

カクテルパーティ効果実現のための音響ストリーム分離の検討

2R-6 II. 残差駆動型アーキテクチャの提案とモノラル音への適用

中谷 智広 川端 豪 奥乃 博
日本電信電話(株)・NTT 基礎研究所

1 はじめに

カクテルパーティ効果のように、実環境での柔軟な音の理解(音環境理解)機構を計算機上で実現するには、混合音から何らかの特徴を一貫して持つ連続した音である音響ストリームの分離が必要である [3]。我々は、音源数未知の入力音に対するマルチエージェントによる音響ストリーム分離システム(以下、『従来システム』と呼ぶ)を構築し、その有効性を示した [1]。しかし、従来システムには、多様な音を含む実環境入力処理のための課題が残っていた。本稿では、これらを解決するために、より一般化した計算モデルとして、残差駆動型アーキテクチャ(Residue-Driven Architecture, 以下、RDAと略す)を提案し、モノラル入力の音響ストリーム分離システム設計へのRDAの適用についても報告する。

2 従来システムおよびその残された課題

音響ストリームの分離は、局所的な連続音である部分ストリームの分離と、部分ストリームを時間的につなぐグループ化の二つの処理からなる。入力音はそこに含まれている音響ストリームが重なってできたものである。音響ストリーム分離では、逆に、音の重なりを取り除いて、重複しないようにストリームに分配しなければいけない。これを分離の排他性と呼ぶ。

従来システムでは、調波構造を一貫性要因にして各追跡エージェントが一つの音響ストリームを専属的に追跡することによって、音響ストリームの分離を行なっている。従来システムは、モノラル入力の男女の混合音声を分離することができた [1]。また、グループ化が部分ストリーム分離と全く同じ手法でできることも分かった。一方、従来システムには2つの主な問題点があった。(1) 分離の排他性が充分ではない。この結果、雑音(調波構造のない音)が入ってきたり、音源の数が増えると、同じ音響ストリームあるいは存在しない音響ストリームを追跡する不要なエージェントが発生して、追跡精度が低下することがあった。また、(2) 時間情報の処理が不十分である。その結果、断続音を一つのストリームにグループ化することができなかった。

Sound Stream Segregation for Cocktail Party Effect
II. Residue-Driven Architecture and its application to sound stream segregation for monaural input
Tomohiro Nakatani, Hiroshi G. Okuno and Takeshi Kawabata
NTT Basic Research Laboratories

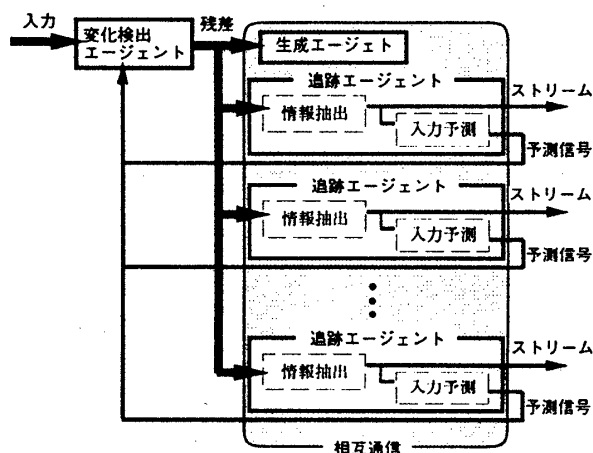


図 1: 残差駆動型アーキテクチャの構成

3 残差駆動型アーキテクチャ (RDA)

前節の問題点を解決するだけでなく、従来システムの基本的枠組を様々な分離システムの設計に利用できるように抽象化した計算モデルとして、残差駆動型アーキテクチャ(RDA)を提案する。RDAは3種類のエージェント: 変化検出エージェント, 生成エージェント, 追跡エージェントからなり(図1), それぞれ、何らかの一貫性要因にもとづき、各時刻ごとに次の動作を繰り返す。

- 各追跡エージェントは、一つのストリームを追跡・分離し、ストリームの次の状態を予測する。ストリームの終了を検知すると、自動的に消滅する。
- 変化検出エージェントは、各時刻で、追跡エージェントが予測した信号を実入力から除去し、その残差を生成エージェントと各追跡エージェントに出力する。
- 生成エージェントは、残差が大きい時、一貫性要因の属性値を計算し、もし、適切な属性値が得られれば追跡エージェントを生成する。そうでなければ、雑音とみなし雑音追跡エージェントを高々1個生成する。

生成エージェントが新しい音を検知して追跡エージェントを生成すると、以後、残差から追跡エージェントが分離する音は取り除かれるので、次に別の音が発生するまで、システムは安定状態になる。別の音が入力されれば、生成エージェントが、残差中にその音を検出して別の追

跡エージェントを生成するので、システムは再び安定化する。一方、二つ以上の追跡エージェントが生成されている時には、残差にはそれぞれの予測信号が取り除かれているので、各音響成分は排他的に分配される。また、各追跡エージェントは、ストリームの終了時に自動的に消滅するので、その数は入力音の数に応じて増減する。

