

科学研究費助成事業 研究成果報告書

平成 27 年 5 月 28 日現在

機関番号：17401

研究種目：基盤研究(C)

研究期間：2012～2014

課題番号：24560523

研究課題名(和文) マルチモーダルインタラクションに向けた音声画像融合計測インタフェースの応用開発

研究課題名(英文) Development of integrated speech and image measurement interfaces for multimodal interaction

研究代表者

緒方 公一 (OGATA, Kohichi)

熊本大学・自然科学研究科・准教授

研究者番号：10264277

交付決定額(研究期間全体)：(直接経費) 4,200,000円

研究成果の概要(和文)：本研究は、マルチモーダルインタラクションに向けた、音声画像融合計測インタフェースの開発を目的とするものである。音声生成過程に基づく音声合成手法と、視線検出や制御、モーションキャプチャなどの画像処理技術を融合し、マルチモーダルインタラクションのための種々のインタフェース開発を行った。声道音響管マッピングインタフェースが、手指の運動様態の巧緻性を判断するツールとなり得ることや、音声からの声道形状を計測(推定)する試みが有効であることが明らかになった。さらに、視線計測によるラジコンカー制御等の有効性も確認された。

研究成果の概要(英文)：The aim of this study is to develop synthesis and measurement interfaces for multimodal interaction. Speech synthesis technique based on speech production process and image processing technique such as eye-gaze estimation, eye-gaze-driven control, and motion capture were combined to create various interfaces which support multimodal interaction. For example, it is revealed that a vocal-tract mapping interface can be a useful measurement tool to judge the dexterity of the finger-nose movement. An attempt to estimate the vocal tract shape from vowel sounds via the vocal-tract mapping interface resulted in successful estimation of the vocal tract shape and visual trajectories on the mapping interface window. Moreover, an eye-gaze interface enabled us to control the motion of a radio-controlled car by eye-gaze.

研究分野：工学

キーワード：信号処理 音声合成 計測 インタフェース

1. 研究開始当初の背景

近年、人の動きをセンシングしたゲームなど、センサ融合応用計測技術の進展が目覚ましい。音声、姿勢情報、視線情報などを活用した、効果的なインタフェース技術融合は、視覚的情報を音で把握する、あるいは音を視覚的な情報に置き換えるなどのメディアの変換活用や、代替手段の提供等、利用者が主体的に外界(環境)とかかわる機会の提供を可能にするものとして、その可能性が期待される。このような背景を踏まえ、申請者が得意としてこれまで行ってきた音声生成過程に基づく音声合成技術と、3次元計測データ解析技術、視線検出インタフェースによるデータ解析技術等をさらに発展進化させ、マルチモーダルインタラクションに向けたインタフェースの開発を目指した。

2. 研究の目的

本研究は、申請者が得意としこれまで行ってきた生成過程に基づく音声合成技術と、3次元計測データ解析技術、視線検出インタフェースによる画像処理および制御技術等をさらに発展進化させ、制御、メディア変換や、補助システムなどの、マルチモーダル・インタフェースの開発を目的とするものである。

具体的には、

- ・ 上肢運動の画像計測データから得られる速度-距離特性パターンを、声道音響管マッピングインタフェースを用いて音声化する試みの有効性の定量評価
- ・ 声道音響管マッピングインタフェースによる母音生成の有効性と、軌跡情報を応用した実音声からの声道形状推定や話者識別の可能性の検討
- ・ 可視光方式による視線インタフェースの使いやすさの向上や視線応用による制御の実現

などを通して、

- ・ データの音声化による情報推定
- ・ 手指の動作等に基づく音声合成
- ・ 実音声からの声道形状推定と軌跡表現
- ・ 視線による音声合成や運動体の動作制御など、種々のメディアを融合したマルチモーダルな融合計測インタフェースを創出する。

