

共通 AM 変調に基づく音源分離

Sound Source Separation based on Common Amplitude Modulation

村岡 秀哉

Hideya MURAOKA

木下 智義

Tomoyoshi KINOSHITA

田中 英彦

Hidehiko TANAKA

東京大学大学院 工学系研究科

Graduate School of Engineering, The University of Tokyo

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、その計算機上での工学的実現の研究が注目されている [1]。この聴覚的情景分析に対応する処理は、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、分離された個々の音を判定する処理 (音源同定処理) とに大別される。特定の入力音を仮定しない音一般を対象とした聴覚的情景分析を考えた場合、音声や楽音に依存した知識/ルール (知識源) を処理システムに含む事は一般性に欠けるため望ましくない。ここでは、音源分離処理と音源同定処理で知識源の利用を分離する [3] ことで解決を試みる。

本稿では、知識源に依存しない音源分離処理について提案し、更に楽音への応用について述べる。

2 音源分離処理

2.1 単音の表現

分離される単音の表現としては、音響パワーの時間/周波数表現における線スペクトル (周波数成分) の集合と捉える手法が挙げられる [2] が、本稿では、周波数成分および周波数成分を抽出した残差の音響パワー値 (白色雑音で近似) の集合で単音を表現するものとする。

ところで、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、AM 変調、FM 変調などの特徴量の共通性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。また各々の周波数成分は、振幅が最小となる時刻における分離される確率 (切断確率) を保持する。

2.3 周波数成分のクラスタリング

続いて周波数成分から単音へクラスタリングを行う。共通 AM 変調を用いるため、同一の単音に属する任意の二つの周波数成分の AM 変調の相関は高いという仮定をおく

(以下共通 FM 変調に関しても同様)。周波数成分間の相関度の計算には正規化された内積を用いる。

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

以上の計算より、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 らの確率統合理論から任意の周波数成分集合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 [3] (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、この他に枝刈りが行われる)。

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/周波数成分から構成されない単音の処理

楽音の和声のように周波数成分が複数の単音間で重なる場合、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、必要に応じて後処理で知識源を利用した照合処理を施す。

また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、残差を白色雑音で近似して出力する。

3 楽音への応用

応用例として、複数楽器のアンサンブル演奏を入力とする場合を考える。後処理の音源同定処理には、個々の楽音毎に単音をカテゴリ分けする処理、および前項の単音集合仮説に関する分離処理が必要となる。

前者に関しては振幅時間変化や時間的排他処理、また後者に関しては整数倍の調波構造を仮定して分離処理を行う手法が考えられる。

4 終わりに

音一般を対象とする聴覚的情景分析の前処理として共通 AM 変調に基づく音源分離処理について提案し、さらに楽音への応用について述べた。今後は楽音以外の応用について提案を行っていく予定である。

参考文献

- [1] 奥乃, 中谷, 川端: “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”, 情報処理学会論文誌, Vol.38, No.3, pp.510-523, Mar.1997.
- [2] 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).
- [3] 村岡, 木下, 田中: “知識源に依存しない音源分離処理の提案”, 情報処理学会 第 55 回 全大, 3J-03 (1997).