

審査の結果の要旨

氏 名 橋 秀幸

歌声は音楽信号の中でも特に一般聴衆からの関心が集まりやすい領域であり、音楽信号からの情報抽出（音楽信号への自動タグ付けなど）、音楽加工（自動カラオケ製作など）などの応用を考えた際には、歌声に関するものがとりわけ重要性が高い。従って、そのような応用のための基礎として、混合音楽信号から歌声を抽出する技術の開発が重要な課題となる。本論文は歌声のゆらぎに着目することで、混合音楽信号中の歌声を強調（抽出）する方法を議論したもので「**Music Signal Processing Exploiting Spectral Fluctuation of Singing Voice Using Harmonic/Percussive Sound Separation**（調波音打楽器音分離による歌声のスペクトルゆらぎに基づく音楽信号処理の研究）」と題し、7章により構成されている。

第1章は序論であり、最初に述べたような音楽信号処理の研究そのものの意義がまず述べられ、次に特に歌声に注目することの重要性が述べられている。

第2章は、まず、その前半で、歌声の性質と、本論文で歌声のゆらぎに着眼する理由が述べられている。歌声には(1)整数倍音の存在、(2)特徴的スペクトル包絡形状、(3)各倍音成分に周波数や振幅のゆらぎ（ヴィブラート、トレモロ）がある、などの性質があり、これらを利用することが歌声強調を考える上では自然であり、従来から、(1)倍音、(2)スペクトル包絡、を用いて歌声を抽出する方法については多数検討がなされているが、これらには、(1)歌声以外の楽器音も整数倍音を持つ、(2)楽器音と歌声が混ざった混合信号の場合にはスペクトル包絡を求めることが困難、などの難点もあったことが指摘されており、他方、(3)歌声のゆらぎに関しては、実際の音楽信号（主にポピュラー音楽）のパワースペクトログラムを観察した際には、歌声の特徴的な性質として視覚的に非常に際立つものであるにもかかわらず、従来歌声強調のための特徴としての利用はあまり検討されてこなかったことが指摘されている。

次いで、後半では、ゆらぎが従来それほど利用されなかった一因として、ゆらぎのある信号はスペクトログラム上で複雑かつ多様な形状をとりうるために定式化が難しいという点を挙げ、本論文ではこのような困難を避けるための着想として、歌声のゆらぎはスペクトログラムの時間周波数解像度（分析時間スケール）次第では非常にシンプルな形状として表現されうることを指摘している。さらにそれに基づいて、スペクトログ

ラムを画像処理的に「縦成分」と「横成分」に分離する手法「HPSS (Harmonic/Percussive Sound Separation)」を異なる解像度で2段階で適用することにより歌声を抽出する手法「Two-stage HPSS」が提案されている。MIR-1Kデータベースによる様々な歌声/伴奏比条件下での実験の結果、歌声/伴奏比の改善値(NSDR)が平均しておよそ4dB程度と、既存の歌声強調手法と比較しても高い水準の値を示すことが確認され、提案手法の歌声強調手法としての有効性が実験的に実証されている。

第3章では、「音楽信号からの情報抽出」に関する一連の問題の中でも、特に音楽信号中のメロディの基本周波数系列を推定する問題(AME, Audio Melody Extraction)へのTwo-stage HPSSの応用について述べている。本章では、比較的多くの音楽では、メロディは歌声により演奏されることが指摘され、前処理としてTwo-stage HPSSを利用して歌声をある程度強調(伴奏を抑圧)することにより、比較的簡易な基本周波数推定手法によって、ある程度の精度でメロディが推定できるはずであると論じており、このことを実証するために以下の二つの実験を行なっている。すなわち、(1)LabROSAデータセットを用いた予備の実験によって、9曲中8曲でTwo-stage HPSSの前処理の有効性の確認を行ない、また、(2)第三者による大規模評価実験(MIREX, Music Informaiton Retrieval Exchange)では、とくに伴奏音がメロディと比較して音量が大きい場合(歌声/伴奏比が-5dBの場合)に、他手法よりも高い性能を示す傾向にあることの確認を行なっており、本章では、これらの実験結果はTwo-stage HPSSによる歌声強調が、「音楽信号からの情報抽出」への前処理としても有効であることを示していると論じている。

第4章では、はじめに、音楽データの流通の環境の変化を背景に、近年はアマチュアが制作しインターネットで公開している楽曲の重要性が増していることを踏まえ、それらの楽曲を想定した自動カラオケ生成システムの開発の重要性の指摘がなされている。また、これを踏まえ、第2章で論じられたTwo-stage HPSSを歌声強調ではなく歌声抑圧のための手法として利用した、混合音楽信号から自動的に擬似的なカラオケを生成するシステム「Euterpe」の設計について述べられている。このシステムには、Two-stage HPSSのほかピッチ変換のエンジンも組み込まれており、両者が協調して動作するように設計されていること、また、本システムで採用された両手法の計算効率の良さにより、このシステムが一般的なパソコンやネットブック上でもリアルタイムで動作することが確認されていることが述べられている。

第5章では、第2章のアイデアの拡張について述べられている。第2章では二種類の時間周波数解像度のスペクトログラムを考えていたが、本章ではそれをさらに拡張し、使用する時間周波数解像度の数を増やすことで、信号を特徴的な時間スケールごとに分離できることが論じられ、それに基づく音響特徴量「CFTS (Characteristic Fluctuation Time-scale)」が提案されている。実験により、朗読音声と歌声ではCFTSの形状に大きな違いが見られることが確認され、また、両者の識別問題を考えた際にも、CFTSを用いた場合は、MFCC(音響分野の標準的な特徴量の一つ)を用いた場合よりも高い識別性能

(92.1%)を示したことが確認されている。

第6章では、ここまでのすべての章で基盤となっていたHPSSについて、その導出過程や、改良の可能性、計算時間と分離性能のトレードオフ、関連手法との性能比較、などが、歌声に関する議論とはやや独立して詳しく述べられている。

第7章は「結論」であり、2章から6章までの成果を各章ごとにまとめるとともに、今後の研究の展望を述べている。

以上、本論文の大きな成果は、(1)歌声のようにゆらぎのある信号を時間周波数表現すると、その形状がスペクトログラムの時間周波数解像度に依存して大きく異なることを指摘し、その着眼点に基づいてシンプルな歌声抽出方法「Two-stage HPSS」を提唱したこと、(2)音楽検索への応用の一例としてTwo-stage HPSSを歌声メロディ抽出問題の前処理として利用することの提案、および、(3)音楽加工アプリケーションの一例としてTwo-stage HPSSに基づくリアルタイム自動カラオケ生成システムを実際に制作したこと、に要約できる。これらの成果は、従来は必ずしも取り扱いが容易ではなかった歌声のゆらぎが、非常にシンプルな方法で捕捉できることを示しており、本論文において、実際にその方法の有効性の実証もなされている。このことから、本論文で提案された手法は、歌声信号処理において大きな意義をもち、音楽検索や音楽加工などの応用領域分野の発展にも貢献するとともに、システム情報学の進歩に対して寄与するところが大きいと認められる。

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。