，プロ奏者とアマチュア奏者の演奏技術の差を見ることが出来る．

これらの結果から提案法は，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が示唆される．今後は提案法を，演奏技術の自動評価や，コンテキストと対応付けた生成モデルなどに応用し，有効性を大規模に評価する必要がある．

5.5 おわりに

本稿では，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．

本稿では，休符による無音区間の音量変動を取り扱わなかったため，分解精度が低下した．今後，楽譜情報をモデルに組み込むことで，休符の扱えるよう生成モデルを改善する必要がある．また，演奏技術の自動評価や，コンテキストと対応付けた生成モデルなどに応用し，有効性や応用を大規模に評価する必要がある．

5.6 関連研究

連続励起振動楽器の音量変化の解析およびモデル化は，その動特性の複雑さから深く議論されていない．楽音合成では，ADSR により音量軌跡を数個の関数で近似する手法がとられているが，自由度を固定したモデルではあらゆる演奏表現を説明することができず，合成音の品質に問題がある．ピアノの自動演奏では，楽譜情報と音量変化をガウス過程により関連付け合成する手法 [66] が提案されているが，打弦楽器のモデル化であるため音符内の動特性を説明するには至っていない．大石らは歌唱音量変動のモデル化を混合ガウス過程で表現する試み [16] を行ったが，この手法ではダイナミクスとアーティキュレーションを独立して制御しないため，歌詞（音素）によるアーティキュレーション変動を説明することができない．

また，音量軌跡のモデル化が問題となっている音楽情報処理タスクに，多重音解析と楽器音分離が挙げられる．これらの手法では，振幅スペクトログラムを行列とみなし，非負値行列因子分解（NMF: Non-negative Matrix Factorization）[67] で処理するものが一般的である．NMF の問題点として，振幅スペクトルのアクティベーション[2]の時間方向への連続性が保障されない．この問題を解決するために，音量変化になんらかの制約条件を与える必要がある．そこで音量変動をガウス関数の組み合わせで表現する HTC（Harmonic temporal structured clustering）[68] や非負値調波時間構造因子分解法（NHTF: Nonnegative Harmonic-Temporal Factorization）[69] が提案されているが，連続励起振動楽器の音量変動の時間構造を模倣するには至っていない．

本研究は，連続励起振動楽器の音量変動のモデル化とその解析法を提案した点で，これらの研究の発展にも寄与できると考える．

[2]各楽器音の音量軌跡と等価である

第6章　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

楽音合成と合成を用いた演奏解析では，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．擦弦楽器の代表的な楽音合成方式には，力学的センサを取り付けた楽器から奏法情報を数値的に取得し，楽音を合成する物理モデル方式 [35] や事前に用意した楽音コーパスから，合成したい音色に近いコーパスを取り出し意図表現を実現する素片接続方式 (e.g., *Vienna Symphonic Library*[1]) が挙げられる [36]．しかし，前者の合成法では専用の機材や演奏技術が要求されるため合成自体が困難であり，後者は作成可能な演奏表現がコーパスに依存するため演奏表現制御が困難である．よって，両者を達成するためには，楽器の物理的な特性や奏法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．それを実現するためには，対象となる楽器の物理機構を考慮した演奏分析法が必要となる

そこで本章では，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，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，定常区間の非調波成分を制御する．モデル評価のために，複数の演奏表現に対して合成音を生成し，提案手法の有効性を主観評価によって評価する．また本章では，正規分布，ガンマ分布，ポアソン分布をそれぞれ，$\mathcal{N}, \mathcal{G}, \mathrm{Poisson}$ と表記する．

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]．奏者は，演奏表現のために，弦を抑える左手でビブラートをかけ，弦を擦る右手で様々な奏法を駆使し弦振動を制御する．擦弦の位置・圧力・速度の相対的な強度差により，弦の各モードの振動の強さや，非周期成分の強さなどが制御され，特徴的な音色が生成される．本章では，実演奏音から擦弦振動の変化を解析するために，擦弦楽器の物理現象ついて考える．

6.1.1　擦弦振動

擦弦中の基本的な弦の運動は，Helmholtz により，駒と枕により固定される弦が描く放物線上をなぞる三角波として知られている．この振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，擦弦の Stick-Slip 運動は，弓に弾かれた弦が臨界点まで引っ張られ，臨界点に到達すると滑り，また摩擦により弓に引っ張られるという現象である．この運動により生成されるヘルムホルツ振動 $h(x,t)$ は，調波成分にのみパワーを持つ振動であり，変位は以下の式で求められる．

$$h(x,t) = \sum_{n=1}^{\infty} s(x,n,t)\sin(\omega_n t + \theta_n) \tag{6.1}$$

$$s(x,n,t) = G(t)\frac{\sin(n\pi\psi)}{n^2},\ \ \psi = \frac{x}{l},\ \ \omega_n t = 2\pi n F_0(t),\ \ \theta_n = n\pi$$

[1] http://vsl.co.at

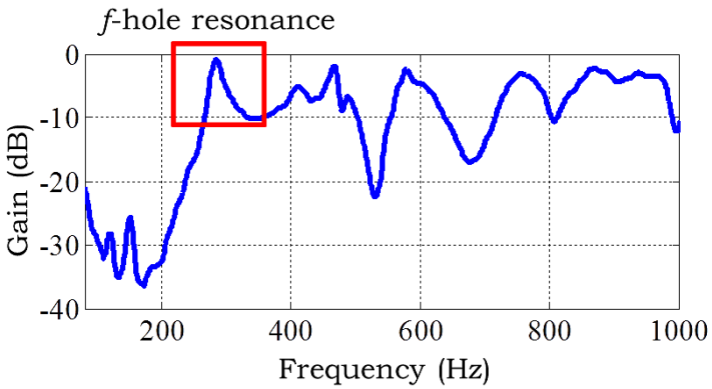

図 6.1: バイオリンの共鳴特性と f 字孔共鳴

ここで x は観測位置，l は弦長，$G(t)$ は時刻 t での振幅，$F_0(t)$ は時刻 t での基本周波数である．

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，一周期中に複数回の slip が起きる multiple-flyback や，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．また，prolonged periods では “choked/creaky sound” と呼ばれるカリカリとしたノイズが発生する．これは，*marcato*（はっきりと）や *feroce*（荒々しく）などの，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]．

6.1.2 楽器の共鳴

擦弦楽器音は，擦弦振動が駒を通して楽器へ伝達され，楽器内で共鳴し，音色変化することによって生成される．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]．

バイオリン属やヴィオール属（e.g. コントラバス）の共鳴特性の特徴として “f 字孔共鳴 (f-hole resonance)” ある [37]．f 字孔とは，楽器の表板の中央にあけられたイリック体の f に似た形の穴のことである．バイオリンの f 字孔共鳴の特徴として，300Hz 付近に共振特性を持つ．図 6.1 は先行研究 [44] で計測されたバイオリンの共鳴特性である．300Hz 付近に，鋭い共振のピークを持つことが確認できる．

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，式 (6.1) で表現される擦弦振動からの乖離と考えられる．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，発音ノイズや定常区間の非調波成分は，スティフネスなどの非線形な要因で生成されるため，線形な伝達系としては記述出来ない．さらに，発音区間と定常区間の非調波成分は発生

