

7B-3

音源分離システムにおける時間的統合

— Old-Plus-New Heuristic の導入 —

柏野 邦夫

田中 英彦

東京大学 工学部

1 はじめに

人間は、複数の音源からの音が同時に存在した場合にも、ある音だけを選択的に聴くことができる場合が多い。音源分離システムは、このような機能を工学的に実現することを目標としたものである。これまでに、われわれは、複数種類の楽器音の混在したモノラルの音響信号を入力とし、音色モデルの事前登録なしに含まれる楽器音を分離し、楽譜に類似した形式および MIDI データの形で楽器ごとの演奏情報を出力するシステムの構成を提案し、実験システムを実装して評価実験を進めてきた[1]。

このシステムにおける問題点のひとつは、異種の楽器音に含まれる周波数成分どうしが高調波関係にあり、かつほぼ同時に立ち上がっている場合に、これらを分離することができないことであった。このような問題点は、システムが、スペクトルの瞬時的な情報のみに基づいて処理を行っていることに由来すると考えられる。そこで本稿では、スペクトルの継続的な情報を音源分離処理に導入する一方法を提案する。音源分離システムにおいて、スペクトルの継続的な情報の利用に関しては、情報の時間的なつながりから音源分離を行うという側面と、音源分離の結果から情報の時間的なつながり(例えばオーディトリ・ストリーム)を形成していくという側面の、2つの相補的な処理が考えられる。これらをまとめて、スペクトルの「時間的統合」と呼ぶことにするが、本稿ではこのうち前者に関する検討の第1歩として、いわゆる "Old-Plus-New Heuristic" の導入を考える。

2 Old-Plus-New Heuristic

Old-Plus-New Heuristic は、人間の聴覚情報処理に関する仮説の一つであって、いくつかの実験結果に矛盾しない処理過程として、Bregman によって名付けられたものである[2]。これは、次のように考えることができる。

「Old-Plus-New Heuristic とは、時間を追っていくつかの複合音が表示されているとき、ある時点において存在している周波数成分の部分集合が、過去にひとつの音色として解釈された周波数成分の集合と、性質が類似しているならば、その部分集合を過去の音色と同じひとつの音色として解釈すること、さらに、該当する周波数成分から、ちょうどその解釈に用いられた分だけのパワー量を差し引き、その残余について解釈(各音色への分離)を続行することである。」

ここでは、実際に人間がこのような処理を行っているかどうかの議論には立ち入らないことにするが、工学的観点からは、この種の継続的な情報の導入によって、前節に挙げた問題点を緩和することができると考えられる。しかし、上に挙げた説明はきわめて概念的であって、実際にこのような処理を音源分離システムに導入するためには、少なくとも次のような点を検討する必要がある。

- 性質の類似度をどのように定義するか
- 時間の経過をどのように評価するか
- 差し引くパワー量をどのように決定するか

Old-Plus-New Heuristic を導入する方法は、これまでに提案した評価ルールと統合ルールに基づくシステムに、「音」を事例とした事例ベースの処理を加えたものと見ることでもできる。また、本稿の方法と、事前に対象とする音色を登録して音源分離に用いる方法とを比較すると、本稿の方法では、いわば未知の音色の学習機能を含むことが新規な点である。ただし、これは、一度はその音色が他の音色と分離できる場面(現在のシステムでは、周波数成分が他の音色の周波数成分と高調波関係にないか、またはそれらの立ち上がり時点が異なっているかのいずれか)で発音されていることが前提である。なお、実用面から見れば、対象とする楽器の具体的な音色モデルが事前に特定できるかどうかによって処理の適性が異なり、これが特定できない場合には本稿の方式が適しており、特定できる場合には、音色の学習の誤りが生じ得ないという点で事前登録方式が有利であると考えられる。

3 処理の概要

提案するシステムの概要を、図1に示す。このうち「ルールに基づく音源分離」までは、これまでに提案したシステムと同じ内容の処理を行う。すなわち、楽器音を構成する周波数成分としてスペクトログラム上のローカルピークを抽出した後、これらの高調波関係および立ち上がりの同時性を評価するためのルールと、評価値を統合するルールとによって、任意の周波数成分間の距離を定め、これに基づいてクラスタリングを行う。評価ルールは、聴覚実験の結果を参考にして定めており、統合ルールとしては、Dempster の結合法則を用いている。この段階までで、人間が同じ音源からの音として聞く確率の高い周波数成分がクラスタ化される。

従来の処理では、これに続いて、各クラスタの特徴量に基づいて更にクラスタリングを行うことにより音源の同定を行っていたが(なお本稿では、音源の同定とは楽器名の同定ではなく同種の楽器音であることの同定を指す)、本稿に提案する処理では、この部分に Old-Plus-New Heuristic を導入して、

