

自動編曲のための音楽的特性に基づく旋律クラスタリング Melody Clustering Based on Musical Characteristics to Automatic Arrangement System

田中 大貴* 伊藤 克亘†
Daiki Tanaka Katunobu Itou

1 はじめに

編曲とは、原曲を何らかの形で変更することである。多様な編曲の在り方があるのに応じて、その意義も様々だ。代表的な意義として、芸術的側面、新たな創意工夫の機会、楽譜販売市場の拡大による経済的意義、作品の演奏用途の変更、レパートリーの拡大といった演奏実績の広がりなどがあげられる。

一方で課題もある。編曲は手間がかかり、素人には困難な作業である。そのため、どんなに良い曲でも人気がなかったり、おのおのの楽器の演奏者が少なかったりすると、編曲されずに楽譜が販売されないことがある。しかし、芸術的側面を維持するためには、様々な楽器で多くの人に演奏してもらうことが重要である。そこで、編曲の負担を減らし、様々な楽器に対応した楽譜を簡単に生成できる編曲システムを提案する。本研究では、編曲の手間がなくなることで各楽器に対応した楽譜が作りやすくなり、楽曲の芸術的側面の維持につなげることを目的とする。

芸術的側面を維持するためには、オリジナル曲の特徴を保持して編曲する必要がある。そこで、リダクションと呼ばれる編曲方法を使用する。リダクションとは、可能な限り元の楽譜に忠実になるよう、オーケストラなどの演奏で使われる複数の楽器パートで構成される総譜(図1)から主要な旋律を残し、それ以外を削る編曲方法である。本研究では、楽曲全体を表現して演奏できる楽器として、電子オルガンを対象にした編曲を行う。

図 1. 総譜

2 電子オルガン譜へのリダクション

2.1 電子オルガンの特徴

電子オルガンの演奏方法は、他の楽器と比べて独特である。ピアノは右手と左手を使って演奏するのに対して、電子オルガンは右手、左手、足を使って演奏する。また、ピアノは一つの鍵盤を使うのに対して、電子オルガンでは三つの鍵盤を使って演奏する。そのため、図2のように楽譜が三段組になっている。

電子オルガンの右手パートの旋律は、主旋律のような旋律を奏でる傾向がある。左手パートの旋律は、和音や和音を含む旋律を奏でることが多く、同時発音数が4音以下である特徴がある。足パートの旋律は、低音の旋律や和音の根音を主に奏で、必ず単音である。右手パートと左手パートの同時に発音する時の音域はそれぞれ1オクターブ以内である。

電子オルガンの大きな特徴として、約900種類の楽器の音色を出せることである。そのことから、演奏時は一つの鍵盤に複数の楽器の音色を重ねて演奏する。また、音色の自動切り替えが可能であるため、異なる楽器の旋律を続けて演奏できる。電子オルガンでは打楽器などのリズムボックスが用意されているので、楽曲のリズムを予め用意しておき、それに合わせて演奏するのが基本的な演奏スタイルである。

図 2. 電子オルガン譜

2.2 関連研究

複数パートの旋律を圧縮して、他の楽器の楽譜へ編曲する方法に関する研究が複数ある。例えば、ポッターらの研究[1][2]ではギターアレンジの手法を提案している。これはギターの演奏上の制約を考慮して、重要な音符の集合を選んでギター譜へ編曲している。しかし、この方法では旋律のフレーズ単位ではなく音符単位に処理される。したがって、選択された音符が異なる楽器から選択された場合、旋律のフレーズが失われる可能性がある。

ピアノ譜へのリダクションに関する研究として藤田らの研究[3]がある。これはピアノの右手は主旋律を主に奏で、左手は伴奏を奏でるという観点から、平均音高、発音時間、音高・リズムパターンに着目して、総譜の旋律からメロディとベースラインを推定している。そして、それらの旋律をピアノの右手パートと左手パートに採用するという方法である。しかし、この方法ではポリフォニックのような曲の場合、メロディを一意に決めることができないので不向きである。更に、ピアノが常にメロディとベースラインを引き続けているとは限らないので、この方法では不十分である。

他にも中村らの研究[4]がある。これは、可能な限り元の楽曲の旋律を使用して、鍵盤上の運指を考慮しながら総譜を圧縮してピアノ譜へ編曲する手法を述べている。しかし、音楽的な特徴が考慮されていないことから、演奏できない理由から楽曲の重要な旋律が消える可能性がある。