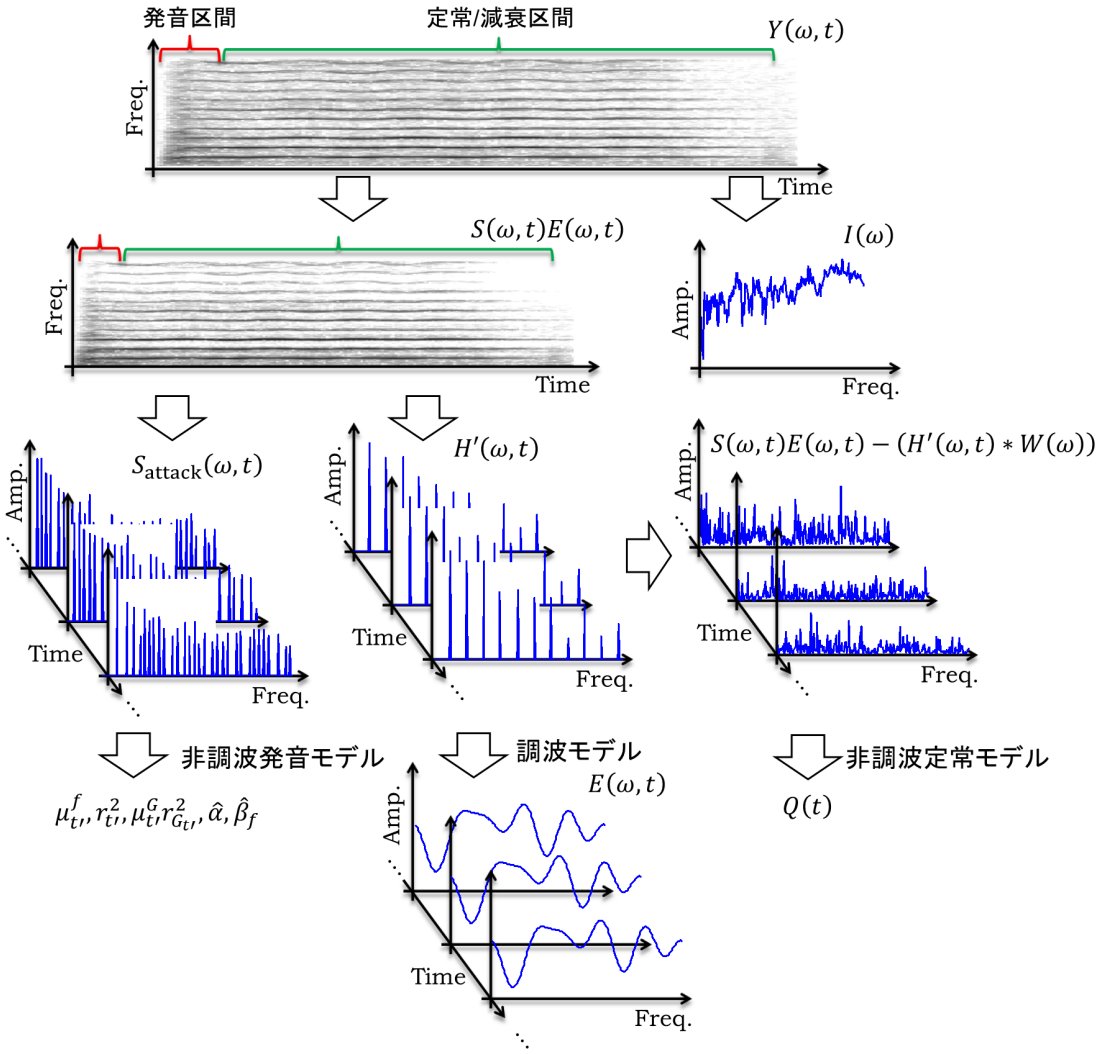

図 6.2: 奏法モデルの構築手順

原因が異なる．よって非調波成分は調波成分とは分けて考え，発音区間の非調波モデルと定常区間の非調波モデルは個別の確率モデルとして構築する．図 6.2 に奏法モデルの構築手順を示す．

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 $e_h(t)$ で記述すると，式 (6.1) は以下のように書き換えられる．

$$\begin{aligned} h'(x,t) &= h(x,t) * e_h(t) \\ &\approx \sum_{n=1}^{\infty} s(x,n,t) E(nF_0(t),t) \sin(\omega_n t + \theta_n) \end{aligned} \tag{6.2}$$

ここで $E(\omega,t)$ が奏法による各モードの振動の強度の比率の変化を表す．すると観測信号の調波成分の振幅スペクトログラム $Y_{\mathrm{harm}}(\omega,t)$ は，簡単のために窓関数の影響を無視すると以下のように書ける．

$$Y_{\mathrm{harm}}(\omega,t) = S_{\mathrm{harm}}(\omega,t) E(\omega,t) I(\omega) \tag{6.3}$$

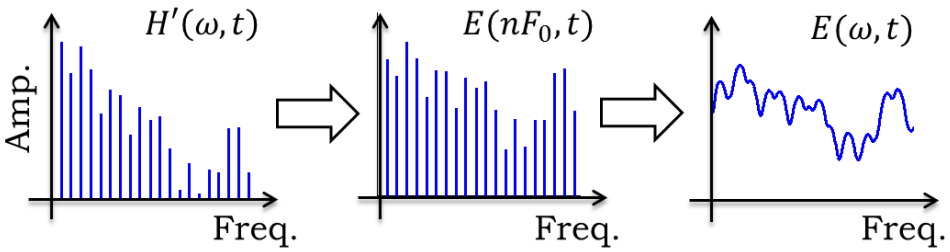

図 6.3: 調波モデルの推定

すなわち，ヘルムホルツ振動のスペクトル $S_{\mathrm{harm}}(\omega,t)$ 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 $I(\omega)$ がかかったものと解釈する．本節では，$E(\omega,t)$ を調波成分における奏法モデル（調波モデル）とし，これを推定する手法を考える．

まず，調波成分への窓関数の影響を低減するために，ピッチ同期分析 [70] により観測振幅スペクトログラム $Y(\omega,t)$ を求める．このスペクトログラムには非調波成分も含まれるため，$Y(\omega,t)$ から調波擦弦振動スペクトログラム $H'(\omega,t)$ を抽出する．

$$\begin{aligned} H'(\omega,t) &= \sum_{n=1}^{F} \frac{Y(\omega,t)}{I(\omega)} \delta(\omega, nF_0(t)) \\ &= S_{\mathrm{harm}}(\omega,t) E(\omega,t) \end{aligned} \tag{6.4}$$

ただし $\delta(i,j)$ はクロネッカーのデルタ，F はナイキスト周波数までに含まれうる倍音数 $F = \lfloor f_s/F_0(t) \rfloor$，$f_s$ はサンプリング周波数，$\lfloor\cdot\rfloor$ は床関数を表す．

ここで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．すると式 (6.4) は式 (6.1)(6.2) より以下のように書き換えられる．

$$H'(\omega,t) = G(t) \sum_{n=1}^{F} \frac{\sin(n\pi\psi)}{n^2} E(\omega,t)\delta(\omega, nF_0(t)) \tag{6.5}$$

ただし

$$G(t) = \sqrt{\frac{\sum_{n=1}^{F} |H'(nF_0,t)|^2}{F \sum_{n=1}^{F} |\sin(n\pi\psi)/n^2|^2}} \tag{6.6}$$

である．式 (6.6) は $n\psi$ が整数の時に零除算となるため，$\psi = 1/31 + 10^{-5}$ に固定する．

すると n 番目の調波成分における調波モデルの伝達特性は

$$E(nF_0,t) = \frac{H'(nF_0,t)}{G(t)\sin(n\pi\psi)/n^2} \tag{6.7}$$

で求まる．しかし一般に伝達特性は式 (6.7) のように，ある一点のみに値を持つことはなく，周波数方向に滑らかに変化する．そこで，窓関数の周波数特性を式 (6.7) の計算結果に畳み込むことにより周波数方向に補間を行い，さらに基本周波数の幅で振動する成分を平滑化により取り除くことにより [71]，周波数方向に連続で滑らかな調波モデル $E(\omega,t)$ を推定する（図 6.3）．

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．よって，発音区間の非調波モデル（非調波発音モデル）は，スペクトルパラメータを直接生成する確率的ソースフィルタモデルとする．確率モデルのパラメータの変化により，奏法による生成スペクトルの変化を記述する．

発音時の非調波成分は，押弦位置で決定する基本周期 F_0 と異なる周期で発生する slip 現象に起因する．この毎振動ごとに波長が変わる Stick-Slip 現象により分析区間内に様々な周波数成分が混在し，結果として観測スペクトルはピンクノイズのように見える．しかし発音区間であっても波形は Stick-Slip 現象により生成されているため，1 波長分のスペクトルは調波スペクトルと同様に周波数軸方向にスパースである．そのため人間の発声機構のソースフィルタ表現のように，ノイズの励起源をピンクノイズで近似すると，分析合成が劣化する（研究業績 J-3 参照）．

そこで周波数軸を F 個の区間 $f = \{1, 2, .., F\}$ 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 $G^f_{\text{attack}}(n_f, t)$ で立ち上ったものとして以下のように記述する．

$$S_{\text{attack}}(\omega, t) = G(t) \sum_{f=1}^{F} \sum_{n_f=1}^{N_f(t)} \int_{fF_0 - F_0/2}^{fF_0 + F_0/2} \exp\{G^f_{\text{attack}}(n_f, t)\} \delta(\omega, \omega_f(n_f, t)) d\omega \tag{6.8}$$

これは，ゲイン $G(t)$ のスパースな励起源にフィルタの周波数特性 $\exp\{G^f_{\text{attack}}(n_f, t)\}$ が乗じられるソースフィルタ表現とみなすことが出来る．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 $\omega_f(k, t), G^f_{\text{attack}}(k, t), N_f(t)$ 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．

$\omega_f(k, t), G^f_{\text{attack}}(k, t)$ は連続変数，$N_f(t)$ は非負の整数である．$\omega_f(k, t)$ 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前後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ぐ確率変数を表現でき，かつ解析的に扱いやすい分布である正規分布でモデル化する．$G^f_{\text{attack}}(k, t)$ はヘルムホルツ振動の対数ゲイン $\ln\left(\sin\left(f\pi\psi\right)/f^2\right)$ が奏法により上下に揺らぐことにより発生する．よって $\omega_f(k, t)$ と同様に正規分布でモデル化する．$N_f(t)$ は区間 f に何本のビンが立ち上がるかを示す変数であるが，このような確率変数を表現するのにはポアソン分布が適している．よって，それぞれの確率分布を以下のように定義する．

$$\omega_f(k, t) \sim \mathcal{N}(\mu_f, \sigma^2) \tag{6.9}$$

$$G^f_{\text{attack}}(k, t) \sim \mathcal{N}(\mu^G_f, \sigma^2_G) \tag{6.10}$$

$$N_f(t) \sim \text{Poisson}(\lambda_f) \tag{6.11}$$

ただし σ^2 は，各区間ごとの分布の過剰なオーバーラップを避けるために，$\sigma^2 = (F_0/3)^2$ で固定とし，σ^2_G は推論の安定のために実験的に $\sigma^2_G = 1$ とする．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，分布パラメータの事前分布を共役事前分布である以下の分布に設定する．

$$\mu_f \sim \mathcal{N}(\mu^f_0, r^2_0) \tag{6.12}$$

$$\mu^G_f \sim \mathcal{N}(\mu^f_{G_0}, r^2_{G_0}) \tag{6.13}$$

$$\lambda_f \sim \mathcal{G}(\alpha_0, \beta_0) \tag{6.14}$$

ただし $\mu^f_0 = fF_0$，$\mu^G_f = \log\left(\sin\left(f\pi\psi\right)/f^2\right)$ であり，α_0, β_0 は有限な密度関数を持ちかつ無情報な事前分布を設計するために，$\alpha_0 = \beta_0 = 1$ とする．また事前実験より，$r^2_0 = \sigma^2$，$r^2_{G_0} = \sigma^2_G$ とした．

