

**学位論文題目** 多様な歌唱様式を予測・説明する歌声音響信号の分析合成モデルとその応用に関する研究

**取得年月** 2009年3月 **学位種別** 博士(情報科学) **大学** 名古屋大学

**氏名** 大石 康智(日本電信電話(株)NTTコミュニケーション科学基礎研究所 研究員)

**推薦研究会** 音楽情報科学

**推薦文** 本研究は、音楽の大切な要素の1つである歌声が通常の話声とどう異なるかを明らかにし、ビブラートのような、時間的に複雑なダイナミクスを統計的信号処理のアプローチによってモデル化することに取り組んでいる。その着想から評価に至る一連の研究成果は高く評価でき、独創的な成果を生んだ論文として推薦する。

人間の口から発せられる音響事象の1つである歌声は、歌詞に込められた感情や想いを旋律に乗せて伝達する最も身近な音楽演奏手段である。歌声は、これまで、音響学、生理学、解剖学、心理学、歌唱学といった多角的な立場から研究され、歌声知覚や生成機構の解明に向けたさまざまな知見が得られた。たとえば、歌声の声の高さを表す基本周波数( $F_0$ )には、通常の話声に見られない、ビブラートやオーバーシュートなどの動的な変動成分が存在する。また、歌声の音色を表すスペクトルの3kHz付近には、顕著なピーク(歌唱フォルマント)が存在し、このピークはビブラートに同期して振幅変調される。さらに、これらの変動成分は歌唱者ごとに多様であり、個人性知覚に寄与する特徴である。

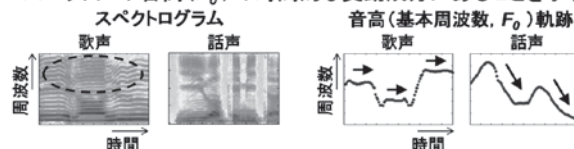
このように複雑で多様な変動をする歌声を計算機が理解できれば、音楽情報検索システムをはじめとして、さまざまな用途で有用である。同じ歌声であっても、人間はその歌唱様式や技術の差を容易に区別できるように、計算機も歌声の多様性を理解し特徴づけた上で、歌声の認識や合成を行うべきであると考え。そこで本論文では、これまで解明された歌声の物理的性質を踏まえ、歌声を工学的に利用するためのモデル化手法に焦点を当てる。そのために、従来のように特定の歌唱者や歌唱法に限定するのではなく、一般人を含め、さまざまなレベルの歌声を対象とした大規模データベースを用いて研究を進める。

まず、聴取実験と自動識別実験に基づいて、歌声と通常の話声を人間と同程度の性能で識別し得る信号特徴尺度を構成する。話声との識別の観点から、あらゆる歌声に共通する音響的特徴として、スペクトルや音高の時間的な変動成分の重要性を示す。

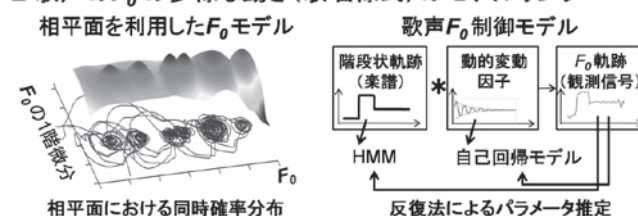
次に、歌声の $F_0$ に焦点を当て、その時間的な動きをモデル化し、歌唱者ごとにどのように動きが異なるか、歌唱様式を特徴づけることに取り組む。本研究では、 $F_0$ が歌唱様式を特徴づけるなんらかの微分方程式に従って生成されると想定し、観測される $F_0$ からその微分方程式を明らかにする問題と位置づけ、2つのモデルを提案する。1つ目のモデルでは、 $F_0$ とその時間微分によって構成される相平面もしくは相空間上に $F_0$ 軌跡を描き、その同時確率分布によって $F_0$ の動きをモデル化す

#### □ 歌声と話声の特徴分析と自動識別

聴取実験と自動識別実験から、歌声と通常の話声との違いは、スペクトルや音高( $F_0$ )の時間的な変動成分にあることを示す



#### □ 歌声の $F_0$ の多様な動き(歌唱様式)のモデルリング



る。相平面は、複雑な微分方程式の解の性質を調べるための便利な手段であり、この平面の同時確率分布が近似的に微分方程式を表現すると考える。このモデルによって、 $F_0$ 軌跡を階段状の音高軌跡に整形し、これを鼻歌検索の検索クエリに利用したところ、従来法と同等以上の性能が得られた。また、相空間の同時確率分布によって特徴づけられる動的変動成分から、歌唱様式(声楽家、ポップス歌手、素人)を自動分類できることを確認した。

2つ目のモデルは、 $F_0$ 制御モデルである。入力に階段状の音高軌跡(楽譜に記される旋律を表す)であり、これを隠れマルコフモデル(HMM)で表現する。この入力にさまざまな動的変動成分が複雑に重ね合わされて、 $F_0$ が出力されると想定し、この入出力関係を自己回帰モデルで表現する。このモデルパラメータが、 $F_0$ の動的変動成分を表現する。そして、観測される $F_0$ だけから、反復法によって、HMMのパラメータと自己回帰モデルパラメータを同時推定するアルゴリズムを提案する。評価実験より、アルゴリズムの収束性と推定性能を確認した。また、ある歌唱者の $F_0$ 動的変動成分を別の歌唱者の動的変動成分と取り換え、歌唱様式だけを自由に交換可能な歌声合成手法も提案する。HMMと効率的な学習アルゴリズム(Baum-Welchアルゴリズム)によって音声認識の性能が劇的に向上したように、提案手法による、 $F_0$ の動きを学習する枠組みは、歌声分析合成の研究分野の発展に大きく貢献できると考えている。

(平成22年3月31日受付)