

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

小泉 悠馬^{1,a)} 伊藤 克亘²

概要: 連続励起振動楽器の演奏音の音量変化(音量軌跡)を、音量記号などに起因するベースラインの大局的变化であるダイナミクスと、発想記号などに起因する局所的变化であるアーティキュレーションに分解する手法を提案する。本稿では、奏者はフレーズに統一感を持たせるために、楽曲の小区間では、いくつかのアーティキュレーションや奏法を選択的に組み合わせて演奏すると仮定する。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する。評価実験は、MIDI データを用いて作成した人口データと、プロ奏者とアマチュア奏者の音量軌跡の分解実験を行った。人口データの分解実験からは、標準絶対誤差が 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。また実演奏音の分解実験では、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た。

キーワード: ダイナミクス, アーティキュレーション, ディリクレ過程, 線形動的システム

Decomposition of an Amplitude Contour into a Dynamics and Articulations for Excitation-continuous Musical Instruments

YUMA KOIZUMI^{1,a)} KATUNOBU ITOU²

Abstract: This paper proposes a method which decomposes an amplitude contour into a dynamics contour and an articulation contour. The dynamics is a global fluctuation arise from dynamic indications, and the articulation is a local fluctuation arise from expression marks. In this paper, the dynamics and the articulation are modeled by random walk and auto-regressive (AR) model, respectively. These models are embodied by using a switching linear dynamical system (SLDS), and an amplitude contour is decomposed by the SLDS. In our experiments, the proposed method successfully decomposed amplitude contours into dynamics and articulation contours.

Keywords: Dynamics, Articulation, Dirichlet process, Linear dynamical system.

1. まえがき

音符内の音量の時間的变化は、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。2 種類の変動量は、奏者が楽譜に記載された演奏記号や音高の昇降などから決定した演奏プラン [1] に基づく。そして、より高度で抽象的な演奏プランを生成する能力や、演奏プランを実際の演奏に正確に変換する能力が「習熟度」や「技巧」と呼ばれる。

本研究の目的は、このような奏者の演奏表現や習熟度による音量軌跡の変化を解析することである。また奏者ごと

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

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

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

の軌跡変動パターンや、音符列とパターンの依存関係を学習することで、自動熟達度評価 [2] や自動演奏 [3]、奏者認識 [4] などへの応用が期待できる。しかし、従来の音量軌跡を用いた熟達度評価や自動演奏は、音符内の音量変動要因が励起振動の減衰のみで、アーティキュレーションが固定とみなせる撥/打弦楽器が主流であった。これは、擦弦楽器などの連続励起振動楽器は、音符内で励起源にロードを加え続けるため、動特性が 2 成分のどちらに起因するかの判別が困難なためである。

よって、連続励起振動楽器の音量軌跡の特徴抽出は、ダイナミクスを持続部の平均音量 [5] や、持続部を直線で結んだもの [6] とし、アーティキュレーションは、発音時刻の間隔 (IOI) と音符の継続時間 (DR) の比率とする [7] などの簡易な手法がとられてきた。また、これらの特徴量は楽音生成に直接応用することが困難なため、生成では、ADSR (Attack, Decay, Sustain, Release) 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 [8] がとられてきた。

本稿では、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。

2 章で述べる、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 [9] (HDP-SLDS) 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。

2. 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 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 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。

3. 音量軌跡の生成モデル

3.1 音量軌跡の線形動的システム表現

観測演奏音をフーリエ変換して得られるスペクトログラムを $X_{\omega,t} \in \mathbb{C}^{\Omega \times T}$ 、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。

$$y_t = 20 \log_{10} \sum_{\Omega} |X_{\omega,t}| = f_t + g_t \quad (1)$$

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。

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

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。

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 (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。

