

科学研究費助成事業 研究成果報告書

平成 30 年 6 月 20 日現在

機関番号：34506

研究種目：基盤研究(B) (一般)

研究期間：2013～2017

課題番号：25280066

研究課題名(和文)平均声道モデルに基づく声質の多様性の身体的基盤に関する研究

研究課題名(英文) Study on physical basis of diversity of voice quality based on averaged vocal tract model

研究代表者

北村 達也 (KITAMURA, Tatsuya)

甲南大学・知能情報学部・教授

研究者番号：60293594

交付決定額(研究期間全体)：(直接経費) 13,000,000円

研究成果の概要(和文)：本研究では、話者の声質の多様性を生み出す身体的な基盤について、生体観測技術と数値音響解析技術を用いて検討を行い、以下の成果を得た。(1) 発話運動の形態と動態を3次元MRI同期撮像法により計測し、対応する音声と組み合わせ、データベースを構築した。(2) このデータベースに基づいて発話器官のモデルを開発した。(3) 声質への寄与が大きい鼻腔・副鼻腔の音響特性を解析した。(4) プロの物真似タレントの物真似発話時、および声優の発話時の声道形状を観測し、声質の多様性を生み出す原理を調査した。(5) 磁気センサシステムのセンサを改良し発話への影響を軽減するセンサを開発した。このセンサは市販されている。

研究成果の概要(英文)：This research project explored physical basis of diversity of voice quality using observation methods and computational acoustics that the project members have developed. This project achieved the followings; (1) The three-dimensional (3D) shape and movements of the speech organs were measured by an MRI scanner using a synchronized sampling method during continuous speaking and a database for 11 adults was constructed. (2) A tongue model was developed based on the database. (3) The acoustic characteristics of the nasal and paranasal cavities, which contribute largely to the voice quality, were investigated by an acoustic measurement method and a numerical acoustic analysis method. (4) The vocal tract of a professional voice impersonator and three voice actors were measured and analyzed. (5) The five-degrees-of-freedom sensors of an electromagnetic articulography system was improved to reduce interference in articulation. The proposed sensors have been made available commercially.

研究分野：音声科学

キーワード：MRI 声道 音声の個人性 鼻腔 副鼻腔 数値音響解析 音声科学 音響計測

1. 研究開始当初の背景

話者の声質(個人性)の多様性は我々のコミュニケーションを豊かにしている一方で、音声認識の精度向上を妨げる要因となっていた。そのため、この多様性が生じる原理の解明が必要である。話者の声質を決定づけるのは声道音響特性と声帯音源波の特性である。声道は咽頭腔、口腔、鼻腔などから成り、その形状の違いが声質に大きく寄与する。従って、声質の多様性の理解にとって声道形状の多様性とその音響的影響の把握が本質的な課題であった。この問題に対して、研究代表者らが蓄積してきた生体観測技術と数値音響解析技術を用いれば、話者の声質の身体的な基盤を解明できると考えた。

2. 研究の目的

(1) 3D-MRI・音声データベースの構築

音声の個人性が生み出される身体的な基盤を明らかにするためには、発話時の発話器官の形態・動態を観測し、その結果として出力される音声との対応を明らかにする必要がある。しかも、従来行われてきたような単母音などではなく連続音声を対象にしなければならない。また、世界的に見ると、南カリフォルニア大学のチームがMRI動画・音声データベースの構築を精力的に行っているが、彼らのMRI動画は技術的制約から2次元にとどまり、発話器官の形態・動態と音声との関係を調べるには限界がある。そこで、本研究では新たな複数断面MRI同期撮像法を用いて、3次元MRI動画を収集し、対応する音声と組み合わせたデータベースを構築する。

(2) 舌形状モデルの開発

平均声道モデルの作成にあたり、発話器官の中で最も変形が大きい舌を対象にしてモデリングを行った。弾性体であり形状変化の大きい舌形状のモデルが開発できれば、発話器官全体のモデリングが可能であると考えられるためである。

