

オクターブ関係の推定と階層構造に着目したパート形成

A System for Figuration of Parts by an Estimation Method
based on Octave Relations and Layered Structures

毛塚 学*
Manabu Kezuka

延澤 志保†
Shiho Nobesawa

太原 育夫‡
Ikuo Tahara

1 はじめに

近年、音楽と計算機を相互に結び付けた研究が数多く行われている。その中で、音響信号から楽譜を作成する自動採譜に関する研究がある。音響信号を細分化することにより、音楽家でないユーザでも、気軽に音楽を扱えるようにすることが自動採譜の目的の一つである。自動採譜の中で、音響信号の個々の構成音を分離・抽出することを音源分離という [1]。精度の高い音源分離システムが実現すれば、自動採譜に役立つほか、楽曲理解全般に利用できるだけでなく、楽曲検索への応用も考えられる。

2 音源分離システム

音源分離システムの多くは、入力された音響信号から周波数解析を行い単音を形成し、単音から音色類似度などによりパートを形成する。

周波数解析の段階でいかに楽器音の波だけを变化なく取り出すか、またその中から高調波関係を探し出し1つの単音に由来する組み合わせをいかに見つけるか、単音の音色類似度をどのように扱いパートに分類するかなど、それぞれの処理がうまくいかなければ認識精度が落ちるのは明らかである。

2.1 単音形成とパート形成における問題点

周波数解析が高い精度で実現しても、単音形成、パート形成のそれぞれの段階において認識を妨げる問題が多くある。

単音形成において

- 他の音の倍音との重複
基音が他の音の倍音と重複して同時に発音されたとき、すべての高調波成分が他の音の成分と重複し、認識できなくなってしまう。完全に重複している場合は当然であるが、周波数がわずかに違う場合においてもFFTの周波数分解能を超えてしまうと分離できない。

パート形成において

- 音高による倍音の変化
同じ楽器でも、音高によって倍音のパワーの割合が変化する。楽器音データベースを用いれば、このような倍音変化をデータベースに加えマッチングすることにより認識することができるが、楽器音データベースを用いない場合は誤認識する可能性が高い。
- 演奏条件による特徴量の変化
同じ楽器、同じ音高であっても、演奏方法等の条件により各単音の特徴量が変動する。これは楽器音データベースを用いても対応することはできない。

*東京理科大学大学院 理工学研究科 情報科学専攻

†武蔵工業大学 知識工学部 情報科学科

‡東京理科大学 理工学部 情報科学科

2.2 提案手法

2.1節で挙げた問題点を解決するため、認識できなかった単音の「うなり」等の情報を用いた分離と、中間的なフレーズ形成を用いたパート形成の2段階に処理を分ける手法を提案する。システムの概要を図1に示す。

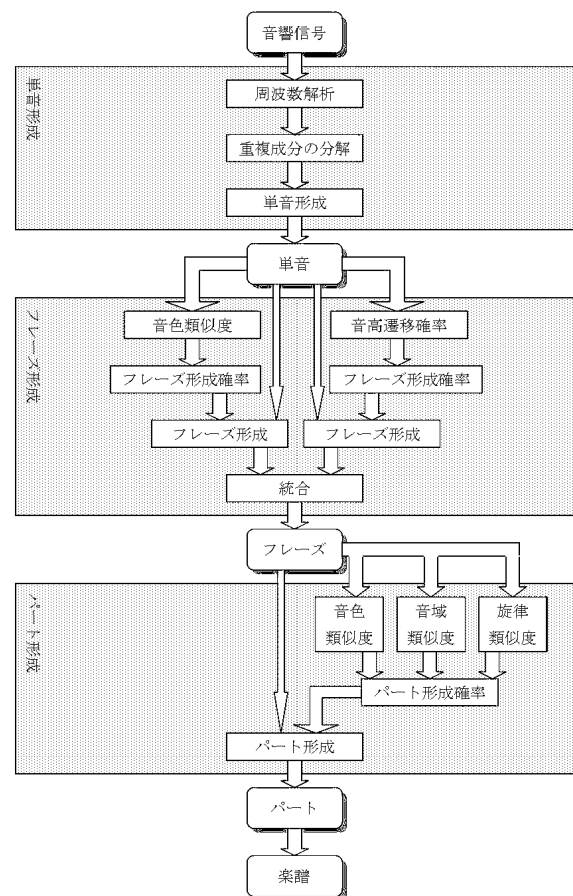


図1 システムの処理の流れ

うなり等を用いた重複成分の分解 他の音の倍音に基音が重複し認識できなかった単音を分離するために、周波数にわずかな違いがある場合「うなり」が発生することを利用し、うなりを探し重複している基音を見つけ、分解能の高いFFTを用いて基音を分離する手法 [2] を用いる。重複した2つの成分の周波数が完全に一致、もしくはうなりを認識できないほど周波数が近い場合は、重複されている音が正しくパートを形成していることを前提として、重複されている音とその音が分類されて

