音響信号の特徴量の類似性に基く楽曲からのストリーム抽出

5G－7　木下 智義* 半田 伊吹 武藤 誠 坂井 修一 田中 英彦
{kino,handa,muto,sakai,tanaka}@mtl.t.u-tokyo.ac.jp
東京大学大学院 工学系研究科†

1 はじめに

筆者らは既に音楽情景分析の処理モデル OPTIMA を提案し [3, 4]、その実験システムを構築した。しかしながら、その処理精度は実用上十分であるとは言えず、改善が課題となっている。これまでいくつかの手法の改善が試みられてきたものの、処理精度の劇的な向上は得られていない [2, 1]。

一般に、楽曲の演奏を録音したものを処理の対象として考えると、複数の楽器に由来する周波数成分が同時刻、同周波数に共存することが多い。そのため、干渉によりそれぞれの周波数成分の形状等が変化するという問題が生じる。これによって情報が欠落するため、このような場合における単音の認識は困難なものとなっている。この問題に対し、周波数成分の重なりを考慮した手法等も考えられるが [1]、既に欠落している情報の復元には限界があると言える。

また、音響信号から周波数成分を抽出することなく、信号波形の状態での音源同定を試みた例もある [6]。この研究では、楽器ごとに波形テンプレートを用意し、それと入力信号を比較することで音源同定を行う。ここで、同一楽器における楽器個体間の差や音の変動を吸収するために、適応処理を追加している。しかし、この適応処理を用いても十分な音源同定精度は得られていない。

これらの例では、いずれも単音が存在する各時点で処理が行われている。ところが、人間が実際に音楽を聴く場合には、各単音を意識して聴くことは少なく、メロディーや伴奏といった各パート全体を一つのまとまりとして聴くと考えられる。実際、演奏のある部分 (1 単音に相当する程度の長さ) を聴いた場合、音高と音源名をともに認識することは難しい。

2 ストリーム

このような背景から、人間がひとつながりの音であると知覚するエネルギーの集合 (ストリーム) を取り出す処理を用いることで、音楽認識の精度向上が期待できる。そこで本研究では、音符列を対象としたストリームを想定し、楽器演奏を録音した音響信号から、そこに含まれるストリーム構造を抽出することを試みる。

音楽演奏におけるストリーム構造に注目した処理として、単音連繋確率ネットワークを用いた手法が提案されている [5]。この手法では、時間的に近接する二つの単音に対して、1) 統計的に得られた単音の遷移確率、2) 最高音部、最低音部などのパートとしての「役割」、3) 単音の音色の類似度、の三つを元にストリームを形成している。

しかし、これらのうち音色の類似性を求めた効果は低く、課題が残されている。また複数のパートが近接した音域を推移する場合など、1) や 2) の効果が期待できないケースを考えると、音色に基づいてストリームを抽出する必要性は高い。

そこで本研究では、音色として周波数成分の物理的な特徴量を用い、隣接する単音間での特徴量の類似性に注目して、ストリーム形成を試みる。

3 処理の流れ

本研究では、以下に示す方法でストリーム構造を抽出する。

3.1 時間・周波数解析と周波数成分形成

入力された音響信号に対し、IIR フィルタバンクを用いた方法で、時間周波数解析を行う。また、そのパワー値のピークを時間方向に追跡することで、周波数成分を形成する [3]。

3.2 単音形成

得られた周波数成分の集合に対し、一つの単音に相当する周波数成分ごとにクラスタを形成する。ここでは、立上り時刻のずれや調波構造のずれなどを用いてクラスタリングを行う [3]。また同時に、複数の単音に属する周波数成分を「重なりパターン」として抽出する [1]。

---

*日本学術振興会特別研究員

†"Music stream extraction based on similarity of acoustic signal feature"
Tomoyoshi Kinoshita, Ibuki Handa, Makoto Muto, Shuichi Sakai and Hidehiko Tanaka
University of Tokyo, Graduate School of Engineering, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

3.3 特徴量の抽出

各単音ごとに、それに属する周波数成分から、物理的な特徴量を抽出する。これらの特徴量は、周波数成分の重なりにより、変形を受けている可能性がある。そこで、文献 [1] で用いた方法を応用し、一部の特徴量は次段のストリーム形成確率の計算に用いないようにする。

