

機関番号：17401

研究種目：基盤研究(C)

研究期間：2008～2010

課題番号：20560398

研究課題名(和文) 調音音声合成インタフェースの開発と信号計測変換コンバータへの応用

研究課題名(英文) Development of articulatory speech synthesis interface and its application to signal measurement converter

研究代表者

緒方 公一 (OGATA KOHICHI)

熊本大学・大学院自然科学研究科・准教授

研究者番号：10264277

研究成果の概要(和文)：本研究は、音声生成過程に基づく音声合成手法をシステムとして開発するとともに、その融通性を活用することで応用インタフェースの創出を目的とするものである。本研究で開発されたシステムにより、合成に関連するパラメータの適切な値の評価が可能となり、その成果に基づいてパラメータの自動調整機能を有するGUIインタフェースが開発できた。さらに、応用インタフェースとして、声道形状マッピングインタフェースを開発した。このインタフェースにより、声道形状を介して入力信号に応じた母音合成が可能なコンバータが構築できた。

研究成果の概要(英文)：The aim of this study is to develop a speech synthesis system based on the speech production process and to create related applications by taking the advantage of its flexibility. A speech synthesis simulation system developed here allowed us to evaluate the adequate ranges of related parameters. Based on the results, a system with GUI(graphical user interface) has been developed for an automatic parameter adjustment function. Moreover, a vocal tract mapping interface has been developed as a signal measurement converter. This interface is applicable to a vowel sound generator driven by an input signal through the vocal tract shape.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2008年度	1,500,000	450,000	1,950,000
2009年度	1,100,000	330,000	1,430,000
2010年度	900,000	270,000	1,170,000
年度			
年度			
総計	3,500,000	1,050,000	4,550,000

研究分野：工学

科研費の分科・細目：電気電子工学・計測工学

キーワード：音声合成，声道，インタフェース，計測，画像処理

## 1. 研究開始当初の背景

現在主流となっている波形接続方式の音声合成では、蓄積された音声信号を利用するため部分的な合成音の品質は優れているものの、連続音での不自然性、声質変換や話速制御などの融通性に問題がある。

一方、音声生成過程に基づいた手法は、人

間の発話メカニズムを模倣した声帯や声道のモデルに基づいて音声波形出力をシミュレートするため、音声信号の蓄積を必要としない利点がある。また、調音器官の運動の連続性による滑らかな連続音の生成の他、音声生成の源となる声帯や声道のパラメータを種々に設定することで、声の高さや話速の変

更、男声や女声の合成など、バラエティー豊かな合成音の生成が期待され、その融通性に大きな特徴がある。

このように融通性が期待される方式であるが、声道断面積などの関連する多数のパラメータの適切な設定や、音声の自然性の向上、システムとしての取り扱いの容易さの実現等、その実用化には種々の解決すべき課題がある。

## 2. 研究の目的

本研究は、音声生成過程に基づく音声合成手法をシステムとして開発することで、その特徴である融通性に優れた合成を実現するものである。さらに、その融通性を積極的に活用することのできるインタフェースの創出を目的とするものである。

申請者は、音声生成過程に基づく手法の融通性に着目し、システムの開発研究を行ってきたが、より確実に音声生成のできる断面積の検討、パラメータの自動設定機能を有したプログラマブルな駆動機構の実現、合成音の自然性の向上のための声帯振動実現、応用展開に向けての効率的な声道形状設定手法の検討、その設定手法を活用した信号変換コンバータの有効性の検討、などを目的として研究を行う。

また、申請者は、画像処理を応用したデータ計測評価についても研究を行っており、視線検出インタフェースの開発、上肢運動の計測と評価などにも取り組んでおり、上記信号変換コンバータへの応用展開を目指して、個々の要素技術についても研究を進め、これらを融合した応用システムの創出を目的としている。

## 3. 研究の方法

ここでは、取り組みのいくつかについて、研究方法の概要を述べる。

子音の合成では、閉鎖等の生じる位置の設定をはじめとする断面積の与え方の他、その形状を時間的にどのように変化させるかという動的要素の設定が重要である。適切な断面積については、音響管の形状に対する声道伝達特性を計算できるシミュレータを活用して検討を行う。本システムでは、声道を20個の音響管の接続で近似表現しているが、縦続1次系関数のステップ応答に基づいて断面積の時間変化を表現し、破裂子音の/p/, /t/および/k/のような合成音を生成する際の、調音運動と声帯振動との相対的な時間関係について検討する。本研究では、これらの時間差を検討するためのGUIシステムを開発しシミュレーションを行う。また、その知見を用いて、音声持続時間の自動調整機能

の実現を検討する。

音声生成過程に基づく音声合成システムの応用展開に向けて、少ない手間で効率よく声道形状を生成する手法について検討する。本方式では声道形状を介して音声を生成するので、声道形状を表現する20個の音響管の断面積や長さを一度に設定できるアルゴリズムの検討を行う。日本語の典型的な母音空間を考慮し、5母音の配置を基準とした座標系を用いる。また、母音間の中間的な位置に対応する音響管に対しては、補間操作によりその形状の表現を検討し、音声生成時の特徴である声道形状の連続性を擬似的に模擬できるように配慮する。

