

音源分離システムにおけるパターン照合モジュールの 動的負荷分散を用いた並列実装

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、モノラルの音楽音響信号を対象とした音源分離システムについて研究を進めている。音源分離とは、複数の音源からの音が混在した音響信号から、音源ごとの情報を抽出することであるが、一般状況下においては、音源分離問題は不良問題であり、解析的に解くことは困難である。そこで、この問題に対する一つのアプローチとして、音記憶を用いたパターンマッチ処理による音源分離システムの評価および実装を行ってきた [1]。このシステムは、われわれがこれまでに実装および評価を行ってきたボトムアップベースの音源分離システムと比べ、和音などの混合音認識に効果的であった。しかし、パターンマッチの際に大量の混合音仮説を生成し、処理時間がかかってしまうため、AP1000¹ に並列実装し処理速度の向上を目指してきた [2]。並列計算機への実装において、従来の仮説生成法では、木構造の解析を比較的容易に行うことができたため、静的な木の分割が可能であった。しかし、今回、さらに速度向上を実現するため、仮説生成パラメータの刻値を動的に変化させることにより、枝刈りを行った。これにより、仮説生成数を抑制することができたが、静的に仮説生成木の構造を解析することが困難となった。そこで、動的に木を分割し、負荷を分散することにより並列実装を行った。

2 混合音仮説生成とマッチング

混合音仮説の生成とマッチングの過程は、次の3段階に大きく分けることができる。

1. 入力処理単位に含まれる可能性がある音記憶を選び、それを組にした組合せ音仮説を生成する。現状では、組合せ音仮説に含まれる最大音記憶数(最大混合音数)は3音としている。
2. 組合せ音仮説に含まれる音記憶に対し時間シフト、位相パラメータを用いて混合音仮説を生成する。
3. 2によって生成された混合音仮説と入力処理単位の距離を計算し、最も距離の混合音仮説を回答として出力する。

ここで、入力処理単位は、周波数成分を発音時刻、倍音関係に注目してクラスタリングを行ったもので混合音に相当する。また、周波数成分はスペクトログラム²における周波数方向のローカルピークを時間方向に接続したものである。

¹富士通研で開発された分散メモリ型のメッセージ通信を基本とした並列計算機、セル数は64台のものを使用。

²時間、周波数、パワーの3軸からなるグラフ。入力音響信号を周波数解析することによって得られる

次にそれぞれの過程について、従来の方法および枝刈りのために行った変更について述べる。

2.1 組合せ音仮説生成

組合せ音仮説は、組合せ音仮説に含まれる音記憶と入力処理単位の周波数成分の周波数値の比較を行い、周波数成分の対応がよいもののみを生成する。具体的には、入力処理単位の周波数成分の個数を N_{in} 個、組合せ音仮説から生成される混合音仮説の周波数成分の個数を N_{hypo} 個、両者の周波数値が対応している周波数成分の個数を N_{match} 個とした場合、

$$N_{hypo} - N_{match} > Nd \quad (1)$$

$$N_{in} - N_{match} > Nd \quad (2)$$

の2式を満たす組合せ音仮説のみを生成する。

Nd に関しては、従来は3と固定していたが、

$$Nd = Const \cdot (N_{hypo} + N_{in} - N_{match}) \quad (3)$$

と周波数成分数の多い組合せ音仮説に対しては、 Nd を大きく、周波数成分数の少ない組合せ音仮説に対しては、 Nd を小さくするよう変更を行った(ただし $Const = \frac{1}{3}$ としている)。

2.2 混合音仮説生成

組合せ音仮説から混合音仮説を生成する際、時間シフトおよび位相パラメータを使用する。これらのパラメータの個数は、仮説生成木の深さを決定するが、組合せ音仮説ごとにパラメータの個数が異なる。具体的には、時間シフトパラメータの個数は組合せ音仮説に含まれる音記憶の数(1~3個)である。位相パラメータは音記憶同士が互いに重なり合う周波数成分を持っている場合に使用する。このため、位相パラメータ数は最大同時混合音数が3音のもとでは、音記憶の周波数成分の対応具合により0~2個に変化する。また、個々のパラメータの変域は、時間シフトパラメータの場合、入力処理単位に対して $-20ms \sim 20ms$ 、位相パラメータの場合、他の重なり合う周波数成分に対し $0 \sim 360$ 度である。パラメータの刻値は、従来は、時間シフトパラメータが $10ms$ 、位相パラメータは 30 度であり、この場合の組合せ音仮説あたりの混合音仮説数は最大で $18,000$ 個となる。

今回、このうちの位相パラメータの刻値 ($Pshift$) を以下のように変更した。この変更は、重なり合う周波数成分数の少ないものはマッチングの距離に与える影響が少ないものとして位相パラメータの刻を粗くとするようにしたものである。

$$Pshift = MaxDeg \cdot \frac{N_{match}}{N_{mix}} \quad (4)$$

ただし、 $MaxDeg$ は30度、 $Nmatch$ は、生成される混合音仮説と入力処理単位で周波数値が対応する周波数成分数、 $Nmix$ は、混合音仮説生成の際、重なり合う周波数成分のうち $Nmatch$ に含まれる周波数成分の数を示している。

