

# 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

木下 智義 柏野 邦夫 中臺 一博 田中 英彦

東京大学工学部

## 1 はじめに

OPTIMA では、複数の独立したモジュールに確率をもった仮説の組を出力させ、これを確率伝搬によって統合することによって外界の音響的事象に関する最尤推定像を求める。本稿では OPTIMA において利用される音楽シーン情報として、拍位置および和音の情報の抽出と利用について議論し、実験システムに対する評価実験の結果を示す。

## 2 音楽シーン情報の抽出

音楽シーン情報は、本稿では複数の単音を特徴づける記号表現のことを指している。本稿では、継時的な性質を持つ音楽シーン情報の例としてリズムを、また同時的な性質を持つ音楽シーン情報の例として和音の情報を扱う。

### 2.1 拍位置情報の抽出

音楽には通常リズムが存在し、その拍はほぼ一定間隔毎に現れる。そこで、単音・周波数成分の立上り時刻を、Desain の方法 [2] によって量子化した上で、立上り時刻間隔の統計をとり、最頻値をもってリズム間隔とする [1]。

### 2.2 和音情報の抽出

和音名を得るには、その時刻に存在する単音仮説の組と、和音毎の構成音の組を比較し、その類似度に応じて和音仮説の確率を与える。この時、和音の基音を含む単音仮説については、類似度が大きくなり、逆に和音の基音が含まれない単音仮説については類似度が大幅に低下することを利用する。

## 3 音楽シーン情報の利用

### 3.1 拍位置情報の利用

OPTIMA では、周波数成分抽出部 (スペクトログラム→周波数成分) の出力する周波数成分情報に、端点候補点を設けている。端点候補点とは、周波数成分上の点であって、周波数成分の端点となる確率を持つ点である。その確率値は、統計的情報に基づくベイズ的統合により与えられている。端点候補点の状態を確定 (切断 or 接続) させるためにリズムを利用する。

リズムの拍がある位置には、音の立ち上がりが存在する確率が高く、それ故、拍の位置にある端点候補点は端点として認識することができる。逆に、拍位置から離れた位置の端点候補点は単なる周波数成分の一部として認識する。

周波数成分抽出部では、ローカルピークなどの局所的な情報から端点を決定しており、これにリズムという時間方向にわたる大局的情報を統合することによって端点

認識率の向上が期待できる。情報の統合には、Dempster の結合規則を用いている (図 2)。

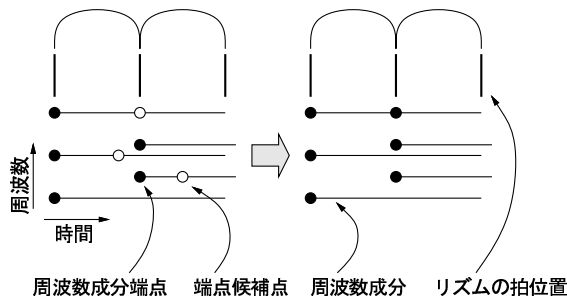


図 1: 端点の認識

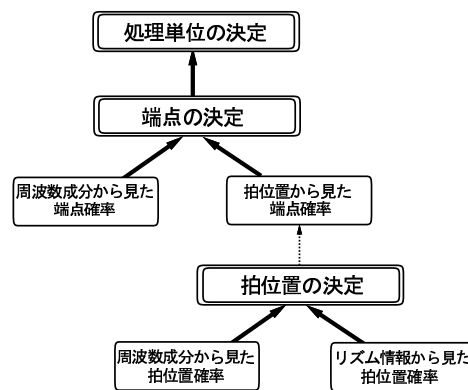


図 2: 確率値の統合による端点候補点の確定

### 3.2 和音情報の利用

和音を時間方向に見た場合、その並びには統計的に一定のパターンがある。そこで、和音遷移のパターンについて確率を与えることにより、時間軸方向に情報を統合してシステム全体の処理精度の向上を図る。

例えば、単音抽出に雑音が含まれていた場合、和音抽出の認識率が低下するが、和音遷移パターンからの統計を用いて和音遷移確率を求め、和音仮説そのものの確率と統合することでその誤りを補完することができる。

また、和音名から単音の存在確率を得ることができ、この確率を単音抽出にフィードバックさせている。

ここで、和音遷移確率情報と単音存在確率情報は、印刷された楽譜から和音遷移頻度・単音存在確率を取得してデータベースとして保管したものを参照している。

現時点において、統計情報の収集に使用した曲数は 206 曲、遷移サンプルは 3570 個。構成音統計データを得るのに用いたデータは 2237 の音符からなる。

## 4 評価実験・結果

### 実験 1: 拍位置情報

拍位置情報の利用の有無による、周波数成分の端点認識率の変化を調べた。

自然楽器音を収録したサンプラーを用いてサンプル曲（「蟹の光」の室内楽アンサンブル）を演奏したモノラルの音響信号に対して処理を行い、拍位置情報の抽出・利用をしなかった場合とした場合の端点認識率を比較した。端点認識率とは、端点候補点のうち、切断・接続が正解と一致したものの割合をいう。

