

演奏音の音量時系列からの奏者の意図表現成分の推定*

小泉 悠馬 (法政大学大学院 情報科学研究科), 伊藤 克亘 (法政大学 情報科学部)

1 まえがき

楽器の奏者は、その曲想や個性を活かすために、楽譜情報を元に楽音をイメージし、奏法に変換する。その過程で、実際の演奏では音の三要素の音量、音色、音高と、音楽特有の要素であるリズムに意図表現による楽譜からの変動成分が含まれる（ビブラートなど）。計算機による、これら変動成分に基づく演奏意図のモデル化は、人間の感性に関する情報処理として、演奏の自動表情付けなどの分野で盛んにおこなわれている [1, 2]。

音量変化からの意図表現推定は、音量系列から、個々の音符の振幅を近似するパラメータ (ADSR) の推定問題などとして扱われている [3, 4]。しかし、この音量変化のモデル化では、音量系列の変化が演奏記号などの楽譜情報による変動なのか、奏者の意図表現による変動なのかは特定できず、またクレシェンドなどの大局的な変動成分も局所変動として扱われるため、意図表現情報の解析技術としては不十分なものであった。

本稿では、音量変化から、フレーズによる大局的な意図表現と音符ごとの局所的な意図表現を個別に推定するために、音量時系列を、楽譜情報、大局的変動成分、局所的な変動成分、奏法誤差に分解する手法を提案する。

2 音量変化の意図表現のモデリング

音響信号の時刻 k における音量を次式で定める。

$$y[k] = 10 \log_{10} \sum_{\omega} |X_{\omega,k}|^2 \quad (1)$$

ここで $|X_{\omega,k}|$ は音響信号をフーリエ変換して得られる振幅スペクトログラムである。

求められた長さ K の音量系列 $y_k = (y[1], \dots, y[K])$ には、演奏記号などの楽譜情報の成分 s_k （以下、楽譜成分）、奏者の意図した変動成分、および楽器の制御ミスによる誤差 e_k が加算されていると仮定する。また、意図表現による楽譜からの変動成分は、奏者がフレーズを理解し演奏することで決まる大局的な変動成分 t_k （以下、フレーズ成分）と、局所的なスラーやアクセントなどのイメージによる、音符ごとの局所的な変動成分 p_k （以下、局所変動成分）の 2 種類の変動成分の和により構成されるとする。すると、音量系列は以下のように書ける。

$$y_k = s_k + t_k + p_k + e_k \quad (2)$$

本稿では音量系列からの意図表現の推定問題を、観測系列の式 (2) の右辺への分解問題として解く。

2.1 音量時系列の状態空間表現

観測系列を各成分へ分解するために、式 (2) を状態空間表現する。楽譜成分 s_k は、楽器の種類および演奏記号によって音量変化の概形が規定され、一つのフレー

ズ内では概形が変化しないものとする。また音符ごとのエンベロープは、音符の長さによって伸縮する。こうすると、楽譜情報は季節調整モデルとして記述できる。フレーズ成分 t_k は、音楽に連続性を持たせるためにフレーズ内ではなめらかに変化すると仮定し、トレンドモデルとして表現する。また、局所変動成分 p_k は AR モデルによって生成されたものとする。これにより、全ての音符の継続時間長が ℓ の場合、式 (2) は以下の状態空間表現として記述することができる。

$$\begin{aligned} \mathbf{x}_k &= \begin{bmatrix} \mathbf{A}_1 & & \\ & \mathbf{A}_2 & \\ & & \mathbf{A}_3 \end{bmatrix} \mathbf{x}_{k-1} + \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \mathbf{b}_3 \end{bmatrix} \mathbf{v}_k \\ y[k] &= [\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3] \mathbf{x}_k + e[k] \end{aligned} \quad (3)$$

ただし、 $\mathbf{x}_k = (s[k], s[k-1], \dots, s[k-\ell+2], t[k], \dots, t[k-m+1], p[k], \dots, p[k-n+1])^T$ 、入力成分 $\mathbf{v}_k = (v_1, v_2, v_3)^T$ 、 $(v_i \sim \mathcal{N}(0, \sigma_i^2))$ であり、 \mathbf{A}_1 は楽譜成分のための周期が ℓ の季節調整状態行列、 \mathbf{A}_2 はフレーズ成分のための m 次のトレンドモデルの状態行列、 \mathbf{A}_3 は局所変動成分のための n 次の AR モデルの状態行列である。また、 \mathbf{b}_i は各成分の入力制御ベクトル、 \mathbf{c}_i は各成分の出力制御ベクトルである。また、 $e[k] \sim \mathcal{N}(0, \sigma_e^2)$ とする。例として時刻 k での楽譜成分は、状態ベクトル \mathbf{x}_k と出力ベクトル $[\mathbf{c}_1, \mathbf{0}_2, \mathbf{0}_3]$ を掛け合わせるにより得ることが出来る。但し $\mathbf{0}$ は要素がすべて 0 のベクトルである。

