

音楽音響信号を対象とする音源分離システム ～音モデルに基づくアプローチ～

中臺 一博 柏野 邦夫 田中 英彦
東京大学 工学部

{nakadai, kashino, tanaka}@mtl.t.u-tokyo.ac.jp

あらまし 音モデルに基づく音源分離システムについて述べる。本システムは、複数種類の楽器演奏を含むモノラルの音響信号を入力とし、これを楽器種ごとに分離して、演奏情報を MIDI データおよび画面表示として出力するものである。音源の分離同定は、システムに蓄積された音モデルを用いて混合音仮説を生成し、これと入力音響信号のスペクトログラムとを照合することによって行われる。評価実験の結果、最大同時発音数 3 音という制限の下で、平均 85 % 以上の認識率を得ることができた。本システムでは、音モデルは予めシステムに与えるものとしている。なお本稿では、スペクトログラムから周波数成分抽出を行う方法に関しても新たな提案を行っている。

A Sound Source Separation System for Polyphonic Music Based on the Tone Models

Kazuhiro Nakadai Kunio Kashino Hidehiko Tanaka

H.Tanaka Lab., Department of Electrical Engineering,
Faculty of Engineering, The University of Tokyo
7-3-1 Hongo, Bunkyo-Ku, Tokyo 113, JAPAN.

Abstract A system configuration, implementation and evaluation of a sound source separation system are described. Input of the system is assumed to be a monaural audio signal of ensemble music, and output is MIDI data which has several MIDI channels, each of which is assigned to one kind of musical instruments. The present approach is based on the matching between registered tone models and sound spectrogram derived from the input signal. Experimental results show that more than 85 % of the notes are correctly identified by the system on average, under the condition that the number of simultaneous notes in the input is three or less.

1 まえがき

音源分離とは、一般に、複数の音源からの音が混在した音響信号をもとに、音源ごとの情報を抽出することである。このうち本稿では、複数種類の楽器演奏を含むモノラルの音響信号を入力とし、これより楽器種ごとの演奏情報を分離抽出して、MIDI (Musical Instrument Digital Interface) データおよび楽譜に類似した形式の画面表示として出力するシステムを扱う。

一般的な条件の下において、音源分離問題はいわゆる不良問題であり、これを解くことは困難である。しかし、音声処理の分野では応用的に極めて重要であるため、この種の問題に対する工学的検討は早くから行われてきた。また近年では、音楽情報処理の分野において、自動採譜システムの実現や演奏された音楽の計算機への効率的な入力が望まれることから、音楽音響信号を対象とする音源分離の研究も行われるようになった。これらの研究のうち、複数のマイクロホンを用いて集音し、音源の位置の情報をを用いる方法については、比較的研究が進んでいる [1] [2] [3] [4]。しかし、これらの方法ではシステムの物理的規模が大きくなりがちであり、音源の位置に対する制約が大きといった問題点もある。この他に、くし型フィルタを用いる方法 [5] や、音色を表現する特徴パラメータを用いたパターン照合に基づく方法などのアプローチ [6] も見られるが、これらを総括すれば、問題にそれぞれ独自の制約条件を設けた上で、少しでも有効に機能する方式を模索している段階と見ることができよう。

このような状況にあって、われわれはこれまでに、聴覚実験によってスペクトル上の特徴と人間の音源分離知覚の確実性との関係を明らかにするとともに [7] [8]、これをモデル化することによって、ボトムアップ処理に基づく音源分離システムを実装した [9]。これは、スペクトログラム上の複数の特徴を音源分離に寄与する確実性という観点から独立に評価し、次いでその評価値を統合するという処理方式に基づいており、モノラルの入力に対し、音源に関する特別の (音源依存の) 先験的情報を一切利用することなく音源分離同定処理を行うことができる点が特徴であった。

しかし、評価実験の結果によれば、処理精度が必ずしも十分ではなく、特に和音など、高調波関係にある周波数成分を持つ複数の音が同時に立ち上がるような場合において、誤った出力を生じがちであるという問題点があった。本稿に述べるシステムは、このような場合に対処するための第一歩となることを目的としたものである。