次に観測スペクトルから予測分布を求める．まず発音区間の観測擦弦スペクトル $Y(\omega, \tau)/I(\omega)$ からピーク検出をし，各区間 f ごとにピーク数 $M_f(\tau)$，ピーク位置 $o_f(m_f, \tau)$ およびその対数ゲイン $g_f(m_f, \tau)$ を求める．ここで $\tau \in \{1, 2, ..., t'\}$ は，実際の計算上のスペクトログラムの時間イ

ンデックスである．そして，ω_f と N_f の予測分布を以下のように求める．

$$\hat{\omega}_f(k,t) \sim \mathcal{N}(\mu^f_{t'}, \sigma^2 + r^2_{t'}(f)) \tag{6.15}$$

$$\hat{G}^f_{\text{attack}}(k,t) \sim \mathcal{N}(\mu^f_{G_{t'}}, \sigma^2_G + r^2_{G_{t'}}(f)) \tag{6.16}$$

$$\hat{N}_f(t) \sim \hat{\alpha}\hat{\beta}^{\hat{\alpha}}_f(\hat{\beta}_f + \hat{N}_f(t))^{-(\hat{\alpha}+1)} \tag{6.17}$$

ただし，

$$\mu^f_{t'} = \frac{\sigma^{-2}\sum_{\tau=1}^{t'}\sum_{m_f=1}^{M_f(\tau)} o_f(m_f,\tau) + r_0^{-2}\mu^f_0}{\left(\sum_{\tau=1}^{t'} M_f(\tau)\right)\sigma^{-2} + r_0^{-2}} \tag{6.18}$$

$$r^2_{t'}(f) = \frac{1}{\left(\sum_{\tau=1}^{t'} M_f(\tau)\right)\sigma^{-2} + r_0^{-2}} \tag{6.19}$$

$$\mu^f_{G_{t'}} = \frac{\sigma_G^{-2}\sum_{\tau=1}^{t'}\sum_{m_f=1}^{M_f(\tau)} g_f(m_f,\tau) + r_{G_0}^{-2}\mu^f_{G_0}}{\left(\sum_{\tau=1}^{t'} M_f(\tau)\right)\sigma_G^{-2} + r_{G_0}^{-2}} \tag{6.20}$$

$$r^2_{G_{t'}}(f) = \frac{1}{\left(\sum_{\tau=1}^{t'} M_f(\tau)\right)\sigma_G^{-2} + r_{G_0}^{-2}} \tag{6.21}$$

$$\hat{\alpha} = \alpha_0 + t \tag{6.22}$$

$$\hat{\beta}_f = \beta_0 + \sum_{\tau=1}^{t'} M_f(\tau) \tag{6.23}$$

である．

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特に未解決部分の多い問題であるが，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ことが分かっている．物理モデル [45] では物理パラメータにより生成された擦弦振動 $h'(t)$ に対し，以下の確率モデルでノイズが付与された擦弦振動 $\tilde{h}(t)$ 生成している．

$$\tilde{h}(t) = \begin{cases} (O + Q(t)u(t))h'(t) & (\text{slip}) \\ h'(t) & (\text{stick}) \end{cases} \tag{6.24}$$

ここで $u(t)$ は 0 から 1 の範囲の一様分布の乱数，O はノイズ強度のための任意の定数，$Q(t)$ は時間変化するノイズの強度である．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）として扱う．

観測された擦弦振動スペクトルの非調波成分 $S_{\text{noise}}(\omega,t)$ は，観測擦弦振動のスペクトログラム $Y(\omega,t)/I(\omega)$ から，調波擦弦振動スペクトル $H'(\omega,t)$ 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で求められると仮定する．よって非調波定常モデル $Q(t)$ は

$$Q(t) = \sqrt{\frac{\gamma^2}{2P}\int |S_{\text{noise}}(\omega,t)|^2\, d\omega} \tag{6.25}$$

で求める．ここで P は STFT 時の切り出しフレーム点数，γ は一様乱数の大きさを調整する正の定数である．

6.3 奏法モデルを用いた楽音合成実験

構築した生成モデルの有効性の評価のために，奏法モデルを用いた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．本章ではまず，奏法モデルを用いた楽音合成法について述べ，さらに実際の擦弦楽器演奏者を対象とした意図表現と品質の聴取実験を行う．

6.3.1 楽音の合成

Step1: 調波擦弦振動の生成

まず，基本擦弦振動のスペクトログラム $S_{\mathrm{harm}}(\omega, t)$ を作成する．

$$S_{\mathrm{harm}}(\omega, t) = G(t) \sum_{n=1}^{F} \frac{\sin(n\pi\psi)}{n^2} \delta(\omega, nF_0(t)) \tag{6.26}$$

ここで，基本周波数 $F_0(t)$ と振幅 $G(t)$ を制御することにより，音高の操作や楽音全体の音量を制御する．次に，$S_{\mathrm{harm}}(\omega, t)$ を窓関数を用いて周波数方向にで補間する．そして，調波モデル $E(\omega, t)$ を用いて，式 (6.4) に基づき任意の調波擦弦振動のスペクトログラム $H'(\omega, t)$ を生成する．最後に $H'(\omega, t)$ に Ovarlap-Add 法で調波擦弦振動 $h'(t)$ を合成する．

Step2: 非調波発音擦弦振動の生成（任意）

非調波発音擦弦振動が楽音知覚に重要となる発想記号（e.g., *marcato*, *feroce*）は非調波発音モデルを用いて発音擦弦振動を合成する．

まず，式 (6.15)(6.16)(6.17) から，周波数成分数 $N_f(\tau)$ とその周波数位置 $\{\omega_f(1,\tau), ..., \omega_f(N_f(\tau),\tau)\}$ および対数ゲイン $\{G^f_{\mathrm{attack}}(1,\tau), ..., G^f_{\mathrm{attack}}(N_f(\tau),\tau)\}$ を生成する．次に式 (6.8) に基づき，発音非調波成分の励起振動 $S_{\mathrm{attack}}(\omega, t)$ を生成する．ここで，振幅 $G(t)$ を制御することにより音量を制御する．そして $S_{\mathrm{attack}}(\omega, t)$ を窓関数を用いて周波数方向にで補間し，Ovarlap-Add 法で波形合成する．最後にここで合成された発音擦弦振動と，**Step1** で合成された調波擦弦振動 $h'(t)$ を加算合成する．

Step3: 持続擦弦振動雑音の付与

式 (6.24) に基づき，**Step2** で合成された擦弦振動にノイズを付与する．slip 区間判定には，slip 区間の擦弦振動の微分値は負の値となる特性を用いる．

Step4: 楽器の共鳴特性のフィルタリング

生成された擦弦振動に対し，楽器の共鳴特性をフィルタリングする．本稿では，先行研究 [72] で計測された共鳴特性と同様の物を用いた．

6.3.2 評価実験

構築された奏法モデルの有効性の検証のために，分析合成音の主観評価実験を行う．実験では，発想記号よる自身の演奏表現を，楽器を用いて忠実に再現することが可能な熟練度を持つ奏者が，3 種類の発想記号（*feroce*: 荒々しく，*marcato*: はっきりと，*dolce*: やわらかく）で演奏したバイオリンの 4 種類の音高の単音（G 線の G 音: 197Hz，G 線の B 音: 234Hz，A 線の A 音: 442z，A 線の C 音: 525Hz）を用いた．全ての演奏音は，IC レコーダーを用いて，防音室で録音した．収録条件は，標本化周波数は 48kHz，量子化 bit 数は 24bit とした．発想記号は，物理モデルでの奏法パラメータの変化が特徴的かつ，*marcato* と *dolce* はオーケストラや室内楽の楽曲に頻出する，*feroce* は音色の迫力やアクセントが特徴的で，*dolce* と極端な音色の差を持つ，という理由で選択した．

表 6.1: 実験条件

発想記号	*feroce*, *marcato*, *dolce*
音高	197Hz，234Hz，442Hz，525Hz
音長	*feroce*, *dolce*：二分音符（BPM = 120）
	marcato：四分音符（BPM = 120）
演奏者	バイオリン歴 12 年の大学生
収録機材	TASCAM DR-07 内蔵マイク使用
収録条件	48kHz, 24bit
収録部屋	空調を切った防音室
比較音	実演奏音，提案法,
	音高操作音（短 3 度上昇，下降）,
	調波モデル無し，非調波発音モデル無し
被験者数	5 名
スピーカー	BOSE Companion 2 series II

本実験では，実演奏音（ORG），提案法（PRO），提案法を用いて音高を短 3 度上昇させた合成音（INT+3）と下降させた合成音（INT-3），奏法モデルを用いない合成音（-EMF），発音時の非調波成分が近くに大きく影響を及ぼすとされる発想記号である *feroce* と *marcato* で，合成 **Step2** を省略し，非調波発音モデルを用いない合成音（-ATT）の計 68 種類の楽音を用いた．ただし *dolce* は，PRO で非調波発音モデルを合成に用いないため-ATT を評価しない．