いる楽器音の特徴量を成分ごとにマッチングを行い、マッチングしなかった成分に他の音の成分が重複していると推定する。これにより単音認識の精度を上げる。

フレーズ形成による階層的パート抽出 周波数成分の重なりによる特徴量の変化に対応するため、単音から中間的にフレーズを形成し、それらの連結によりパートを形成する手法 [3] を用いる。これは音高遷移確率等を求めることによりフレーズを形成し、フレーズ毎の特徴量の類似度、及び旋律の類似度を計算した結果によりパートを形成するものである。この方法により、パート抽出の手がかりとして大局的な情報をいきなり与えてしまうと計算量の爆発が起ってしまう問題を解決しながら、パート形成の精度を上げることが可能になる。

3 評価実験

システムの評価実験を行い、提案手法が音源分離の認識精度において有効であることを検証する。

3.1 評価方法

入力条件と評価値は以下のとおりである。

入力条件 実験には 2~3 パートで、クラシックなど比較的旋律の単純な楽曲を用い、使用される楽器はピアノ、ヴァイオリン、トランペット、クラリネットの 4 種類の中から 2~3 種類を選択、楽曲は 44.1kHz, 16bit, モノラルの WAV 形式に変換しシステムの入力とする。楽器数は既知とする。

評価値 単音形成の段階において重複成分の分解を行う場合と行わない場合、フレーズ形成を行ってからパート形成を行う場合と単音から直接パートを形成する場合とに分け、それぞれにおける単音形成率、パート形成率、再現率を比較し評価する。ただし、

$$\text{単音形成率} = \frac{\text{正しく単音を認識できた数}}{\text{楽譜上の単音数} + \text{誤認識された単音数}}$$

$$\text{パート形成率} = \frac{\text{正しいパートに分類できた数}}{\text{正しく単音を認識できた数}}$$

$$\text{再現率} = \frac{\text{正しく音高・パートを認識できた数}}{\text{楽譜上の単音数} + \text{誤認識された単音数}}$$

とする。

3.2 実験結果

認識結果に対する評価値は表 1 に示す値となった。

3.3 評価及び考察

単音形成について 単音形成率において、重複成分の分解を行った場合は、行わなかった場合に比べ 3~4 ポイント程度の精度の向上が見られた。この結果より、単音形成の段階において重複成分の分解は有効な手法であるといえる。特に他の音の倍音と重複する基音の数が多い場合には、重複成分の分解を行うことで認識精度は大きく向上する。

また、再現率においても楽器数が 2 の場合、3 の場合ともに重複成分の分解を行うことで精度が上がっている。これは、認識できなかった単音が認識できるようになるというだけでなく、この手法による分解が行われた後に再び楽器音に分類する段階でも適切な処理が行われていることを示している。単音形成率が上がっても、分解された単音が正しいパートに分類されなければ再現率は上がらないからである。

表 1 実験結果 (上:楽器数 2, 下:楽器数 3, 単位:%)

重複成分の分解 フレーズの形成	なし	あり	なし	あり
単音形成率	84.76	88.57	84.76	88.57
パート形成率	94.38	94.62	95.51	95.70
再現率	80.00	83.81	80.95	84.76

重複成分の分解 フレーズの形成	なし	あり	なし	あり
単音形成率	79.53	82.68	79.53	82.68
パート形成率	90.10	90.48	92.08	91.43
再現率	71.65	74.80	73.23	75.59

パート形成について パート形成率において、フレーズの形成を行った場合は、単音から直接パートを形成した場合に比べ 1~2 ポイント程度の向上が見られた。単音から直接パートを形成した場合でも高い認識率が得られていたため、それほど大きく精度は向上しなかったが、パート形成においてフレーズの形成を行うことは有効であった。

フレーズ同士の類似度を求める際に用いた尺度として、音域類似度がある。パートによって音高の差が大きいほど、この音域類似度の信頼性は高くなる。逆にパートによる音高の差が小さい場合には、音域類似度のパート形成確率に与える影響は少ない。つまり、メロディと伴奏など、パートによって楽曲中での音域の分担のはっきりしている曲ほど、音域類似度は有効であるといえる。

4 おわりに

本研究では、自動採譜における音源分離システムの精度向上を目的とし、単音形成の段階で重複成分の分解を行い、単音から中間的にフレーズを形成し、それらを連結することでパートを形成した。実験により、重複成分の分解は主に単音形成率に、フレーズ形成は主にパート形成率に対し有効であることが示された。結果としてシステム全体の認識精度が向上し、提案した手法は音源分離システムにおいて有効な手法であるといえる。

今回のシステムでは、ユニゾン関係にある音の削除や、楽器数、認識する楽曲の調性の入力などを人手により行った。今後はこれらの情報も自動で処理するシステムを目指し、入力条件の制約を減らすことに注目したい。

参考文献

- [1] 後藤真孝, “音楽音響信号を対象としたメロディーとベースの音高推定,” 電子情報通信学会論文誌, Vol.J84-D2, No.1, pp.12-22, 2001.
- [2] 中村裕樹, “楽器音の特徴を用いたパート形成とオクターブ関係の推定,” 東京理科大学大学院 平成 16 年度修士論文, 2005.
- [3] 木下智義, 半田伊吹, 武藤誠, 坂井修一, 田中英彦, “自動採譜処理における知覚的階層に着目したパート分離処理,” 電子情報通信学会論文誌, Vol.J85-D2, No.3, pp.373-381, 2002.