上述のように、発話時の調音動態の3次元形状は複数断面MRI同期撮像法により実現できるが、リアルタイムの観測はできない。その一方で、2次元の調音動態であれば、リアルタイムMRI、超音波断層装置、磁気センサシステムなどによりリアルタイムかつ高精度に観測可能である。しかも、後者の2手法はMRIと比べて低コストで大量のデータを得やすい。これらの手法で得られた2次元調音動態から3次元調音動態を推定することができれば有益である。そこで、本研究では3次元MRIデータに基づいて正中面の舌断面形状から外側の舌断面形状を推定する手法を開発する。これが実現できれば、超音波診断装置や磁気センサシステムの2次元データから3次元発話器官の形状を推定することが可能となる。

(3) 鼻腔・副鼻腔の音響特性の解析

音声は口腔と鼻孔から放射されている。

我々はこれらを分離した音声を用いて聴取実験を行い、鼻腔・副鼻腔が音声の個人性に大きく寄与することを示した(Aminoら, 2014)。しかし、従来の研究では鼻腔・副鼻腔を単純なモデルで表現するにとどまっている。そこで、本研究では、3次元X線CTデータから抽出した詳細な鼻腔・副鼻腔形状に基づいてその音響特性を明らかにする。

(4) 声質制御メカニズムの調査

物真似タレントや声優は、複数の話者の音声に良く似た声質を出し分けることができる。この技巧を解明することができれば、発話器官の形状と声質との関係や個人性知覚に関する有益な知見が得られる。このような背景の下、近年、海外にて物真似発話の研究が増えている。しかし、物真似音声の分析が行われている一方で、その生成過程を観測した研究はほとんどない。そこで、本研究では物真似タレントの物真似発話時の声道形状をMRIにて観測した。

(5) 磁気センサシステムのセンサの改良

磁気センサシステムは舌や口唇などの数点に小型(3mm×3mm×3mm)の有線センサを貼り、磁気を利用してその位置をリアルタイムで追跡する装置である。観測点数が限られるものの、MRIよりも低コストで多数の被験者のデータ収集に適している。しかし、このシステムのセンサのケーブルが硬く、被験者の発話を阻害することが多々あることが問題となっていた。そこで、本研究では、計測精度やセンサの耐水性・耐久性を維持しつつ発話への影響を低減するセンサを開発する。

3. 研究の方法

(1) 3D-MRI・音声データベースの構築

本研究ではMasakiら(1999)により開発されたMRI同期撮像法を改良した手法にて撮像を行った。この手法は、ストロボ撮影の原理を応用し、時間分解能の低いMRIにて高フレームレートかつ高画質の動画を得る手法である。本研究ではさらに複数断面の撮像を実現し、発話時の3次元MRI動画を得た。

撮像はATR-Promotions脳活動イメージングセンサに設置されたSiemens社製MAGNETOM Verioを利用した。主要な撮像パラメータの一例を示す。スライス厚3mm、スライス間隔3mm、エコー時間1.62ms、繰り返し時間33mm、フリップ角15度、ピクセルサイズ1mm×1mm、対象領域256mm×256mm、平均回数1回。

撮像後にMRI装置内で同じ姿勢で発話させた音声データの収集も行った。MRI撮像中は大きな騒音が発生するためである。この収録にはMRI用の光マイクロフォン(北村ら, 2006)を用いた。

なお、この実験は「甲南大学におけるヒトを対象にした研究審査」により承認されている。

(2) 舌形状モデルの開発

上記のMRI-音声データベースに収録された男性1名のMRIデータを用いた。発話資料は連続母音/aieuo/, /oeuia/である。このMRIデータには正中矢状面を含む5枚の矢状面が含まれているが、本研究では正中面 (S_m)、その左外側の断面 (S_{m+1})、さらにその左外側の断面 (S_{m+2}) の計3断面を使用した。

このデータに対し、まず各断面の舌形状を目視により決定し、その境界を舌断面形状とした。次に、舌断面形状を200点にて等間隔に再サンプリングし、平滑化を施した。以下では、/aieuo/系列、/oEUia/系列の舌断面形状セットを D_t 、 D_e と呼ぶ（それぞれモデルの学習用、評価用データに対応する）

次に、顔画像の主成分分析(永田ら, 1996)と同様の方法により、舌断面形状を表現するパラメータを削減した。この処理により、ある舌断面形状は、平均形状、固有形状ベクトル、記述パラメータで表される。