本研究では、楽曲の旋律が持つ特徴と電子オルガンの特徴を考慮して編曲を行う。まず、音楽的な特徴を考慮して、同じ役割を持つ旋律を小節単位にグループ化する。次に、既存の電子オルガン譜から右手パート、左手パート、足パートの特徴と各パートが演奏する旋律の特徴、更に小節間のつながりを考慮して、各パートに対応する旋律グループを選択する。最後

* 法政大学大学院 情報科学研究科

† 法政大学 情報科学部

に、電子オルガンで演奏できるように楽譜の修正を行う。この手法により、旋律を一つのフレーズとして考えられる。また、メロディを一意に決めず、電子オルガン譜の特徴を考慮した選択をするため、ポリフォニックのような曲にも対応できるようにする。

3 電子オルガン譜へのリダクションシステム

本システムは、MusicXML[5] で表現された総譜の情報を入力として処理を行う。MusicXML から楽器名、小節番号、オクターブ、音符の種類、音高、最小の音符に合わせた持続時間、拍子情報を取得する。そして、これらの情報から楽曲の特徴量を生成する。これらの特徴量を用いて、総譜の旋律のクラスタリングを行い、類似した特徴を持つ旋律のグループを作成する。クラスタリングの後には、電子オルガン譜の右手パート、左手パート、足パートに対応するグループを選択する。そして最後に電子オルガンで演奏できるように楽譜の修正を行う。

3.1 旋律クラスタリング

総譜の各々のパートは主旋律、ハーモニー、装飾などの役割を持っている。そのため、総譜が約 30 パートで構成されていたとしても、旋律の役割を考慮して分類することで 5 パート程に集約できる。本研究では楽曲の特徴量として、リズム、音高の変化、和声、音価、楽器の音色を使用し、総譜の旋律のクラスタリングをして総譜のパート数を削減する。

3.1.1 旋律の特徴量

リズム、音高の変化、和声、音価の特徴を考慮する。これらの特徴量は、スコアリーディング支援を目的とした松原らの研究 [6] に利用されていた旋律クラスタリングの特徴量から用意したものである。各特徴量は、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。図 3 は特徴量ベクトルの例である。ここでは、リズム (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 、音価 (発音持続パターン) を SR として表現する。



図 3. 旋律の特徴量ベクトルの例

3.1.2 音色の特徴量

3.1.1 で用意した特徴量は、視覚的に旋律が似ているかを判断しているだけである。そのため、視覚的に旋律は似ていないが、同じ役割を持つような旋律をまとめることができない。今回は、似た音色の楽器の旋律は同じような旋律を担当する傾向がある、という音響的な特徴を考慮する。新たに音色を元にした楽器間の距離の特徴量を追加する。

ローズの文献 [7] によると楽器の音色は、音の振幅エンベロープ、フォルマント構造、時間的な周波数成分の変動などで

あると述べられている。本稿ではフォルマント構造に着目して音色を表現する。フォルマント構造を分析する手法として、LPC(Linear Predictive Coding)[8]、LPC ケプストラム [9]、ケプストラムやメル周波数ケプストラム係数 (MFCC)[10] などが挙げられる。ここで対象とするのは楽器音であり、各楽器において発音可能な音域が異なる。従って、同じ音域で比較することが難しい。また、楽器音は基本周波数の整数倍の周波数が強く表れる倍音構造の性質が見られる。そのため、周波数が強く表れている部分も含めたホルマント構造を求める LPC は、ここでは適していない。そこで、音程の影響をあまり受けにくい分析方法として、本稿では MFCC を用いて音色の分析を行う。

音色の特徴抽出をする際、オーケストラや吹奏楽の演奏で用いられる約 30 種類の楽器音を使う。これらの楽器音は RWC データベース [11] の実音源を使用している。楽器音は音程によって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用いる。また、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 秒間の音を使用した。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 でフォルマント構造を求め、それを楽器の音色の特徴量とする。