2 音モデルに基づくアプローチ

では、どのような方法によれば、この問題を解決できるであろうか。周波数成分の高調波関係や立ち上がりの時間差といったスペクトログラム上の特徴に基づいたボトムアップの処理を行う限り、解決は容易ではないと考えられる。そこで人間の処理にヒントを求めれば、このような状況における人間の処理方式のモデルとして Old-Plus-New Heuristic と呼ばれるものが提案されている [10]。これは、ある時点における音が、過去に出現した音 (Old) の重ね合わせとして解釈できる場合にはその解釈を適用し、その解釈の残差が新たな音 (New) として認識されるという仮説である。このような処理方式をシステムに導入することによって、前述の問題点に対する処理精度の改善が期待できる [11]。これを、本稿では「音モデルに基づくアプローチ」と呼ぶことにする。

ここで、工学的観点から、処理に関しさらにいくつかの前提および仮定を設ける。即ち、音モデルとしては、高調波比率やエンベロープなど音の特徴を抽出してパラメータを蓄積することも考えられるが、ここでは、そのような抽象化を行わず、次のようなデータを音モデルとして蓄積する。

$$\begin{aligned} T_k &= \{\bar{a}_{ij}\}, \\ \bar{a}_{ij} &= (p_{ij}, f_{ij}). \end{aligned} \quad (1)$$

即ち、ある単音のモデル T_k は、パワー値 p 、周波数値 f を要素とする 2 次元ベクトルを要素とする行列である。行列の各行はその音に含まれる周波数成分に対応し、各列は時間のサンプル点に対応するものとする。

本稿のアプローチは、入力音響信号のある時間区間について、蓄積された音モデルを用いて混合音仮説を生成し、これと入力との照合を行い、予め定義された距離が最小のモデルをもって混合音の分離同定結果とするものである。これは、単音を「事例」とみなし、記憶に基づく推論 (MBR: Memory Based Reasoning) を適用していると見ることができる。

MBR は、ルールベースの推論とは対照的に、過去の事例と回答のペアを大規模なデータベースに蓄積しておき、入力事例に対し最も距離の近い事例を検索して、事例とペアになっている回答をそのままシステムの回答とする推論方式である。このような推論方式は、従来のルールベース型システムの持つ知識獲得のボトルネックや知識の脆弱性といった問題点を回避できることから、近年注目を集めるところとなった。蓄積事例は通常数千から数万件に達するため、逐次型計算機によって実用的なシステムを構築することは難し

いが、並列計算機によれば大幅に処理速度を高めることが可能である。これまでに、分類集計システム、自然言語処理、画像理解システム [12] など、様々な研究開発事例が報告されているが、本稿のアプローチは、これを音源分離に適用しようとするものである。

なお、本稿の範囲では、蓄積する音モデルはあらかじめシステムに与えられるものとし、その自動獲得の方法については扱わないものとする。また、システムの実装は、現状では逐次型計算機 (Sun ワークステーション) 上に行われている。

3 システムの構成と実装

3.1 システムの全体構成

具体的なシステムの全体構成を図1に示す。入力は複数種類の楽器演奏が混在したモノラルの音響信号である。この信号に対し AD 変換および周波数解析を行い、スペクトログラムを作成する。AD 変換は 48kHz、16bit で行い、周波数解析は 2 次の IIR デジタルフィルタから成るフィルタバンクによって行った。

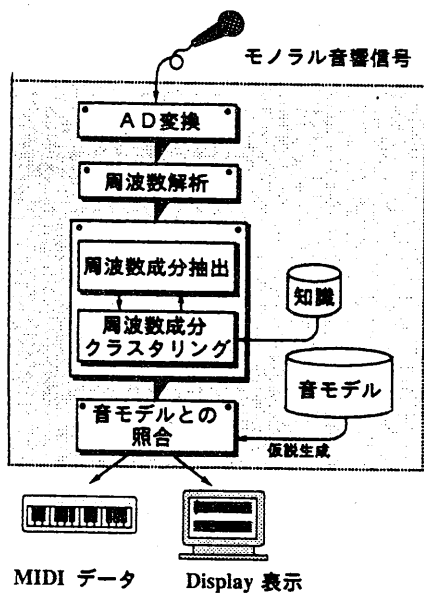


図 1: システムの全体構成

出力は、楽器種ごとにチャンネル分けされた MIDI データである。この MIDI データはインタフェースを介して電子楽器に送ることによりこれを演奏させることができ、処理結果を耳で確認することができる。また、楽譜に類似した (音符の種類や小節線の表示を省いた) 形式の画面表示によって処理結果を確認することも可能である。