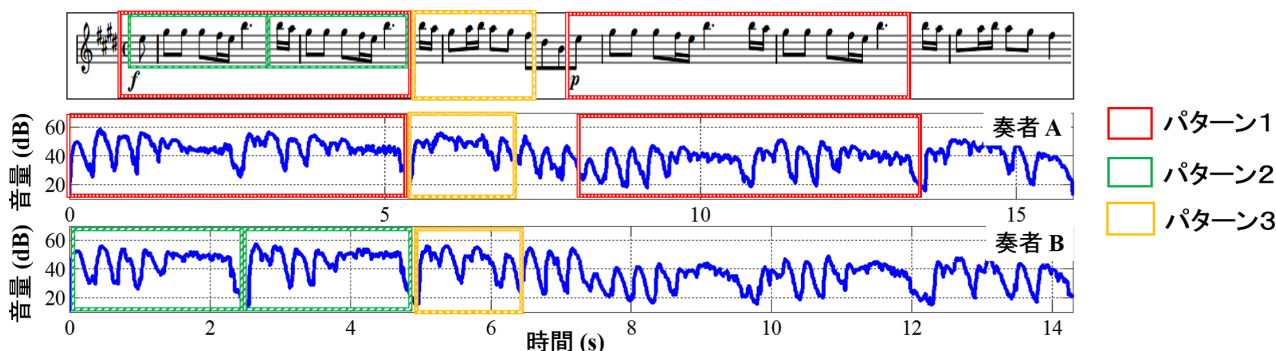


図 1 バイオリンの音量軌跡例．楽曲は A. Vivaldi の“四季”より“春”の 1 楽章冒頭．

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クレ過程隠れマルコフモデル (HDP-HMM) [10] を適用する．

$$z_t \sim \pi_{z_{t-1}}, \quad \pi_k \sim \text{DP} \left(\alpha + \kappa, \frac{\alpha \beta + \kappa \delta_j}{\alpha + \kappa} \right) \quad (3)$$

$$\beta_k = \nu_k \prod_{l=1}^{k-1} (1 - \nu_l), \quad \nu_k \sim \text{Beta}(1, \gamma) \quad (4)$$

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”とみなすことができる．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．

$$g_t = g_{t-1} + e_t^g, \quad e_t^g \sim \mathcal{N}(0, \sigma_g^2) \quad (5)$$

よって，式 (1)(2)(5) より時刻 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．

$$\mathbf{x}_t = \begin{bmatrix} A^{z_t} \\ 1 \end{bmatrix} \mathbf{x}_{t-1} + \mathbf{e}_t^{z_t}, \quad y_t = \mathbf{U} \mathbf{x}_t \quad (6)$$

ただし， $\mathbf{x}_t = (f_t, f_{t-1}, \dots, f_{t-r+1}, g_t)^\dagger$ ， $\mathbf{e}_t^{z_t} = (e_t^f(z_t), 0, \dots, 0, e_t^g)^\dagger \in \mathbb{R}^{(r+1) \times 1}$ であり， A^k は k 番目の奏法に対応する VAR(r) 行列， $\mathbf{U} = (1, 0, \dots, 0, 1) \in \mathbb{N}^{1 \times (r+1)}$ 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 $\mathbf{x}_{1, \dots, T}$ 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 $\text{IG}(\nu, \psi)$ 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．

3.2 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 $z_{1, \dots, T}$ を効率的に求めるために，音色の音響特徴量も用いる．

楽音の音色知覚には，倍音の強度比や発音区間の非調波性が重要と知られている [11]．よって音色の音響特徴量は，非周期性とスペクトル包絡に関するものを用いる．

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 の振幅スペクトルで調波成分の支配的度を数値化したい．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 [12] を用いる．また楽音のスペクトル包絡を表す特徴量は，スペクトルのセントロイドやカートシスが有効といわれている [13]．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 つのモーメントに対し主成分分析を行う．そして寄与率の高い順から 3 次元 ($\mathbf{c}_t = (c_t^1, c_t^2, c_t^3)^\dagger$) を特徴量として用いる．

そして時刻 t で観測される音色特徴量 \mathbf{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 [14] から出力されたものとみなす．

$$\mathbf{c}_t^k \sim \mathcal{N}(\mu_k^c, \Sigma_k^c) \quad (7)$$

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 2 に示す．

4. 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， $z_t, \pi, \beta, \Upsilon, \theta, \alpha, \kappa, \zeta, \mathcal{H}, \mathbf{x}_t$ の順に行う．基本的なアルゴ

リズムについては文献 [9][14][15] 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。

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 (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、

$$p(z_t|z_{t-1}, \boldsymbol{\pi}, f_{1-r:T}, \mathbf{c}_t, \Theta, \Upsilon) \quad (8)$$

$$\propto p(z_t|\boldsymbol{\pi}_{z_{t-1}})p(f_t|\mathbf{f}_{t-1}, \Theta_{z_t})p(\mathbf{c}_t|\Upsilon_{z_t})m_{t+1,t}(z_t)$$

となる。ここで $\mathbf{f}_t = (f_t, f_{t-1}, \dots, f_{t-r+1})^\dagger$ であり、 $m_{t+1,t}(z_t)$ は、遷移 $z_t \rightarrow z_{t-1}$ 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

$$z_t \sim \sum_{k=1}^K \pi_{z_{t-1},k} P_k(\mathbf{f}_t, \mathbf{c}_t) \delta(z_t, k) \quad (9)$$

