

OPTIMAにおける統計的単音仮説生成処理

中臺 一博 柏野 邦夫 木下 智義 田中 英彦

東京大学工学部

1 はじめに

われわれは、音楽情景分析における処理モデルとして OPTIMA を提案し、これに基づく音楽情景分析の実験システムの実装・評価を行った [1, 2]。

本稿では、実験システムのうち、周波数成分レベル、単音レベル間の処理を行う単音仮説生成処理部の実装および、評価について述べる。

2 単音仮説生成処理部の実装

単音仮説生成処理部は、以下の 3 つの部分から構成される。

- 単音形成クラスタリング部
- 音源同定部
- 抽象音記憶によるパターン照合部

2.1 単音形成クラスタリング部

単音形成クラスタリング部は、立上り情報、高調波情報を利用して、処理単位内の周波数成分が単音を表すようにクラスタリングし、音色なし単音仮説を生成する。ここで、音色なし単音仮説とは音高情報のみからなる音源同定を行っていない状態の単音仮説を表すものとする。また、音高がオクターブの整数倍違う仮説を生成するなど、重複周波数成分に対応するために、各周波数成分が複数のピッチ仮説に属することを許している。なお、実験システムでは、1 つの処理単位内に含まれる音色なし単音仮説は数は 3 音までという上限をおいている。

2.2 音源同定部

音源同定部では、予め主成分分析によって生成された音色空間上で、音色なし単音仮説を判別分析することによって音源を同定し、単音仮説を生成している。判別分析の際の距離関数には、各カテゴリの重心だけでなく、分散まで考慮にいたれた距離を算出することができるマハラノビスの汎距離を使用した。

また、音色空間の生成法は以下の通りである。まず、各単音サンプルの周波数成分から、過去の関連研究、および楽器の構造を考慮して、楽器の特徴を良く表していると考えられる特徴量 (表 1 参照) を 41 次元のベクトル

ルとして抽出する。次に、抽出した特徴量に対し主成分分析を行い、空間軸を決定する。最後に、各カテゴリをこの空間上に n 次正規分布表現することによって音色空間を生成する。

音色空間には、楽器毎にカテゴリを生成したフラットな構造のものと、発音形態が近い音色を同一カテゴリとみなし、さらに各カテゴリ内で階層的に音色空間を生成した構造のものを併用した。

表 1: 抽出特徴量

1. エンベロープに関する特徴

- 立上り時間、傾き
- エンベロープ重心
- 持続音かどうか
- AM、FM の度合

2. 高調波に関する特徴

- 高調波パワー比
- 高調波の立上り時間の遅れ
- 最高次高調波の次数、周波数
- 高調波数

2.3 抽象音記憶によるパターン照合部

抽象音記憶とは、単音を構成する周波数成分の集合を基本周波数値が 1.0、最大パワー値が 1.0 となるように正規化し、音源名および、適用できる音高範囲情報を付加したものである。

パターン照合部では、まず、単音仮説の組の情報にしたがって事前登録された抽象音記憶を組み合わせて、混合音仮説を生成する。この際、抽象音記憶では、各周波数成分が周波数、パワー情報に関して正規化されているため、そのままでは抽象音記憶を組み合わせて、合成することができない。このため、処理単位内の対応する周波数成分を検出し、その周波数値、パワー値に合わせて抽象化された情報の復元を行っている。また、混合音仮説生成の際には、各抽象音記憶の時間のずれ、および重複周波数成分がある場合には、位相のずれをパラメータとして、各単音仮説の組から複数の仮説を生成する。次に、混合音仮説と処理単位のマッチングを行い、距離を計算し、音楽情景分析システムへの出力である条件付確率を算出する。マッチング距離から確率を算出する際には、各混合音仮説に対して、処理単位との距離の分布が正規分布であることを仮定し、距離に対する確率の算出を行う。また、OPTIMA における統合を行うことにより 2.2 節で述べた音源同定部で同定できなかった音源の補完同定も行うことができる。

3 評価実験

以下の評価実験を行った。

実験 1 音色空間の作成

フルート、ピアノ、トランペット、バイオリン、クラリネットの 5 種類の楽器で演奏された単音サンプルを使用した。音源には自然楽器音を収録したサンプラーを使用し、MIDI のノート番号でピアノは 48~83、それ以外は 62~83 の範囲の音が含まれている。総サンプル数は 124 音である。

