

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-5

I. 音環境理解によるモデル化

奥乃 博 中谷 智広 川端 豪

日本電信電話（株）・NTT 基礎研究所

1 はじめに

最近、複数の音が存在する実環境で音一般を分析し、理解しようという音環境理解の研究が始まった。従来の音響研究では、音声や楽音などの個別の音だけを扱い、かつ、「研究室環境」という理想的な音場を想定しているものが多く、実環境への適用は容易ではなかった。言い換えると、現在の計算機を持つ聴覚機能は、聴覚に障害のある人のレベルに留まっており、少しでも雑音が入ると適切な処理ができない。一方、健康な聴力を持つ人は、入力音に含まれるさまざまな音の中からどれかの音に注目して聞くことができる。このような適応的な聴覚機能は「カクテルパーティ効果」と呼ばれる。本稿では、カクテルパーティ効果を取り上げ、音環境理解からそのモデル化を行い、計算機上でそのような機能を実現するための課題を指摘する。

2 音環境理解研究とは

聴覚心理の一分野である「聴覚的情景分析」(Auditory Scene Analysis)¹⁾は、音一般に対する人間の聴覚モデルの解明を目指している。実際、一般の音を扱うために手がかりとなる特徴が数多く発見されている。(例、調波構造(ピッチ)、立ち上がり(onset)、立ち下がり(offset)、AM変調、FM変調、音色、ホルマント、方位同定など。)しかし、そのような特徴を用いてどのように音を分離するかについてはほとんど知見が得られていない。一方、最近、計算機科学や人工知能の立場から研究が始まった「音環境理解」(Computational Auditory Scene Analysis)は、計算機に、音の分離を通じて音響事象を分析し理解させることを目指している^{2), 3)}。さらに、その一環として、音楽における情景分析・理解を目指し、柏野らは楽音の生起確率をベイズネットワークで処理し、楽音の認識を行っており⁴⁾、後藤らは並列処理によるビートの実時間追跡を行っている⁵⁾。

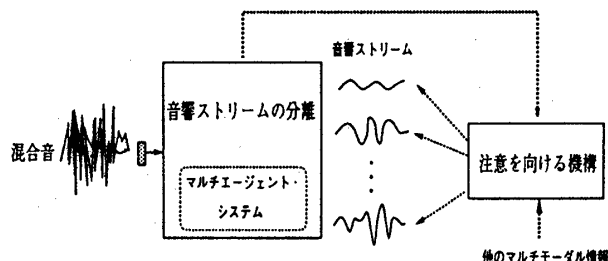


図1: カクテルパーティ効果のモデル化

3 音響ストリームによる表現

音環境理解のためには音の表現が不可欠である。我々は音の表現として、ある一貫した特徴を持つ音のまとまりである「音響ストリーム」を用いる。音響ストリームは、特徴のとらえ方で階層構造をなす。たとえば、オーケストラ全体を一つの音響ストリームととらえてもよく、各パートを別々の音響ストリームととらえてもよい。ストリームが形成される階層のレベルは、注意の焦点や理解と大いに関連する。

我々は音環境理解の応用として、次のような2つの一般環境下での音響処理を設定している:

1. カクテルパーティ効果 — 混雑したパーティ会場で選択的に会話が聞ける機能である⁶⁾。
2. 聖徳太子効果 — 相互に関連のない会話や音を同時に聞ける機能である(名前は7人の訴えを同時に聞いたという聖徳太子の故事による)⁷⁾。

これらの課題を音響ストリームでモデル化してみよう⁶⁾。カクテルパーティ効果は、音響ストリーム分離で得られた複数の音響ストリームに対して、動的に注意を切替える機構から構成されるとモデル化できる(図1参照)。注意を切替えるには、音だけでなく、画像情報などのマルチモーダル情報も必要となろう。聖徳太子効果のモデル化は、注意の切替え機構の代わりに、分離された音響ストリームから音声ストリームだけを選別し、それらの音声ストリームを入力として受けとる別々の音声理解システムから構成される。人間には難しい聖徳太子効果が、計算機では自然にモデル化が行える。

Sound Stream Segregation for Cocktail Party Effect

I. Modeling by Computational Auditory Scene Analysis

Hiroshi G. Okuno, Tomohiro Nakatani, and Takeshi Kawabata NTT Basic Research Laboratories

