

クラシック音楽理解力拡張インタフェースを目指して： 複数の演奏家による解釈共通旋律と解釈相違旋律の推定

前澤 陽[†] 後藤 真孝[‡] 高橋 徹[†] 駒谷 和範[†] 尾形 哲也[†] 奥乃 博[‡]

[†] 京都大学大学院 情報学研究科 知能情報学専攻

[‡] 産業技術総合研究所 (AIST)

1. はじめに

クラシック音楽では、同一の楽曲の譜面を、様々な演奏家がそれぞれの独自の解釈で演奏する点が特徴的である。鑑賞者は、こうした同一楽曲の異なる演奏における共通点や差異を聞き取り、その背後にある解釈の違いを理解する。これは、クラシック音楽を鑑賞する際に特有な楽しみ方の一つと言えるが、クラシック音楽を聞きなれない鑑賞者にとっては、そうした共通点や差異を聞き取ることが困難で、深い理解が得られにくい問題があった。

本研究では、こうした同一楽曲に対する異なる演奏録音の共通点や差異を可視化し、音響信号操作を可能にすることで理解しやすくするシステムを実現することで、クラシック音楽を聞きなれない鑑賞者でも、より豊かに楽曲を楽しめ、より興味を抱けるようにすることを目的とする。こうした音楽的な要素の可視化や操作を可能にすることで人間の音楽理解力を拡張する研究アプローチは、既に後藤によって「音楽理解力拡張インタフェース」[1]として提唱されている。しかし、これまでポピュラー音楽を対象にしたシステムしか実現されていなかった。そこで本研究は、クラシック音楽、特にオーケストラによる演奏を対象とした音楽理解力拡張インタフェースの実現を目指し、その第一歩として、複数の演奏録音における旋律の音量バランスを推定し、それらを比較することで、異なる演奏家による解釈が共通な旋律（解釈共通旋律と名付ける）と、相違している旋律（解釈相違旋律と名付ける）を推定する。前者はその旋律が共通に重要だと解釈されている可能性が高く、後者は演奏者ごとに異なる解釈で演奏されている可能性が高い。ただし、単に共通なだけでなくと重要とは限らないため、音量がある程度大きい旋律だけを対象とする。将来的に解釈共通旋律と解釈相違旋律を可視化すれば、オーケストラ演奏に対する音楽理解力を拡張可能なことが期待できる。

そのための音量バランスを推定するには、オーケストラ演奏の混合音中の個々の音の音量を推定する方法が考えられる。混合音中の音量推定では、文献 [2] や [3] の先行研究があるが、オーケストラ演奏のような多数の楽器による混合音まで扱うことは困難であった。しかし、本研究の目的では、旋律の個々の音を高品質に音源分離をして再生するような処理までは不要で、あくまで旋律の解釈の違いがわかる程度に、旋律全体としての音量バランスが求めればよい。またクラシック音楽では既知の楽譜を演奏することが多いことから、混合音に対応した楽譜表現として標準 MIDI ファイル (SMF) を利用して推定することとする。そこで本稿では、オーケストラ演奏の混合音の音響信号と、それに対応した SMF を入力とすることで、SMF 中の旋律の音量バランスを推定する手法を提案し、さらに、異なる演奏録音での音量バランスの違いを分析することで、解釈共通旋律と解釈相違旋律を推定する手法を提案する。

2. 旋律音量バランスに基づく解釈共通旋律と解釈相違旋律の推定

オーケストラ用楽曲の譜面 (SMF) と、それに対する複数の異なる演奏録音から、ある旋律の音量バランスが演奏録音間でどう違うかを求め、音量が大きい重要な旋律の中で、解釈が共通している旋律と相違している旋律を推定する。事前の準備として、各演奏録音はテンポが違いうことから、既存手法 [4] を使って、SMF と対応付けておく。

2.1 旋律音量バランス推定手法

演奏録音から FFT で求めた振幅スペクトルを観測したときに、各時刻 (フレーム $t \in [1, T]$) における、SMF 上の各楽器音の音符の音量を求める問題を解く。 ω は周波数であり、フレーム t での振幅スペクトル $A_t = \{a_{t,w} | \omega \in [1, W]\}$ は等ラウドネス曲線を用いて人間の聴感上の印象に近づける工夫をしておく。T はフレームの総数、W は周波数ビンの総数であり、各ビン w に対応する周波数を f_w とする。

我々の目標は、フレーム t における振幅スペクトル A_t を楽器音ごとに分離することである。本研究では、 A_t をヒストグラムであるとみなす。すなわち、周波数 f_w は $a_{t,w}$ 回観測されたと考える。また、フレーム t において K 個の楽器音が存在し、各楽器音は M 個の倍音から構成されているとき、周波数 f_w がどの楽器音のどの倍音から生じたものであるかを示す潜在変数を $z_{t,w}$ とする。 $z_{t,w}$ は対応する次元のみが 1 で他は 0 の KM 次元のベクトル $[z_{t,w,1,1}, \dots, z_{t,w,K,M}]^t$ で表せる。 $Z_t = \{z_{t,w} | w \in [1, W]\}$ としておく。以降各時刻 t の処理は独立しているため、時間の添字は省略する。