以下、特に重要なモジュールについて詳細に述べる。

3.2 周波数成分抽出

このモジュールでは、スペクトログラムをもとに、これに含まれる周波数成分を抽出する。これは、閾値処理による単純な方法では、実用的な処理精度を得ることは到底不可能である。そこで本稿では、スペクトログラムにおける高精度のピークトラッキング法として、双方向挟平面回帰法を提案する。また、周波数成分抽出においては、周波数成分がどこまでひと続きであり、どこで切れるのかを正確に判定することが必要であるが、複数音が混在する場合には各周波数成分の振幅値は大きく変化するため、これをボトムアップ処理のみによって行うことは困難な場合がある。そこで、ボトムアップ処理においては周波数成分が途切れる候補の点 (切断候補点) を抽出しておき、実際に周波数成分が連続するか中断するか判断はモデルとのマッチング時に行うものとした。ここで、これら 2 つの手法について説明する。

3.2.1 双方向挟平面回帰法

周波数成分抽出は、スペクトログラム上のピークの抽出と、その接続によって行われる。最も簡単には、単にパワー値の周波数軸方向での極大点をピークとして抽出し、これを時間軸方向に接続することが考えられる。しかしこの方法では、近接した周波数成分が存在する時に接続を誤るなど、周波数成分の開始から終了までを正確に抽出することが難しい。そこで、周波数成分抽出におけるピーク接続法として挟平面回帰法を考案し、これを時間的に双方向に適用した。

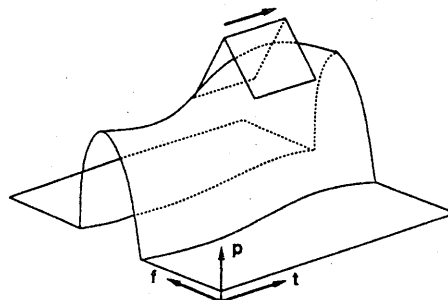


図 2: 挟平面回帰法

挟平面回帰法は、図2に示すように、スペクトログラム上のピークを 2 枚の平面で挟み込み、その 2 平面の交線の方向ベクトルを利用し、次ピークの位置をあらかじめ予想することによってピーク接続の方向を定

めるものである。この挟み込みに使用する平面(これを挟平面と呼ぶ)は、最小2乗法により定めたピーク近傍の回帰平面である。平面の方程式を

$$z = at + bf + c, \quad a, b, c : \text{const.} \quad (2)$$

とし、スペクトログラムの (t_i, f_j) でのパワー成分を z_{ij} とすると、法線ベクトル $(a, b, -1)$ は、

$$\begin{pmatrix} n \sum_{i=1}^m t_i^2 & \sum_{i=1}^m \sum_{j=1}^n t_i f_j & n \sum_{i=1}^m t_i \\ \sum_{i=1}^m \sum_{j=1}^n t_i f_j & m \sum_{j=1}^n f_j^2 & m \sum_{j=1}^n f_j \\ n \sum_{i=1}^m t_i & m \sum_{j=1}^n f_j & mn \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^m \sum_{j=1}^n t_i z_{ij} \\ \sum_{i=1}^m \sum_{j=1}^n f_j z_{ij} \\ \sum_{i=1}^m \sum_{j=1}^n z_{ij} \end{pmatrix} \quad (3)$$

を解くことによって与えられる。こうして得られた2枚の挟平面 α, β の法線ベクトルを法線ベクトルを \vec{n}_1, \vec{n}_2 とすると、方向ベクトル \vec{l} 、2枚の挟平面のなす角 φ はそれぞれ、

$$\vec{l} = \frac{\vec{n}_1 \times \vec{n}_2}{|\vec{n}_1 \times \vec{n}_2|} \quad (4)$$

$$\cos \varphi = \frac{\vec{n}_1 \cdot \vec{n}_2}{|\vec{n}_1| |\vec{n}_2|} \quad (5)$$