3 状態推定による音量時系列の分解

式 (3) の状態空間表現を実際の演奏に適用するためには、演奏音中の音符長が、全ての音符で等しくなければならない。しかし、実際の楽譜は音符ごとに音長が異なり、また音長が同じであっても、演奏表現による音長の変動が含まれる。そこで本稿では、入力音響信号から発音時刻を推定し、音符長を適当な定数 L に正規化してから状態推定を行う。

3.1 発音時刻検出

我々は聴覚特性を考慮した発音時刻の検出に、メルスケールのサブバンド振幅スペクトルの KL 情報量 (KLD) を特徴量として提案してきた [5]。しかし、この手法では擦弦楽器などの発音検出に有効とされる位相情報 [6] を用いていなかった。本稿では、位相情報を考慮できるメルスケールの複素スペクトル特徴量を提案する。微小時間 τ 前から予測される複素スペクトルとの乖離度として、KLD を複素領域へ拡張した尺度を用いる。

$$D_{\text{KL}}(k) = \sum_{\omega} \left| \hat{S}_{\omega,k} \log \frac{\hat{S}_{\omega,k}}{S_{\omega,k}} \right| \quad (4)$$

ここで、 $S_{\omega,k}$ は時刻 k において観測されたスペクトル、 $\hat{S}_{\omega,k}$ は時刻 $k - \tau$ から予測される時刻 k のスペクトル

* Decomposition of the Amplitude Envelope by Considering Players Musical Intentions.: Yuma Koizumi (Hosei Univ.) et al.

である．各スペクトルは以下の式で求められる．

$$S_{\omega,k} = \text{mel} \left[\frac{|X_{\omega,k}|}{\sum_{\omega} |X_{\omega,k}|} \right] e^{j \text{mel}[\phi_{\omega,k}]} \quad (5)$$

$$\hat{S}_{\omega,k} = \text{mel} \left[\frac{|X_{\omega,k-\tau}|}{\sum_{\omega} |X_{\omega,k-\tau}|} \right] e^{j \text{mel}[\hat{\phi}_{\omega,k}]} \quad (6)$$

ここで $\phi_{\omega,k}$ は音響信号をフーリエ変換して得られる位相スペクトログラム， $\hat{\phi}_{\omega,k}$ は先行研究 [6] の手法により τ 時間前から予測される位相スペクトログラム， $\text{mel}[\cdot]$ はスペクトルのメルスケールへの変換を表す．

ここで，KLD の値が局所的に大きくなる時刻が発音時刻であると考え， $D_{\text{KL}}(k)$ からピーク値を検出することにより，発音時刻候補集合を生成する．さらに，発音時刻の最適化法 [5] を用いて発音時刻を決定する．

3.2 カルマンフィルタによる音量時系列の分解

まず各状態行列 A_i および，入力成分の分散 σ_i^2 を決定する．季節調整の次数を 1 次とすると，季節調整の周期 p およびトレンドモデルの次数 n を決定することにより，楽譜成分とフレーズ成分の状態行列 A_1, A_2 が決まる [7]．局所変動成分の状態行列 A_3 は，観測系列と，その κ での移動平均 \bar{y}_k との差分を用いて，最小二乗法により AR 係数を推定し決定する．各入力成分の分散は事前実験により，楽譜成分にトレンド成分より柔軟な変化を許すために $\sigma_1^2 = 1$ で固定とし， σ_2^2 は区間 $[0, 0.2]$ で線形探索を用いた最尤推定により決定し， σ_3^2 は AR 係数の推定時の値で固定とし， σ_4^2 は最尤法により決定する．

次に推定された発音時刻を用いて各音符における音量時系列の長さを L に伸縮する．伸縮した時系列を決定されたパラメータを用いて，カルマンフィルタおよび固定区間平滑化のアルゴリズムにより，長さが正規化された各変動成分へと分解する．最後に分解された時系列を元の時間長へ伸縮し， s_K, t_K, p_K を得る．

