

## 旋律特徴と楽器特徴に基づく総譜の旋律クラスタリング\*

田中 大貴<sup>†</sup>, 伊藤 克亘<sup>‡</sup>

### 1 自動リダクション

アンサンブルという音楽の演奏形態がある。アンサンブルは楽器編成の自由度が高いため、希望の編成の楽譜が存在しないことがある。この問題を解決する方法として、リダクションという編曲方法がある。リダクションは、複数パートで構成されている総譜から、旋律を収縮したり主要な要素の旋律以外を削ったりして、別の楽器の楽譜へと編曲する方法である。この編曲には演奏技術とは別に、編曲知識が必要になる。

本稿では、音楽的な特徴を考慮して、総譜から各々の楽器に合わせた楽譜へ編曲する自動リダクションシステムを提案する。本稿では、電子オルガンを対象として編曲する。

### 2 旋律のグループ化

#### 2.1 旋律の特徴量

旋律の分類をするために、発音時刻、音高の変化、和声、発音持続パターンの特徴を考慮する。

発音時刻、音高の変化、和声、発音持続パターンの四つの特徴量は、松原ら [1] の方法で求めた。これらの特徴量は、一番短い音符の長さを基本単位として、単位時間ごとに  $-1, 0, 1$  の要素を持つベクトルである。下図は特徴量ベクトルの例である。ここでは、発音時刻の特徴量ベクトルを  $RA$ 、音高の変化を  $MA$ 、和声を  $CA$ 、発音持続パターンを  $SR$  として表現する。



#### 2.2 音色の特徴量

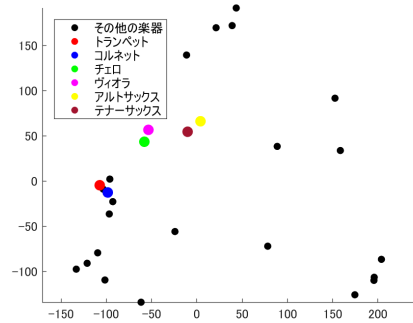
同じような音色の楽器は、似た旋律を奏でることがある。そこで、各楽器の音色の特徴を考慮する。

音色の特徴抽出を行う際、オーケストラや吹奏楽の演奏で用いられる約 30 種類の楽器の MIDI 音源を用意した。各楽器の MIDI 音源は、約 3 秒間発せられた音である。また、音程によって楽器の音色が変化することがあるので、ここでは総譜で用いられる約 2 オクターブ分の音域の音程を使用している。今回は、音色の変化が少なく、比較的安定しているサステインの区間の 1.5 秒間の特徴量に使用した。ここでは、フレーム長を 256 点、シフト長を 128 点として、15 次元の MFCC [2] でフォルマント構造を求め、それを特徴量として使用する。

求めた特徴量を用いて、音色のモデル化を行う。モデル化には、混合ガウスモデル (GMM) [3] を利用する。混合数を 8 とした。

音色の類似度は確率分布の差を測る尺度として、KL ダイバージェンス [4] で比較する。本稿では互いの KL ダイバージェンスの値の平均を取る。KL ダイバージェンスの値はペアワイズな値になっており、楽器ごとに求めている他の特徴

量と表現が異なる。そこで、多次元尺度構成法 (MDS) [5] を用いて、KL ダイバージェンスの値を楽器ごとの値へと変換を行う。ここでは 5 次元の特徴量へと変換を行っている。ここで得られる音色のベクトルを  $IT$  とする。Isomap により  $IT$  ベクトルを 2 次元で示した結果を下図に示す。



#### 2.3 旋律のクラスタリング

クラスタリングに用いる尺度を定義する。求めた各特徴量は、 $\cos$  類似度を用いて計算を行う。1 からコサイン類似度の値を引いたものを尺度として利用する。各尺度の値に足して 1 になるような重み  $w_k (k = 1, 2, 3, 4, 5)$  を掛け、足し合わせていく (式 1)。この  $D$  を用いて、クラスタリングを行う。

$$D(i, j) = w_1 D_{RA}(i, j) + w_2 D_{SR}(i, j) + w_3 D_{MA}(i, j) + w_4 D_{CA}(i, j) + w_5 D_{IT}(i, j) \quad (1)$$

五つの特徴量を用いて、k-means 法でクラスタリングを行う。本研究では、二小節単位に見ていき、同時発音数が五つ以上にならないようにクラスタリングをする。総譜の旋律のクラスタリングを以下の手順で行う。