被験者は，擦弦楽器を 5 年以上経験し，発想記号による音色の変化をイメージできる 5 名とした．音圧は，被験者の聴きやすいレベルとなるよう，事前に調節した．詳細な実験条件を表 6.1 に示す．

評価は，提示音の，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（A）と，音質（B）を，それぞれ 5 段階で評価する MOS（Mean Opinion Score）で行った．各評定は，1 が非常に悪い，2 が悪い，3 が普通，4 が良い，5 が非常に良い，を表す．刺激の提示順序はランダムとし，被験者にはどの刺激が合成音であるかは伝えずに評価した．各刺激の間には 3sec の間が空けられる．

MOS により算出された各合成法の平均値と標準誤差を，発想記号ごとに図 6.4 と図 6.5 に示す．図の横軸は合成法の種類，縦軸はイメージに対する合致度および音色の自然さ (図 6.4) と音質 (図 6.5) を示す．

自身の音色のイメージとの合致度および自然さ（A）の評価結果から，提案法の評点は実演奏音と比べ若干の低下がみられるが，ほぼ等価である．Dunnett の多重比較検定により実演奏音との有意差を検定した結果，提案法による合成音は，全ての発想記号で危険率 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でも，全ての発想記号で危険率 5%で有意差が認められなかった．奏法モデルを用いなかった合成音は，全ての発想記号で危険率 5%で有意差が認められ，非調波発音モデルを用いなかった合成音は，*marcato* で危険率 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と分かる．また，発音非調波モデルを用いない分析合成では，*feroce* では有意差は認められなかったが，*feroce*，*marcato* 共に音高操作を行った楽音よりも評点が下がったことから，非調波発音モデルが発想記号による音色のイメージの分析合成に効果を持つと考えられる．また，奏法モデルを用いない合成音に有意差が認

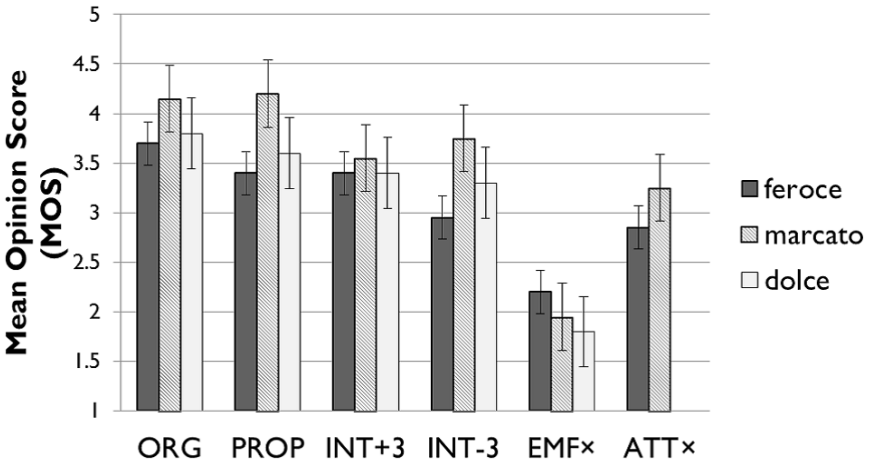

図 6.4: 主観評価の結果（イメージに対する音色の合致度および音色の自然さ）

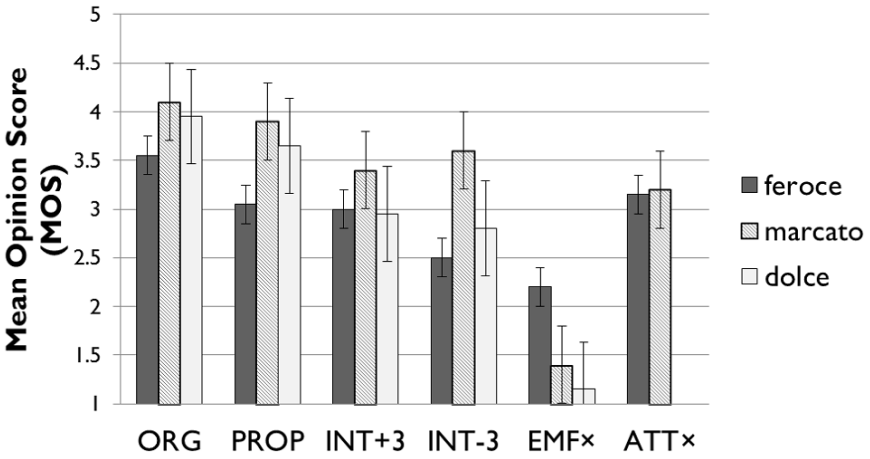

図 6.5: 主観評価の結果（音質）

められたことから，演奏表現の知覚で音色の変化が重要な要素であることが示された．

音質（B）の評価結果を，Dunnett の多重比較検定により実演奏音との有意差を検定した結果，提案法による合成音は，全ての発想記号で危険率 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では，*feroce* の短 3 度上昇音と *marcato* で危険率 5%で有意差が認められなかった．奏法モデルを用いなかった合成音は，全ての発想記号で危険率 5%で有意差が認められ，非調波発音モデルを用いなかった合成音は，*marcato* で危険率 5%で有意差が認められた．

この結果から提案法を用いた楽音合成は，音高操作を行わない分析合成では実演奏音の音質を保ったまま分析合成することが可能である．しかし，音高操作を行う場合は，いくつかの発想記号において音質の劣化が生じる．(A) において有意差が認められず，(B) において有意差が認められることから，これは楽音合成の際，音高操作によって生じる部分的な位相の不整合に起因する問題であると考えられるため，今後合成法の改良が必要である．

本評価実験では，単音のみを用いて評価を行ったが，提案法は奏法による音色変化を演奏音と同等の音質が得られる分析合成が可能なため，フレーズであっても同様の性能が達成されると考えられる．

また本評価実験は，バイオリンの楽音のみを対象とした．提案法は，実演奏音を擦弦振動と楽器の共鳴特性に分解し，擦弦振動から奏法によるスペクトルパラメータの変動を推定する手法であるため，バイオリンと類似した物理機構を持つバイオリン属やヴィオール属の分析合成でも，同等の性能が達成されると考えられる．また発想記号は，提案法は，聴衆の耳になじみ深い *marcato*, *dolce* 等や，特徴的な音色を持つ *feroce* 等の発想記号に関して有効な合成手法であると考えられる．

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と音質の評価では，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，構築した奏法モデルを固定した音高制御では，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は，元音声と同等の品質で楽音制御を行えることを示した．

しかし，本研究で扱った評価実験は，バイオリンの楽音のみを対象とし，発想記号は *feroce*, *marcato*, *dolce* の 3 種類のみを扱ったものであった．今後は，他の擦弦楽器や，網羅的に実験することは困難であるとしても他の多くの種類の発想記号においても分析合成を行い，評価実験を行う．

6.5 関連研究

ソースフィルタ表現は，人間の発生機構のモデル化である Vocoder（ボコーダ）と関連が深い．Vocoder では声帯振動を周期パルス[2]とみなし，それを，声道特性を模倣したフィルタで制御し，音色を制御する．提案法は，Vocoder の駆動源を擦弦振動の物理特性に合わせて変化させたものとみなすこともできる．

Vocoder の考え方は，スペクトル素片接続方式 [36] による歌声合成方式 VOCAOID や，HMM 音声合成 [74]，声質変換 [73] などの音声合成/変換技術に適用されている．また励起源やフィルタの周波数特性を適切な重みで重ね合わせることで，声質や感情表現のモーフィングが行えることが示されている [75]．

歌声合成にモーフィング技術を適用した手法として，初音ミク Append を利用した声質による演奏表現転写 [76] がある．この手法では初音ミク Append に含まれる 6 種類の声質（フィルタの周波数特性）の混合比を演奏データから決定する．

提案法は，擦弦楽器音の物理特性を破壊せずに演奏表現情報を直接制御できるのもの，制御パラメータがスペクトル情報を数値的に示したものであるため，適用範囲は分析合成および数値的な制御のみに限られた．今後，演奏表現に応じたスペクトル情報が張る音質空間 [76] を提案法の奏法モデルで構築し，パラメータ制御に適用することにより，直観的なパラメータによる楽音制御が可能になると考えられる．

[2]調波成分にのみ等しいパワーを持つ波形

第7章　真のテンポ曲線の推定に基づく演奏音の伸縮修正

本章では，テンポ変動を対象に，観測演奏音から求めた逸脱量を，演奏表現に由来する逸脱と，奏法誤差に由来する逸脱（奏法誤差成分）に分解する手法を考える．また，演奏音から奏法誤差成分を除去することにより，演奏音を自動修正する応用技術を提案する．

図 7.1 にプロ奏者とアマチュア奏者のテンポ変動例を示す（ただしテンポは，各音符の持続時間長を音価[1]で割ることにより求めた）．実際の演奏では，テンポは一定ではない．熟練した奏者[2]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 7.1 左）．これは，多くの先行研究のテンポ曲線である．一方，熟練度の低い奏者[3]のテンポは滑らかに変化せず，ばらつく（図 7.1 右）．ここで，作曲家がテンポ変動を指定しないフレーズでは，アマ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と仮定する．本章では，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