いま、楽器音 k の基本周波数を μ_k とすると、その第 m 次倍音から周波数 f_w が観測される確率は、平均 $m\mu_k$ および精度 λ_k をもつガウス分布に従うことを仮定する。すなわち、各倍音周辺の急峻なスペクトル形状に対し、ガウス分布をフィッティングさせる。また、楽器音 k の混合比は π_k であり、楽器音 k における第 m 次倍音の混合比は $\tau_{k,m}$ であるとする。求めたいのは、 $\mu = \{\mu_1, \dots, \mu_K\}$ 、 $\lambda = \{\lambda_1, \dots, \lambda_K\}$ 、 $\pi = [\pi_1, \dots, \pi_K]^t$ 、 $\tau = \{\tau_1, \dots, \tau_K\}$ 、 $\tau_k = [\tau_{k,1}, \dots, \tau_{k,M}]^t$ の値である。まず、 A および Z の確率分布は

$$p(A|Z, \mu, \lambda) = \prod_{w=1}^W \prod_{k=1}^K \prod_{m=1}^M \mathcal{N}(f_w | m\mu_k, \lambda_k^{-1})^{a_w z_{w,k,m}} \quad (1)$$

$$p(Z|\pi, \tau) = \prod_{w=1}^W \prod_{k=1}^K \prod_{m=1}^M (\pi_k \tau_{k,m})^{a_w z_{w,k,m}} \quad (2)$$

で与えられる。さらに、混合比 π および τ に対してディリクレ事前分布を導入する。

$$p(\pi) = \mathcal{D}(\pi | \alpha(c\pi' + (1-c)e/K)) \quad (3)$$

$$p(\tau) = \prod_{k=1}^K \mathcal{D}(\tau_k | \tau_{0,k}) \quad (4)$$

ここで、 e は単位行列で、 c は 1 つ前のフレーム $t-1$ における楽器音の混合比 π' と、全楽器音について等価な混合比 e/K との重み付けを制御する。 α は集中度で、大きくなるほど π はディリクレ分布の期待値 $c\pi' + (1-c)e/K$

Toward Augmented Music-Understanding Interface for Classical Music: Estimation of Melodic Lines with Similar/Dissimilar Interpretations by Different Performers Akira Maezawa (Kyoto Univ.), Masataka Goto (AIST), Kazunori Komatani, Tetsuya Ogata, and Hiroshi G. Okuno (Kyoto Univ.)

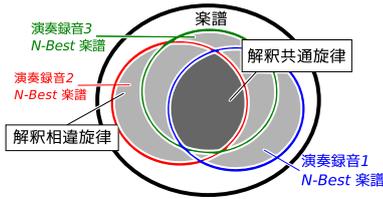


図 1: 解釈共通旋律と解釈相違旋律

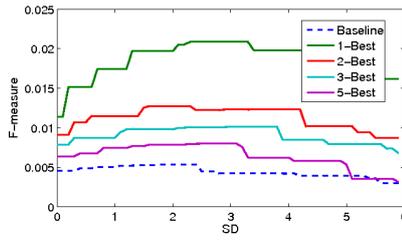


図 2: 単一の演奏録音を用いた場合の「N-Best 楽譜」の N の減少による主題検出性能の向上

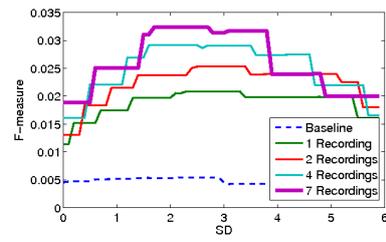


図 3: 同一楽曲に対する演奏録音の数の増加による主題検出性能の向上 (1-Best のみ使用)

付近に分布するようになる。超パラメータ $\tau_{0,k}$ は楽器音 k の基本周波数に依存して設定する。 μ および λ に関しては、ガウス・ガンマ事前分布を導入する。

$$p(\mu, \lambda) = \prod_{k=1}^K \mathcal{N}(\mu_k | \mu_{0,k}, (b_{0,k} \lambda_k)^{-1}) \mathcal{G}(\lambda_k | a_0, b_0) \quad (5)$$

ここで、超パラメータ $\mu_{0,k}$ および $b_{0,k}$ は楽器音 k におけるガウス分布の平均と精度のスケール、 a_0 および b_0 はガンマ分布の形状および尺度母数である。本研究では、音響信号に対応する SMF (楽譜情報) が利用可能であるという前提のもと、楽器音数や超パラメータを適切に設定する。これらを用いて完全な同時分布は

$$p(A, Z, \pi, \tau, \mu, \lambda) = p(A|Z, \mu, \lambda) p(Z|\pi, \tau) p(\pi) p(\tau) p(\mu, \lambda) \quad (6)$$

で与えられる。本研究では、事後確率最大化原理に基づき、EM アルゴリズムを用いて最適な π, τ, μ, λ を推定し、 π_t を時間 t で推定された音量バランス π とする。

