# 3J-03 知識源に依存しない音源分離処理の提案*

村岡 秀哉　木下 智義　田中 英彦
{hideya,kino,tanaka}@mtl.t.u-tokyo.ac.jp
東京大学大学院 工学系研究科

## 1 はじめに

複数の音が混在する音響信号から、外界の状況を認識理解するという人間の聴覚における情報処理を聴覚的情景分析と呼び、その計算機上での実現の研究がさかんに行われている。

従来の研究の多くは、音声あるいは音楽といった特定の入力の混合音を前提として分離・認識するものであり、音一般に関する研究はあまり見られない [1, 2]。これは、高精度な認識処理を行うために特定の音にあてはまるルール/知識を処理システムに導入している点が一因にあると考えられる。

さて混合音の認識においては、入力信号から個々の音に相当する部分を分離抽出する処理 (音源分離処理) と個々の音を判定する処理 (音源同定処理) が課題として挙げられる。聴覚的情景分析における知識源の利用には、大別して (1) 処理精度の向上 (2) 音源の同定処理への利用が挙げられるが、これらの処理 (音源同定を含む) を後段に配置し、知識源を利用しない音源分離処理を前段に置く事で一般の入力音響信号に対する認識処理が可能になると考えられる (図 1)。

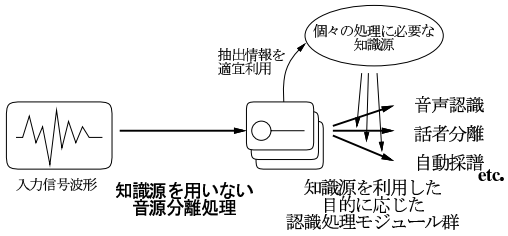


図 1 : 知識源に依存しない音源分離処理

本稿では、音一般を対象とする聴覚的情景分析の前処理としての、知識源に依存しない音源分離処理を提案し、具体的処理内容を述べる。

*Proposal of Sound Source Separation without using Knowledge Source*
Hideya Muraoka, Tomoyoshi Kinoshita and Hidehiko Tanaka
University of Tokyo, Graduate School of Engineering,
7-3-1 Hongou, Bunkyo-ku, Tokyo 113, Japan

## 2 音源分離処理

### 2.1 単音を構成する周波数成分に共通する性質

音源分離処理の実現としては、整数倍の調波構造、立上り時間、共通 AM 変調、共通 FM 変調などの特徴に基づく手法が研究されている。ここでは有声音や管弦楽器のみではなく一般音 (整数倍の調波構造を持たない打楽器等を含む) を対象とするため、単音を構成する周波数成分に共通する性質として AM 変調/FM 変調の共通性を用いる。

即ち、個々の周波数成分における振幅/周波数の時間変化の類似度を用いて周波数成分集合から単音にクラスタリングを行う。

### 2.2 周波数成分の抽出

まず入力信号波形から周波数成分の抽出を行う。

これには信号波形の周波数解析、及び得られたサウンドスペクトログラムからの線スペクトルの抽出処理が含まれる。

### 2.3 周波数成分から構成される単音の分離処理

2.3節では周波数成分の集合から構成されるような単音を分離する処理を示す。有声音や打楽器を含む楽器音等がこれに当てはまる。

抽出された周波数成分を単音と対応づけるのがこの処理の目的であるが、音楽の和声では複数の単音間で周波数成分が重なるなどの問題のため一般には多対多の対応付けとなり、有効な手法を提案した研究は見られない。

#### 2.3.1 周波数成分の融合と分離

単音にクラスタリングを行う前に、個々の周波数成分に初期存在確率を付与する処理を行う。

ところで、一つの周波数成分が複数の単音に対応する可能性としては以下の点がある。

1 時間的に連続した同じ周波数の音の間で前後の周波数成分が接続した場合
2 音楽の和声 (harmony) の場合

また複数の周波数成分が一つの単音に対応する場合は以下の点が考えられる。

3 急激な振幅の時間変化により周波数成分が時間的に分離されて認識される場合

4 高調波を持つ音の場合

1 項と 3 項は本質的には同じ問題で、周波数成分の抽出処理において一方のみになるようパラメータ調整する事が可能である。ここでは 3 項の場合がないように予め周波数成分の終了検出の閾値を低く設定する。