- I クラスタ数  $K = 5$ 、パート数  $N$ 、二小節分の各パートの音符データを取得。
- II 取り出した各パートの音符データから定義した特徴量を用意して、式 (1) を用いてクラスタ数  $K$  の k-means 法を使用。
- III クラスタリングをした結果、各グループで同時に異なる音名が五つ以上でない場合、または  $K = N$  であるときクラスタリングを終了。そうでなければ  $K = K + 1$  して I に戻る。

初期値を初回のみランダムに決め、二回目以降はひとつ前のラベル番号を初期値として使用する。クラスタ数は、小節ごとに旋律の役割の数が異なる事があると考えられるので、クラスタ数 5 からパート数分まで増やしてクラスタリングを行う。初期のクラスタ数は実験的に調べた結果、5 の時が良い結果だった。

#### 2.4 分類したグループの選択

分類したグループから電子オルガン譜の右手、左手、足に対応するものを選ぶ。この中で、一番演奏方法が異なる足パートを初めに選び、次に右手パート、最後に左手パートの順に選ぶ。

まず、足パートに対応するグループを選択する。足パートは主に曲全体を低音の旋律で支えている役割がある。そこで、分類された各グループの中で、音高が C3 以下の音符の割合

\* : Melody Clustering for Multi-part Musical Scores Based on Melody and Instrument Characteristics Daiki Tanaka (Hosei Univ.) et al.

<sup>†</sup> 法政大学大学院 情報科学研究科

<sup>‡</sup> 法政大学 情報科学部

が一番大きいグループを足パートとして選択する。低音の割合は、C3以下の音数を単位時間あたりの音数で割って計算する。

次に、右手に対応するグループを選択する。右手は他の鍵盤に比べて高い音が出せるので、高音の割合を考える。また、主旋律のような発音数が多い旋律を主に担当することが多いので、発音数の割合も考慮する。音高の割合は、C5以上の音数を単位時間あたりの音数で割って計算して、この時の値を  $HR$  とする。発音数の割合は、小節内の発音数を最小の音符を1とした時の単位時間あたりの時間で割って計算して、この時の値を  $RR$  とする。これらの重みづけ和の一番値が大きいグループを右手パートとして選択する。

最後に、左手に対応するグループの選択を行う。左手は和音を奏でることが多いので、和音の使用数の割合が大きいグループを選択する。和音の割合は、和音に含まれている音数を音数で割って計算する。この値が一番大きいものを左手パートとして選択する。

### 2.5 電子オルガンの制約に基づいた修正

グループ内の音符は各楽器に対応した音高のままであるため、実際に鍵盤で演奏することができないことがある。そこで、電子オルガンで演奏できるように修正する。楽譜は小節単位に同じ時刻の右手、左手、足パートを見て修正する。

まず、足パート旋律の修正する。足パートは二音同時に演奏することがないので、二音同時に音が存在してしまう箇所を修正する。重なっている二音が同じ音名の音である時、一つの音へまとめる。この時、発音楽器数が多い音を基準にしてまとめる。二音が異なる音名の音である時、和音の根音に対応する音を残して他の音を消す。また、和音の根音に対応する音がない場合は、発音楽器数が多い旋律を残す。以上の修正の他に、足鍵盤の音域は約1.5オクターブ(C2からG3)であるため、その音域内になるように修正する。

次に右手と左手の修正である。右手と左手に関しては、同時に鳴っている最小の音高と最大の音高の差が1オクターブ以上ある時に修正をする。足パートと同様に、同じ音名の音の場合は発音楽器数が多い音高を基準に一つの音にまとめる。異なる音名の音の時は、1オクターブ未満になるように音の集約を行う。左手に関しては、右手と交差して演奏しにくいことがあるので、C3からC4の間の音域を基準にして集約を行う。

## 3 評価

### 3.1 評価方法

電子オルガンの特徴が考慮された適切な楽譜ができていのかを評価する。ここでは既存の電子オルガン譜を編曲の正解として扱う。評価値の計算は式(2)で行う。

$$\text{評価値} = \frac{\text{全体の音符数} - \text{不足数} - \text{挿入数} - \text{誤り数}}{\text{全体の音符数}} \quad (2)$$

編曲に使用する総譜は、パート数が異なり、曲全体の小節数も異なるものを利用している。評価対象の楽曲は『モテット「アヴェ・ヴァルム・コルプス」K.618』、『組曲「展覧会の絵」より「プロムナード」』、『愛のあいさつ』、『交響曲第36番「リッツ」K.425より第3楽章メヌエット』の4曲である。