求めた特徴量を用いて、音色のモデル化をする。本研究では、混合数 8 の混合ガウスモデル (GMM)[12] を使用してモデル化を行う。

次に、音色のモデルの距離を測る。本稿ではバタチャリヤ距離 [13] を使用して、各モデルの分布間の距離を計算する。パート i, j の音色の確率分布を P, Q としたとき、バタチャリヤ距離は式 1 のように表される。この時、 $p(x), q(x)$ は音色の確率密度関数である。バタチャリヤ距離は、値が 0 に近い時は二つのモデルの距離が近いということである (式 2)。

$$D(i, j) = -\ln \left(\sum_{x \in X} \sqrt{p(x)q(x)} \right) \quad (1)$$

$$0 \leq D(i, j) \leq \infty \quad (2)$$

$D(i, j)$ はペアワイズな値である。それに対して 3.1.1 で求めた特徴量は、楽器ごとに求めているため表現が異なっている。本稿では、多次元尺度構成法 (MDS)[14] を用いて、 $D(i, j)$ の値を楽器ごとの値へと変換する。ここでは 16 次元の特徴量へと変換し、この音色のベクトルを IT とする。

図 4 は、Isomap により IT ベクトルを 2 次元の楽器空間で示したものであり、各楽器の音色の距離をわかりやすくするために可視化したものである。例えば、ユーフォニウムとチューバは似たような楽器であり、音色も似ている。そのため、近いところに位置していることが確認できる。その他の金管楽器のトランペットとホルネット、弦楽器のヴァイオリンとヴァイオラもそれぞれ似た音色を持っており、近い場所に位置している。従って、作成した音色のモデルは、適切なモデルであることが確認できる。

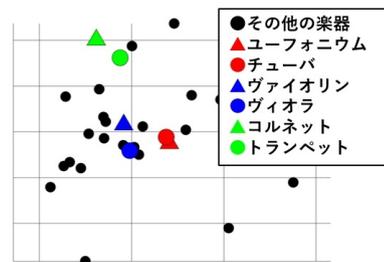


図 4. 2次元の楽器空間で音色の距離を可視化した結果

3.1.3 旋律のクラスタリング

五つの特徴量を用いて、小節単位にクラスタリングを行う。本研究ではファジィ・クラスタリング [15] を使用する。ファジィ・クラスタリングとは、k-means 法のようにクラスタを 0,1 で分類するのではなく、あるクラスタに 0.8 の割合で属し、もう一方のクラスタには 0.2 で属す、というように曖昧に分類できるクラスタリングである。ここでは、初期の重心の数を 5 つとして、各重心に近い旋律を 4 つ取ることでクラスタを生成する。クラスタリングは手順で行う。

- I 重心の数 $K = 4$ 、パート数 N 、一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 3.1.1, 3.1.2 で定義した特徴量を用意して、重心数 K のファジィ・クラスタリングを実行。
- III クラスタリングをした結果、各重心から近い旋律を 4 つ取得して、 $K = K + 1$ して I に戻る。 $K = N$ であるときクラスタリングを終了。

3.2 旋律クラスタの選択

旋律クラスタを選択する時に、隠れマルコフモデル (HMM)[16] を使用する。小節単位の場合、右手パート、左手パートが入れ替わる問題が起きるので、旋律のつながりを考慮した HMM を利用してクラスタを選択する。本研究で使用する HMM は、図 5 のようなトポロジーになっている。8 分音符を一つの状態としていて、状態 1,2,3,4 で開始され、状態 5,6,7,8 で終了する。また、各状態では C,C#,...,B, 休符の 13 次元の音高情報のベクトルを出力する。

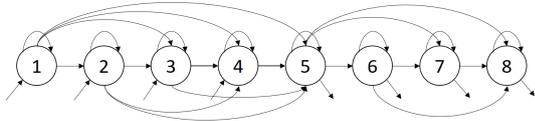


図 5. 旋律のつながりを考慮した HMM

HMM の学習では、既存の電子オルガン譜を利用して旋律のつながりを学習させる。また、既存の電子オルガン譜を利用することで、各パートが演奏する旋律の特徴を学習させる。学習には図 6 にあるように、小節の区切りの前後 2 拍 (計 4 拍) の旋律を用いる。そして、右手パートの旋律を使用して右手の HMM、左手パートの旋律で左手パートの HMM、足パートの旋律で足パートの HMM をそれぞれ生成する。旋律クラスタを選択する時は、各クラスタ間の旋律を右手パート、左手パート、足パートの HMM に対して尤度を計算して、値が大きい旋律クラスタを選択する。



図 6. 既存の電子オルガン譜を用いた HMM の学習

3.3 楽譜の修正

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パート、左手パート、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。

次に右手パートと左手パートの修正である。右手パートと左手パートは、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の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。左手パートは、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。