で与えられる。この方向ベクトルによって、次ピークの位置をあらかじめ予想する。実際の処理においては、処理速度向上のために、挟平面回帰法を時間的に双方向に行っている。即ち、パワー値について、ピーク検出のための閾値 θ_p と、有効な周波数成分であると認めるための閾値 θ_e (ただし $\theta_p \leq \theta_e$) の2つのパラメータを設け、パワー値が θ_e を上回るピークが存在した場合に、時間的に双方向に挟平面回帰処理を開始し、パワー値が θ_p となるまでピークを接続するのである。これを双方向挟平面回帰法と呼ぶ。

3.2.2 切断候補点

双方向挟平面回帰法は、単純な閾値処理に比較して高精度の周波数成分抽出が可能であるが、複数音を含むスペクトログラムは複雑な振幅変化を含んでいるために、実際の処理結果を見ると、ひと続きのものとして抽出された周波数成分が実は時間的に連続する複

数音に由来するものであったり、その逆であったりという場合が見られた。ボトムアップの処理のみによってこの問題に対処することは困難であると考えられる。このため、その周波数成分の振幅エンベロープから、ひとつの周波数成分として抽出された一連のローカルピークの中で、ふたつの周波数成分に分かれる可能性のある点をマークしておき、より上位の処理からのフィードバックによって周波数成分の時間的連続性を判断する処理方式を導入した。

具体的には、周波数成分を横軸時間、縦軸パワーとして見たとき、ある時点において

- パワー値が $m \cdot p_{max}$ 以下であること
- その前が立ち下がりであること
- その後 t_d 間に、傾き g 以上の立ち上がりがあること

という条件が満たされる場合、その時点を切断候補点としてマークする。ここで、 p_{max} は、その周波数成分におけるパワーの最大値である。切断候補点における切断の確実度 c_c は、実験的に

$$c_c = \frac{v_{max} - v_{min}}{(p_c/p_{max}) + \varepsilon} \quad (6)$$

と定めた。ここで、 p_c は切断候補点におけるパワー値、 v_{min} は切断候補点より前における立ち下がりの傾きの最大値、 v_{max} は切断候補点より後の t_d 間における立ち上がりの傾きの最大値である。また g は、切断候補点の時刻を t_c としたとき、

$$g = p_{top} / (t_{top} - t_c) \quad (7)$$

で与えられる値である。ただし p_{top}, t_{top} は、 t_c 以後の t_d 間におけるパワーの最大値とその時刻である。

3.3 音モデルとの照合

本システムでは、対象とする曲に含まれる可能性のある全ての単音について、予め周波数成分抽出を行い、その結果を式1の形で音モデルとして蓄積しておく。動作時には、音モデルをもとに混合音仮説を生成し、これと入力音響信号から得たスペクトログラムとの照合によって音源分離同定結果を得る。ここでは照合の詳細について述べる。

3.3.1 処理単位の決定

音モデルとの照合に際しては、入力音響信号から得たスペクトログラムについて、まず時間的な処理単位を決定する。これは固定長の時間窓にはならず、周

波数成分の開始時点に着目したクラスタリングによっている。クラスタリングには凝集型平均値アルゴリズムを使用し、クラスタ間距離は、クラスタに含まれる周波数成分の立ち上がり平均時刻と別のクラスタの立ち上がり平均時刻との差の絶対値と定義している。

また、このクラスタリングを利用して、前述のように周波数成分抽出モジュールへのフィードバックも行っている。即ち、以下の切断条件を満たす周波数成分が存在する場合、これを切断する。

- 切断候補点と一番近いクラスタが、距離 D 以下であること
- 切断された周波数成分をそのクラスタ内に入れた時、クラスタ幅が W_{cl} 以下になること
- クラスタ内に、その周波数成分と高調波関係にある周波数成分が存在すること