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

$$P_k(\mathbf{f}_t, \mathbf{c}_t) = \mathcal{N}\left(f_t \left| \sum_{i=1}^r a_i^k f_{t-i}, \sigma_f^2(k) \right.\right) \mathcal{N}(\mathbf{c}_t | \boldsymbol{\mu}_k^c, \Sigma_k^c) m_{t+1,t}(k) \quad (10)$$

$$m_{t,t-1}(k) = \sum_{j=1}^K \pi_{k,j} \mathcal{N}\left(f_t \left| \sum_{i=1}^r a_i^j f_{t-i}, \sigma_f^2(j) \right.\right) \mathcal{N}(\mathbf{c}_t | \boldsymbol{\mu}_j^c, \Sigma_j^c) m_{t+1,t}(j) \quad (11)$$

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

まず、文献 [10] の手法を用いて $\boldsymbol{\pi}, \beta$ をサンプリングする。次に、文献 [14] の手法を用いて Υ をサンプリングする。

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 $A^k, \sigma_f^2(k)$ 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。

$$\sigma_f^2(k) \sim \mathcal{IG}\left(\nu + \frac{N_k}{2}, \psi + \frac{S}{2}\right) \quad (12)$$

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

$$\mathbf{a}_k \sim \mathcal{N}\left(\mathbf{S} \mathbf{m}, \sigma_f^2(k) \mathbf{S}\right) \quad (13)$$

となる [9]。ただし \mathbf{m} と \mathbf{S} は以下となる。

$$\mathbf{m} = \mathbf{F}_k \bar{\mathbf{F}}_k^\dagger, \quad \mathbf{S} = \left(\Sigma_A^{-1} + \bar{\mathbf{F}}_k \bar{\mathbf{F}}_k^\dagger\right)^{-1} \quad (14)$$

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

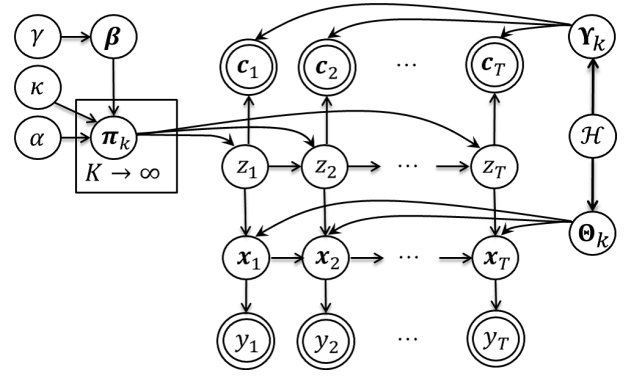


図 2 提案法のグラフィカル表現。二重丸が観測データを表す。

まず、文献 [10] の手法を用いて $\alpha, \kappa, \varsigma$ をサンプリングする。次に、文献 [14] の手法を用いて \mathcal{H} の iGMM に関するパラメータをサンプリングする。 \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: 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 $\mathbf{x}_1, \dots, \mathbf{x}_T$ の推論は、カルマンスムーザで行うことができる。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 [16] とも呼ばれる。本稿では推論の高速化のために、Fox らの Block Sampling による手法 [9] で推論を行う。

5. 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準化周波数 48kHz にリサンプリングした。音量 $y_{1,\dots,T}$ は式 (1) から求め、短時間フーリエ変換 (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 [14] 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。

5.1 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 1)。本実験では、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 (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. *mp* から *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。

5.2 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、標準化周波数 192kHz で録音した。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の、プロ奏者の演奏を模倣するように演奏した。図 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 回繰り返す、その後 4 拍の補充を入れるという旋律パターンを、

表 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

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ズ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つ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。

6. 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

*1 <http://ocmnet.com/saxguru/Timidity.htm#sf2>
 (2014/01/24 アクセス)