上記で開発するインタフェースの応用展開を目指し、上肢運動の計測データの評価方法の検討や、視線検出インタフェースの開発と改良などを並行して行う。

## 4. 研究成果

本研究による主要な成果を以下に述べる。

(1) 図1にパラメータの自動設定機能を実行するGUI音声合成システムのスナップショットを示す。図の例では、/takaipasu/という合成音を生成した際の、ある音響管の断面積の時間変化(上段)、有声無声の音声生成に関与する声帯の声門中立面積の時間変化(中断)、生成された合成音声のサウンドスペクトログラム(下段)を示している。上段の各音素に対応した声道断面積の決定については、音響管の形状に対応した声道伝達特性を計算できるシミュレータを活用することで、妥当と思われるデータセットを構築した。/p, t, k/のような破裂音の生成では、知覚に影響を与えるエネルギー分布の再現が重要であり、シミュレータの活用により、そのような特性をもつ断面積関数の調整が可能となった。これらの断面積の時間的変化、すなわち調音運動に対応した動きと、声帯振動との相対的な時間関係については、GUIシステムでシミュレーションを行い、適切なパラメータ値の設定範囲が明らかとなり、これに基づいたパラメータ設定自動化機能を実現し、より容易で確実に子音の合成ができる音声合成システムの構築が可能となった。

(2) 本研究では、(1)で述べたような音声生成過程に基づく音声合成システム開発の他に、応用展開に向けて、少ない手間で効率よく声道形状を生成する手法について検討した。図2にインタフェースにおける母音配置とイメージを示すが、声道形状を表す音響管の形状を、ディスプレイ上のマウス操作で生成できるインタフェースとして実現している。このインタフェースは、マウス操作に連

動して音響管による声道形状が変化していく一種のマッピングインタフェースとなっており、簡単な操作で形状の構築が可能な特徴を有している。声道形状は20個の音響管から構成されているが、本インタフェースでは、2次元平面上の1点を指定するのみで、その点に対応した音響管の生成が可能な仕組みとなっている。また、その点を時間的に移動させれば、連続的な声道形状変化を表現することが可能であり、連続音の生成が可能となる。本インタフェースでは声道形状に基づいて音声生成されるので、その形状に応じて、音声の特徴である第1、第2、第3等のホルマントが自ずと定まる。音声における第1、第2ホルマントを独立に生成して音声生成を行う他の研究に比べて、声道形状が自然な拘束条件として働き、従属して定まるホルマント周波数をより自然な形で生成できる利点がある。

(3) (2)におけるインタフェースの特徴、すなわち、2次元平面上の一点を指定することで、その点に対応した音響管が生成できることを利用して、2次元軌跡の音声化など的一种のコンバータとしての応用展開が可能となった。応用例では、指鼻試験と呼ばれる運動失調症の診断基準により得られる生体の運動の計測データを例に音声化を試みた。図3に、指鼻試験における示指の運動軌跡(速度)にもとづく音声信号への変換例を示す。この図は、運動(速度)軌跡を、(2)のインタフェース上に投影し、それに基づいて生成される音響管を介して得られる合成音のサウンドスペクトログラムを示したものである。(a)が健常者、(b)が脊髄小脳変性症(SCD)患者のパターンにそれぞれ対応している。運動は5往復の運動のうち、中央の3往復分を投影したものとされている。/ieaou/の母音の並びで3往復分のサウンドスペクトログラムが得られているが、(a)の健常者の場合は、黒色で示されるホルマントが規則的に変化し、3往復とも類似したパターンとなっている。(b)に示すように、SCD患者のサウンドスペクトログラムにおいては、下から2番目の黒線で示される第2ホルマントの揺らぎが/u /→/ i /→/ e /→/ a /の区間に特に顕著に観測され、/ u /→/ i /においてはその持続時間も長い。また、往復に要する時間も一様ではなく、運動が滑らかに行われていないことを反映したものとされている。生成された合成音の試聴においても、揺らぎの違いが確認でき、コンバータとしての可能性を確認できた。

(4) 本研究ではもう一つの応用例として、視線を介した音声合成について取り扱った。視

線検出システムは肢体不自由者のパソコン操作など種々の応用が研究されている。(2)で述べた2次元平面上の一点を指定することで、その点に対応した音響管が生成できることを利用して、母音合成への応用を検討した。(2)で開発したインタフェース上の各頂点をターゲットとして、/aieuo/の並びで視線により操作した結果に基づく合成音声のサウンドスペクトログラムを図4に示す。各母音に対応したホルマントの時間遷移の様子から各母音の生成が確認できる。このシステムでは、画面上のある点で視線操作のマウスを停留させれば、その点に対応した音響管の形状を時間的に保つ操作となる。例えば、頂点(図2)における停留時間を長くすれば、ターゲットに対応した母音の定常部を長くする効果につながる。提示例では、頂点における停留時間を長くしているので、サウンドスペクトログラムに示すように、母音の定常部が明確に現れている。