ここで、クラスタ幅とは、クラスタ内の周波数成分の一番早い開始時点と一番遅い開始時点との時間差である。

3.3.2 混合音仮説の生成

このようにして定められた処理単位ごとに、混合音仮説を生成する。これは、式 1 に示した音モデルのうち、発音していると考えられるものを混合させることによって行う。混合音仮説を生成する際、そのパラメータとしては、最大同時発音数、音の立ち上がり時刻のずれ及び位相差、及び音量などが挙げられる。本稿のシステムでは、これらを以下のように扱った。

- 最大同時発音数
本システムでは、最大同時発音数は 3 音とした。
- 立ち上がり時間のずれ
音モデル間の時間のずれを 10ms 間隔で -20ms から 20ms まで、また混合音仮説全体を処理単位に対し 10ms 間隔で最大 50ms までシフトすることにした。
- 周波数成分の干渉時の位相
同一の周波数に、複数の音源に由来する周波数成分が重なって存在する時、それらの干渉が起きるため周波数成分の振幅が大きく変化する(これを overlapping harmonics の問題という)。このため本システムでは、周波数成分の位相差を考慮して混合音仮説を生成している。即ち、パワー値が p_1, p_2, p_3 である 3 つの周波数成分 C_1, C_2, C_3 が

あって、 C_1 と C_2 の位相差が φ_{12} 、 C_2 と C_3 の位相差が φ_{23} であるとするとき、3 つの周波数成分を合成した時のパワー値 p_a を

$$p_a = \{ (p_1 + p_2 \cos \varphi_{12} + p_3 \cos \varphi_{23})^2 + (p_2 \sin \varphi_{12} + p_3 \sin \varphi_{23})^2 \}^{\frac{1}{2}} \quad (8)$$

という式によって求めている。

● 音量

本システムにおいては、音モデルの音量と入力音響信号の演奏に含まれる単音の音量とは等しいという前提がある。

3.3.3 混合音仮説の数

このように混合音仮説の生成を行うと、パラメータ数の増加にともない仮説の生成数が爆発的に増加する場合がある。そこで、混合音仮説の生成数について検討する。

表 1 は、総当たりの仮説生成を行った場合の 1 処理単位あたりの混合音仮説の生成数を表している。これに示されるように、音数、モデル数が増えれば指数関数的に仮説数が増加する。これを解決するため、仮説生成に次のような条件をつけ、無意味な仮説生成を削減した。

- 処理単位に含まれる周波数成分を持っている音モデルのみ仮説生成の対象とした。
- 処理単位に含まれる周波数成分を 2 種以上持っているモデルが必ず 1 つは含まれると仮定した。
- 2 つ以上周波数成分を持っているモデルがない場合、その処理単位は仮説生成を行わないものとした。
- 仮説と処理単位の周波数成分において、周波数の対応が良くない仮説は、パラメータのシフトをせずに、棄却することにした。

この条件によって、無意味な仮説生成はかなり抑制されている。実際の仮説生成数は処理単位内の周波数成分数や音モデルの性質に左右されるため、定量的な判断は難しいが、無条件に仮説生成する場合に比べ、平均的には十数分の一になっている。

3.3.4 混合音仮説の照合

生成した混合音仮説は、処理単位と照合してその距離を求め、距離が最小のものをもって音源分離同定結果とする。距離の定義としては、含まれる周波数成分ごとに、時間方向のパワー誤差の絶対値の和をとつ

表 1: 混合音仮説の生成数

音数	音数による仮説数	位相	時間シフト	仮説数 (10)	仮説数 (20)	仮説数 (30)
1	nC_1	1	5	10	20	30
2	約 nC_2	12	25	13,500	57,000	130,500
3	約 nC_3	144	125	2,160,000	20,520,000	73,080,000

表 2: 評価実験に用いたパラメータ値

θ_p	100
θ_e	1600
m	0.5
t_d	50ms
W_{cl}	100ms
D	0~50ms

て、それに周波数、パワー値をかけたものの和を用いた。即ち距離 D_i を、

$$D_i = \sum_{i=1}^F \sum_{j=1}^T |p'_{ij} - p_{ij}| \cdot f'_{ij} \cdot p'_{ij} \quad (9)$$