追跡エージェントの信号予測には、通常、誤差があるので、残差からすでに追跡しているストリームに関連する音を完全には取り除けない。この結果、残差誤差は分離性能を低下させかねない。RDAでは、残差誤差の影響を排除するために、追跡エージェントどうしが、各ストリームの状態にもとづき、相互通信する。そのような相互通信の一例は、追跡エージェントどうしが共有する音の感度に関する閾値を動的に変更し、残差誤差に対する感度を低くすることである [2]。

RDAによる音響ストリーム分離の特徴は、1) 入力音の数に応じた適応的なシステム構造、2) 追跡エージェントの高度化による分離の排他性の向上、特に、雑音追跡エージェントの追加による耐雑音性向上、3) 入力と与えられる順に分離を行なう漸次的分離、4) 既存システムの部品化による高い拡張性、などである。

4 背景雑音下でのモノラル音の分離

背景雑音の下で、モノラル音の入力に対して、一貫性要因として調波構造を用いて音響ストリーム分離を行なうという従来システムの拡張をRDAに基づいて設計しよう。追跡エージェントの機能の内、調波構造を追跡する機能は従来システムとほぼ同じである [1]。生成エージェントは、一定レベル以上の背景雑音を検出すると、雑音追跡エージェントを生成する。以後、雑音追跡エージェントは、スペクトル強度の時間平均を計算することによって、背景雑音の強度を推定・予測する。ただし、残差中の誤差を除外するために、雑音の推定値の更新は、他の追跡エージェントが生成されていない時のみ行なう。雑音追跡エージェントは、予測信号をスペクトル減算を用いて入力から除去し、各調波構造への背景雑音の干渉を低減させる。また、雑音追跡エージェントは、他の調波構造追跡エージェントの雑音への感度を低減させるために、各エージェントが感度調整のために利用している閾値を、雑音推定値に比例して上昇させる。

5 実験結果と考察

白色雑音下で男女の声(共に「あいうえお」と発声)の混合音を分離する実験を行なった。男/女の声のSN比は、-1.6 dBであり、男の声/雑音のSN比は、0 dBである。雑音追跡エージェントなし/ありの場合の分離結果を図2に示す。雑音追跡エージェントありの方が、図3に示されているように不要な調波構造追跡エージェントの生成が抑制され、2人の声の基本周波数も安

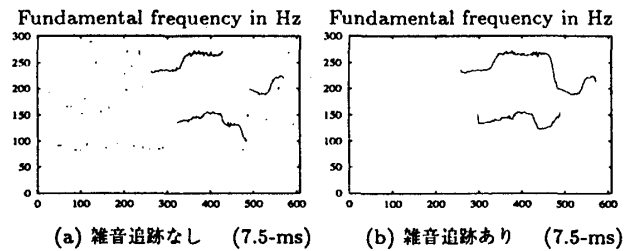


図2: 0 dB 白色雑音下の分離音声(男女)の基本周波数

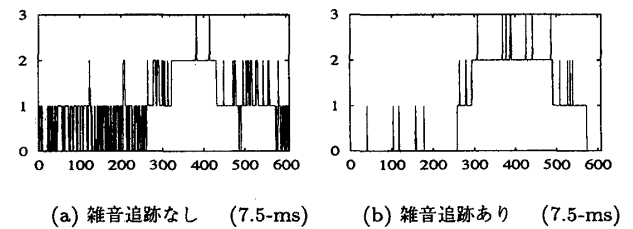


図3: 図2で生成された調波構造追跡エージェントの数

定している。生成された調波構造追跡エージェントの延べ数は105と39である。図2(b)、3(b)より、部分ストリーム自身が最終的な音響ストリームであるので、本実験ではグループ化は行なわなかった。

調波構造以外の一貫性要因を利用する追跡エージェントを設計して、分離の排他性を向上させることが考えられる。実際、音源の方向情報も同時に追跡することにより、調波構造だけでは区別できないストリームを区別できるようになることが示されている [4]。また、そのグループ化は、RDAによってモデル化されている [4]。

6 まとめ

本稿では、音響ストリーム分離のための計算モデルとして、残差駆動型アーキテクチャ(RDA)を提案した。RDAを用いることにより、従来システムの問題点であった、分離の排他性の向上、時間情報の扱いが可能になる。実際に、RDAを適用して背景雑音下でのモノラル音分離システムを実装し、定常雑音下での男女の混合音声の分離性能が向上することを通じ、RDAの有効性を示した。今後の予定としては、音声分離システムを構築し、音声認識システムと統合して、複数の音声を同時に認識するシステムを構築していきたい。

参考文献

- [1] 中谷 他: 音環境理解のためのマルチエージェントによる調波構造ストリームの分離, 人工知能学会誌, Vol.10, No.2, 1995.
- [2] Nakatani, et al: A Computational Model of Sound Stream Segregation with Multi-Agent Paradigm, ICASSP-95.
- [3] 奥乃 他: I. 音環境理解からのモデル化, 第51回情処全大, 2R-5.
- [4] 後藤 他: III. 両耳聴による音響ストリーム分離, 同上, 2R-7, 1995.