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。

声道音響管マッピングインタフェースでは、上肢運動の計測データから得られる速度-距離特性パターンに基づいて、そのパターンを音声化する研究を更に進め、パターンを直接用いる場合、その時間差分に基づくパターンを用いた場合、パターンの真円からの差分を累積したパターンを用いる場合等について、それらを声道形状の表現に反映させた場合と声帯の緊張パラメータに反映させた場合とで、どの程度の判別が可能かを対比較法等を用いて検証する。

声道音響管マッピングインタフェースは、本来、インタフェース画面上でマウスを動かすことで音声を生成するものであるが、これを逆方向に利用し、音声からの声道形状の逆推定を試みる。インタフェース上の座標と声道形状、ホルムント周波数の関係の情報を蓄積し、これを逆方向に探索する手法を検討し、声道形状の計測(推定)システム、また、インタフェース画面上への音声情報の投影、すなわち視覚化の可能性について検討する。

視線インタフェースに関する研究では、これまで開発してきたシステムの実用性向上のための自動設定等について検討しつつ、視線によるマウスポインタの制御によるインタラクティブな応用の可能性について検討する。視線の軌跡を利用する方法、あるいは視線によるメニュー選択の方法などを実現するインタフェース開発を行い、それぞれによる制御実験を通して、有効性を検討する。

手指の動作に基づく音声合成を検討するため、データグローブを用いて声道形状を生成し、それに基づいて合成音を生成する。本研究で使用するデータグローブは、右手の5本指のまげ具合に応じて信号が出力される5チャンネル仕様のものであるが、20区間からなる声道音響管のどの部分にそれらに対応させるか検討する。まずは、音響管の長さ方向に制御点を設け各音響管の断面積を直接制御する方式について、生成される形状や合成される音声のホルムントの観点から検討を行う。

4. 研究成果

本研究による主要な成果を以下に述べる。

(1)マッピングインタフェースでは、その有効性を検証する研究を進め、ワンクリックで音響管を生成できる簡易さにもかかわらず、実際の調音でも観測される"pivot"の生成、典型的なホルムント空間の生成、マップ上でのマウス移動による連続母音生成、声道の形状変化の連続性に基づくパラメータの連続性の保証などの特徴や利点を有するインタフェースであることが明らかとなった。

萌芽的な応用展開として、上肢運動の計測データから得られる速度-距離特性パターンに基づいて、そのパターンを音声化する研究を更に進め、合成音声のひずみとして元の運動の巧緻性を判定できることを明らかにした。図1に、変換元の信号を声帯緊張パラメータに対応させた場合における一対比較法による実験結果を示す。図では3種の変換方法を示すが、いずれも大局的には横軸の評価値(ひずみの度合い)の増加とともに、尺度値が上昇することを示している。すなわち、運動の巧緻性の低下を音声として順位付けできることを意味している。種々の比較実験の結果、声道形状に投影した場合は、速度-距離特性パターンの真円からのずれの累積に基づく方法において、声帯緊張パラメータ

に投影した場合には、真円との差分に基づく方法において、それぞれ、速度-距離特性パターンの評価値と一対比較法での尺度値との相関が 0.98 の高い値を示し、合成音を介しての変換元の信号の様態の推定に有効な表現方法であることが明らかとなった。

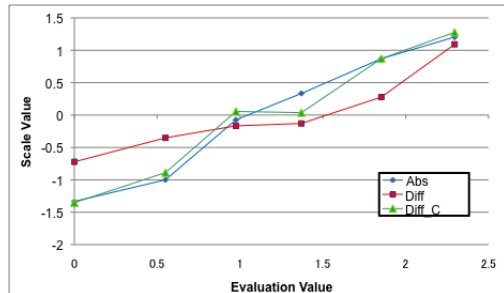


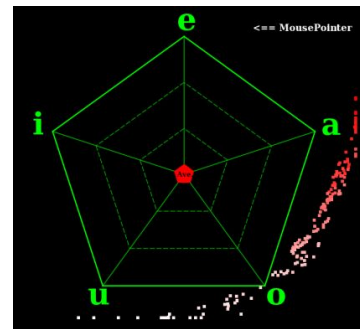
図 1 変換元の信号を声帯緊張パラメータに対応させた場合における一対比較法による実験結果