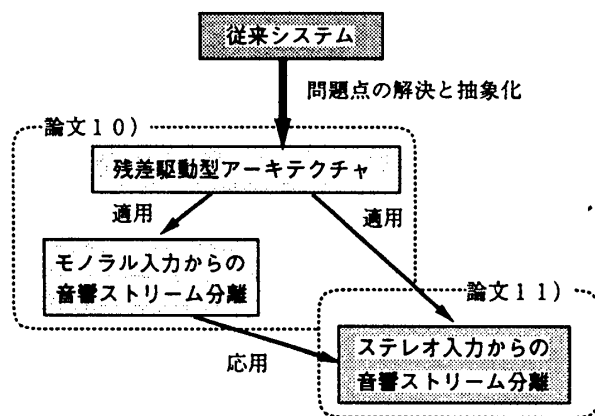


図2: 研究の流れと関連論文 10), 11) との関係

4 音響ストリーム分離の手法

音響ストリーム分離は、2段階に分かれている。まず、入力音から何らかの特徴を一貫性として持った部分ストリームを抽出し、次に、部分ストリームのうち同じ特徴を持ったものをグループ化する。音響ストリーム分離方法は、『ボトムアップ方式』と『トップダウン方式』に大別できる。ボトムアップ方式では、音響ストリームを分離するのに用いる手がかり以外には入力音中に含まれる音についての仮定を設けないのに対して、トップダウン方式では、入力音中に含まれる音についての情報が外部から与えられる。たとえば、発話者が分かっていたり、その音や声を聞いたことがある場合、あるいは、画像情報から音響事象が推定でき、その結果こういう音が聞こえるはずであるという予断ができる場合である。

5 マルチエージェントによる分離

我々はボトムアップ方式の限界を明らかにするためにそれによる音響ストリーム分離に取り組んできた³⁾(これを『従来システム』と呼ぶ)。分離の手がかりとして、調波構造¹⁾という音の特徴を使用する。一般に音の分離では、入力音中に含まれる音の種類と個数が予め与えられており、さらに、その個数が途中で変化しない場合が多い。我々はそのような仮定を行わずに、音響ストリーム分離を試みた。音の個数が動的に変化するので、個々の音響ストリームを一つのエージェントが専属的に追跡するというマルチエージェントシステムで構成することにし、一般的な黑板モデル⁸⁾は採用しなかった。

¹⁾基本周波数とその整数倍の倍音とからなる構造。

これまでの研究経過を図2に示す。従来システムでは、無雑音下で混合音から調波構造ストリームを分離することができた。また、このマルチエージェント構成法が、他の音の特徴を用いて分離する場合にも適用できることが分かった。一方、入力音をどれかの音響ストリームに排他的に割り当てることが不十分なために、音の個数が増えたり、雑音が入ったりすると分離性能が低下することも分かった。そこで、これらの問題点を解決し、さらに、音の特徴に依存しないように抽象化したモデル化として『残差駆動型アーキテクチャ』を提案した^{9), 10)}。これに基づいて構築したモノラル入力の調波構造ストリーム分離システムの雑音下での分離精度は、従来システムより大幅に向上した¹⁰⁾。また、ステレオ入力から方向情報を利用する調波構造ストリーム分離システムでは、さらに分離精度の向上が得られた¹¹⁾。

6 おわりに

本稿では、実環境での一般の音を分析し理解する音環境理解の研究には、音響ストリーム分離が不可欠なことを指摘し、それによるカクテルパーティ効果のモデル化を示した。また、音響ストリーム分離のマルチエージェントによる構成を概観した。これまでに得られた知見から、カクテルパーティ効果を計算機上で実現するための糸口がつかめたと考えられる。最後に、御討論いただいた柏野牧夫氏、柏野邦夫氏、後藤真孝氏、萩田紀博氏に感謝致します。

参考文献

- 1) Bregman: *Auditory Scene Analysis*, MIT Press, '90.
- 2) Brown: *Computational auditory scene analysis: A representational approach*, *PhD, Univ. Sheffield*, '92.
- 3) 中谷他: *Auditory Stream Segregation in Auditory Scene Analysis with a multi-agent system*, *AAAI-94*.
- 4) 柏野他: *Organization of Hierarchical Perceptual Sounds: Music Scene Analysis ...*, *IJCAI-95*.
- 5) 後藤他: *リアルタイムビートトラッキングシステムの並列計算機への実装 — AP1000 による ...*, *JSP-95*.
- 6) 奥乃他: *Cocktail Party Effect with Computational Auditory Scene Analysis — Preliminary Report*, *HCI-95*.
- 7) Cooke 他: *Computational Auditory Scene Analysis*, *Endeavour*, 17:4, '93.
- 8) Lesser 他: *IPUS: An Architecture for Integrated Signal Processing and Signal Interpretation ...*, *AAAI-93*.
- 9) 中谷他: *Residue-driven architecture for Computational Auditory Scene Analysis*, *IJCAI-95*.
- 10) 中谷他: *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回情報学会全大, 2R-6, 1995.
- 11) 後藤他: *III. 両耳聴による音響ストリーム分離*, 第51回情報学会全大, 2R-7, 1995.