

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-7

III. 両耳聴による音響ストリーム分離

後藤 真孝[†] 中谷 智広[‡] 奥乃 博[‡]

[†]早稲田大学 理工学部

[‡]NTT 基礎研究所

1. はじめに

本稿では、残差駆動型アーキテクチャ[3] に基づいた音響ストリーム分離システムを、音源方向を扱えるように拡張した結果について報告する。音響ストリームとは、一貫した特徴を持った意味のある音の連続である。複数の音源による混合音をこの音響ストリームへと分離する処理は、音環境理解の課題の一つであるカクテルパーティ効果を、計算機上で実現する上で重要である[2]。我々は従来のシステム[1] に対し、1) ステレオ入力による音響ストリームの音源方向の同定、2) 部分ストリーム (分断された音響ストリーム) の継時的グルーピング、の二点で拡張をおこなった。その結果、音響ストリーム間の排他性が向上し、方向属性がグルーピングに有効であることを実験により確認した。

2. 従来の問題点と解決法

従来の音響ストリーム分離システム[1] には、以下のような問題点があった。

- モノラル入力であったため、音源方向が扱えなかった。
- 調波構造に基づく分離だけでは、周波数成分が近接した場合や基本周波数が交差した場合に、適切に分離できないことがあった。
- 音響ストリーム中で調波構造が一時的に壊れている場合にそこで分断され、複数の部分ストリームとして抽出されていた。また、断続音を一つの音響ストリームとして抽出できなかった。

そこで、我々は以下のようにシステムを拡張することで、上記の各問題点を解決する。

- ステレオ (バイノーラル) 入力に対応し、音源方向を同定する。調波構造中の各倍音に対し、位相差と強度差を用いて方向を同定する。
- 追跡エージェント [3] が追跡する一貫したストリームの特徴 (一貫性要因) に方向を加える。これにより、従来分離が困難だった状況でも音源方向が異なれば適切に分離できる。
- 音響ストリームが分断された部分ストリームを、音響ストリームへと継時的にグルーピングする。グルーピングは、部分ストリームの音源方向と音程に基づいておこなう。

3. 部分ストリームに対する方向同定

音源方向の同定は、単一音源であれば左右の波形を比較する方法などがあるが、混合音には適用できない。そこで、一貫性要因を基に入力を各ストリームへ分離しながら、同時に調波構造中の各倍音に対して方向同定をおこなう。以上の処理は、残差駆動型アーキテクチャ[3] では追跡エージェントが担当する。つまり追跡エージェントは、音の一貫性要因として調波構造に加え音源方向の連続性に基づいて分離する。

方向同定は以下の手順でおこなう。方向同定の手がかりとして、左右の入力音の位相差、強度差、発音時刻の差などが考えられるが、現在の実装では位相差と強度差だけを扱う。

- 各倍音の左右の波形の位相差 $\Delta\phi$ から、その倍音の方向を求める。ここで、方向同定の精度を整数値 $-10 \sim 10$ とし (-9 が真左, 0 が中央, 9 が真右に相当), 各方向 k における左右の耳に到達する時間差を τ_k ($k = -10 \sim 10$) とすると、

$$\tau_k - \lambda \leq \frac{\Delta\phi + 2n\pi}{2\pi f} \leq \tau_k + \lambda \quad (1)$$

を満たす整数 n が存在するとき、その倍音の方向が k であると判断する。ただし、 λ は τ_k の分解能であり ($\tau_k = k\lambda$)、現在の実装では $0.0833[\text{msec}] (= 1 / 12[\text{kHz}])$ である¹。位相差に $2n\pi$ の自由度があるため、 $0.6[\text{kHz}] (= 1 / (20\lambda[\text{msec}]))$ 以上では、複数の方向 k が候補にあがる。

1. で求めた各倍音の方向のヒストグラムを作る。ヒストグラムには各倍音の強度を加える。その際、右側の方向なのに左の強度が大きい場合のような、左右の強度比が反転しているものは除外する。
- ヒストグラム中の最大ピークを、追跡している音源の方向とする。たとえ各倍音において 1. で述べたような複数の方向の候補が得られても、正しい方向は多くの倍音に支持されるため適切に求まる。

こうして求めた方向は、倍音構造が重なったときなどに瞬間的に不安定になることがあるため、安定時の方向を用いて音源方向をロックする。ただしこれは完全な固定式ではなく、音源の移動にも追従できる。音源方向がロックされると、追跡エージェントはその方向の倍音成分だけを用いて基本周波数を決定する。

4. 部分ストリームのグルーピング

追跡エージェントによって得られた部分ストリームを、継時的にグルーピングして音響ストリームとする。

¹ 現在の実装では、標準化周波数は $12[\text{kHz}]$ である。

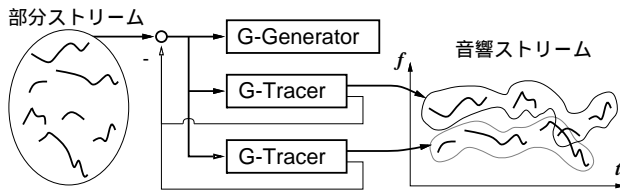


図 1: 部分ストリームのグルーピング

グルーピングの手がかりとして、部分ストリームの音源方向、音程（基本周波数）、音色（調波構造、倍音比）などの属性が考えられるが、現在の実装では方向と音程だけを扱う。