4 動作実験

プロ奏者とアマチュア奏者の演奏録音を用いて提案法の動作を確認した．各パラメータは，季節周期 $\ell = L$ ，トレンド次数 $m = 1$ ，AR 次数 $n = 2$ ，移動平均長 $\kappa = 50\text{ms}$ ，正規化長 $L = 50$ 点，フーリエ変換点数 1024 点とした．また，音響信号は全て 48kHz にリサンプリングし，音量系列は 1ms ごとに求めた．楽曲はバイオリンのソロ曲¹とした．

音量系列の分解結果を図 1 に示す．楽譜成分には，スタッカート記号による山型の振幅変動が反映されている．またフレーズ成分には，上昇音階からイメージされたと考えられるクレシェンドが反映されている．局所変動成分には，プロ演奏は 3 秒目付近にのみスラーによる局所的に大きな変動成分が推定されているが，アマチュアの演奏は全ての音符において局所変動が推定されている．これは，低熟練度による音符ごとの音量のばらつきよるものであると考えられる．提案法は，奏法誤差による音量変化に各変動成分と無相関な正規雑音を仮定しており，音量のばらつきのような奏法行動

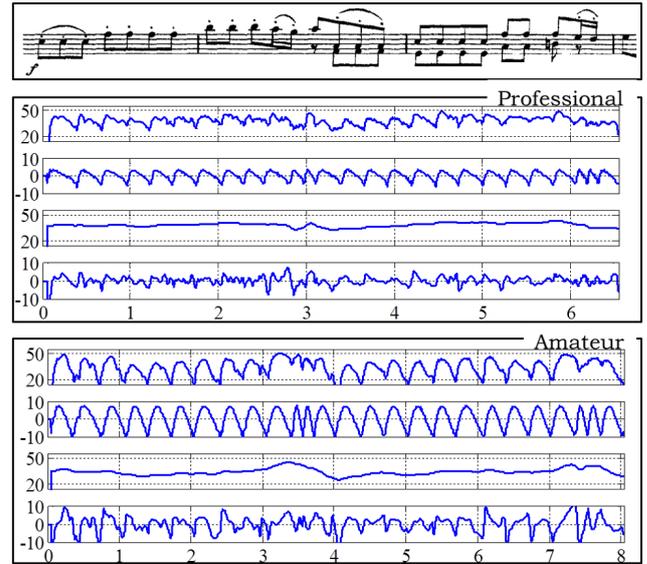


図 1. 演奏に用いた楽譜および音量系列の分解結果．図は上から，観測系列 y_K ，楽譜成分 s_K ，フレーズ成分 t_K ，局所変動成分 p_K であり，縦軸は振幅 (dB)，横軸は時間 (s) を表す．

に起因して発生する誤差は局所変動成分として推定される．今後，誤差成分を楽譜成分と連動して推定する手法を検討するなど，局所変動成分と誤差成分の取り扱い方について検討しなければならない．

これらのことから，提案法は音量系列から楽譜成分とフレーズ成分を推定できるといえる．

5 あとがき

本稿では，音量変化における奏者の意図表現情報の推定のために，音量系列を，楽譜情報，大局的な変動成分，局所的な変動成分，奏法誤差へと分解する手法を提案した．楽音データから発音時刻および音量系列を算出し，状態空間モデルを用いて時系列を分解した．

今後，局所変動成分と誤差成分の取り扱いを検討した上で，提案法の楽音修正や音色の時間的変化のモデル化に応用する．

参考文献

- [1] Yuichi Uwabu, et al., “A Structural Analysis Tool for Expressive Performance,” of International Computer Music Conference, pp.121-124, 1997.
- [2] Osamu Ishikawa, et al., “Extraction of Musical Performance Rules Using a Modified Algorithm of Multiple Regression Analysis,” In Proc. International Computer Music Conference, pp.348-351, 2000.
- [3] Kristoffer Jensen, “Envelope Model of Isolated Musical Sounds,” In Proc. DAFx99, Dec. 1999.
- [4] Esteban Maestre, et al. “Automatic characterization of dynamics and articulation of expressive monophonic recordings” In Proceedings of the AES 118th Convention, May 2000.
- [5] 小泉 悠馬, 他, “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.
- [6] Bello, J.P., et al., “Phase-based note onset detection for music signals,” In Proc. ICASSP 2003, vol.5, 6-10 Apr. 2003.
- [7] Will Gersch, et al., “The Prediction of Time Series with Trends and Seasonalities” Journal of Business & Economic Statistics, Vol. 1, No. 3 pp.253-264, 1983.

¹A. Corelli: Sonata for Violin and Piano in F major, 2 楽章