1B−4

音楽単音記号列生成システムの処理モデル OPTIMAにおける単音仮説生成処理

中臺 一博　柏野 邦夫　田中 英彦
東京大学工学部

1 はじめに

われわれは、音楽単音記号列生成システムにおける処理モデルとして OPTIMA を提案した。[1] OPTIMA では、モジュールが確信度をもった仮説の組を出力する場合、これを確率伝搬によって統合することができる。したがって、音楽単音記号列生成システムのように複数の情報を統合する必要がある場合には、有用な処理モデルであるということができる。 OPTIMA の処理のうち本稿で扱う単音仮説生成モジュールでは、各仮説に確信度を与えなければならないため、確信度の与え方が問題である。このような確信度を与える単音仮説生成モジュールとして、音記憶を使用したモジュールが実装されている。このモジュールは音記憶から生成した混合音仮説と入力とのマッチングを行うモジュールであり、和音などの混合音の認識に効果的であった。しかし、一音ごとに音記憶が必要であること、および混合音数の増加にともない計算量が爆発してしまうことなど、効率、精度の面で音記憶だけでは限界があった。

そこで、これらの問題を解決するために音色としての本質的な特徴を抽出し、音色空間上に表現を行った。このような音色空間を利用した楽器の類別、認識の研究はニューラルネットワークを使用したものなどがあり [2]、単音に関しては良好な結果が得られている。そこで、本稿では音色空間の利用により、確信度をもった仮説の組を出力し、混合音に対しても認識を行うことができる単音仮説生成法を提案する。

この手法では、各単音仮説の確信度は統計的手法により算出することができ、知識は音色ごとに与えられるため、音数に対する知識量の爆発、計算量の爆発を抑えることができる。

2 音色空間の生成

2.1 音色空間の軸の決定

音色空間を生成するためには、音色空間の軸を決定する必要がある。音色空間の軸は、次のようにして定めた。

1. 単音の周波数成分データから表1に示す特徴量を抽出する。表1に挙げた特徴は、文献 [3]、および楽器音の構造を考慮して選択したものである。
2. 抽出した特徴ベクトルに対し主成分分析を行い、音色空間の軸を得る。

主成分分析は、各特徴量から、サンプルの分布をよりよく表すような互いに直交する新しいパラメータを導く手法であるが、寄与率を導入することによりパラメータ数を減らし、以後の処理で計算量を抑えることができる。

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
   - 倍音パワー比
   - 倍音の立上り時間の遅れ
   - 最高次倍音の次数、周波数
   - 倍音数

2.2 音色カテゴリの表現法

音色空間上で音色カテゴリを表す場合、様々な表現法が考えられるが、本稿では、各音色は音色空間上で、音色空間の次数を n とした場合、 n 次元の正規分布として表すことができると仮定する。この場合、各音色のカテゴリは音色重心および分散・共分散行列で表現することができる。音色 A に属する単音サンプルを A_iと表し(サンプル数:n)、A_iの j 番目の特徴量を x_{ij}とすると音色重心$\bar{x}_j$、分散・共分散行列 V はそれぞれ、

$$\bar{x}_j = \frac{1}{n}\sum_i x_{ij} \tag{1}$$

$$V_{jk} = \frac{1}{n-1}\sum_i (x_{ij}-\bar{x}_j)(x_{ik}-\bar{x}_k) \tag{2}$$

と表すことができる。このように、音色ごとに音色重心、分散・共分散行列を知識として与えるだけであるので、音記憶に比べ知識量が少なくてすむという利点がある。また、このような表現を用いれば、音色同定の際、次節で述べる判別分析を行うことによって、統計的に各音色に属する確率を算出することができる。

Creation of Single Note Hypotheses in OPTIMA
Kazuhiro Nakadai Kunio Kashino Hidehiko Tanaka
University of Tokyo, Department of Electorical Engineering
7-3-1 Hongo, Bunkyo-Ku, Tokyo, 113, Japan.

