

審査の結果の要旨

氏名 亀岡 弘和

本論文では、様々な音が混在する中から、目的音の音響的性質を特徴づける情報（基本周波数やスペクトル包絡など）を分離推定する方法（多重音解析と呼ぶ。）について論じられている。多重音解析技術は実用性が高く、ロボット聴覚、音声認識、音源分離、自動採譜、オーディオコーデックのための分析（符号化）手段、音楽コンテンツの自動メタデータ化とそれによる高機能な音楽検索システムなど、実に広範囲にわたるアプリケーションへの応用が期待される。論文ではこの多重音解析の問題を解くため、(1)多重ピッチ推定法を確立、(2)音源数推定法を確立、(3)音脈分凝プロセスを定式化、することを目的とし、次の4つのアプローチを検討している。

多重音中の各音源の基本周波数は、各音源のスペクトルが既知であれば、高い精度で推定できる。一方で、多重音スペクトルは、各音源の基本周波数が既知であれば高い精度で分離できる。そのため、多重音スペクトルを音源ごとに分解することと各音源の基本周波数は同時最適化問題として解かれるべきであるという考えが本論文の出発点となっている。第2章では、対象とする混合音のパワースペクトルを、音源ごとに対応するようにクラスタ化する分配関数と、各音源の基本周波数を反復的に最適推定するアルゴリズム（調波構造化クラスタリング）を打ち出し、この原理を基礎としながら、第3章では、これをベイズ的枠組で別解釈および再定式化を行い、その結果からさらにモデル構造選択規準により音源数を推定するための規準が作れることを示している。調波構造化クラスタリングによる多重ピッチ推定法の先行研究に対する優位性と音源数自動推定法の有効性がいずれも評価実験を通して示された。

第4章では、第3章の原理を拡張し、人間の聴覚機能における音脈分凝プロセスに相当するアルゴリズムを定式化している。各時刻で独立に調波成分を見つけ出すための処理（周波数方向の群化）と、抽出された調波成分特徴量の時系列を時間方向にスムージングする処理（時間方向の群化）を多段処理的に行う多くの従来法に対し、これらを協調し合う同時最適化問題として定式化すべきであるとの問題意識のもと、Bregmanの分凝要件と呼ぶ能動的に外界を把握するための人間の優れた聴覚機能をヒントにししながら、個々の音源の時間周波数全域に渡ったパワースペクトル構造を一挙に推定できる新しい方法論（調波時間構造化クラスタリング）を開発した。評価実験により、混合音声信号および音楽音響信号の基本周波数推定精度が、それぞれの分野における最先端の従来法を上回ることを確認した。

第5章では、従来まで独立な問題として扱われがちであったスペクトル包絡推定と基本

周波数推定は本来相補関係にあることを論じ、その問題意識のもと、これらを同時最適化問題として定式化することで、個々の音源のスペクトル包絡推定も同時に行える多重音解析法への応用可能性を示した。また、単一話者音声を対象としたピッチ推定、合成分析、分析合成に関する各評価実験を通して、提案法のようにピッチ周波数とスペクトル包絡を同時推定することが、いずれの推定の精度に対しても良い効果をもたらすことを確認した。

第 5 章までの手法は、パワースペクトルの加法性が近似的に成り立つ(周波数成分間の干渉項を無視できる)という仮定のもとで、観測パワースペクトルから所望の情報を得るためのアプローチであった。しかし、2 音以上の近接する基本周波数の分離推定や、近接周波数成分の分離を高精度に行うためには、調波構造だけでなく各信号の位相差が分離の手がかりになる。さらに、将来的に複数センサ入力の多チャンネル信号を対象とした音源分離手法を視野に入れるのであれば、位相推定を含んだ複素スペクトル領域での解析が不可欠である。第 5 章までの方法論の有効性と上記のような展望を踏まえ、第 6 章では、第 3 章のアプローチをその本質を損なうことなく複素スペクトル領域に拡張できないかということがテーマとなっている。第 6 章での中心的な議論の対象は、正弦波重畳モデルの最尤パラメータを求めるための非線形最適化アルゴリズムを開発することであり、正弦波重畳モデルのパラメータ推定問題の核である周波数推定ないし基本周波数推定の難しさの本質が、正弦波重畳モデルが周波数パラメータに関して非線形である点にあることに着目し、補助関数を用いた新しい反復推定アルゴリズムを導いている。この考え方は、EM(Estimation-Maximization)アルゴリズムをヒントに、その本質が何であるかを理解した上で明らかになったものである。シミュレーション実験により、第 5 章で提案するパラメータ最適化アルゴリズムは勾配法を用いる多数の従来法よりも局所解回避能力と収束速度の面で優れていることを示した。また、実音声を用いた 1 チャンネル混合音声分離実験を行い、提案法の基本性能を確認した。

本論文の大きな成果は、広範囲な用途がありながら解決が難しいとされてきた多重音解析の問題の本質を明確に整理し、その解決策となる独創的な方法論を着実に築き上げていった点にある。

よって本論文は博士(情報理工学)の学位請求論文として合格と認められる。