と定義する。ここで、 F は照合する周波数成分の数(仮説と処理単位において周波数が対応する周波数成分はひとつと数える)、 T は混合音仮説の時間軸方向のサンプル点の数、 p'_{ij} 、 f'_{ij} は処理単位の周波数成分のパワーおよび周波数、 p_{ij} は混合音仮説の周波数成分のパワーである。

4 評価実験

本システム対し、3つの観点から評価実験を行った。1つは音長変化に対する頑健性の測定、2つめは音モデル導入による和音認識の成果の確認、3つめは無意味な仮説生成を抑制した効果の検証である。これらの評価実験においては、RS-PCM音源(Roland U-220)の各音色を用いてサンプル曲を演奏したものを入力とした。前節に述べたパラメータの値は表2のように定めた。また認識率 R [%] は、

$$R = \frac{right - miss}{total} \cdot \frac{1}{2} + 50 \quad (10)$$

と定義した。ここで、 $right$ は正しく分離同定された音の数、 $miss$ は、出力に存在する音のうち、正しく音源分離同定されなかったり、入力に対応するものがない音(余計な所で出力された音)の数、 $total$ は入力に含まれる総音数である。式10において、 $1/2$ をかけて50を加えているのは、値を0%から100%に正規化するためである。

4.1 実験1: 音長変化に対する認識率変化

実験1は、合唱用にアレンジされた4パートから成る“Angels We Have Heard on High”という曲(Traditional French, 以下 Angels と略称する)のうち2パートを選び、1パートをフルートで、もう1パートを各音色で音の継続時間を変えながら演奏したものである。この結果を図3に示す。

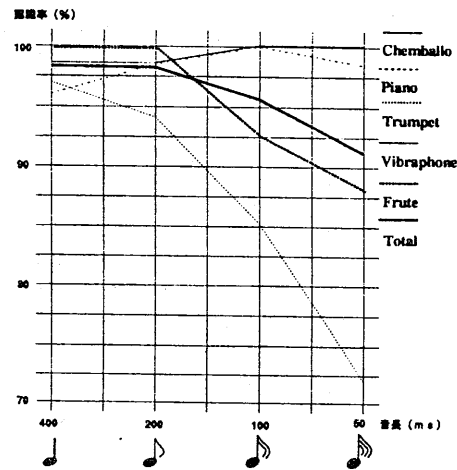


図 3: 音長と認識率

図3から、100ms以上の音長であれば、85%以上の認識率で分離同定できることが分かる。これは、4分音符=120という演奏速度では、スタッカート等を除外すればおおむね16分音符の音長に相当する。そこで、評価実験に使用する曲は16分音符より短い音がない曲を選定した。また、音色によって、音長変化に対する頑健性はかなり違ったものとなっているが、これは、音のエンベロープ特性によるものと考えられる。

4.2 実験2: 音数・音色変化に伴う認識率変化

実験2は、Angelsのうち1, 2, 3パートを選び(即ち最大同時発音数がそれぞれ1, 2, 3音)、それぞれを各音色で演奏したものである。この結果を図4から図6に示す。また、これらの結果をまとめたものが図7である。

この結果から分かるように、複数パートの演奏で

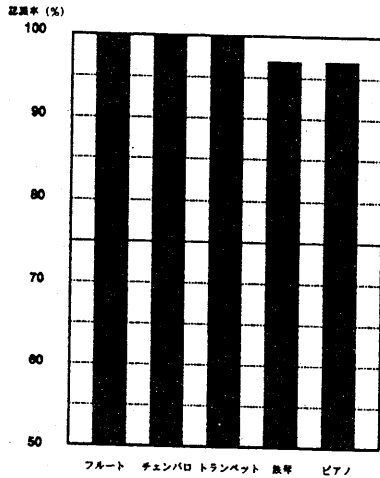


図 4: 音色と認識率 (最大同時発音数 1)