3 音色空間を使用した単音仮説生成処理

単音仮説の生成は判別分析によって行う。判別分析はサンプルがどのカテゴリに属するかを判別する手法で一般的には、2 カテゴリのものを扱うために用いられる。判別分析を行う方法としては、

- 線形判別関数
- マハラノビスの汎距離

の二通りが考えられる。マハラノビスの汎距離は、ユークリッド距離と比べ、カテゴリの分散を考慮した距離ということができ、線形判別関数と比べて、

- カテゴリ数が 3 つ以上の場合へ自然に拡張することができる。
- どのカテゴリに属するかを確率として算出することができる。

といった利点があるため、マハラノビスの汎距離を使用して判別分析を行った。i 番めのサンプルと音色 A とのマハラノビスの距離 D^2は、音色 A の音色重心を$\bar{x}_j$、分散・共分散行列の逆行列を $S = V^{-1}$とすると

$$D_i^2 = \sum_j \sum_k (x_{ij} - \bar{x}_j) S_{jk} (x_{ik} - \bar{x}_k) \tag{3}$$

と表すことができる。また、音色の分布は正規分布を仮定しているので、マハラノビスの汎距離を使用して、i 番めのサンプルが音色 A に属する確率 p を式 4で算出することができる。

$$p = \frac{1}{(2\pi)^{m/2}\sqrt{|S|}} e^{-\frac{1}{2}D_i^2} \tag{4}$$

ただし、m は音色空間の次数、Sは分散・共分散行列を示す。同様の操作を他の音色カテゴリに対して行うことにより、入力サンプルがそれぞれの音色カテゴリに属する確信度を算出することができる。

4 評価実験・結果

前述したモジュールのプロトタイプを実装し、評価実験を行った。

1. 音色空間の作成実験

   フルート、ピアノ、トランペット、チェンバロからなるサンプル音に対し、特徴抽出、主成分分析を行い音色空間を作成した。サンプル音は PCM 音源を使用しており、MIDI のノート番号で 79〜65 の範囲の音が含まれている。総サンプル数は 90 音である。結果を表 2に示す。

2. 音色同定実験

   - 音色空間作成データに対する音色同定

     実験 1 で音色空間を作成したサンプル 90 音に対して判別分析による音色同定を行った。音色空間は、寄与率 95%で作成したものを使用した。第 1 候補を回答としたところすべて正しく認識することができた。

   - ベンチマークデータに対する音色同定

     フルートとピアノの 2 音からなるベンチマークに対して音色同定を行った。単音同定クラスタリングによって、正確にクラスタリングされたもののうち 95%が正しく認識された。

表 2: 実験結果 1

寄与率	特徴ベクトル数	音色空間次数
95%	41	11
99%	41	19

5 考察

実験 1 において、41 次元の特徴ベクトルを寄与率 95%で音色空間の次元を 11 次元まで減らすことができた。ただし、サンプル数、音色数の増加にともない、音色空間の次数が増加することが考えられる。次数が増加すると、モジュールがノイズに対して弱くなることが考えられるので、同定結果に悪影響を及ぼすことが可能性がある。これに対応するためには、音源の発音機構を参考に音色の階層化を行い、音色空間の次元を減らすことなどが考えられる。実験 2 において、音源同定結果としては、良好な結果が得られている。この場合も、サンプル数、音色数が増加がどの程度結果に影響を及ぼすかを調べる必要がある。また、評価実験は、PCM 音源の音を使用しているため、実音データで音色空間がどの程度有効であるかもあわせて調べる必要があると考えられる。

6 おわりに

音色空間を使用した単音仮説生成処理について述べ、プロトタイプを実装・評価することにより音色空間の有効性を示すことができた。この手法では、各単音仮説の確信度を統計的手法に基づいて導出することができ、かつ音記憶を使用する場合と比べ知識量が少なくなるということができる。今後は、実音に対応した音色空間の生成、サンプル数、混合音数が増加した場合への対応が必要であると考えられる。

参考文献

[1] 柏野邦夫, 中臺一博, 田中英彦 : “OPTIMA: 音楽音響信号から単音記号列を生成するシステムの処理モデル”, 第 49 回情処全大,1B-3 (1994).

[2] Talor, I. and Greenhough, M.: An Object Oriented ARTMAP system for Classifying Pitch, *Proc. of ICMC (International Computer Music Conference)* (1993).

[3] 安部素嗣 : “楽器音の学習認識システム”, 卒業論文, 東京大学, 工学部計数工学科 (1991).