

# 論文の内容の要旨

論文題目 Statistical Approach to Multipitch Analysis  
(統計的手法による多重音解析に関する研究)

氏 名 亀岡 弘和

本研究は、様々な音が混在する中で目的音の情報（基本周波数やスペクトル包絡など）を分離推定する方法を提案するものである。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。本論文では、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た。

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最適な分配関数を推定する原理を提案し、この原理を基礎としながら、第3章では、これをベイジックの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す。評価実験により、本章で提案した2つの要素技術である、調波構造化クラスタリングによる多重ピッチ推定法と音源数自動推定法の有効性がいずれも示された。

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する。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）を提案する。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べきとの問題意識のもと、これらを同時最適化問題として定式化し、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したことを確認した。

第5章までの手法は、パワースペクトルの加法性を近似的に仮定する(周波数成分間の干渉項を無視した)アプローチであったが、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にはパワースペクトルからは一般に難しく、この場合、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く。この考え方は、EM アルゴリズムの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。