これらのパラメータを3断面 (S_m, S_{m+1}, S_{m+2}) について求めた。そして、 S_m の記述パラメータおよび舌断面形状の重心の加速度を入力として与えると外側の断面 (S_{m+1}, S_{m+2}) の舌断面形状を推定する線形モデルを作成した。

(3) 鼻腔・副鼻腔の音響特性の解析

3次元X線CTデータから鼻腔・副鼻腔領域を抽出し、コンピューターモデルを作成した。このデータはノイズやムラが多かったが、本研究では機械学習を利用し、鼻腔・副鼻腔領域の抽出を効率的に行う手法を開発した。

得られたコンピューターモデルから3Dプリンタを用いて精密な実体模型を製作した。この模型の声門部から白色雑音を入力し、鼻腔・副鼻腔の音響特性を計測した。実験の様子を図1に示す。



図1: 鼻腔・副鼻腔実体模型の音響特性の計測の様子

その一方で、同じコンピューターモデルに対して時間領域差分法(FDTD法)による数値音響解析を実施し、音響計測の結果と比較した。さらに、副鼻腔(左右の上顎洞、前頭洞、篩骨

洞)を除去したモデルを対象に解析を行い、それぞれの副鼻腔の音響的働きを調査した。

(4) 声質制御メカニズムの調査

実験協力者はプロの物真似タレント1名(男性)である。実験協力者はMRI装置内にて物真似をしていない状態、および目標話者12名(うち1名は女性)の物真似をした状態にて母音/a/, /e/を持続発話した。発話中に正中矢状面のMRIデータを計測し、図2に示す5種の距離を計測した。

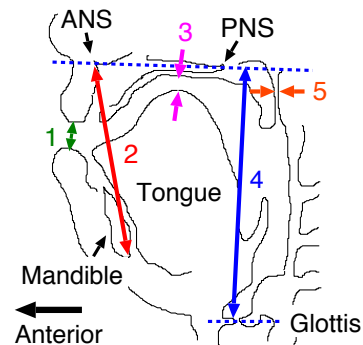


図2: 物真似タレントの正中矢状面MRIデータ上にて計測した距離(5種)

(5) 磁気センサシステムのセンサの改良

磁気センサのケーブルを細く(直径0.1mm)、柔軟性の高いものと交換した。センサから約10mmの部分を持ち、新しいケーブルをハンダ付けした。ケーブルを擦る(ツイスト)ことによってノイズを抑制している。

この改良による計測精度への影響を調べるため、クランク-ロッカー機構によりセンサを複数の速度で往復運動させ、センサ間の距離の計測誤差とデータ欠損率を調査した。なお、計測用の磁場に影響を与えないため、クランク-ロッカー機構ではプラスチック製のアームを使用した。

加えて、実際に被験者4名を対象に発話器官(舌、歯茎、口唇)にセンサを貼り付け、各種の発話タスク実行時のデータ欠損率を調査した。

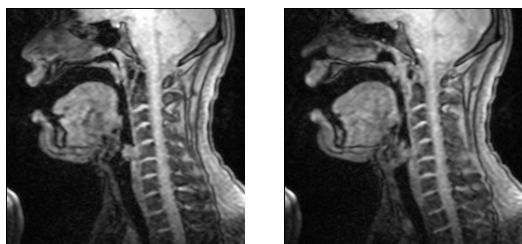
4. 研究成果

(1) 3D-MRI・音声データベースの構築

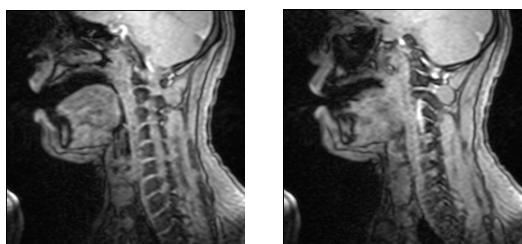
本研究では男性5名、女性6名の3次元MRI動画およびそれに対応する音声データを収集しデータベース化した。世界的に見てもこの種のデータが一定人数分収集されたことはこれまでになく、本研究の大きな成果である。得られたデータのうち、男性1名の複数断面のMRIデータを図3に示す。MRI同期撮像を採用しているため、画質が非常に高い。

これらのデータは下記の舌形状モデルの開発に利用された。さらに、今後の発話器官全体のモデルの開発にも利用される。

また、得られたデータのうち実験参加者から許諾を得られているものについては、学術研究目的での利用に限り、要望に応じて外部に提供する。



(a) 正中矢状面 (b) 正中面の 1 枚外側



(c) 正中面の 2 枚外側 (d) 正中面の 3 枚外側

図 3 : 3 次元 MRI データの例.