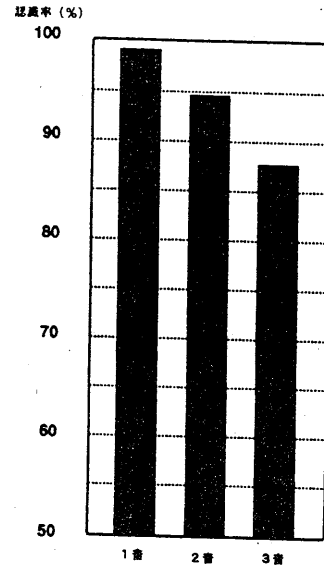


図 7: 同時発音数と認識率

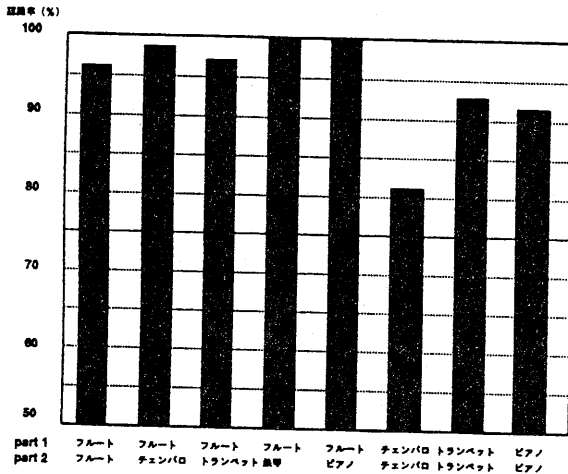


図 5: 音色と認識率 (最大同時発音数 2)

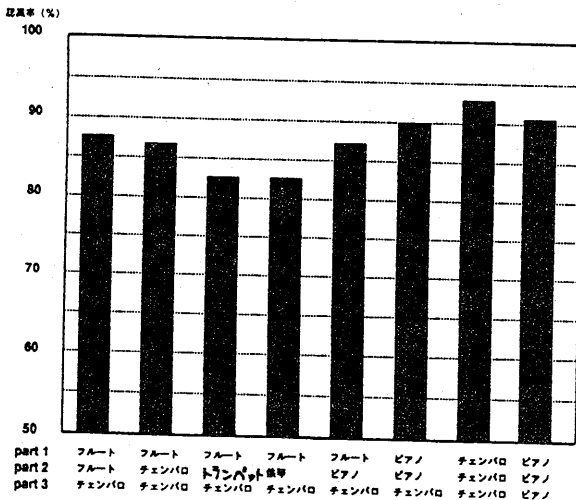


図 6: 音色と認識率 (最大同時発音数 3)

は、音色の組合せによって認識率にかなりの差が生じる。また、同時発音数が増えるにつれて認識率が低下する様子が示されている。これは、処理単位を決定するクラスタリングの精度に問題があると考えられる。

4.3 実験 3: 曲・音モデル数に対する認識率・処理時間

実験 3 は、Angels の他、“元気を出して”(竹内まりや曲、適宜アレンジ) 及び“ビバルディのバイオリンコンチェルト”(L'estro Armonico Op.3, No.6)、の一部をサンプル曲として用い、それぞれの 3 パートをフルート、ピアノ、及びチェンバロの音色で演奏させたものを入力として(最大同時発音数は 3 音)、認識率と処理時間を調べたものである。結果を表 3 に示す。

表 3: 曲・音モデル数に対する認識率・処理時間

曲名	曲長	音モデル数	処理時間	認識率
Angels	8.1s	15	1.1h	88 %
元気を	16.7s	23	2.8h	89 %
Vivaldi	33.3s	45	17.4h	90 %

実験 3 からは、混合音仮説の生成数を抑制した効果を確認することができる。3 つの曲で認識率があまり変わらないことを考えると、この抑制処理によって無意味な仮説のみを抑制できたと判断できる。また、仮説生成数抑制が処理時間短縮に及ぼす効果を 図 8 に示す。モデル数 45 では抑制を行った場合と行わない場合で 7 倍以上の時間短縮が実現されている。しかし、