4 評価実験

4.1 実験概要

原曲の特徴が保たれ、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るかを評価する。従来の評価は、編曲した楽譜を聴き比べたり、編曲した楽譜が演奏できるか楽器演奏の上級者に演奏してもらったりして評価している。しかし、前者の評価だと演奏可能か評価されていないし、後者の評価だと十分な評価ができる程の楽器演奏者を集めることが困難である。そこで、本研究では元の楽曲の特徴を保って編曲され、演奏可能な楽譜になっているのか、既存の電子オルガン譜との比較で評価した。評価値の計算は式 3 で行う。右手パート、左手パート、足パートごとに評価値を計算し、その結果を確認する。

$$\text{評価値} = \frac{\text{全体の音符数} - \text{不足数} - \text{挿入数} - \text{誤り数}}{\text{全体の音符数}} \quad (3)$$

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「リンツ」K.425 より 第 3 楽章メヌエット』の 4 曲である。

4.2 実験結果

『モテット「アヴェ・ヴァルム・コルプス」K.618』を編曲した結果を示す。図 7 は、曲中の 4 小節分を示した楽譜である。図 7(a) は正解となる既存の電子オルガン譜で、図 7(b) は編曲した楽譜だ。緑色が不足している音符で、赤色が誤って挿入された音符である。この曲の評価値は、表 1 に示されていて、右手が 0.79、左手が 0.76、足が 0.71 となっている。編曲者によって編曲した楽譜は変わるため、評価値を 1 にすることは難しい。しかし、評価値と編曲後の楽譜を聴いたり見たりしたところ、評価値が 0.8~1.0 の時は、演奏可能な楽譜ができていて、元の曲の特徴を保った編曲であることが確認できた。0.5~0.8 の時は、演奏しにくい楽譜であるが、聴いたところ曲の特徴を保った編曲がされている。0.5 以下の方は、演奏しにくい楽譜であり、演奏も違和感があるような編曲になっていた。このことから、楽曲『モテット「アヴェ・ヴァルム・コルプス」K.618』は、右手、左手、足パートの評価値がすべて 0.7 以上であるので、演奏しにくい箇所もあるが、曲の特徴を保った編曲ができた例になっている。

5 考察

誤った編曲の約 6 割が、足パートへの音の挿入であった。ファジィ・クラスタリングは、クラスタ内に重複した旋律を



図 7. 編曲結果 (a):既存の電子オルガン譜, (b):システムでの編曲結果

表 1. 評価結果

曲名	パート数	小節数	右手	左手	足
アヴェ・ヴェルム・コルプス	13	46	0.82	0.75	0.70
展覧会の絵	20	24	0.79	0.79	0.75
愛のあいさつ	6	99	0.78	0.73	0.73
交響曲第36番 K.425 メヌエット	14	57	0.80	0.78	0.68
平均	—	—	0.79	0.76	0.71

含めるため自由度が高い。そのため、左手パートで担当するはずの旋律が、足パートにも含まれるようになってしまった。これは、クラスタリングの時に足パートの旋律が、右手パート、左手パートのクラスタの旋律と同じものを含んでいるかを判断することが必要である。もし、含んでいるのであれば、足パートのクラスタの旋律を他の旋律のクラスタへ移動させることで解決できると考えられる。

次に多かった誤りは左手パートの旋律の不足である。これは編曲の誤りの約3割を占めていた。左手パートは、単旋律、ハーモニー、和音のように様々な旋律を担当するため、時に同時に演奏することもある。今回の誤りは、同じ単旋律を担当しているパートが4つ以上ある時、もう一方のハーモニーフレーズを採用できず音が不足してしまった。これは、選択されなかったクラスタの旋律を考慮する必要がある。現在は、3つのクラスタを選んで編曲をしているため、重要な旋律が不足あったとしても不足してしまうことがある。したがって、3つのクラスタを選択した後、もしくは選択する前に、同時に演奏できる旋律クラスタがあるのであれば、一つのクラスタにまとめると良いと考えている。これにより、重要な旋律を失うことがなく、可能な限り元の旋律を使用した編曲ができるようになる。

今回の編曲結果では、0.7以上の評価値を得ることができ、元の楽曲の特徴を保った演奏可能な楽譜へ編曲することができた。しかし、元の楽曲の特徴を保つという点で考えると、楽曲の持つ特徴しか考えていないため、人間が編曲後の演奏を聴いた時に、印象が保たれているかについては考えられていない。よって、人間がどのような旋律を目立って聴いているのか、重要に聴取しているのかを楽曲分析および音楽心理学の方面から考える必要がある。これにより、音楽的な根拠に基づいた編曲が行えるようになる。