2.3 混合音仮説のマッチング

従来は、すべての混合音仮説において完全にマッチング計算を行っていたが、距離計算の途中で距離が大きくなり過ぎた場合は、計算を中止し無駄な計算を行わないように変更を行った。

3 動的負荷分散による並列実装

前述したような仮説生成とマッチング方法の変更の結果、入力処理単位に対する混合音仮説生成本は、深さに関しては、従来と同様に組合せ音仮説ごとに特定することができるが、それぞれのノードから派生する枝の数については、動的に変化するため、静的に木を解析することが困難となった。そこで、動的に木を分割することにより負荷分散を行った。

動的に木を分割し粒度を調節するために、LTC(Lazy Task Creation)[3]を使用した。LTCを実現するために、LTQ(Lazy task queue)を用いている。LTQに格納されるノード情報は、組合せ音仮説、時間シフト、位相パラメータ、および入力処理単位情報である。また、実装に使用した並列計算機 AP1000 はメッセージパッシング型の並列計算機であるため、以下のメッセージを用意し LTC を実現した。

- タスク要求メッセージ
ロードの低いセルが、タスク要求のために送信するメッセージ。
- タスク分配メッセージ
タスク要求メッセージを受信したセルのロードが高い場合、LTQ から dequeue されたノード情報とともにタスク要求メッセージ送信元セルに対し送るメッセージ。
- タスク要求拒否メッセージ
タスク要求メッセージを受信したセルのロードが低い場合、タスク要求メッセージ送信元セルに送るメッセージ。

タスク要求メッセージの受信は、LTQ に enqueue するたびにポーリングをすることによって行っている。

4 評価実験

評価実験用データとしては、音源分離用ベンチマークを使用した。このベンチマークは、複数種の楽器音によるランダムノートパターンであり、PCM 音源によってモノラル録音されている。混合音数は2音と3音のものがああり、楽器はフルートとピアノのものをを使用した。評価実験として次の3種類を行った。

- 実験1 有意な枝を刈っていないかを調べるために、従来方法との認識率の比較を行った。結果を表1に示す。
- 実験2 仮説生成数をカウントすることにより、枝刈りの効果を測定した。結果を表2に示す。
- 実験3 台数効果の測定を行った。1プロセッサで実行した場合の処理時間を1としている。結果を図1に示す。

実験結果より、認識率はほとんど変わらず、仮説生成数が平均で、30%程度に抑えられていることがわかる。台数効果も線形性を示しており、効率的な実装ができていたといえる。セル台数が増えると、台数効果が落ちるのはセル台数が増えることにより、無駄なメッセージの送信が増え、パフォーマンスに影響を与えているためであると考えられる。これを解決するためには、タスクの要求が拒否されるようなタスク要求メッセージの送信を少なく抑える工夫をする必要がある。

表1: 認識率の比較

ベンチマーク	枝刈りなし(従来版)	枝刈りあり
2音	89.4%	88.8%
3音	84.5%	82.6%

表2: 枝刈りの効果(仮説数は入力処理単位あたりの平均)

ベンチマーク	仮説数(従来版)	仮説数(枝刈り)
2音	130,631	44,497(34%)
3音	123,257	32,458(26%)

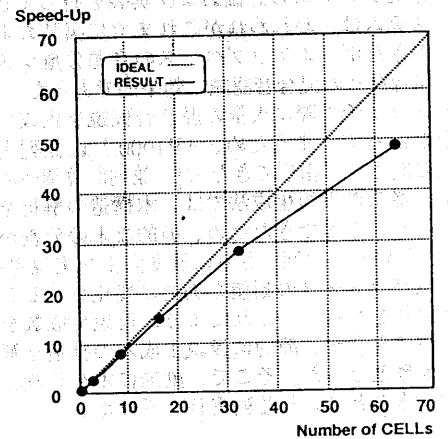


図1: 台数効果

5 おわりに

音記憶ベースの音源分離システムの仮説生成において、速度を向上するために、枝刈りを行った。これにより、仮説生成木の静的な解析が困難となったため、動的に木を分割することにより負荷分散を行った。台数効果では、線形性が得られており、将来的にプロセッサ数を増やすことによりリアルタイム処理を期待することができる。

参考文献

- [1] 中臺一博, 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] 中臺一博, 柏野邦夫, 田中英彦: “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”, 第48回情報大, AT-6 (1993).
- [3] Feeley, M.: A Message Passing Implementation of Lazy Task Creation, in *Parallel Symbolic Computing: Languages, Systems, and Applications, US/Japan Workshop Proceedings*, LNCS 748, pp. 94-107, Springer-Verlag (1992).