7.1.1　真のテンポ曲線の推定

音符の持続時間の定義は楽器の系統によって様々だが，本稿では一般化のために，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（IOI: intra-onset interval）．すなわち休符は考慮せず，8 分音符と 8 分休符を一つの 4 分音符として扱う．音価についても同様の定義を行う．すると，奏法誤差成分を含まない n 音目の発音時刻は，1 音目から $(n-1)$ 音目までの持続時間の和となり，音価とテンポ（beats/min）を用いて以下のように書ける．

$$\text{発音時刻}\,[n] = \sum_{m=1}^{n-1} \frac{60}{\text{テンポ}\,[m]} \times \text{音価}\,[m] \tag{7.1}$$

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すると，n 音目の観測発音時刻 $y[n]$ は以下のように書ける．

$$y[n] = \sum_{m=1}^{n-1} \frac{60}{b[m]} h[m] + e[n] \tag{7.2}$$

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．

[1]本章では，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．
[2]以降，本来の定義とは異なるが，“プロ奏者” と呼ぶ．
[3]以降，本来の定義とは異なるが，“アマチュア奏者” と呼ぶ．

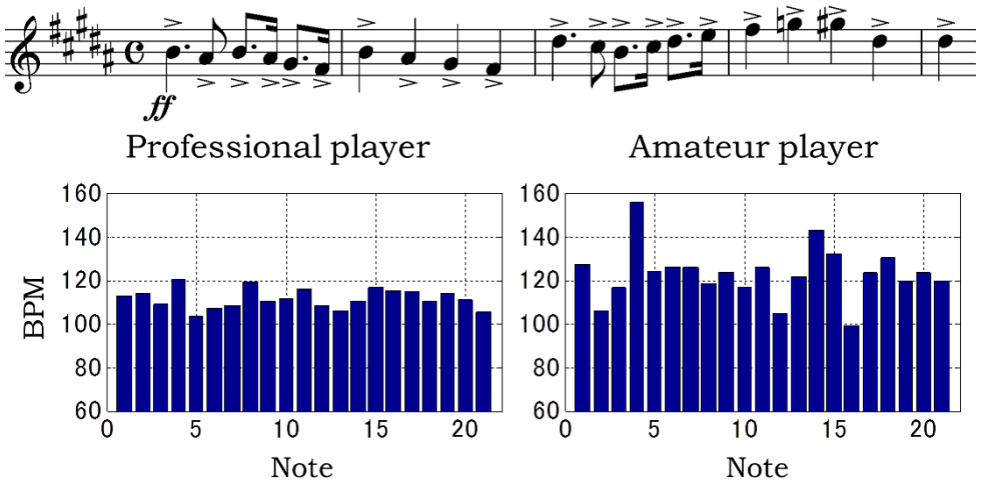

図 7.1: プロ奏者とアマチュア奏者のテンポ変動例

さらに，テンポ変動を曲線として推定するために，武田らのテンポ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定義する．

$$b[n]^{-1} = \sum_{p=0}^{P} w_p g[n]^p \tag{7.3}$$

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，$g[n] = \sum_{m=1}^{n-1} h[m]$ で求める．よって，式 (7.2)(7.3) より，n 音目の持続時間 $\Delta y[n]$ は以下のようになる．

$$\begin{aligned}\Delta y[n] &= y[n+1] - y[n] = \frac{60}{b[n]} h[n] + e[n+1] - e[n] \\ &= 60 \sum_{p=0}^{P} w_p g[n]^p h[n] + e[n+1] - e[n]\end{aligned} \tag{7.4}$$

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 $\boldsymbol{G}_{n,p} = \{g[n]^{(p-1)} h[n]\}$ と置くことにより，音符の持続時間ベクトル $\Delta \boldsymbol{y} = (\Delta y[1], ..., \Delta y[N])^T$ は以下のように書ける．

$$\Delta \boldsymbol{y} = 60 \boldsymbol{G} \boldsymbol{w} + \Delta \boldsymbol{e}, \tag{7.5}$$

ここで $\boldsymbol{w}$ は回帰係数を並べたベクトル $\boldsymbol{w} = (w_0, ..., w_P)^T$ であり，$\Delta \boldsymbol{e}$ は奏法誤差成分のデルタベクトル $\Delta \boldsymbol{e} = (e[2] - e[1], e[3] - e[2], ..., -e[N])^T$ である．

ここで先刻研究 [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，$\Delta \boldsymbol{e}$ 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 $\boldsymbol{w}$ を求めることで，式 (7.3) より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．

$$\sigma^2_{\Delta \boldsymbol{e}} = \frac{1}{N} \sum_{n=1}^{N} \left(\Delta y[n] - 60 \sum_{p=0}^{P} w_p g[n]^p h[n] \right)^2 \tag{7.6}$$

$$\mathrm{AIC} = N \log(2\pi \sigma^2_{\Delta \boldsymbol{e}}) + N + 2(P+2) \tag{7.7}$$

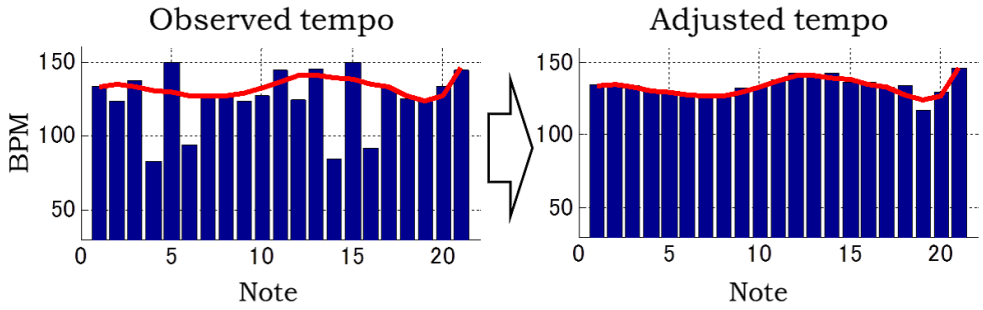

図 7.2: 観測音と修正音のテンポ変動例．左図のバーが観測音のテンポ変動，右図のバーが修正音のテンポ変動，両図の赤線が観測音から推定された真のテンポ曲線（$P = 5$）．

7.1.2 音響信号の伸縮修正

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 $\boldsymbol{b}$ を用いて以下のように書ける．

$$\hat{z}[n] = \frac{60}{b[n]} h[n] \tag{7.8}$$

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 音目の持続時間を以下の式で表される伸縮係数 $\alpha[n]$ 倍することとなる．

$$\alpha[n] = \frac{\hat{z}[n]}{y[n+1] - y[n]} \tag{7.9}$$

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．

図 7.2 に修正結果の例を示す．左図は奏法誤差を含む観測テンポ変動を示し，右図が修正された音響信号から求めたテンポ変動を示す．提案法の修正により，テンポ変動がテンポ曲線に近づいていることが確認できる．修正後の一部の音符のテンポ変動が真のテンポ曲線に一致しないのは，修正前，または修正後の発音時刻検出で誤差が生じたためである．

7.2 評価実験

提案修正法により，音響信号が奏者の意図したテンポ変動に修正されているかを，聴取実験で評価した．テンポ変動は，楽器の種類による逸脱の差異が小さいため，連続励起振動楽器に限定せず，撥/打弦楽器も評価対象に含める．対象とした楽器は，連続励起振動楽器からバイオリンとチェロ，撥/打弦楽器からエレキギター（エフェクトなし）とした．

本研究で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．そこで本実験では，目標とするテンポ変動として，プロ奏者の演奏を用いた．楽器の演奏を 3 年以上経験しているアマチュア奏者が，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に，メトロノームを用いずに演奏した．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．

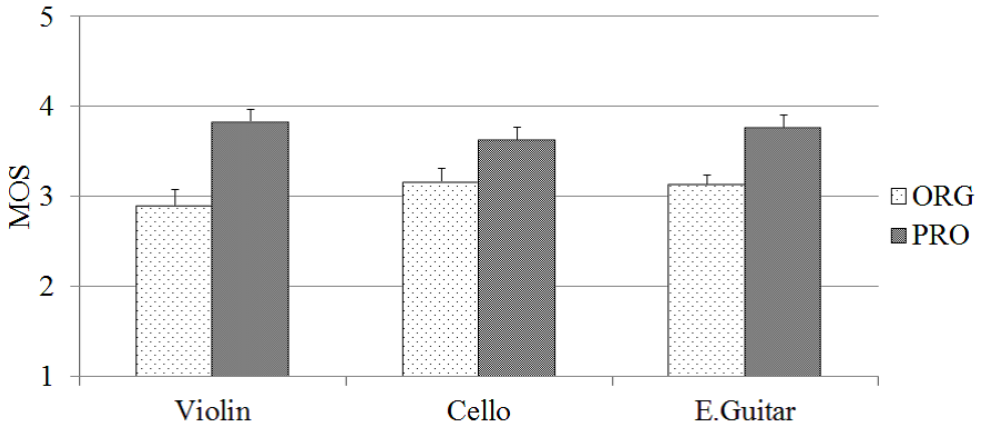

図 7.3: 主観評価結果

表 7.1: アマチュア奏者の収録楽曲

作曲者	楽曲名	小節番号
<Violin>		
A. Dvorak	Symphony No. 8 Mov.1	244–250
R.Wagner	Tannhauser - Grand March	40–44
R.Wagner	Tannhauser - Grand March	64–68
<Cello>		
A. Dvorak	Symphony No. 8 Mov.1	1–6
A. Dvorak	Symphony No. 8 Mov.1	165–169
A. Dvorak	Symphony No. 8 Mov.4	26–33
<E.Guitar>		
LUNKHEAD	ENTRANCE	5–12
MONKEY MAJIK	アイシテル	52–56
松本孝弘	Thousand Dreams	2–9