(2) 舌形状モデルの開発

学習用データ、評価用データに対する舌断面形状の推定結果の例を図 4 に示す。前者ではほぼ正確に推定できている。後者でも概形は一致しているものの、咽頭部、下顎部、舌尖部などに局所的に約 3 mm の誤差が見られる。

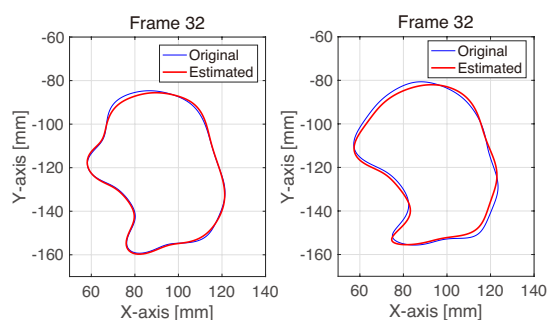
次に、学習用データ、評価用データに対する面積一致率を図 5 に示す。学習用データ、評価用データに対してそれぞれ約 95%、90% の高い精度で推定できている。

以上により、本研究で開発した舌形状モデルは有望な手法であることを確認できた。学習用データの追加や統計モデルの導入により性能がさらに向上すると見込まれるので、引き続き検討を行う。その際、推定誤差による音声への影響を考慮した評価法も検討する。

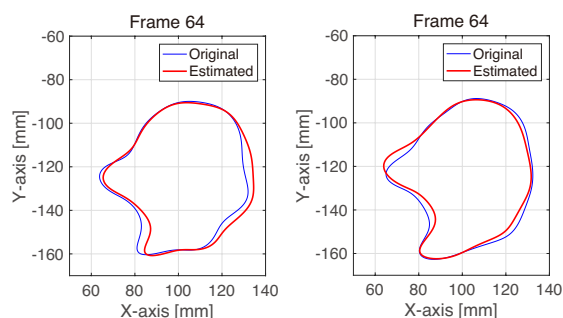
(3) 鼻腔・副鼻腔の音響特性の解析

音響計測および数値音響解析による鼻腔・副鼻腔の音響特性を図 6 に示す。これらはおおむね一致しているため、2 つの手法は信頼できると判断できる。

また、副鼻腔を 1 つずつ取り除いたコンピュータモデルの数値音響解析により、各副鼻腔の音響的働きを明らかにした。例えば、左の上顎洞を取り除くと 1.55 kHz と 3.95 kHz のピークが消失し、1.50 kHz と 3.80 kHz に極零対が出現した。先行研究では副鼻腔は分岐管として考えられてきたが、本研究により共鳴腔としても働くことが明らかになった。



(a) 断面 S_{m+1} (b) 断面 S_{m+2}



(c) 断面 S_{m+1} (d) 断面 S_{m+2}

図 4 : 舌断面形状の推定結果の例。(青線)原形状、(赤線)推定結果。(a)、(b)学習用データ、(c)、(d)評価用データ。断面 S_{m+1} 、 S_{m+2} はそれぞれ正中矢状断面の 1 枚外側、2 枚外側の断面を表す。

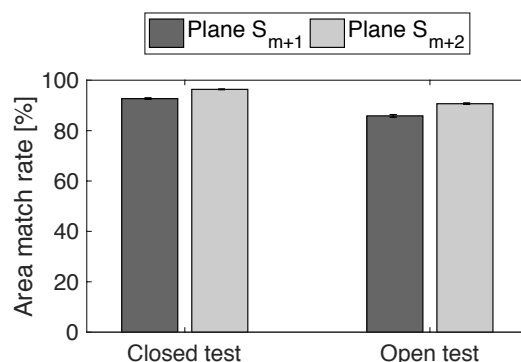


図 5 : 舌断面形状モデルの精度。(左)学習用データ (Closed test)、(右)評価用データ (Open test) に対する結果。

本研究では昭和大学の協力を得て 11 名の鼻腔・副鼻腔形状データを抽出できた。今後はこのデータを用いて鼻腔・副鼻腔の形状の個人差およびそれが音声に及ぼす影響に関する検討を続ける。