(2) インタフェース画面上でマウスを動かして音声を合成するマッピングインタフェースを、逆方向に利用して、音声からそれに対応する声道形状を推定する手法について検討した。

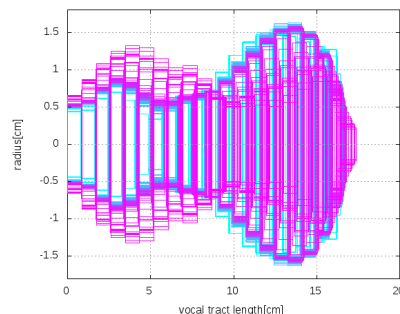
マッピングインタフェースにより、予め、インタフェース上の座標とホルムント周波数のデータセットを用意しておき、音声のホルムント周波数の情報から、Coarse-to-Fineの考え方をういて該当する声道形状を探索できることを明らかにした。図 2(a)に、母音/a/の発話状態から次第に口唇部を狭める発話動作についての逆推定結果を、インタフェース画面上の軌跡として示す。図では、時間の推移を表現するため、推定点を赤色から白色に変化させて表示している。また、声道形状表示機能を用いて得られた、逆推定時の声道形状の変化を図 2(b)に示す。図では、声道を 20 区間の等長音響管の接続で表現しており、時間の推移を表現するため、形状の描画色を青色から赤色に変化させて表示している。図 2(a)を見ると、頂点/a/付近から頂点/o/および/u/方向に推定結果が推移していることが分かり、実験時の発話動作が反映されていることが確認できる。このときの声道形状の変化を示す図 2(b)では、大きく口を開く/a/の調音の様子が形状(青色)に反映され、次第に口唇部をすぼめていく動作の結果、図右側の口唇部が右側に移動し声道長が長くなる様子が示されている(赤色)。すなわち、口唇部の突き出しが表現されており、母音生成の現象に符合した結果が確認でき、推定の妥当性が示唆される。その他、同一被験者の複数回の発話に対する逆推定結果の再現性、複数の被験者での推定結果の個人性など、興味深い結果も得られた。

音声信号がインタフェース画面上の点の軌跡として得られる特徴に基づき、音声を画

像として捉える、複合的、融合的インタフェースの可能性が見い出された。また、それらの軌跡は、マウスをどのように動かせば類似した音声が生産できるかを示す有益な経路情報としての意味を持ち、更なる応用展開の可能性が開けており、継続して研究を進めている。



(a)



(b)

図 2 母音/a/の発話状態から次第に口唇部を狭める発話動作についての逆推定結果,(a)インタフェース画面上の軌跡としての表現,(b)音響管としての表現

(3) 視線検出インタフェースの開発研究では、虹彩検出処理における閾値設定の自動化を目的として、彩度・明度のヒストグラムの特徴に着目した適切な閾値設定手法を提案し、その有用性の検討を行った。彩度に関しては、判別分析法による閾値設定が有効であること、その結果に基づいて明度の閾値設定が可能であることを明らかにし、自動設定の実現による実用性の向上が図られた。

応用開発として、視線インタフェースと声道音響管マッピングインタフェースとをネットワークで接続融合した視線計測による母音合成システムを実現した。視線操作は、被験者の習熟度や個人差により、その制御の巧緻性に違いがあるが、大局的な操作は可能であること、ユーザのスキルに応じて補助機能を有効化し、スキルの不足分を補完する方法が、インタフェースの利便性カスタマイズ性の向上に有効であることを確認した。

他の応用開発の例として、視線によるメニュー選択によりラジコンカーの動作を制御する試みも実施した。個人差はあるものの、

初心者でも，ラジコンカーの前進後退，右折左折などの基本動作の制御が可能なユーザも存在すること，熟練したユーザでは，ラジコンカーをコースに沿って動かすことが可能なことや，小型電動車いすの遠隔制御が可能なことが明らかとなった。

