JAIST Repository
https://dspace.jaist.ac.jp/

Title	Studies on Spectral Modification in Voice Transformation
Author(s)	Nguyen, Binh Phu
Citation	
Issue Date	2009-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/8002
Rights	
Description	Supervisor:instruction Prof. Masato Akagi, 情報科学研究科, 博士

JAIST
JAPAN
ADVANCED INSTITUTE OF
SCIENCE AND TECHNOLOGY

Japan Advanced Institute of Science and Technology

音声変換のためのスペクトル変形法に関する研究

Binh Phu Nguyen

北陸先端科学技術大学院大学

2009 年3 月

論文の内容の要旨

本論文は，音声変形において，変形された音声の品質を向上させるためのスペクトルモデリングとスペクトル変形手法を提案することを目的とする．特に，変形された音声の自然性をどのように回復するか，また，多種多様な音声データベースを構築するために多様な発話スタイルをどのように創造するかに焦点をあてる．

音声変形は，他の性質を保存しながらある特定の知覚的性質を変形するプロセスであり，多くの場で応用されている．たとえば，Text-to-Speech システムにおいては，予め録音されたデータから様々な声質を作り出すために使われる．外国語学習システムでは，発話スピードを遅くすることでより聞き易くする．また，聾者に対して聴取能力を高めるために，残存聴力を使える帯域に音声を移動させる．

スペクトル処理は，人の音声知覚に深く関係しているので，スペクトルを変形することは知覚に多大な影響を及ぼす．このため，音声変形の中で，スペクトル変形は中心的な役割を演じている．スペクトル処理の基本は，時間波形を時間－周波数領域で表現することである．今までに，多くの時間－周波数領域でのスペクトル変形手法が提案されてきた．しかし，時間領域では，多くの手法は音声信号をフレーム毎に処理しており，パラメータ値の時間変化をうまく記述できていない．このために，変形後の音声では滑らかさが保たれておらず，変形音声の品質低下の原因となっている．一方，周波数領域では，ルールベースあるいは統計的な手法でスペクトル変形が行われている．ルールベースのスペクトル変形手法の中で，線形予測による方法あるいは周波数軸伸縮法では，ホルマントの重要な性質である周波数，振幅，帯域幅を独立に制御することができず，スペクトル変形の自由度が小さい．また，統計的な変形手法では，数学的記述が使えることは利点であるが，音響現象としての制約を組み込むことが難しい．このため，有効なスペクトル変形が行える新しい手法の出現が望まれている．

スペクトルをどのような特徴によってモデル化するか，そして，変形をどのようなアルゴリズムによって行うかは，大きな問題である．本論文では，これらを考慮し，有効なそして自由度が大きい分析／合成手法を提案する．提案手法では，パラメータ値の時間変化をうまく記述するために temporal decomposition (TD) 手法を採用する．また，スペクトル包絡を記述するために spectral-GMM パラメータを用いる．これらの手法を用いることにより，TD によって分解されたイベント関数をモデル化し，時間領域でのパラメータ値の変化を記述できることとなる．さらに，spectral-GMM パラメータ制御することにより自由にスペクトル包絡を変形できる．

特に，TD を用いれば，変形音声の滑らかさが保たれるため，変形音声の自然性を向上させることができる．本論文では，波形片結合音声合成法に TD を用いた平滑化法を適用し，結合部でのスペクトル，F0，ゲインのミスマッチを補償する手法を新たに提案する．

さらに本論文では，多様な発話スタイルを創造するために，スペクトル変形のための新しいアルゴリズムを提案する．提案法は，ルールベースのアルゴリズムであり，これを感情音声合成と男女声変換に適用する．このアルゴリズムでは，spectral-GMM パラメータを制御することによりスペクトル包絡の変形を行うが，新たに，GMM パラメータ抽出の精度向上とフレーム間の滑らかさを保つための方策を適用することで，アーティファクトを発生させない構造に改善されている．

多数の実験の結果，変形音声の品質を向上させるという目的に対して，提案法は有効であることが示された．本論文の貢献を要約すれば，まず，パラメータ値の時間変化をモデル化したこと，および，スペクトル包絡のモデル化の精度を向上させたことである．このモデル化により，変形された音声の滑らかさが保たれ，有効なスペクトル変形が実施されることとなった．そしてさらなる貢献として，波形片結合音声合成法の自然性を向上させる手法を提案したこと，そして，様々な発話スタイル（感情音声とか男女声間変形）を生成できる手法を示したことがあげられる．

キーワード: スペクトル変形, 音声変換, temporal decomposition, spectral-GMM