図1 音声合成用 GUI システム

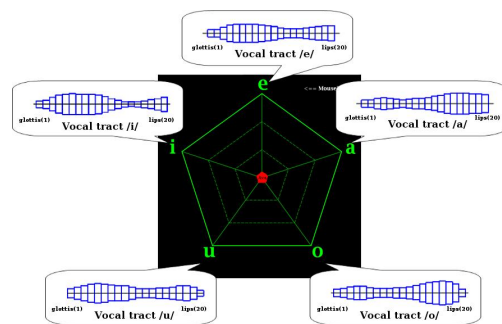
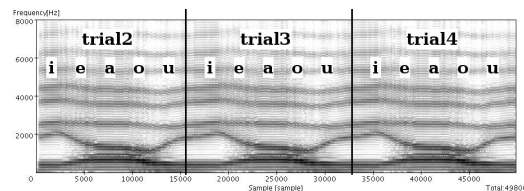
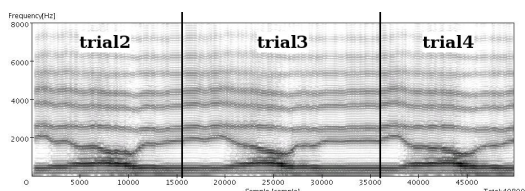


図2 声道形状マッピングインタフェースにおける母音配置とイメージ



(a) 健常者



(b) SCD患者

図3 指鼻試験における示指の運動軌跡(速度)に基づく音声信号への変換例

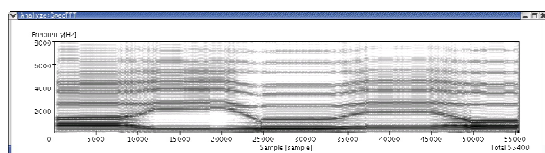


図4 視線操作により生成された合成音声 /aiueo/のサウンドスペクトログラム

このように本研究では、音声生成過程に基づく音声合成手法をシステムとして開発し、そのパラメータ設定の効率化を実現した。また、与える形状次第で合成音を生成できる融通性を踏まえ、2次元平面上の一点を指定することで、その点に対応した音響管が生成できる応用インタフェースを創出し、2次元軌跡の音声化など的一种のコンバータとしての応用展開が可能となった。今後も、引き続き応用開発を進める予定である。

## 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計7件)

① Kunihisa NAKASHIMA and Kohichi OGATA, "Representation of changes in the vocal tract shape by superposition model and speech synthesis," Proceedings of the Kyushu-Youngnam Joint Conference on Acoustics 2011, Kumamoto, Japan, 査読有, 2011, 161-164

② 掛谷拓史, 緒方公一, 中西亮二  
三次元運動計測システムによる指鼻試験の定量化-楕円フーリエ記述子による速度-距離特性パターンの解析-, 電子情報通信学会論文誌(D), 査読有, Vol. J93-D, No. 10, 2010, 1969-1976

③ Kohichi Ogata and Mamio Hokazono  
Development of speech synthesis simulation system and study of timing between articulation and vocal fold vibration for consonants /p/, /t/ and

/k/, Proceedings of 20th International Congress on Acoustics, Sydney, Australia, 査読有, Vol. CD-ROM, 2010

④ 米沢徹也, 緒方公一, 松本耕平, 平瀬 賢, 白谷和幸, 城戸大輔, 西村仁志  
視線インタフェースシステムのための虹彩中心検出誤差軽減手法の検討, 電気学会論文誌(C), 査読有, Vol. 130-C, No. 3, 2010, 442-449

⑤ 外園真実雄, 緒方公一  
音響管モデルに基づく音声合成-音声持続時間の調整機能の実現-, 電子情報通信学会技術研究報告, 査読無, SP2008-58, 2008, 13-18

[学会発表] (計21件)

① 緒方公一, 山下健太郎, 掛谷拓史, 平瀬賢, 中島邦久, 声道形状マッピングインタフェースのコンバータとしての応用, 日本音響学会2011年春季研究発表会, 2011年3月10日, 早稲田大学(東京)

② 中島邦久, 緒方公一, 重畳モデルによる声道形状変化の表現と音声合成実験, 日本音響学会2010年秋季研究発表会, 2010年9月16日, 関西大学(大阪)

③ 松村洗平, 緒方公一, 母音合成のための手の動きを利用した声道形状設定ツールの開発, FIT2010 第9回情報科学技術フォーラム, 2010年9月9日, 九州大学(福岡)

④ 松隈宣裕, 緒方公一, 母音合成のための声道形状マッピングインタフェースの開発-声道形状の設定と音響特性の検討-, 平成21年度電気関係学会九州支部連合大会, 2009年9月29日, 九州工業大学(福岡)

⑤ 松隈宣裕, 緒方公一, 音響管に基づく母音合成のための声道形状マッピングインタフェースの開発, 日本音響学会2009年秋季研究発表会, 2009年9月17日, 日本大学工学部(福島)

## 6. 研究組織

(1) 研究代表者

緒方 公一 (OGATA KOHICHI)

熊本大学・大学院自然科学研究科・准教授

研究者番号: 10264277