グルーピングには、1) 新しいグループの生成、2) 各グループの追跡および消滅検知、3) グループ間の競合の解消、の三つの処理が必要になる。これらは基本的に、残差駆動型アーキテクチャで複数のエージェントによって解決した課題と同一なため、同じアーキテクチャによって実現できる。その構成を図 1 に示す。本処理は、一つのグループ生成エージェント (G-Generator) と、動的に生成・消滅する複数のグループ追跡エージェント (G-Tracer) からなる。G-Generator は 1) 新しいグループの生成を担当し、G-Tracer は 2) 各グループの追跡および消滅検知を担当する。そして、グループ追跡エージェント同士の相互作用によって、3) グループ間の競合を解消する。

初期状態では G-Tracer は一つもなく、まず入力された部分ストリームを追跡する G-Tracer を G-Generator が生成する。以後 G-Tracer は、入力の中から後述する同一グループ条件に合致する部分ストリームを受けとる。どの G-Tracer にも合致しない部分ストリーム（残差駆動型アーキテクチャの残差に相当する）があれば、新たな G-Tracer が生成される。

同一グループ条件を以下に示す。

1. 音程差 Δf [cent]² が閾値 ν 以下である。
2. 方向差 Δd [単位は前節の方向同定の分解能] が閾値 ρ 以下である。

G-Tracer は現在の基準音程、基準方向を持っており、条件判定はすべてこれらとの比較に基づいておこなう。これらの基準値は、現在の部分ストリームの音程・方向か、以前追跡した部分ストリームが消滅した時の音程・方向である。現在の実装では、追跡中の他の部分ストリームがある場合には $\nu = 350$ [cent]、ない場合には $\nu = 600$ [cent] である。また、 $\rho = 2$ である。

競合は、ある部分ストリームが複数の G-Tracer の同一グループ条件に合致する場合に起きる。その場合、距離 $K = \alpha|\Delta f| + (1 - \alpha)|\Delta d|$ ($\alpha = 0.003$) (2) が最小の G-Tracer が、その部分ストリームを受けとる。

5. 実験結果

ダミーヘッドでバイノーラル録音した音声の混合音を対象に、分離実験をおこなった。その結果、方向属

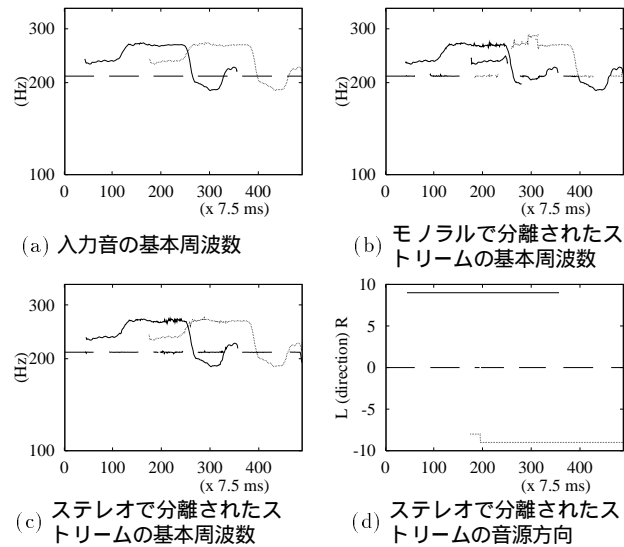


図 2: 実験結果

性が部分ストリームの抽出・分離に有効だけでなく、グルーピングにおいても有効であることを確認した。モノラルの場合には分離できなかった二つの音響ストリームの基本周波数が近接した入力に対しても、方向属性を用いることにより音源方向が異なれば正しく分離できた。

実験結果例を図 2 に示す。これは、同一の女性の声（あいうえお）が 1 秒差で発音する混合音（最初の音が右 90 度、次の音が左 90 度）に、さらに中央から断続的な三角波が発音する場合の結果である。モノラルの場合 (b) では音響ストリーム間で干渉が起きてしまっているが、ステレオの場合 (c)(d) では三つのストリームが適切に分離されていることがわかる。

6. おわりに

本稿では、位相差と強度差を用いた各部分ストリームに対する方向同定と、方向と音程を用いた部分ストリームのグルーピングの二点で、音響ストリーム分離システムを拡張した結果について述べた。両耳聴により方向属性を得ることで、部分ストリームの抽出精度が上がるだけでなく、音源方向を利用したグルーピングもおこなえるようになった。今後は、今回扱わなかった方向同定における発音時刻の差や、グルーピングにおける音色属性にも対応していく予定である。現在はボトムアップ処理だけで構成しているが、音声固有の知識等を用いたトップダウン処理と統合することにより、実環境下での音声認識へと応用していきたい。

参考文献

- [1] 中谷 智広, 奥乃 博, 川端 豪: 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, pp.232-241 (1995).
- [2] 奥乃 博, 中谷 智広, 川端 豪: 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 I. 音環境理解からのモデル化. 第 51 回情処全大, 2R-5 (1995).
- [3] 中谷 智広, 川端 豪, 奥乃 博: 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 II. 残差駆動型アーキテクチャの提案とモノラル音源への適用. 第 51 回情処全大, 2R-6 (1995).

² 音程の対数表記で、1 オクターブの音程の周波数差は 1200cent.