アマチュア奏者は各楽器 2 名ずつとし，楽曲は各楽器に対して 3 曲ずつとした（表 7.1）．これらのフレーズは，楽譜上の BPM は 60–180，平均音符数は 22 個，演奏時間は 9–16 秒である．また，各演奏から求めた多項式の次数は $2 \leq P \leq 5$ であった．

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（ORG）と修正音（PRO）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 mean opinion score (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．音圧は，被験者の聴きやすいレベルとなるよう事前に調節した．

各楽器ごとの MOS と標準誤差を図 7.3 に示す．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ていることが確認できる．t-検定で有意差を検定した結果，全ての楽器の評価で，バイオリンとギターは危険率 1%で，チェロは危険率 5%で有意差のある上昇が認められた．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といえる．

一方で，リズムとテンポの知覚には，本稿で扱った音符の発音時刻の間隔だけでなく，アクセ

ントなどに起因する音符の強弱も関係する [77][78]．今後修正の精度を向上させるために，音符の強弱変化に関する奏者の意図を推定し，楽音修正する手法を検討する必要がある．

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．従って提案法は，奏者の意図したテンポ変動を推定し，それに基づき楽音修正を行えるといえる．

今後の展望として，本稿で推定した真のテンポ曲線は，奏法誤差成分を含む音響信号からの，演奏表現の特徴の抽出とみなせる．演奏表現情報抽出技術は，奏者認識 [4]，合成音への表現力付与 [79] などに応用されている．本手法もこれらの分野への応用法を検討していく．

7.4 関連研究

テンポ変動に限らず，奏者の意図しない逸脱を除去する研究は，いまだ深く検討されていない．これは，意図した逸脱と意図しない逸脱をアルゴリズミックに見分けるのが困難なためである．よって奏者の意図しない逸脱を扱う研究では，従来の解析研究を参考に熟練した奏者の逸脱モデルを立て，そのモデルで表現できない逸脱を意図しない逸脱とみなしている．

木立らは，熟練した歌唱者の発声法に着目した歌声修正を提案した [80]．腹式呼吸による正しい発声法の歌声は，第 3 フォルマントが大きく盛り上がるという特徴がある．これは，腹式呼吸で発声を行った場合，声帯振動の高調波減衰が 1 オクターブごとに 12dB 以下になるという特徴に起因する．一方で，正しい発声法を身に着けていないアマチュア奏者の歌声は，声帯振動の高調波成分が 1 オクターブごとに 12dB 以上減衰し，“ハリのない歌声” となる．そこで，第 3 フォルマント付近の調波成分を強調する処理を行うことで，歌唱修正を行っている．

熟練した奏者の逸脱モデルからの乖離度を，楽器演奏の習熟度評価に適用した例として，三浦らのピアノ演奏習熟度評価法 [81] が挙げられる．この手法では，熟練した奏者のピアノの 1 オクターブの音階練習は，音量，テンポの逸脱ともに滑らかに変動するという仮定を置く．そして，実測の逸脱変動のなめらかさを数値的に評価し，習熟度を推定している．

第8章　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量の解析法を提案した．3 章では，聴衆の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ることにより，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，連続励起振動楽器の楽音制御の不確定性を内包する楽音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より，A–to–S が 10.6%，S–to–R が 51.2 %減少した．5 章では，演奏表現に起因する音響的変動は時間方向に統計的な一貫性を持つという観点から音量軌跡の生成モデルを立てることにより，従来困難とされていた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が 0.75dB で分解可能であることを示した．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．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，高品質な楽音合成が出来ることを示した．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，楽音修正を行えることを示した．

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と考える．音響信号を “音楽” とみなし，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により，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の展望として，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関解析が挙げられる．例えば擦弦楽器や歌唱の音高変動は，音量や音色に影響を及ぼすことが知られている．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．

謝辞

まず始めに，付属高校 3 年次の成果発表会から 6 年間指導していただいた恩師であり，本論文の主査を務めていただいた，伊藤克亘教授に敬意をこめて深く感謝します．先生からは，議論のたびに刺激的かつ的確なアドバイスを頂きました．研究以外にも，インターンシップや海外の大学への訪問，また就職活動でのアドバイスなど，数えきれないほどの助言や成長の機会を与えていただきました．

お忙しい中副査を快く引き受けていただいた，尾花賢教授と小池崇文教授に感謝します．他分野の視点から見た提案法に対する有益なご助言により，修士論文の完成度と可読性が高まりました．

English Corner でお世話になったマイケル・マクドナルド教授と劉少英教授に感謝します．お二方から教えていただいたテクニカルライティングやスピーキングの技術は，英語論文を読む際や国際会議の討論など，様々な場面で活かされました．また，マクドナルド教授には，数多くの英文とスライドを添削していただきました．

NTT コミュニケーション科学基礎研究所の柏野邦夫博士，亀岡弘和准教授，大石康智博士，中野允裕氏には，学外実習でお世話になりました．同実習ではモデル提案時の数理的な扱いの厳密さ・厳格さの重要性を教えていただきました．

旭化成情報技術研究所の庄境誠博士，中川竜太博士，山川暢英氏には，インターンシップでお世話になりました．同インターンシップで学んだ技術が，3 章で提案した複素メルスペクトル KL 情報量のアイディアとなりました．

ポンペウファブラ大学のザビエル・セラ准教授，エステバン・マエストレ博士には，様々なアドバイスを頂きました．4 章で提案した音符内状態推定は，同氏らから重要性を説かれて開発したものです．

情報処理学会音楽情報科学研究会の皆様には様々なアドバイスを頂きました．特に嵯峨山茂樹教授，北原鉄朗博士，森勢将雅助教授，阪上大地氏に頂いたご助言や激励のお言葉は筆者の宝物です．

最後に，大学院の研究を何不自由なく行えるよう常に陰から惜しみなく支え続けてくれた家族に感謝します．

付 録A　アライメントデータセット

アライメント評価実験のデータセットは，Music Information Retrieval Evaluation eXchange（MIREX）の発音時刻検出データセット [82] からサキソフォン，クラリネット，トランペットの独奏音を各 1 フレーズずつ（表 A.1），RWC 研究用音楽データベース [83] からフルート，トランペットの独奏音を各 1 フレーズずつ（A.2），我々が収録したバイオリンの *legato*，*marcato* などの，様々な奏法を含む独奏 5 フレーズ（表 A.3）の計 10 フレーズを用いた．我々の録音は，全て 1^{st} バイオリンの楽譜を用い，音楽スタジオで 196kHz，24bit で録音した．これらは，吹奏楽器と擦弦楽器が半数ずつかつ，ジャズやクラシックの様々な奏法を含むという理由で選択した．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．全ての録音は，処理前にモノラル化し，標本化周波数 48kHz にリサンプリングした．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．

表 A.1: Onset Leveau

楽器名	ファイル名	演奏時間 (sec)
トランペット	trumpet1.wav	14
クラリネット	clarinet1.wav	30
サキソフォン	sax1.wav	12

表 A.2: RWC 研究用音楽データベース: ジャズ音楽

楽曲名	選択時間
No.12 For Two (Flute & Piano Duo)	5:03–5:22
No.22 For Two (Piano Trio & Tp)	6:48–6:57

表 A.3: 収録楽曲

作曲者	楽曲名	小節番号
C. Petzold	Minuet (BWV Anh. 114)	1–8
A. Vivaldi	The Four Seasons - Spring - I	1–7
F. Schubert	Death and the Maiden - III	1–9
E. Grieg	Holberg Suite - II	1–4
R. Wagner	Tannhauser - Grand March	64–68

参考文献

[1] C. Palmer, “Music performance,” Annu. Rev. Psychol., vol. 48, pp. 115–138, 1997.

[2] B. H. Repp, “Diversity and commonality in music performance: an analysis of timing microstructure in Schumann’s “Traumerei”.” J. Acoust. Soc. Amer., vol. 92, pp.2546–2568, 1992.

[3] M. Clynes, “Microstructural musical linguistics: composers’ pulses are liked most by the best musicians,” Cognition: Int. J. Cogn. Sci., vol. 55, pp. 622–641, 1990.

[4] R. Ramirez, E. Maestre and X. Serra, “Automatic performer identification in commercial monophonic jazz performances,” Pattern Recognition of Non-Speech Audio, vol.31, no.12, pp.1514–1523, 2010.

[5] K. Jensen, “Timbre Models of Musical Sounds,” PhD. theses, University of Copenhagen, 1999.

[6] 安部武宏ほか, “音高による音色変化を考慮した楽器音の音高・音長操作手法” 情報処理学会研究報告, MUS-76, 2008.

[7] E. Maestre and E. Gomez, “Automatic characterization of dynamics and articulation of expressive monophonic recordings,” Proc. the 118th Audio Eng. Society Convention, 2005.

[8] M. Caetano J.J. Burred, X. Rodet, “Automatic Segmentation of the Temporal Evolution of Isolated Acoustic Musical Instrument Sounds Using Spectro-Temporal Cues,” Proc. Int. Conf. on Digital Audio Effects, 2010.

