

審査の結果の要旨

氏名 ルルー ジョナトン

音声アプリケーションや通信システムなどの音を媒体としたあらゆるアプリケーションを実現する上で、非定常で強い背景雑音、複数音源の同時観測、あるいは機器の誤作動や欠損パケット等の影響による通信の中断など、我々が身近に直面しうる複雑な現象を扱うことが重要な課題となる。しかし我々人間は周囲の複雑な音響情景を、信号源の知識や所与の観測信号の統計的規則性に基づいて頑健に認識し理解することができる。

本論文の目的は、音環境に内在する規則性のモデルに基づき、かくのごとく複雑な観測信号から音響情景を解析するための統計的アプローチを確立することである。本アプローチでは主として、(1) ある音環境モデルのもとで、何らかの原因で歪み、複数の音が混在し、あるいは部分的に欠損したような観測信号データに対し、どのような音響的現象が生じているかについての最も自然な解釈を与える問題、(2) 上記モデルに基づき欠損した情報を復元する問題、(3) 上記モデルを観測データから自律的に獲得する問題を扱っている。これらすべての問題に対し、数理的定式化を通して目的関数を立てて解決を図ることを基本戦略とし、これにより様々な問題を拘束つき最適化問題として見通しよく解決することができた。

第2章で関連研究をサーベイした後、第3章ではまず音響情景解析を目的とした時間周波数パワー領域における統計的モデルを導入している。平らな基本周波数(F0)の音楽信号を対象に亀岡により導入された周波数時間構造化クラスタリング(HTC)法を、音声のようにF0が滑らかに変動する信号に適応できるように拡張した。時間周波数領域で定式化を行うことにより、人間の聴覚機能における音脈分凝プロセスに用いられている音声の規則性を明白に取り入れた拘束付き混合正規分布の完全にパラメトリックな音声信号モデルを導入できた。滑らかに変動するF0包絡を3次元スプラインにより表現することで、元の手法の最適化アルゴリズムの効率性を保ちながら音声信号へ対応できるようにHTC法を拡張した。また、混合正規分布に基づいた背景雑音モデルを導入し、雑音と音声の混合する音響情景のモデリングを可能にした。第4章では、評価実験により音声信号のクリーン・雑音中・複数話者のF0推定や音声強調・音源分離といった音響情景解析の様々なタスクにおいて提案手法が従来法より優れていることを示した。

第5章では、不完全データから情景解析を行うためにこのモデルを拡張する方法、欠損データを復元する方法を説明している。補助関数法を用いることにより、全般的に分布のフィッティングに基づく手法の不完全データへの拡張が可能であることを証明した。フィッティングの良さを測る尺度がBregmanダイバージェンスという尺度のクラスに属する場合に補助関数法とEM(Expectation-Maximization)アルゴリズムとの関係を明白にした。提案手法により不完全データからの情景解析と欠損データの復元を同時に行うことが可能になることを評価実験で確認した。

続いて、パワースペクトル領域処理の限界を超えること、また、位相情報を有効利用することの重要性に着目し、複素時間周波数領域における位相の無矛盾性規準を第6章で導

入し、時間領域の適応的テンプレートマッチングの枠組を第7章で構築している。第5章までは位相情報を使わずパワースペクトル領域で処理を行ったが、それによるいくつかの課題を説明している。パワースペクトルから信号の再合成が必要な場合は失われた位相情報をパワースペクトルと位相との無矛盾性を考慮しながら推定しなければいけない。また、パワースペクトルの加法性が近似的に成り立つ(音源間の干渉項を無視できる)ことを仮定したが、加法性が成り立っていない場合の処理も考えられなければならない。最後に、音のクラスによって位相情報が有効な特徴になり、それを積極的に利用する手法も考えるべきである。全ての点において、複素時間周波数領域か時間領域での処理が自然な解決となりうることに着目し、各領域での手法を第6章と第7章で導入している。

第6章では短時間フーリエ変換(STFT)で得られる複素スペクトログラムの構造を分析し、パワースペクトルと位相の無矛盾条件を明白にしている。重なり合うフレームのフーリエ変換により構築されるSTFTスペクトログラムが時間信号の冗長な表現であるため、複素時間周波数領域での任意の複素数の集合がかならずしもある時間信号から得られたSTFTスペクトログラムであるわけではない。そういったスペクトログラムを無矛盾スペクトログラムと呼び、時間周波数領域での数理的な無矛盾性拘束を導出し、その下で任意のパワースペクトルから無矛盾な位相を復元する高速かつ柔軟性の高いアルゴリズムを導入した。

第7章では、観測信号が少数の基礎的なテンプレートの任意のオンセット時刻と非負な振幅での重ね合わせからなるという仮定の下で、時間領域で波形を直接モデル化する適応的テンプレートマッチング手法を提案し、効率的な最適化アルゴリズムを導出している。テンプレートが重なっているという困難な状況でも、信号内に繰り返し起こるテンプレートをデータから自動的に取り出し、そのオンセット時刻及び振幅を推定できることを音楽信号や生理的信号での実験により確認した。

最後に、第3章から第6章までは時間周波数領域で作業を行うために、ウェーブレット変換あるいは短時間フーリエ変換という、古典的なフィルタバンクから得られた時間周波数表現を用いたが、第8章では観測データから信号の時間周波数分析に一番適しているフィルタバンクを学習することを検討している。人間の聴覚に重要な役割を果たすとされている信号の変調構造に着目し、信号の変調構造を効率良く抽出できるように聴覚末梢系を適応させる人間の機能についての仮説を数理的な枠組みで定式化し、音声データからの変調エネルギー基準下でのフィルタバンク学習実験で妥当なフィルタバンクが得られることを確認した。

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モノラル入力からの音響情景解析の問題の本質を明確に整理し、代表となるモノラル信号推定・分解・復元・加工の問題を、統一的な着眼点の下で効率的に解決した点にある。

以上を要するに、本論文は博士(情報理工学)の学位請求論文として十分に合格と認められる。

以上