(4)データグローブを用いて声道音響管を生成し母音を合成する方式について研究を進めてきたが，5本の指で声道形状を生成する方法に加え，声道断面積を表すプレートパターンを3本の指の動作で改変して声道音響管を生成する方法を新たに提案した。少ない指の本数ながらも，5母音の生成が可能であり，連続母音の生成も可能であることを明らかにした。

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12件)

Kohichi Ogata and Shingo Niino, "Automatic threshold-setting method for iris detection for brown eyes in an eye-gaze interface system with a visible light camera," *Optics & Laser Technology*, 査読有, Vol. 66, pp.112-121, March 2015 (DOI: 10.1016/j.optlastec.2014.09.001)

Kohichi Ogata, Kensuke Sakamoto and Shingo Niino, "Home position recognition methods using polarizing films for an eye-gaze interface system," *Proceedings of the 11th International Conference on Signal Processing and Multimedia Applications (SIGMAP 2014)*, Vienna, Austria, 査読有, pp.275-280, August 2014.

Hiroki Shojaku, Shingo Niino and Kohichi Ogata, "A study of an eye-gaze interface system for radio-controlled cars," *Proceedings of 2013 ICT-PAMM Workshop on Mobility Assistance and Service Robotics*, Kumamoto, Japan, 査読有, pp.71-77, November 2013.

緒方公一, 山下健太郎, "ワンクリック声道音響管マッピングインタフェースとその信号変換への応用," *電子情報通信学会論文誌(A)*, 査読有, 第J96-A巻, 第8号, pp.529-540, 2013年8月

Daisuke Ito and Kohichi Ogata, "Development of a vocal tract design tool based on a growth curve of the vocal tract length," *Proceedings of Meetings on Acoustics (ICA 2013)*, Montreal, Canada, 060281_1.pdf, 査読有, June 2013. [<http://dx.doi.org/10.1121/1.4799101>]

Kohichi Ogata and Tomohiro Hayakawa, "Vowel synthesis using a vocal tract mapping interface and simulation study of inverse mapping," *Proceedings of Meetings on Acoustics (ICA 2013)*, Montreal, Canada, 査読有, 060167_1.pdf, June 2013. [<http://dx.doi.org/10.1121/1.4800751>]

Kohichi Ogata and Kohei Matsumoto, "Semi-dynamic calibration for eye gaze pointing system based on image processing," *Proceedings of International Conference on Signal Processing and Multimedia Applications (SIGMAP 2012)*, Rome, Italy, 査読有, pp.233-236, July 2012.

[学会発表](計14件)

児玉 揺, 緒方公一: 声道音響管マッピングインタフェースに基づく声道形状の解析, *日本音響学会2015年春季研究発表会講演論文集*, pp.299-300, 2015年3月. 発表日: 3月16日, 場所: 中央大学後楽園キャンパス(東京都)

井上衆人, 緒方公一: 視線操作による母音合成と補助機能の検討, *平成26年度電気・情報関係学会九州支部連合大会論文集*, 09-1A-02, p.63, 2014年9月. 発表日: 9月18日, 場所: 鹿児島大学工学部(鹿児島市)

伊藤大輔, 緒方公一: 成長曲線を考慮した声道音響管デザインツールのための声帯緊張パラメータと基本周波数の関係の検討, *日本音響学会2014年春季研究発表会講演論文集*, pp.301-302, 2014年3月. 発表日: 3月11日, 場所: 日本大学理工学部(東京都千代田区)

井上衆人, 伊藤拓哉, 正尺宏樹, 緒方公一: 視線操作による母音合成システムの開発, *平成25年度電気関係学会九州支部連合大会論文集*, 01-1A-08, 2013年9月. 発表日: 9月24日, 場所: 熊本大学(熊本市)

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