2.2 解釈相違旋律/解釈共通旋律の推定手法

各演奏録音に対して上記の手法で旋律音量バランス π_t を求めた結果を用いて、解釈共通旋律と解釈相違旋律を推定する。その準備として、クラシック音楽では、ある音符の系列を複数の楽器が同じ音高 (ユニゾン) かオクターブ離れて演奏している場合がよくあるが、これは聴取時の印象としては同一の旋律となるので、それらの楽器の推定音量を合計しておく。その合計値は最も音高が低い楽器に割り当て、それ以外は除外しておく。

解釈の共通点や差異を見るために、それぞれの演奏録音から大きい音量の旋律を抽出し、抽出された旋律の共通点や違いを分析する。クラシック音楽では、ある旋律が複数の楽器によって演奏されることがあるため、楽器の違い (MIDI チャネルの違い) は無視し、楽譜を音符の集合として見て、小さい音量の音符を除外して部分集合を抽出する。ここで除外する対象は、各演奏録音の各時刻における上位 N 位に入らない時刻が、音符の時間長全体の $2/3$ 以上となるような音符である。このような過程を経て生き残った音符のみによって構成される楽曲の楽譜の部分集合を、本稿では「N-Best 楽譜」と名づける。

最後に、各演奏録音から求められた N-Best 楽譜を比較して、解釈共通旋律と解釈相違旋律を求める。解釈共通旋律は、複数の演奏録音における N-Best 楽譜の積集合、演奏録音 i の解釈相違旋律は、演奏録音 i の N-best 楽譜と解釈共通旋律との差分と定義する (図 1)。

3. 実験

上記で推定した解釈共通旋律と解釈相違旋律を定量的に評価しようとしても、正解データの用意が困難なため、本稿では、解釈共通旋律がクラシック音楽で大切な主題と一致しているかを評価する。主題は共通に一貫して大きい音量で演奏されていると考えられるからである。そこで、モーツァルト K550 第 1 楽章の演奏録音 7 種類と、音楽学研究用の主題データベース [5] から得た K550

の主題 (音符数 15 程度の 4 小節の単音の系列が二種類) を用意した。

主題との一致度を評価するために、主題と解釈共通旋律のそれぞれをピアノロールで表現した行列 (時刻 t 、MIDI ノート番号 n の要素 (t, n) が発音中は 1、それ以外は 0 となる、楽譜長さの大きさ $\times 127$ の行列) 同士の相互相関 $s(i)$ を求める。主題をラグ i だけずらしたときに、解釈共通旋律と合致すれば相関が高いため $s(i)$ が大きくなる。つまり、 $s(i)$ は解釈共通旋律中に主題が出現する箇所にはピークを持つので、 $s(i)$ がその平均よりも ($s(i)$ の標準偏差) \times (閾値 SD) より大きい箇所のすべてのピークを検出する。その際、高さ方向にもずらして、あらゆる転調の可能性も考慮する。

こうして推定した主題の出現箇所は、SMF による楽譜中に主題が実際に出現する箇所を正解として F 値を評価できる。単一の演奏録音の N-Best 譜面中で主題が検出できるかを、 N を変化させて評価した結果を図 2 に、 M 種類の演奏録音に基づき生成された 1-Best 譜面中で主題が検出できるかを、 M を変化させて評価した結果を図 3 に示す。ベースラインとして、楽譜上の音符をすべて解釈共通旋律 (ただしユニゾンとオクターブを吸収) したときに、主題がどう検出されるかを示した。以上の結果から、N-Best 譜面の N を減少させたり、比較する録音数 M を増やしたりすると、解釈共通旋律が含む音符数は減るが F 値が増加するため、解釈共通旋律に占める主題の音符の割合が増加することが示された。つまり、共通に一貫して大きく演奏される音符が主題と一致し、主題以外の音符をより含まない傾向があることがわかった。

4. おわりに

本稿では、異なる解釈で演奏されたオーケストラの演奏録音の共通点や相違を理解することを可能にするインタフェースの実現を目指し、第一段階として、解釈共通旋律と解釈相違旋律を推定する手法を提案した。今後は、各音のアタック部分が音量の知覚に与える影響も考慮することで音量推定精度の向上に取り組むと共に、ユーザが実際に操作可能なオーケストラ音楽理解力拡張インタフェースを実装する予定である。最後に、本研究の一部は、科研費、GCOE、CrestMuse の支援を受けた。

参考文献

- [1] 後藤 真孝: "音楽音響信号理解に基づく能動的音楽鑑賞インタフェース", 情処研報 音情科 2007-MUS-70-9, pp.59-66, 2007.
- [2] 糸山 克寿 et al.: "楽譜情報を援用した多重奏音楽音響信号の音源分離と調波・非調波統合モデルの制約付パラメータ推定の同時実現", 情処学論, Vol.49, No.3, pp.1465-1479, 2008.
- [3] M. Davy et al.: "Bayesian analysis of polyphonic western tonal music," JASA, 119, pp. 2498-2517, 2006.
- [4] N.Hu, et al.: "Polyphonic Audio Matching and Alignment for Music Retrieval", WASPAA2003, pp.185-188, 2003.
- [5] D. Huron: "ThemeFinder", www.themefinder.org, 2009