

4T-6

音源分離システムにおける
パターン照合モジュールの並列実装と評価

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、モノラルの音楽演奏を対象とする音源分離システムについて研究を進めている。これまでに実装および評価を行ってきた音源分離システムには、音記憶を用いたパターン照合処理が含まれている[1][2]。この照合処理は、システム内に蓄積されている音記憶（音名と楽器音、及び、それに対応するスペクトログラムパターンをもった具体的な音のモデル）から混合音仮説を生成し、入力音楽音響信号との照合を行うことにより、音源の分離同定を行うものである。このような処理では、音記憶数、混合音数が増大すると、大量の混合音仮説が生成され、処理時間がかかるという問題があった。しかし、大量の混合音仮説の生成、照合は、データ並列性が高い。そこで今回、音源分離システムにおけるパターン照合モジュールを並列計算機（AP1000）に実装し、処理時間短縮の効果に関する評価実験を行った。AP1000は、富士通研で開発された並列計算機であり、分散メモリ型のメッセージ通信を基本としている。プロセッシングエレメント（CELL）数は64台のものを使用した。

2 パターン照合モジュールの並列実装

パターン照合モジュールの並列実装の際、次の2点を考慮する必要がある。第1点は、音記憶数、混合音数が増大して、仮説生成数が爆発してしまうと、並列実装を行うことによる処理速度向上の効果が薄れてしまうということである。これについては、3節で述べるものとする。第2点は、現状の並列計算機は、各セルに依頼するタスクの粒度がある程度揃っていて、かつ、ある程度の大きさでないと大きな台数効果が期待できないということである。これについては、4節で述べるものとする。

3 混合音仮説生成数の爆発の緩和

3.1 パターン照合モジュールで扱う

仮説の種類

まず、パターン照合モジュールで扱われる仮説の種類を次のように区別する。

音組合せ仮説 事前登録された音記憶全体の集合を音記憶集合とした時の部分音記憶集合のこと（ただし、要素数は最大混合音数以下となる）。

混合音仮説 音組合せ仮説に含まれる音記憶の組を、時間シフト、位相パラメータにより合成した仮説のこ

と。合成時に使用するパラメータにより、2つのレベルに細分することができる。

レベル1 混合音仮説のうち、時間シフトパラメータのみで合成されるもの

レベル2 混合音仮説のうち、時間シフトと位相パラメータにより合成されるもの

3.2 混合音仮説生成数

次に、枝刈りをしない場合に、最大混合音数を1音～3音、総音記憶数を10～30個に、それぞれ変化させた時の混合音仮説生成数を表1に示す。

表1: 音記憶数と最大混合音数
に対する混合音仮説生成数

		最大混合音数		
		1音	2音	3音
音記憶数	10	50	67,500	10,800,000
	20	100	285,000	102,600,000
	30	150	652,500	3,675,400,000

この表より、混合音仮説数が爆発していることがわかる。つまり、このままでは、並列実装によって、実時間処理が期待できない。そこで、次のような仮説生成を行うことにより、枝刈りを行った。

3.3 混合音仮説生成法

1. 入力処理単位¹から候補となる音記憶を優先度つきで選出する。
2. 選出された音記憶から音組合せ仮説を生成する。
3. 音組合せ仮説に含まれる音記憶同士に互いに重なり合う周波数成分が存在しない場合は、レベル1の混合音仮説、存在する場合は、レベル2の混合音仮説を生成する。
4. 3で生成された、レベル1、またはレベル2の混合音仮説と入力処理単位とを照合してベストマッチした混合音仮説を回答とする。

1において、事前登録した音記憶をすべて使用せず、候補となる可能性のある音記憶のみを仮説生成対象すること、および、2において、優先度の低いもの同士の音組合せ仮説は、入力処理単位に対する回答となる可能性が低いと考え、生成しないものとするにより、枝刈りを行っている。

3.4 枝刈りの効果

上記のような混合音仮説生成法による枝刈りの効果を確認するため、SPARCstation10(1CPU)を用いて処理

¹入力処理単位は、スペクトログラムから抽出した周波数成分をクラスタリングしたもので、入力混合音に対応している。

時間の測定を行った。この結果を図1に示す。この実験で逐次計算機を用いたのは、ホストとCELL間の通信のオーバーヘッドなどのコストを考慮する必要がないからである。図1を見る限り、全解探索に比べ、探索空間は数分の1になり、混合音仮説生成数の爆発は、緩和されていることがわかる。

しかし、依然、音記憶数の増加に伴い、混合音仮説数が増加する傾向がみられる。これと同様の傾向が、最大混合音数を増加させた場合にも見られている。つまり、さらに音記憶数、混合音数が増加した場合、混合音仮説数の爆発を引き起こし、実時間処理が期待できなくなる。