[9] H. Fletcher and L.C. Sanders, “Quality of Violin Vibrato Tones,” J. Acoust. Soc. Am. 41, 1534, 1967.

[10] D. Young, “A Methodology for Investigation of Bowed String Performance Through Measurement of Violin Bowing Technique,” PhD Thesis. MIT., 2007.

[11] G.De Poli, A. Roda and A. Vodolin, “Note-by-note analysis of the influence of expressive intentions and musical structure in violin performance,” Journal of New Music Research , vol. 27, no. 3, pp. 293–321, 1998.

[12] Y. Ohishi, H. Kameoka, D. Mochihashi and K. Kashino, “A Stochastic Model of Singing Voice F0 Contours for Characterizing Expressive Dynamic Components,” In Proc. International Conference on Spoken Language Processing (INTERSPEECH 2012), 2012.

[13] T. Nakano, M. Goto, and Y. Hiraga, “An Automatic Singing Skill Evaluation Method for Unknown Melodies Using Pitch Interval Accuracy and Vibrato Features,” in Proc. of

the International Conference on Spoken Language Processing (INTERSPEECH 2006), pp.1706–1709, 2006.

[14] E. Stamatatosa and G. Widmerb, "Automatic identification of music performers with learning ensembles," Artificial Intelligence, Vol. 165, Issue 1, pp. 37–56, 2005.

[15] S. Canazza, G. De Poli, C. Drioli, A. Roda, "Modeling and control of expressiveness in Music Performance," In Proc. of IEEE, Vo.92, pp. 686–701, 2004.

[16] 大石康智, 持橋 大地, 亀岡 弘和, 柏野 邦夫, "混合ガウス過程に基づく歌声音量軌跡の生成モデル," 情報処理学会研究報告, MUS-100, 2013.

[17] H. Takeda, T. Nishimoto, and S. Sagayama, "Rhythm and tempo recognition of music performance from a probabilistic approach," in Proc. of 5th International Conference on Music Information Retrieval (ISMIR), pp. 357–364, 2004.

[18] K. Hirata and R. Hiraga, "Ha-hi-hun: Performance rendering system of high controllability," ICAD 2002 Rencon Workshop, pp. 40–46, 2002.

[19] G. Widmer, S. Flossmann and M. Grachten, "Yqx plays chopin," AI Magazine, 30, 3, pp. 35–48, 2009.

[20] P. Grosche and M. Muller, "Extracting predominant local pulse information from music recordings," IEEE Transactions on Audio, Speech, and Language Processing, vol. 19, no. 6, pp. 1688–1701, 2011.

[21] P. Desain, and H. Honing, "Tempo curves considered harmful," Contemporary Music Review, Vol. 7, Issue 2, 1993.

[22] 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.

[23] J.P. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies and M. Sandler, "A tutorial on onset detection in music signals", IEEE Trans. Audio, Speech, & Lang. Process., vol.13, no.5, pp.1035–1047, 2005.

[24] M. Goto, "An Audio-based Real-time Beat Tracking System for Music With or Without Drum-sounds", J. of New Music Research, vol.30, no.2, pp.159-171, 2001.

[25] J.P. Bello and M. Sandler, "Phase-based note onset detection for music signals", Proc. Int. Conf. on Acoust., Speech, & Signal Process., pp.49–52, 2003.

[26] J.P. Bello, C. Duxbury, M. Davies and M. Sandler, "On the use of phase and energy for musical onset detection in the complex domain", IEEE Signal Process. Letters, vol.11, no.6, pp.553–556, 2004.

[27] S. Hainsworth and M. Macleod, "Onset detection in musical audio signals", Proc. of Int. Comput. Music Assoc., 2003.

[28] P. Renevey, et al., "Entropy based voice activity detection in very noisy conditions," in Proc. EUROSPEECH, 2001.

[29] G. Peeters, "A large set of audio features for sound description (similarity and classification) in the CUIDADO project", `http://www.ircam.fr/anasyn/peeters/ARTICLES/Peeters_2003_cuidadoaudiofeatures.pdf.`, 2003.

[30] E. B. Fox, E. Sudderth, M. Jordan, and A. Willsky, "The sticky HDP-HMM: Bayesian Nonparametric Hidden Markov Models with Persistent States," Tech. Rep., MIT Laboratory for Information and Decision Systems, 2007.

[31] David M. Blei, M. I. Jordan, "Variational inference for dirichlet process mixtures," Bayesian Analysis, vol. 1, pp. 121–144, 2005.

[32] A. Rodrigueza, D. B. Dunsona and A. E. Gelfand, "The nested dirichlet process," the American Statistical Association, pp. 1131–1154, 2008.

[33] E. B. Fox, E. Sudderth, M. Jordan, and A. Willsky, "Bayesian Nonparametric Inference of Switching Linear Dynamical Systems," IEEE Trans. on Signal Processing, Vol. 59 Issue 4, pp. 1569–1585, 2011.

[34] C. E. Rasmussen, "The infinite gaussian mixture model," in In Advances in Neural Info. Proces. Systems, pp. 554–560, 2000.

[35] M. Demoucron, "On the control of virtual violins: Physical modelling and control of bowed string instruments," Ph.D. thesis, Universite Pierre et Marie Curie (UPMC), Paris, France and Royal Institute of Technology (KTH), Stockholm, Sweden, 2008.

[36] J. Bonada and X. Serra, "Synthesis of the singing voice by performance sampling and spectral models," IEEE Signal Processing Magazine, 24 (2), pp. 67-79, 2007.

[37] L. Cremer, "Physics of the Violin" The MIT Press, Cambridge, MA, pp.201–382, 1984.

[38] 村上智之, "擦弦振動の平均法による解析," 日本機械学會論文集. C-62(598), pp. 2102–2109, 1996.

[39] M.E.Mcintyre, R.T.Schumacher and J.Woodhouse, "Aperiodicity in bowed-string motion," Acustica 49, pp. 13–32, 1981.

[40] K. Popp and P.Stelter, "Stick-Slip Vibrations and Chaos," Philosophical Transactions: Physical Sciences and Engineering. vol.332 No.1624, pp. 89–105, 1990.

[41] K. Guettler, "On the Creation of the Helmholtz Motion in Bowed Strings," Acta Acustica united with Acustica, Vol.88, No.6 ,pp.970-985, 2002.

[42] K. Guettler and A. Askenfelt, "Acceptance limits for the duration of pre-Helmholtz transients in bowed string attacks" J. Acoust. Soc. Amer., vol. 101, 1997.

[43] P.R. Cook and D. Trueman, "A database of measured musical instrument body radiation impulse responses, and computer applications for exploring and utilizing the measured filter functions," in Proc. 1998 Int. Symp. Musical Acoust., 1998.

[44] C.A. Perez, J. Bonada, J. Patynen and V. Valimaki, "Method for measuring violin sound radiation based on bowed glissandi and its application to sound synthesis," J. Acoust. Soc. Amer., 2011.

[45] C. Chafe, “Pulsed Noise in Self-Sustained Oscillations of Musical Instruments,” In Proc. of the International Conference on ICASSP, 1990.

[46] C. Joder, S. Essid and G. Richard, “Hidden Discrete Tempo Model: A Tempo-Aware Timing Model for Audio-to-Score Alignment,” ICASSP-11, pp.397-400, 2011.

[47] H. Akaike, “Information theory and an extension of the maximum likelihood principle,” Proc. the 2nd Int. Sympo. on Information Theory, 1, pp. 267–281, 1973.

[48] 水野 優, 小野 順貴, 西本 卓也, 嵯峨山 茂樹 “パワースペクトログラムの伸縮に基づく多重音信号の再生速度と音高の実時間制御,” 聴覚研究会資料, 39, pp. 447–452, 2009.

[49] D. W. Griffin and J. S. Lim: “Signal estimation from modified short-time fourier transform,” IEEE Trans. Audio, Speech, & Lang. Process., 32, 2, pp. 236–243, 1984.

[50] J.J. Burred and A. Robel, “A Segmental Spectro-Temporal Model of Musical Timbre,” 13th International Conference on Digital Audio Effects (DAFx-10), 2010.

[51] H.H.Hall, “Sound Analysis,” J. Acoust. Soc. Am. 8, 257, 1937.

[52] H.W. Eagleson and O.W. Eagleson, “Identification of Musical Instruments when heard directly and over a public-address system,” J. Acoust. Soc. Am. 19, pp. 338–342, 1947.

[53] W.H. George, “A sound reversal technique applied to the study of tone quality,” Acoustica, 4, pp. 224–225, 1954.

[54] J.C. Risset and M.V. Mathews, “Analysis of musical instruments tone,” Physis Today, 22, 2, 1969.

[55] N. H. Adams, M. A. Bartsch, J. B. Shifrin and G. H. Wakefield, “Time series alignment for music information retrieval,” in Proc. ISMIR-04, pp. 303–310, 2004.

[56] N. H. Adams, M. A. Bartsch and G. H. Wakefield, “Note segmentation and quantization for music information retrieval,” IEEE Trans. Audio, Speech, & Lang. Process., 14, pp. 131–141, 2006.

[57] M. Jeub, M. Schafer, P. Vary “A Binaural Room Impulse Response Database for the Evaluation of Dereverberation Algorithms” Proc. Int. Conf. on Digital Signal Process., 2009.

[58] J. Schluter and S. Bock, “Musical Onset Detection with Convolutional Neural Networks, ’’ In Proceedings of the 6th International Workshop on Machine Learning and Music, 2013.

