

# 音色空間を用いた音源同定システム

2L-02

日比 啓文<sup>†</sup> 半田 伊吹<sup>\*</sup> 坂井 修一<sup>\*\*</sup> 田中 英彦<sup>\*\*</sup>  
 東京大学工学系研究科<sup>\*</sup> 東京大学情報理工学研究所<sup>\*\*</sup>

## 1.はじめに

長年にわたり、コンピュータで音楽から楽音を認識し、楽譜をつくる研究がなされてきた[荻野 96]。それらのシステムは一般的に楽音の周波数成分が重なりあったときに楽音を同定することは難しく適応処理も考案された[木下 00]。しかし、従来の手法は解析空間として確率空間やスペクトログラムを用いているため、時間周波数解析で分離できない楽音に対する音源同定は困難であった。本稿では楽器の種類は有限であるという観点から、楽器が出現する音色空間を解析空間として用い、楽音の周波数成分の重なりを考慮した音源同定システムを紹介する。また、音色空間の研究は未だ確立しているとは言えないため、本稿では音高、音長によって音色には違いがあるという観点から音色空間写像を用いて、一定の音高、音長で音色空間を扱うことを試みた。

## 2.楽器の音色空間[McAdams 99]

音色空間の研究は心理学の分野からはじまり、Grey が楽器の音色距離をアンケート等から測定し、それらの距離を元に 3次元空間を生成したことを発端とする[Grey 77]。

近年では McAdams らが音色特徴量の計算式を提案しており、これらの音色特徴量に従って音色空間を形成した [McAdams 99]。この研究で用いられている音色特徴量は 1) スペクトルの重心、2) スペクトルの立ち上がり時間、3) スペクトルの細かい構造、4) スペクトル流動、の 4 つである。この研究では、アンケートによる観測データを CLASCAL のアルゴリズムを用いて作られた音色空間軸と、提案しているの計算式を用いて求められたの軸の確からしさを潜在構造解析で検証している。しかし、これらは計算式としては音高、音長の考慮はされているが、実際には音高、音長の違いによって、音色空間の位置は大きくずれてしまうなどの問題点があげられる。

Sound Source Identification System using Timbre Space,  
 Hirofumi Hibi, Ibuki Handa, Shuichi Sakai, Hidehiko Tanaka,  
<sup>\*</sup> Graduate School of Eng. Univ. of Tokyo <sup>\*\*</sup> Graduate School of  
 Information Science and Technology, Univ. of Tokyo  
 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113, Japan

## 3.音色空間写像

先述の通り McAdams の提案した音色空間は音高、音長の違いによって大きく値が変化した。その変化の影響をなくすための音色空間写像を紹介する。

まず時間周波数解析の結果からは、それぞれの調波構造の音長と基本周波数を検出することができる。この検出された音長と基本周波数の音高をもちいて、従来の音色空間から音長と音高の差異をなくした仮想音色空間上への写像のことをここでは音色空間写像と呼ぶ。そして、ある音高  $i$ 、音長  $j$  の音色空間上の点  $X$  から固定音色空間上の点  $Y$  への音色空間写像  $F_{ij}$  は式(1)のように定義する。ここで  $n$  はサンプル音源データの楽器数、 $M$  は音色空間の次元数、 $P_{ij}$  はサンプル音源データ、 $I$ 、 $J$  はサンプル音源データの音高数と音長数、表記のベクトルはそれぞれ  $n$  次のベクトルである。

$$P_{ij} = (\bar{p}_{ij1}, \dots, \bar{p}_{ijM}), F_{ij} = (\bar{f}_{ij1}, \dots, \bar{f}_{ijM})$$

$$F_{ij} P_{ij} = \frac{1}{I \cdot j} \left( \sum_{j=1}^J \sum_{i=1}^I P_{ij} \right) \quad \text{式(1)}$$

$$X = (\bar{x}_1, \dots, \bar{x}_M), Y = (\bar{y}_1, \dots, \bar{y}_M)$$

$$Y = F_{ij} X$$

## 4.音色空間を用いた音源同定システム

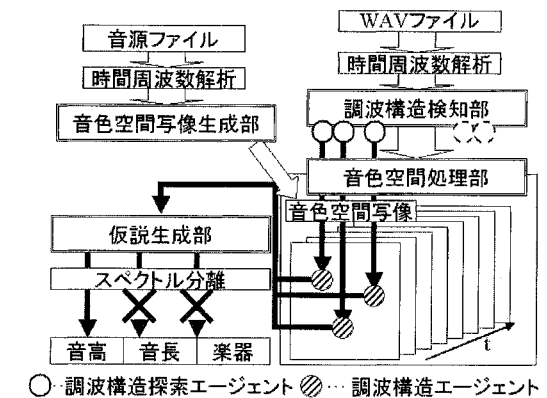


図1:音源同定システムの全体像

本稿で提案する音源同定システムの全体像を図1に示す。システムは入力を音楽WAVフォーマットファイルとし、出力を楽器名、音高、音長とし、大きく音色空間処理部、調波構造検知部、音色空間処理部、仮説生成部の4つからなる。実験の入力WAVファイルとしてでは「蛍の光」の3重奏を用いた。