これに対処するためには、さらに有効な枝刈りを行う必要があるが、これは今後の課題とする。

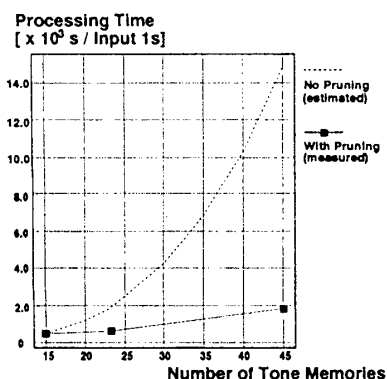


図1: 音記憶数と処理時間 (SPARC station 10 使用)

4 並列実装法

以下のような3種類の並列実装を行った。

- 1つの音組合せ仮説を、1タスクとしてCELLに分配する。
- 粒度が均一化されるように、レベル2の混合音仮説が生成される音組合せ仮説では、混合音仮説数が最大でも144個となるようにタスクを分割し、CELLに分配する。
- すべての処理を1タスクとしてCELLに分配するが、それぞれのCELL内では、混合音仮説に番号を付加し、自分の担当番号の混合音仮説のみ生成、照合を行う。

1の方法は、1タスクあたりの混合音仮説数が5~18,000となり、タスクの大きさにばらつきがある実装法である。また、2の方法は、1タスクあたりの混合音仮説数が5~144となり、それぞれのタスクの大きさは1に比べ、均一化される。また、タスクが分割されるため、タスク数は増加し、1に比べ、粒度が細くなる。3の方法は、1,2の方法とは異なり、1つのCELLには1つのタスクしか渡されない。このため、通信のオーバーヘッドが低く抑えられる。また、混合音仮説に番号を付加することにより、それぞれのタスクの粒度は同じになるという特徴がある。

5 評価実験

評価実験では、サンプルデータ²に対し、パターン照合モジュールにおける処理時間を、CELL台数を1~64

に変化させて測定し、台数効果を求めた。いずれの場合も、分離同定結果に関しては、全く同じ出力を示している。実験結果を図2に示す。

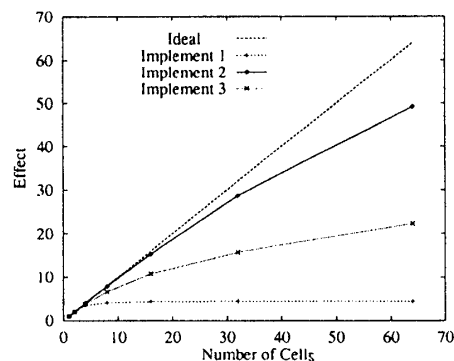


図2: 台数効果

6 考察

1の実装法では、タスクの大きさに3600倍の開きがあり、これが台数効果を低くしている原因となっている。これに対し2の実装法では、かなり台数効果が出ているといえる。これは、タスクの大きさに最大30倍程度の開きしかないためと考えられる。CELL台数が増えると台数効果が落ちてくるのは、データのブロードキャストなどCELLの初期化にかかる時間が一定であるためである。つまり、CELL台数が増え、処理時間が短くなると、全処理時間に対する初期化の時間の割合が高くなり、台数効果が落ちるのである。

3の実装法は、通信のオーバーヘッドが低く抑えられ、粒度が均一であるため、最も良い結果が期待されていた。CELLの稼働状況を示す稼働率は、100%近い値が出ており、1,2の場合よりも高い値となっている。しかし、台数効果においては、2の場合よりも劣っている。これは、混合音仮説に番号をつけるために、無駄なループを通るためであり、番号の付け方を工夫すれば、2と同等、または、それ以上の台数効果をあげることが期待できる。

7 おわりに

本稿では、音記憶ベースの音源分離システムにおいて、パターン照合モジュールの並列実装を行った。評価実験では、64台のCELLを使用した場合、50台近い台数効果をあげ、処理時間短縮を実現することができた。

将来的に、実時間処理を行う場合、さらにCELL台数の多い並列計算機を使用することが考えられる。しかし、このためには、CELL台数が多くなることによるホストマシンの負荷の増大、通信のオーバーヘッドの増大といった問題点を克服することが必要である。また、混合音数、音記憶数の増大による混合音仮説生成数の爆発を防ぐため、さらに効果的な枝刈りを行うことも今後の課題である。

参考文献

- [1] 中憲一博, 柏野邦夫, 田中英彦: "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~", 情報研報 SIGMUS 1-1 (1993).
- [2] 柏野邦夫, 田中英彦: "モノラル楽器音の音源分離のための知覚の手がかりの検討と処理モデルの実装", 日本音響学会 聴覚研資 H-93-84 (1993).

²フルート、チェンバロの2パートからなるAngelという曲