切断確率は、振幅が極小となる時刻 (切断候補点) における周波数成分が分離されるといえる確率であり、立下がりや立上りの傾斜から求められる [3]。切断候補点が $n$ 個 (両端を含む) 存在する周波数成分のうち、開始/終了が切断候補点 $i,j$ である周波数成分 $\mathrm{c}_{i,j}$ の存在確率は切断確率 $\mathrm{P}_k$ から以下の式により求められる。

$$\mathrm{Prob}(\mathrm{c}_{i,j}) = \mathrm{P}_i \cdot (1-\mathrm{P}_{i+1}) \cdots (1-\mathrm{P}_{j-1}) \cdot \mathrm{P}_j$$

ここで周波数成分集合 $\{\mathrm{c}_{i,j}\}$ の任意の時刻における存在確率の総和は 1 となる。

2 項の問題に関してはここでは触れず、2.4節で取り扱う。また以下の単音クラスタリング処理で 4 項を解決する。

### 2.3.2 周波数成分のクラスタリング

存在確率の与えられた周波数成分を用いて単音へクラスタリングを行う。共通 AM 変調を用いるため以下の仮定をおく (以下、共通 FM 変調に関しても同様のため割愛)。

- 同一の単音に属する任意の二つの周波数成分の AM 変調の相関は高い

これにより、周波数成分集合内で任意の二つの周波数成分が高相関である場合にその周波数成分集合を単音の仮説として生成することが妥当であるといえる。

周波数成分間の相関度の計算には正規化された内積が利用できる。

$$\mathrm{Prob}(c_i, c_j) = \frac{\int P(f_{c_i},t)P(f_{c_j},t)dt}{\sqrt{\int P^2(f_{c_i},t)dt}\sqrt{\int P^2(f_{c_j},t)dt}}$$

以上の計算より、それぞれの周波数成分の存在確率、及び任意の二つの周波数成分間の相関度が求められる。ここで、求められた相関度を両周波数成分が同じ単音に属する確率であると仮定すると、Dempster らの確率統合 [4] を用いて任意の周波数成分集合 $\{\mathrm{C}|c_1, c_2, .., c_i\}$ が一つの単音にクラスタリングされる確率が計算できる。

$$\mathrm{P(C)} = \prod_{c_a, c_b \in \mathrm{C}} \mathrm{P}(c_a, c_b) \prod_{c_a \in \mathrm{C} \cap c_b \notin \mathrm{C}} (1 - \mathrm{P}(c_a, c_b))$$

実際には全ての可能性を探索すると処理時間が膨大にかかるため、この他に枝刈りを行う必要がある。

得られた単音の仮説はそれぞれ確率 (確信度) で表現されるので確信度最大となる仮説を出力する。

### 2.4 分離が困難な音集合の処理

前項で単音にクラスタリングされた個々の周波数成分間の相関度の平均が閾値以下である場合、実際は複数の単音が混在していると考えられる。周波数成分が多く重なる音楽の和声や、人間が混合音として分離せずに認識する人混みの音などが例として挙げられる。

知識源を利用しないで分離を行うのは困難なため、単音集合仮説として出力し (図 2)、必要ならば後処理で知識源を利用した照合処理を施す。

### 2.5 周波数成分から構成されない単音の分離処理

無声子音など、明確な周波数成分を構成しない単音に関しては、2.2節で周波数成分を抽出した残差の音響エネルギー値によって判断し、白色雑音に近似して出力する。

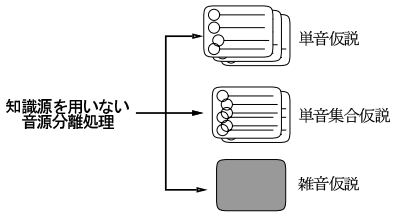


図 2: 音源分離処理の出力

## 3 終わりに

本稿では音一般を対象とする聴覚的情景分析の前処理としての、知識源に依存しない音源分離処理を提案した。周波数成分が時間方向にも周波数方向にも重ならない場合の実験では AM 変調を利用した単音クラスタリングは概ね良い結果を得ている。

今後は実装及び評価、および後処理の処理モデルの提案を行っていく予定である。

## 参考文献

[1] 奥乃 博, 中谷 智広, 川端 豪: “音声ストリーム分離法の提案と複数音声の同時認識の予備実験”, 情処学会論文誌, Vol.38, No.3, pp.510-523, Mar.1997.

[2] 中谷 智広, 柏野 邦夫, 奥乃 博: “音源分離と楽音分離の統合のための音オントロジーの提案”, 人工知能学会 第 11 回 全大, 19-02 (1997).

[3] 柏野 邦夫: “音楽音響信号を対象とする聴覚的情景分析に関する研究”, 博士論文, 東京大学 (1995).

[4] 石塚 満: “Dempster & Shafer の確率理論”, 信学誌, Vol.66, No.9, pp.900-903 (1983).