[59] S. Bock and G. Widmer, “Local Group Delay based Vibrato and Tremolo Suppression for Onset Detection, ’’In Proceedings of the 14th International Society for Music Information Retrieval Conference, 2013.

[60] S. Liu, M. Yamada, N. Collier and M. Sugiyama, “Change-Point Detection in Time-Series Data by Relative Density-Ratio Estimation, ’’Structural, Syntactic, and Statistical Pattern Recognition, Lecture Notes in Computer Science, Vol. 7626, pp 363–372, 2012.

[61] 三浦種敏 監修, “新版 聴覚と音声” コロナ社, 1980.

[62] E. B. Fox, et al., “An HDP-HMM for systems with state persistence,” in Proc. on ICML, pp. 312–319, 2008.

[63] M. Nakano, J.L. Roux, H. Kameoka, T. Nakamura, N. Ono and S. Sagayama, “Bayesian Nonparametric Spectrogram Modeling Based on Infinite Factorial Infinite Hidden Markov Model,” in Proceedings of Applications of Signal Processing to Audio and Acoustics (WASPAA), pp. 325–328, 2011.

[64] E. Fox, “Bayesian nonparametric learning of complex dynamical phenomena,” Ph.D. thesis, MIT, 2009.

[65] P. Jong, “The Simulation Smoother for Time Series Models,” Biometrika, Vol. 82, No. 2 pp. 339–350, 1995.

[66] K. Teramura, H. Okuma, Y. Taniguchi, S. Makimoto and S. Maeda, “Gaussian Process Regression for Rendering Music Performance,” In Proc. ICMPC, 2008.

[67] D. D. Lee and H. S. Seung, “Learning the parts of objects by non-negative matrix factorization,” Nature, Vol. 401, No. 6755, pp. 788–791, 1999.

[68] H. Kameoka, T. Nishimoto and S. Sagayama, “A Multipitch Analyzer Based on Harmonic Temporal Structured Clustering,” IEEE Transactions on Audio, Speech, and Language Processing, Vol.15 , Issue 3, 2007.

[69] D. Sakaue, T. Otsuka, K. Itoyama, H. G. Okuno, “Initialization-Robust Bayesian Multipitch Analyzer based on Psychoacoustical and Musical Criteria,” In Proceedings of 2013 International Conference on Acoustics, Speech and Signal Processing (ICASSP 2013), 2013.

[70] 森勢将雅, 高橋徹, 河原英紀, 入野俊夫, “窓関数による分析時刻の影響を受けにくい周期信号のパワースペクトル推定法,” 電子情報通信学会論文誌. D, 情報・システム, J90-D(12), 3265-3267, 2007.

[71] H. Kawahara, M. Morise, “Technical foundations of TANDEM-STRAIGHT, a speech analysis, modification and synthesis framework” Sadhana, Vol.36, Part 5, pp.713-727, Oct, 2011.

[72] E. Maestre, M. Blaauw, J. Bonada, 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), pp. 855-871, 2010.

[73] T. Toda, A.W. Black and K. Tokuda, “Voice conversion based on maximum likelihood estimation of spectral parameter trajectory,” IEEE Transactions on Audio, Speech and Language Processing, Vol. 15, No. 8, pp. 2222–2235, 2007.

[74] K. Tokuda, T. Yoshimura, T. Masuko, T. Kobayashi and T. Kitamura, “Speech parameter generation algorithms for HMM-based speech synthesis,” In Proc. of ICASSP, pp.1315–1318, June 2000.

[75] H. Kawahara, H. Banno, T. Irino and P. Zolfaghari, “ALGORITHM AMALGAM: Morphing waveform based methods, sinusoidal models and STRAIGHT,” In Proc. ICASSP, pp.13–16, 2004.

[76] T. Nakano and M. Goto, “VocaListener2: A Singing Synthesis System Able to Mimic a User’s Singing in Terms of Voice Timbre Changes as well as Pitch and Dynamics,” In Proceedings of the 36th International Conference on Acoustics, Speech and Signal Processing (ICASSP2011), pp.453–456, 2011.

[77] D. Deutsch 編, 寺西立年ほか監訳, “音楽の心理学 (上),” 西村書店, 1987.

[78] D. Deutsch 編, 寺西立年ほか監訳, “音楽の心理学 (下),” 西村書店, 1987.

[79] T. Nakano and M. Goto: “Vocalistener: A singing-to-singing synthesis system based on iterative parameter estimation,” Proc. SMC-2009, pp. 343–348, 2009.

[80] 木立 真希, 伊藤 克亘, “呼気量のモデル化に基づく歌唱修正システム,” 情報処理学会 第 76 回全国大会, 2014.

[81] 三浦 雅展, 江村 伯夫, 秋永 晴子, 柳田 益造, “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.

[82] P. Leveau, L. Daudet, G. Richard, “Methodology and Tools for the evaluation of automatic onset detection algorithms in music”, In Proc. International Symposium on Music Information Retrieval, 2004.

[83] M. Goto, H. Hashiguchi, T. Nishimura, R. Oka, “RWC Music Database: Popular, Classical, and Jazz Music Databases”, In Proc. International Conference on Music Information Retrieval, 2002.

研究業績

J-1) 小泉 悠馬，伊藤 克亘, “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,” 電子情報通信学会論文誌, Vol.J 97-D, No.3, 2014(in press).

J-2) 小泉 悠馬，伊藤 克亘, “擦弦楽器の意図表現合成のための奏法モデル,” 情報処理学会論文誌，Vol.54，No.4，pp. 1319–1326，Apr. 2013.

J-3) Y. Koizumi，K. Itou, “Performance expression synthesis for bowed-string instruments using “Expression Mark Functions”,” Proceedings of Meetings on Acoustics (POMA). Vol. 15, pp. 035003, Nov. 2012.

I-4) Y. Koizumi，K. Itou, “Intra-note Segmentation via Sticky HMM with DP Emission,” in Proc. of International Conference on Acoustics, Speech and Signal Processing (ICASSP 2014), May, 2014, (accepted).

I-5) Y. Koizumi，K. Itou, “Expressive Oriented Time-Scale Adjustment for Mis-played Musical Signals based on Tempo Curve Estimation,” the 16th International Conference on Digital Audio Effects Conference (DAFx-16), Sept., 2013.

I-6) Y. Koizumi，K. Itou, “Synthesis of performance expression of bowed string instruments using “Expression Mark Functions”,” The Acoustics 2012 Hong Kong Conference and Exhibition, May, 2012.

D-7) 小泉 悠馬，伊藤 克亘, “ディリクレ過程を出力する Nest 型 HMM を用いた音符内状態推定,” 日本音響学会 2014 年春季研究発表会 講演論文集, 2014(in press).

D-8) 小泉 悠馬，伊藤 克亘, “連続励起振動楽器を対象とした音量軌跡のダイナキクスとアーティキュレーションへの分解法,” 情報処理学会研究報告, SIGMUS-102, 2014 (in press).

D-9) 小泉 悠馬，伊藤 克亘, “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 ,” FIT2013 第 12 回情報科学技術フォーラム, Sept., 2013, （船井ベストペーパー賞受賞）.

D-10) 小泉 悠馬，伊藤 克亘, “連続励起振動楽器のためのパワーに基づく音符内状態推定,” 日本音響学会 2013 年秋季研究発表会, 3-3-2, pp. 923–926, Sep. 2013.

D-11) 小泉 悠馬，伊藤 克亘, “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～,” 情報処理学会研究報告, 2013-MUS-99-58, May. 2013.

D-12) 小泉 悠馬，伊藤 克亘, “演奏音の音量時系列からの奏者の意図表現成分の推定,” 情報処理学会 第 75 回全国大会, 3R-7, Mar. 2013, （学生奨励賞 受賞）.

D-13) 小泉 悠馬，伊藤 克亘, “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, 2012-MUS-97-02, Dec. 2012.

D-14) 小泉 悠馬，伊藤 克亘, “意図表現における非周期擦弦振動を考慮した楽音合成手法の検討,” 日本音響学会 2012 年秋季研究発表会, 2-10-3, pp. 923-926, Sep. 2012, （第 6 回 学生優秀発表賞 受賞）.

D-15) 小泉 悠馬，伊藤 克亘, “擦弦時の奏法行動を考慮した意図表現の合成手法:VIOCODER,” 情報処理学会研究報告, 2012-MUS-95-02, Jun. 2012.

D-16) 小泉 悠馬，伊藤 克亘, “合成音への表現力付与のための擦弦楽器の発想伝達関数の推定,” 情報処理学会 第 74 回全国大会, 4S-1, Mar. 7, 2012.

C-17) 安田 沙弥香, 小泉 悠馬，伊藤 克亘, “ラジオ放送話者ダイアライゼーション,” 情報処理学会 第 76 回全国大会,2014.

C-18) 塩出 萌子, 小泉 悠馬，伊藤 克亘, “中間話者コーパスを用いたアニメーション演技音声のための話者変換,” 情報処理学会 第 76 回全国大会,2014.

C-19) 上野 涼平, 小泉 悠馬，伊藤 克亘, “音楽知識を利用したハーモナイザー,” 情報処理学会 第 75 回全国大会,2013.

C-20) 森田 花野, 小泉 悠馬，伊藤 克亘, “教則本を利用したギターフレーズの難易度推定,” 情報処理学会 第 75 回全国大会,2013.