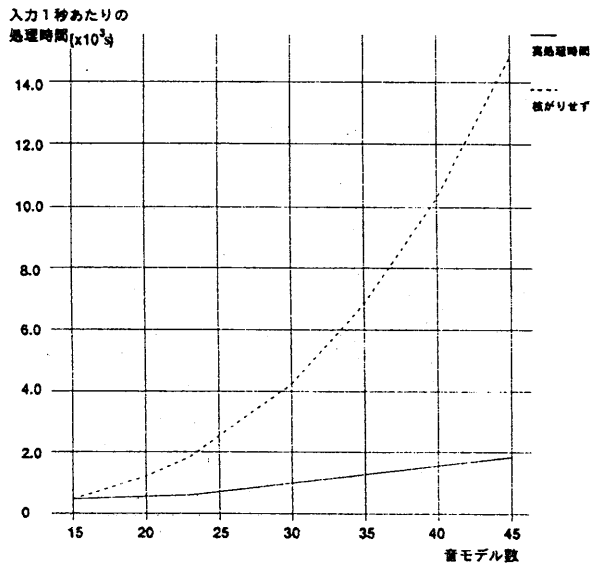


図 8: モデル数と処理時間

行った実験の範囲では、実際の処理時間がモデル数に対して線形なものかどうかまでは、導くことができなかった。今後さらに時間の短縮を行うためには、高次の音楽的知識の利用などによる探索アルゴリズムの改善に加え、仮説生成および照合を並列計算機によって行うことが効果的であると考えられる。これは、音モデル数、最大同時発音数が多い場合、混合音仮説の生成数も大きくなり、高いデータ並列性が得られるためである。

5 むすび

本稿では、複数種類の楽器演奏を含むモノラルの音響信号を対象とする音源分離システムを扱った。本稿では特に、音モデルを用いるアプローチについて、システムの構成と実装、および評価実験の結果を述べた。目的としていた和音への対処については、その可能性が十分に示されたものと考えている。しかし、処理精度や処理時間の点ではまだ実用的と言えるものではなく、また、同時発音数や音量など制限事項が存在している。今後は、これらの制限事項を排除すること、処理精度と処理時間を改善することを課題として検討を進める。また、ボトムアップ処理との協調によって、音モデルを自動獲得する機構についても考察する予定である。

参考文献

- [1] Mitchell O. M. E., Ross C. A. and Yates G. H.: "Signal Processing for a Cocktail Party Effect", *J. Acoust. Soc. Am.*, 50(2), (1971).
- [2] 柳田, 角所, 植田, 野村: "一般逆行列を用いたカクテルパーティ効果の知覚的検討", 信学技報, EA80-69, (1981).
- [3] Flanagan, J. L., Johnston, J.D., Zahn, R. and Elko, G. W.: "Computer-steered microphone arrays for sound transduction in large room", *J. Acoust. Soc. Am.*, 78(5), (1985).
- [4] 永田, 安倍, 城戸: "多数センサによる音源波形の推定", 音響誌, 47(4), (1991).
- [5] Nehorai, A. and Porat, B.: "Adaptive Comb Filtering for Harmonic Signal Enhancement", *IEEE Trans. on ASSP*, 34(5), (1986).
- [6] 長東, 才脇, 井口: "異種楽器を対象とした採譜システム", 信学 '92 春大会, D-499, (1992).
- [7] 柏野, 田中: "音源分離要因に関する一検討 — 共通FMと高調波関係 —", 信学 '91 秋大会, A-107, (1991).
- [8] 柏野, 田中: "音源分離要因に関する一検討: II — 周波数成分の立ち上がりの時間差および傾きの効果 —", 信学 '92 秋大会, A-138, (1992).
- [9] Kashino, K. and Tanaka, H.: "A Sound Source Separation System using Spectral Features Integrated by the Dempster's Law of Combination", *Annual Report of the Engineering Research Institute, Faculty of Engineering, University of Tokyo*, Vol.51 (1992).
- [10] Bregman, A. S.: *Auditory Scene Analysis*, MIT Press, (1990).
- [11] 柏野, 田中: "音源分離システムにおける時間的統合 — Old-Plus-New Heuristic の導入 —", 情処 '92 秋全大, 7B-3, (1992).
- [12] Tucker, L., Feynman, C., and Fritzsche, D.: "Object Recognition using the Connection Machine", *Proc. of IEEE Conf. Computer Vision and Pattern Recognition*, (1988).