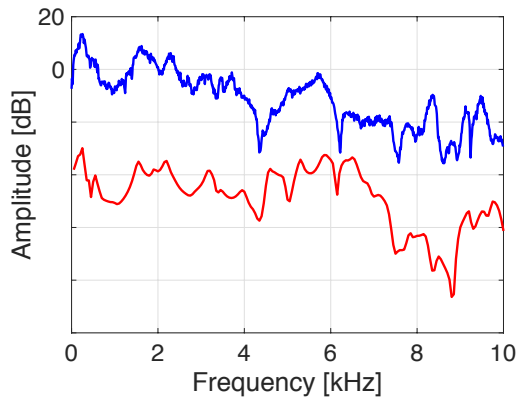


図6: 鼻腔・副鼻腔の音響特性. (青)音響計測による結果, (赤) FDTD 法による結果.

(4) 声質制御メカニズムの調査

得られた MRI データの一部を図 7 に示す. 実験により, 物真似時には舌を多彩に変形していることがわかった. 図 7(b)では喉頭を約 20 mm 下降させ, 声道長を長くし, 低い声を作り出している. 図 7(c)の舌形状は極めて特徴的で, 一般的な発話にはほとんど現れない. /e/発話のためのせばめを作りつつ口腔後方に大きな空間を作り出している. この物真似タレントが独自に会得したテクニックといえる.

距離 1~5 (図 2 参照) の計測結果を表 1 に示す. 最も顕著な変化を示したのは距離 4 (喉頭の高さに対応) であり, 母音/a/, /e/とも約 40 mm の範囲で上下している. これは, 音声の個人性にとって声道長や基本周波数が重要な情報であることと対応している.



(a) 地声 (b) 目標話者 1 (c) 目標話者 2

図 7: 物真似タレントの発話時 (地声および物真似発話) の正中矢状面 MRI データ

表 1: 正中矢状面上の各部の距離 [mm] (距離に関しては図 2 参照)

| 距離 | 地声 | | 物真似 | |
|----|-----|-----|--------|--------|
| | /a/ | /e/ | /a/ | /e/ |
| 1 | 13 | 11 | 8-36 | 9-32 |
| 2 | 88 | 91 | 90-101 | 92-108 |
| 3 | 18 | 7 | 8-23 | 4-16 |
| 4 | 105 | 98 | 77-116 | 76-118 |
| 5 | 0 | 0 | 0-5 | 0-3 |

(5) 磁気センサシステムのセンサの改良

図 8 にオリジナルのセンサと改良型センサの写真を示す. センサは共通でケーブルのみが異なっている. ケーブルが柔軟になったことにより, 被験者の発話を阻害しなくなり, よりリアルな発話運動を計測できるようになった. また, ケーブルの取り回しが良くなったため, センサを被験者の舌などに貼り付ける作業が容易になった.

改良型センサではさらにケーブルと計測装置本体との接続を Molex コネクタで簡単, 確実に行えるよう改良した. これによって実験中にケーブルが外れることもなくなり, 実験の効率が大幅に向上した.

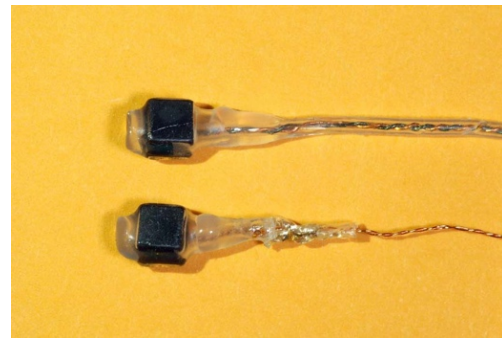


図 8: (上) オリジナルのセンサ, (下) 本研究にて開発した改良型センサ

クランク-ロッカー機構を用いた精度評価の結果, オリジナルのセンサに対して改良型センサは約 2 倍の精度を示した. データ欠損率も改良型センサの方が低かった. これは, 細かいケーブルを擦ったことにより, ノイズ耐性が高くなったためと考えられる.

その一方で, 被験者による発話タスク実行時のデータ欠損率は, 改良型センサの方が高くなった. とはいえ, データ欠損率は 2%未満であり, その影響は小さいと見込まれる.