3.4 ストリーム形成確率の計算

時間的に隣接する二つの単音クラスタに対し、前段で得られた特徴量を比較することで、これらの単音クラスタが同じ音源に由来する確率を計算する。類似度の計算には、文献 [1] と同様の方法を用いる。前項で述べた通り、周波数成分の重なりにより意味をなさなくなったと考えられる特徴量は、確率の計算には用いないものとする。

この確率値を、確率計算の対象となった二つの単音が同一のストリームの一部である確率 (ストリーム形成確率) とみなし、次段にて用いる。

3.5 ストリーム構造の抽出

得られたストリーム形成確率を元に、ストリームの形成を行う。形成には、文献 [5] による手法を応用するものとする。

この手法では、新たな単音が出現した際に、既存の隣接する複数の単音とストリームを形成する確率を求めた上で、それが最大であるものを選んでストリームとして連結する。

4 評価

本研究で提案した手法を用い、予備的な評価実験を行った。本実験では、上行・下行する 2 つのパートからなる楽譜 (図 1) を演奏し、それを録音した音響信号に対して処理を行った。単音の開始時刻と音高はあらかじめ与えるものとし、ストリームの抽出精度をみた。

楽器はクラリネットとフルートを用いた。

図 1に示されている網かけ部分が、処理の結果ストリームとして抽出された箇所である。この結果により、特徴量の類似性のみを用いたストリーム抽出が可能であることが示された。

5 おわりに

本研究では、楽器演奏中に含まれるストリーム構造を、周波数成分の特徴量に基づいて抽出する手法を提案した。

評価実験では、音源の類似性のみを用いてストリームを抽出することに成功した。この他、音高の遷移確率等の情報を用いた処理と組み合わせることで、より精度の高い処理が期待できる。

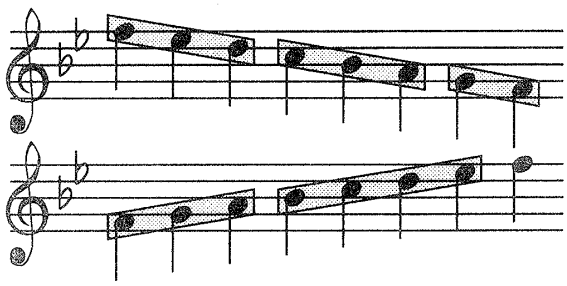

図 1: 実験で用いた楽譜およびストリーム抽出結果

今回提案した処理の後に、各ストリーム毎に音源同定処理を行うことで、従来のような、各時点において音源同定処理を行う場合に比べて同定の対象となる単音数が多くなることから、同定処理が容易になると考えられる。また、周波数成分が重なって、特徴量が変化している場合も、その前後にある周波数成分の重なりがない場合の特徴量を用いて同定処理が行えるため、精度の向上が期待できる。今後はこれらの処理の実装と評価も進めていく予定である。

謝辞

本研究は、文部省科学研究費補助金 (課題番号 09-07628) による研究成果の一部である。また、本研究を進めるにあたり、音響信号データ NTTMSA-P1 の使用許可をいただいた NTT コミュニケーション科学基礎研究所に感謝する。

参考文献

[1] 木下智義, 坂井修一, 田中英彦. 特徴量に注目した複数楽器の演奏における音源同定処理. 電子情報通信学会研究会報告 SP98-136, Vol. 98, No. 611, pp. 1–6, 1999.

[2] 木下智義, 村岡秀哉, 田中英彦. 単音の遷移に注目した単音認識処理. 日本音響学会誌, Vol. 54, No. 2, pp. 190–198, March 1998.

[3] 柏野邦夫, 中臺一博, 木下智義, 田中英彦. 音楽情景分析の処理モデル OPTIMA における単音の認識. 電子情報通信学会論文誌, Vol. J79-DII, No. 11, pp. 1751–1761, 11 1996.

[4] 柏野邦夫, 木下智義, 中臺一博, 田中英彦. 音楽情景分析の処理モデル OPTIMA における和音の認識. 電子情報通信学会論文誌, Vol. J79-DII, No. 11, pp. 1762–1770, 11 1996.

[5] 柏野邦夫, 村瀬洋. 単音連繋確率ネットワークに基づく音楽演奏の音源同定. 人工知能学会誌, Vol. 13, No. 6, pp. 962–970, 11 1998.

[6] 柏野邦夫, 村瀬洋. 適応型混合テンプレートを用いた音源同定 — 音楽演奏 への応用 —. 電子情報通信学会論文誌, Vol. J81-DII, No. 7, pp. 1510–1517, 7 1998.