寄与率の変化による認識率の変化調べるため、各サンプルから特徴を抽出し、主成分分析を行った。主成分分析結果に対し 95% ~ 60% の寄与率で音色空間を生成し、サンプル音の音源同定を行った。結果を図 1 に示す。

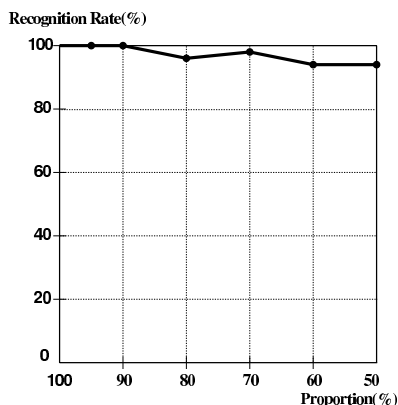


図 1: 寄与率と単音認識率

実験 2 混合音認識テスト

混合音に対する認識率を調べるため、一定時間毎に、2 または 3 音が同時に発音する混合音認識テストデータを用意した。このデータは、周波数成分の重なり方に応じて、以下の 3 種類に分類されている。

- クラス 1
混合音中に、基本周波数成分が必ず他の周波数成分と重なってしまう単音が含まれているデータ。
- クラス 2
混合音中に、基本周波数成分は重ならないが、第 2 高調波成分、または第 3 高調波成分が必ず他の周波数成分と重なってしまう単音が含まれているデータ。
- クラス 3
混合音中に、基本周波数成分、第 2 高調波成分、第 3 高調波成分のいずれにも重なりが生じないデータ。

これらのデータに対し、

1. ボトムアップの単音形成クラスタリングの認識率
2. ボトムアップの音源同定まで含めた認識率
3. パターン照合処理との統合後の認識率

の 3 通りの測定を行った。テストデータの音色は、混合音数 2 音のものは、フルート、ピアノ混合音数

3 音のものは、フルート、クラリネット、ピアノからなるデータを使用した。また、音色空間は寄与率 95 % で作成したものを使用した。

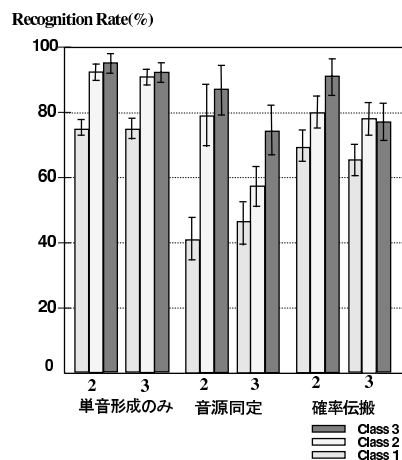


図 2: 混合音認識率

4 考察

実験 1 より、寄与率 90% 以上で作成した音色空間は、100% 単音の音源同定が、達成されている。少なくとも寄与率 90% 以上で作成された音色空間を使用すれば、単音が単独で存在する場合には、正確な音源同定を行うことができるといえる。

実験 2 では、混合音に対する認識率の測定を行った。周波数成分の重なりがあまり見られないクラス 3 のデータの場合、高い認識率を示していることがわかる。しかし、基本周波数成分、または低次の高調波成分が、重なってしまうようなクラス 2、クラス 1 のデータに対しては、単音形成クラスタリング、音源同定処理ともに認識率が悪くなっている。これは、単音形成クラスタリング、音源同定処理が、混合音に対してあまり効果的でないためであるが、確率伝搬後の認識率をみると、OPTIMA によって、和音など混合音の認識に効果的なパターン照合モジュールと協調・統合されていることがわかる。

5 おわりに

音楽情景分析の処理モデル OPTMA における単音仮説生成処理について述べ、その単独評価を行い、単音仮説生成処理の有効性を示した。また、単音仮説生成処理部を音楽情景分析システムの一部として動作させ、確率伝搬によるモジュール統合を行うことにより、単音仮説生成処理の精度を向上を達成するとともに、OPTIMA の有効性を示すことができた。

参考文献

- [1] 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 の実装”, 第 50 回情報大, 6D-2 (1995).
- [2] 木下智義, 柏野邦夫, 中臺一博, 田中英彦: “音楽情景分析の処理モデル OPTIMA におけるシーン情報の抽出と利用”, 第 50 回情報大, 6D-3 (1995).