6 おわりに

本稿では、楽曲の芸術的側面の維持のために、編曲の負担を減らし、様々な楽曲の楽譜を自動編曲できるリダクション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、リズム、音高の変化、和声、音価、音色といった楽曲特徴を使って、ファジィ・クラスタリングによる旋律のクラスタリングを行った。次に、電子オルガン譜の右手パート、

左手パート、足パートの旋律の特徴と旋律のつながりを考慮したHMMを用いて、旋律クラスタを選択した。最後に、電子オルガンで演奏できるように楽譜の修正を行った。評価の結果、右手パート、左手パート、足パート共に0.7以上の評価値を示し、演奏可能で楽曲の特徴を保持した編曲結果を得ることができた。一方で編曲の誤りも確認された。最も多い誤りは足パートへの音の挿入であった。これはファジィ・クラスタリングの自由度が高い故に、余計な他の旋律を含んだクラスタを生成していることが原因であった。したがって、足パートに含まれる音と同じ音高の音が右手パート、左手パートのクラスタにあるかを判定し、あれば他のクラスタへ移動させるといった処理を行うことで解決できると考えられる。二つ目の誤りは、左手パートの音の不足であった。左手パートは様々な旋律を同時に演奏することがある。そのため、現在の一つのクラスタのみ選択する方法だとフレーズが欠けてしまうことがある。よって、クラスタを選択する際、他のクラスタと組み合わせ、演奏可能であれば採用するといった処理を追加することが必要である。最後に、楽曲の特徴を維持するという点についてしっかり考えなければいけない。演奏を聴いた時に原曲の特徴が保持された編曲をするためには、今回考慮した楽曲が持つ特徴のみを使うだけでは不十分であると考えている。人間がどのような旋律を知覚して重要に感じているのか、楽曲分析や音楽心理学の分野から調査し、編曲に取り入れることが今後の目標である。

参考文献

- [1] R. Daniel and W. D. Potter: "GA-based Music Arranging for Guitar," Proceedings of International Congress on Evolutionary Computation, 2006.
- [2] D. R. Tuohy and W. D. Potter: "An Evolved Neural Network/HC Hybrid for Tablature Creation in GA-based Guitar Arranging," Proceedings of the International Computer Music Conference, 2006.
- [3] 藤田顕次, 大野博之, 稲積宏誠, "習熟度を考慮した複数楽譜からのピアノ譜生成手法の提案", 情報処理学会音楽情報科学研報, 2008, pp.47-52.
- [4] E. Nakamura and S. Sagayama, "Automatic Piano Reduction from Ensemble Scores Based on Merged-Output Hidden Markov Model," in Proc. ICMC, 2015, pp.298-305.
- [5] MusicXML (2018). <http://www.musicxml.com/>
- [6] M. Matsubara et al., "Scoreilluminator: Automatic illumination of orchestra scores for readability improvement," Proceedings of the 2009 International Computer Music Conference, ICMC 2009, pp. 113-116.
- [7] Curtis Roads, コンピュータ音楽—歴史・テクノロジー・アート, 東京電機大学出版局, 2001.
- [8] J. Makhoul, "Linear prediction: A tutorial review", Proceedings of the IEEE, Volume:63, Issue:4, April 1975.
- [9] 吉井貞照, デジタル音声処理, 東海大学出版会, 1998.
- [10] 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.
- [11] 後藤真孝, 橋口博樹, 西村拓一, 岡隆一: "RWC 研究用音楽データベース: 音楽ジャンルデータベースと楽器音データベース", 情報処理学会音楽情報科学研究会 研究報告 2002-MUS-45-4, Vol.2002, No.40, pp.19-26, May 2002.
- [12] Dong Yu, Li Deng, Automatic speech recognition: a deep learning approach, Springer-Verlag London, 2015.
- [13] C.H. You, K.A. Lee, and H. Li, "Gmm-svm kernel with a bhattacharyya-based distance for speaker recognition," Audio, Speech, and Language Processing, IEEE Transactionson, vol.18, no.6, pp.1300-1312, 2010.
- [14] J. Edwards and P. Oman, Dimensional Reduction for Data Mapping-A practical guide using R, R News, Vol. 3/3, 2003, pp. 2-7.
- [15] 宮本定明, クラスタ分析入門, 森北出版, 1999.
- [16] Thayabaran Kathiresan, "Automatic melody generation", Master's thesis, KTH Royal Institute of Technology, 2015.