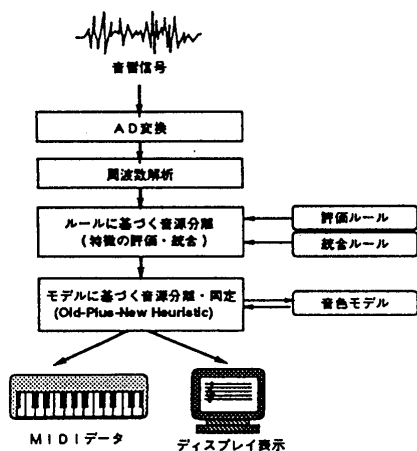


図1: 提案する音源分離システムの構成

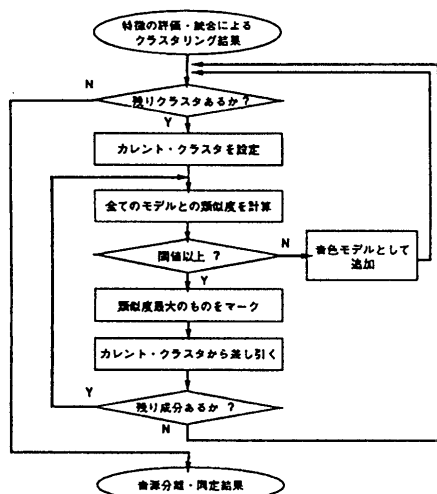
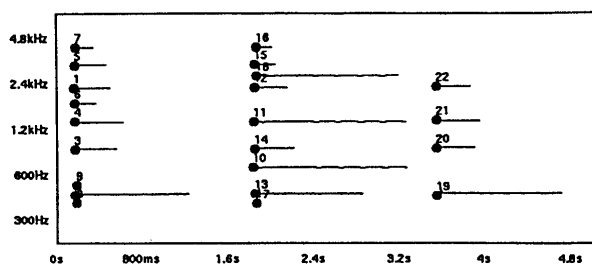


図2: モデルに基づく音源分離・同定の処理の流れ

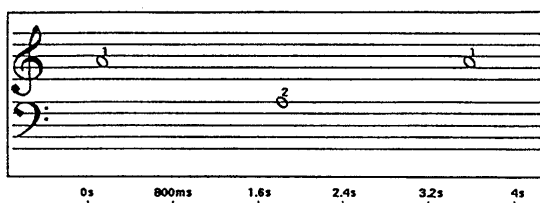
ルールに基づく処理で分離できなかった音源の分離を行うと同時に音源の同定を行っている。図2に、この部分の処理の概略を示す。

ねらいとする動作の説明のために、図3に、簡単な入力に対する本システムの出力例を示す。これは、ピアノとフルートの音（どちらもPCM音源を使用）を題材としたものである。図3(a)に、スペクトログラム上のローカルピークの抽出結果を示すが、ピアノが A_3 （基本周波数440 Hz）の単音を3回弾いており、このうち2回目だけに、ピアノ音の立ち上がりとはほぼ同じ時刻から、フルートの E_4 （基本周波数660 Hz）の音が加わっている様子を表している。図3(b)は、従来方式による処理結果であるが、周期性ピッチを抽出する処理が含まれているために、基本周波数220 Hzのひとつの音として認識されている。一方、図3(c)は、本稿の方式による処理結果であるが、入力に含まれていた演奏情報が再現されていることが分かる。



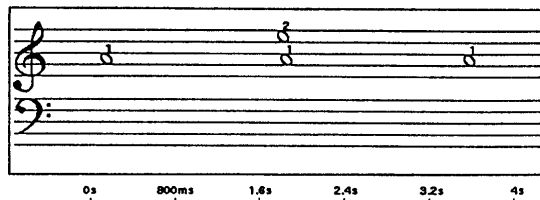
(図中の数字はローカルピークの通し番号である)

(a) ローカルピークの抽出結果



(図中の数字は認識された音色の通し番号である)

(b) 従来の方式による処理結果



(図中の数字は認識された音色の通し番号である)

(c) 本稿の方式による処理結果

図3: 簡単な入力に対する処理結果の例

4 おわりに

本稿では、対象とする音色の事前登録を行わない音源分離システムにおいて、スペクトルの継続的な情報を用いることによって処理精度を向上させる方法を提案した。現在、第2節で挙げた課題についてなお検討している。今後は、それらの検討結果を実装に反映させ、より実的な入力を題材として評価実験を行う予定である。

参考文献

- [1] 柏野邦夫:「音源分離に関する研究」, 東京大学工学部修士論文, (1992).
- [2] Bregman, A. S.: *Auditory Scene Analysis*, MIT Press, (1990).