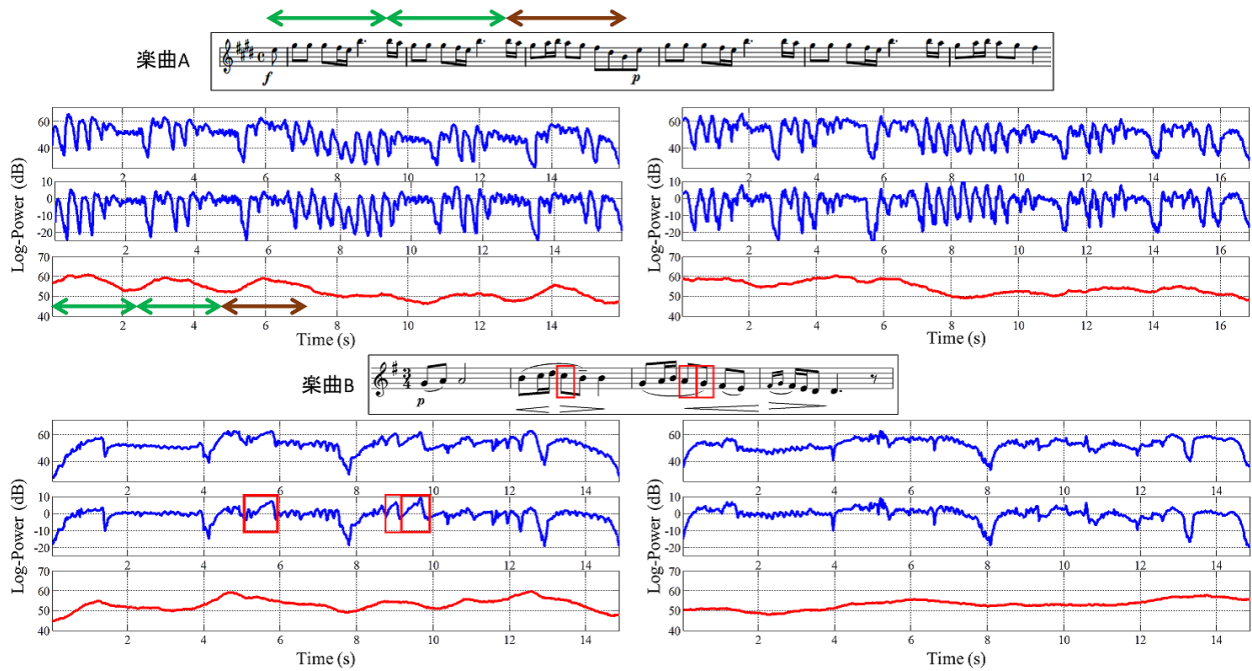


図 3 プロ奏者(左)とアマチュア奏者(右)のバイオリン演奏音への推定結果. 3 つ並んだ音量グラフは上から, 実測音量軌跡, アーティキュレーションの推定結果, ダイナミクス
 の推定結果を示す.

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.

参考文献

[1] D. Deutsch 編, 寺西立年ほか監訳, “音楽の心理学 (下),” pp. 588–609, 西村書店, 1987.
 [2] 三浦 雅展, 他, “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.
 [3] K. Teramura, et al., “Gaussian Process Regression for Rendering Music Performance,” In Proc. ICMPC, 2008.
 [4] R. Ramirez, et al.: “Automatic performer identification in commercial monophonic jazz performances”, Pattern Recognition of Non-Speech Audio, vol.31, no.12, pp.1514–1523, 2010.
 [5] E. Stamatatos, et al., “Automatic identification of music performers with learning ensembles,” Artificial

Intelligence, Vol. 165, Issue 1, pp. 37–56, 2005.
 [6] E. Maestre et al., “Automatic characterization of dynamics and articulation of expressive monophonic recordings,” In Proc. the 118th Audio Eng. Society Convention, 2005.
 [7] S. Canazza, et al., “Modeling and control of expressiveness in Music Performance,” In Proc. of IEEE, Vo.92, pp. 686–701, 2004.
 [8] 大石康智, ほか, “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.
 [9] E. B. Fox, et al., “Bayesian Nonparametric Inference of Switching Linear Dynamical Systems,” IEEE Trans. on Signal Processing, Vol. 59 Issue 4, pp. 1569–1585, 2011.
 [10] E. B. Fox, et al., “The sticky HDP-HMM: Bayesian Nonparametric Hidden Markov Models with Persistent States,” Tech. Rep., MIT Lab. for Info. and Decision Systems, 2007.
 [11] D. Deutsch 編, 寺西立年ほか監訳, “音楽の心理学 (上),” pp. 30–69, 西村書店, 1987.
 [12] P. Renevey, et al., “Entropy based voice activity detection in very noisy conditions,” In Proc. EUROSPEECH, 2001.
 [13] G. Peeters, “A large set of audio features for sound description (similarity and classification) in the CUIDADO project,” http://www.ircam.fr/anasy/peeters/ARTICLES/Peeters_2003_cuidadoaudiofeatures.pdf, 2003.
 [14] C. E. Rasmussen, “The infinite gaussian mixture model,” in In Advances in Neural Info. Proces. Systems, pp. 554–560, 2000.
 [15] E. Fox, “Bayesian nonparametric learning of complex dynamical phenomena,” Ph.D. thesis, MIT, 2009.
 [16] P. Jong et al., “The Simulation Smoother for Time Series Models,” Biometrika, Vol. 82, No. 2 pp. 339–350, 1995.