この改良型センサはすでに市販され, 複数の研究者により利用されている.

5. 主な発表論文等

[雑誌論文] (計 4 件)

- ① T. Kitamura, Y. Nota, M. Hashi, H. Hatano, Replacement of sensor cables for reducing effects on articulation in the Northern Digital Incorporated's Wave electromagnetic articulography system, JASA Express Letters 143, EL154 (2018).
- ② 北村達也, 中間隆正, 大村宙, 川元広樹, ATR 音声データベース内の文音声における知覚的話者間類似度の計測, 日本音響学会誌, 71(10), 516-525 (2015).
- ③ T. Kitamura, Y. Nota, M. Hashi, H. Hatano, Improvement of five-degree-of-freedom sensors for Northern Digital Incorporated's Wave speech research

system, *Acoustical Science and Technology*, 36(4), 347—350 (2015).

- ④ 北村達也, 磁気センサシステムによる調音運動のリアルタイム観測, *日本音響学会誌*, 71(10), 526—531 (2015).

[学会発表] (計 3 1 件)

- ① H. Takemoto, P. Mokhtari, T. Kitamura, (招待講演) Complementary relationship between 1D and 3D acoustic analyses of the vocal tract, *JAIST World Conference 2018* (2018/02/27-28) 開催地: 石川県能美市.
- ② H. Takemoto, T. Kitamura, P. Mokhtari, T. Yamaguchi, K. Maki, (招待講演) Three-dimensional acoustic analysis of the nasal and paranasal cavities, *International Speech Production Seminar 2017* (2017/10/16-19) 開催地: 中国・天津.
- ③ 北村達也, 蒔苗久則, 伊藤仁, 連続母音発話 MRI データに基づく舌断面形状モデルの評価, *日本音響学会秋季研究発表会* (2017/09/25-27) 開催地: 愛媛県松山市.
- ④ T. Kitamura, H. Takemoto, H. Makinae, T. Yamaguchi, K. Maki, Acoustic analysis of detailed three-dimensional shape of the human nasal cavity and paranasal sinuses, *Interspeech 2017* (2017/08/20-24) 開催地: スウェーデン・ストックホルム.
- ⑤ 北村達也, 蒔苗久則, 伊藤仁, 舌断面形状のモデリングに関する予備的検討, *日本音響学会春季研究発表会* (2016/03/09-11) 開催地: 神奈川県横浜市.
- ⑥ K. Amino, H. Makinae, T. Kitamura, Nasality in speech and its contribution to speaker individuality, *Interspeech 2014* (2014/09/ 14-18) 開催地: シンガポール.
- ⑦ 川岸容子, 北村達也, MRI による物真似発話時の声道形状観測, *日本音響学会春季研究発表会* (2015/03/16-18) 開催地: 東京都文京区.

[図書] (計 2 件)

- ① 北村達也, 音声生成, *人工知能学大辞典*, 773—775, 共立出版 (2017).
- ② 北村達也, 音声生成過程の可視化技術, *五感インタフェース技術と製品開発 事例集*, 267—271, 技術情報協会 (2016).

[その他]

ホームページ等

(1) 甲南大学知能情報学部北村研究室:
<http://www.konan-u.ac.jp/hp/kitlab/>

(2) 甲南大学の研究力 2016 年度版「ハナス」を「ハカル」:
<http://www.konan-u.ac.jp/front/research/year/year2016/>

6. 研究組織

(1) 研究代表者

北村 達也 (KITAMURA Tatsuya)
甲南大学・知能情報学部・教授
研究者番号: 60293594

(2) 研究分担者

伊藤 仁 (ITO Masashi)
東北工業大学・工学部・講師
研究者番号: 70446962

齋藤 毅 (SAITO Takeshi)
金沢大学・電子情報学系・助教
研究者番号: 70446962

蒔苗 久則 (MAKINAE Hisanori)
科学警察研究所・法科学第四部・主任研究官
研究者番号: 20415441

網野 加苗 (AMINO Kanae)
科学警察研究所・法科学第四部・研究員
研究者番号: 70630698

(3) 連携研究者

竹本 浩典 (TAKEMOTO Hironori)
千葉工業大学・先進工学部・教授
研究者番号: 40374102