使用した楽器は、ピアノ (p) ・クラリネット (c) ・トランペット (t) ・バイオリン (v) のうちから 2 パートまたは 3 パートである。

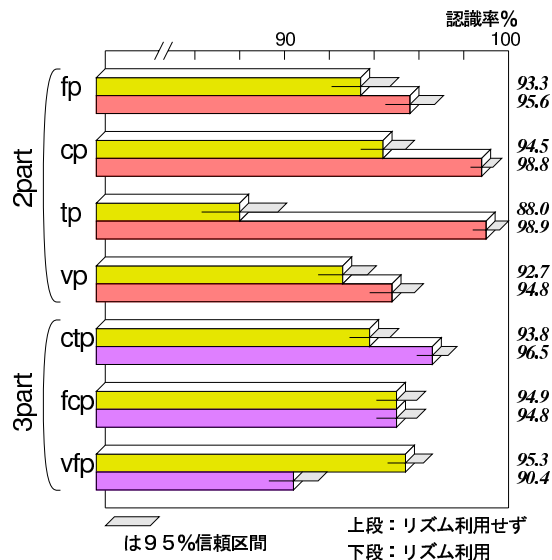


図 3: 楽器別の端点認識率

### 実験 2: 和音情報抽出

和音は単音仮説の組から得られる音高の情報から抽出される。ここでは、C を基音とする和音の構成音からなる仮説にノイズを加えてその和音認識精度の差を見る。

ノイズはランダムに単音を重畳もしくは除去することで加える。精度は出力仮説の上位  $n$  位に含まれる割合で示す。

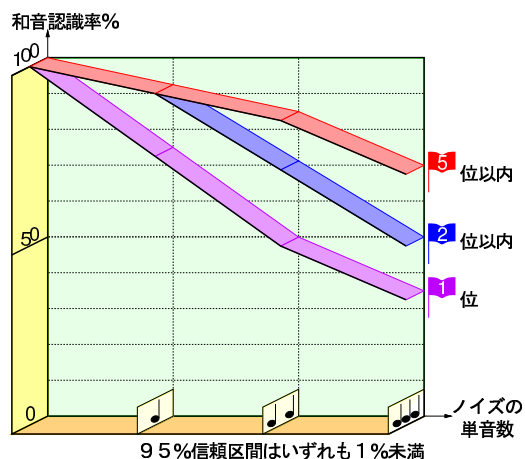


図 4: 和音認識率

### 実験 3: 和音情報利用

正解の和音遷移を実際の楽譜における和音遷移を元に設定し、その構成音からなる単音群データを作成した。そして、それに実験 2 と同様にランダムなノイズを加えて処理を行い、和音遷移情報から得られる和音遷移確率の伝搬を行うことによる和音認識率の変化を調べた。

**実験 3.1** 今回設定した正解は 18 個の和音からなる。18 個ある和音のうち 1 つについて、単音 1 つのノイズを加えて認識率の差を見た。

**実験 3.2** 実験 3.1 のノイズを単音 2 つにして認識率の差を見た。

**実験 3.3** 18 個ある和音の全てについて、単音 1 つのノイズを加えて認識率の差を見た。

**実験 3.4** 実験 3.3 のノイズを単音 2 つにして同じ比較をした。

結果は以下の通りである。

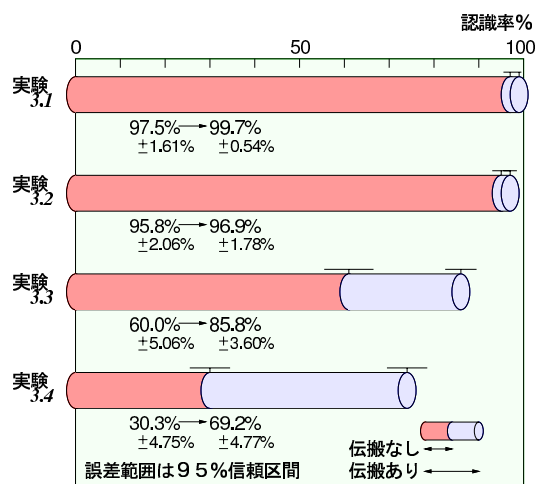


図 5: 和音遷移情報の効果

## 5 考察

実験 1 において、フルートやバイオリンを用いた場合には精度が若干悪くなった。これはこれらの楽器の周波数成分の振幅の変化が激しく、振幅変化の谷が周波数成分の端点として出力されてしまっていることが理由として考えられる。

実験 3 において、ノイズとして加えられた単音の個数が多い実験 3.3, 3.4 では効果が顕著であり、特にノイズの割合が大きい実験 3.4 では最も伝搬の効果が現れた。一方で、実験 3.3, 3.4 では伝搬後の和音認識率が実験 3.1, 3.2 に比べて下がったが、これは加えたノイズが大きいため正解の仮説順位が低くなっているためと考えられる。このことは実験 2 の結果からも明らかである。生成する仮説の順位をもっと下位のものまで広げることで、精度が向上するものと考えられる。

## 6 おわりに

音楽シーン情報を用いた周波数成分端点確定・和音仮説生成モジュールについて述べ、またその実装・評価を行って情報統合の有効性を示した。今後はフルートやバイオリンといった複雑な周波数成分構造を持つ楽器への対応をさらに検討したい。

## 参考文献

- [1] David Felix Rosenthal : Machine Rhythm , Computer Emulation of Human Rhythm Perception , Doctoral Thesis , MIT (1992).
- [2] Peter Desain and Henkjan Honing : The Quantization of Musical Time , A Connectionist Approach , Computer Music Journal 13:pp.56-66 (1989).