### 3.2 結果と考察

実験の結果、4曲の右手、左手、足パートの評価値の平均が0.80、0.71、0.77となった。評価値と編曲後の演奏と楽譜を比較したところ、評価値が0.7~1.0の時は良い楽譜ができていて、元の曲のイメージが保たれた編曲になっていた。0.5~0.7の時は楽譜としては演奏しにくいだが、聴いたところ曲のイメージは保たれている編曲になっていた。0.5以下の場合、楽譜としては演奏しにくく、演奏自体もところどころ違和感があるような編曲になっていた。従って、このリダクション

システムは総譜から電子オルガン譜を用意するのに適しているという結果になった。

生成した楽譜には不自然に編曲されている小節がいくつかあった。本研究では、旋律クラスタの選択を、足、右手、左手の順に行っている。元々総譜の音数、もしくはパート数が少ない場合、旋律クラスタが二つしかないことがある。その際、足、右手、左手の順でクラスタを選ぶと、右手と足に旋律があり、左手に旋律がないような楽譜ができてしまう。電子オルガンでは、足パートの演奏が難しい、右手と左手で旋律を補えるなら、足パートは演奏しない楽譜が理想的である。他にも小節単位でクラスタを決定していることから、小節ごとに旋律が切れてしまったり、右手から左手へと旋律が不自然に跨いでしまったりしていた。これらのことから、旋律のグループの選択方法を検討し直す必要がある。現在の方法では、小節内の旋律だけしか考慮していないので、新たに隠れマルコフモデル(HMM)[6]などを用いて、前後の小節の旋律との繋がりを考慮することを考えている。これにより、適切な旋律を選択することができるだろう。

表3.2は、クラスタリングに使用した特徴量の重みの最適な値である。鍵盤楽器は複数の旋律を一人で演奏するため、同じ動きをしたり同時に発音したりする旋律がまとまった楽譜になっている。そのため、今回は発音時刻の特徴量  $RA$  と音高の変化の特徴量  $MA$  が良く効いた結果となった。特徴量の重みを調節することで、他の楽器の楽譜へ編曲が可能になる。例えば、ギターのようなコード弾きする楽器は、和音の重みを増やすことで編曲できるだろう。

曲名	$RA$	$MA$	$CA$	$SR$	$IT$
アヴェ・ヴァルム・コルプス	0.399	0.4	0.1	0.1	0.001
展覧会の絵	0.39	0.3	0.2	0.1	0.01
愛のあいさつ	0.3999	0.4	0.1	0.1	0.0001
第3楽章メヌエット	0.4999	0.2	0.2	0.1	0.0001

## 4 おわりに

本稿では、リダクションによる編曲を自動化するために、電子オルガンを対象とした編曲システムを提案した。まず、総譜から似た特徴を持つ旋律をまとめるために、発音時刻、発音持続パターン、音高の動き、和声、音色を使って旋律のクラスタリングを行う。次に、右手、左手、足パートに対応する旋律を選び、最後に電子オルガンで演奏できるように修正を行う。クラスタリングの時、新たに音色の特徴量を考慮することで、似た音色の楽器の旋律をまとまりやすくなり、電子オルガンの楽譜に合った編曲が可能になった。

実験の結果、クラスタの選択をする時に、誤りが多く起こることがわかった。これは、同時に演奏する小節内だけで、旋律を選択していて、前後の旋律のつながりを考えずに選んでいることが原因である。今後は隠れマルコフモデルによる旋律のつながりを考慮して、適切なクラスタの選択を行う予定である。

## 参考文献

- [1] 松原正樹, 岡本紘幸, 佐野智久, 鈴木宏哉, 延澤志保, 斉藤博昭, “ScoreIlluminator: スコア色付けによるオーケストラスコアリーダー支援システム”, 情報処理学会論文誌, 2009, pp.2937-2948.
- [2] 小林隆夫, “音声のケプストラム分析, メルケプストラム分析”, 電子情報通信学会技術研究報告, 音声 98(263), 1998, pp.33-40.
- [3] Dong Yu, Li Deng (2015): Automatic speech recognition: a deep learning approach, Springer-Verlag London
- [4] John R. Hershey, Peder A. Olsen, “Approximating the Kullback Leibler Divergence Between Gaussian Mixture Models”, Acoustics, Speech and Signal Processing, pp.IV317-IV320, 2007.
- [5] Edwards, J. and Oman, P (2003): Dimensional Reduction for Data Mapping-A practical guide using R, R News, Vol. 3/3-2-7.
- [6] Thayabaran Kathiresan, “Automatic melody generation”, Master’s thesis, KTH Royal Institute of Technology, 2015.