#### 4.1 調波構造検知部

調波構造検知部は、時間周波数解析の結果から調波構造探索エージェントを用いて調波構造を検出し、調波構造エージェントを生成する部分である。動作として1)時間周波数解析結果から得られる3次元空間上に、十分な数の調波構造探索エージェントを配置する、2)各々の調波構造探索エージェントは自律的に時間方向、周波数方向を移動し、一番近いソワワーの極大値を探索する、3)各々の調波構造探索エージェントは周波数の整数倍の位置に高調波探索エージェントを配置し、各々の高調波探索エージェントは自律的に高調波成分を探索する、4)各々の調波構造探索エージェントは高調波成分の2乗平均平方根が極小値もしくは閾値より小さくなる点を時間方向に探しこれを調波構造の両端点として調波構造を検出、5)調波構造探索エージェントが他の高調波探索エージェントと同じ波形を追跡していないとき、調波構造エージェントを生成する、という手順を経て調波構造エージェントを生成する。また、調波構造エージェントは高調波成分の重なりが検知されたとき、その情報も保持する。

#### 4.2 音色空間写像生成部

音色空間写像生成部は3.で示した音色空間写像関数  $F_{ij}$  をあらかじめ計算してデータとして保持しておく部分である。実験では、音源ファイルとしてサウンド・キャンバス SC-8850 から得られたピアノ、ヴァイオリン、クラリネット、フルート、トランペットの5種類の楽器音を用いた。また、音色空間軸として2.で提唱されている1)スペクトルの重心、2)スペクトルの立ち上がり時間、3)スペクトルの細かみ構造、4)スペクトル流動を用いた。音色空間写像関数  $f_{mn}$  としては2次のx方向B-スプライン曲線関数を用いた。

#### 4.3 音色空間処理部

音色空間処理部は、時間方向に並ぶ仮想音色空間と調波構造エージェントの仮想音色空間上の位置を保持しておく部分である。動作として、1)ある時間の仮想音色空間上で同時に出現する調波構造エージェントの数を元に音源数と音源の仮想音色空間上の位置を特定する、2)音色空間写像をもちいて、入力

音楽ファイルから求められた様々な音高、音長を持つ調波構造エージェントを同じの仮想音色空間上に保持する3)仮説生成部と連動して、仮説の検証を行うときの仮想音色空間上の出現位置を保持する、の3つがある。

#### 4.4 仮説生成部

仮説生成部は音色空間処理部から得られた音源数と音源位置からの距離を用いて、それぞれの調波構造がどの音源に属するかの仮説を生成する部分である。ここで仮説を生成するために用いる距離  $D$  には式(2)で表すマハラノビス凡距離を用いる。なお、 $S$ は分散共分散行列である。

$$\bar{X} = (X - \overline{F_{ij} P_{ij}})$$

$$D^2 = \bar{X} S^{-1} \bar{X}'$$

生成される仮説は、大別して1)周波数の重なりがなく良好な仮想空間の位置関係にある、2)周波数成分の基本周波数同士が一致もしくは整数倍の関係になっている、3)周波数成分の一部同士が重なっていて、お互いの仮想音色空間位置がずれている、と大きく3つに分類される。このうち、2)は基本周波数が同じ、2倍、3倍と仮説を立てその周波数成分を、3)は重なる周波数成分を調波構造エージェントが保持しているためその周波数成分を、「お互いかわらない」、「片方が使う」、「両方が使う」、「両方で1/2ずつ使う」といった場合を考え、スペクトル分離を行いその結果を音色空間処理部に送り、仮想音色空間位置より判断する。

#### 5. まとめ

本稿では音色空間を用いて楽器の重なりを認識するシステムを紹介した。周波数成分が重なったとき、音源同定が非常に困難であるという問題に対し、音色空間を用いた解決策を提案した。

#### 参考文献

- [柏野 96] 柏野, 中臺, 木下, 田中: 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-D II (11), No.11, pp.1751-1761, 1996  
 [木下 00] 木下 智義, 坂井 修一, 田中 英彦: 周波数成分の重なり適応処理を用いた複数楽器の音源同定処理, 信学論(D-II), Vol.J83-D-II, No.4, pp.1073-1081, 2000  
 [McAdams 99] McAdams, S.: Perspectives on the Contribution of Timbre to Musical Structure, Computer Music Journal, Vol.23, No.3, pp.85-102, 1999  
 [Grey 77] J. M. Grey: Multidimensional perception scaling of musical timbres, J. Acoust. Soc. Am., Vol.61, No.5, pp.1270-1277, 1977