

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成 25 年 5 月 23 日現在

機関番号：34506

研究種目：基盤研究(B)

研究期間：2009～2012

課題番号：21300071

研究課題名（和文） 発話動態観測に基づく連続音声生成メカニズムに関する研究

研究課題名（英文） A study on speech production mechanisms on the basis of measuring articulatory movement

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

研究成果の概要（和文）：本研究課題では、連続音声の発話運動の観測法と子音の生成メカニズムに関して研究を行った。子音の発話運動を観測するために高い時間分解能をもつ 3 次元 MRI 動画撮像法を開発したり、子音生成時の声門開閉の様子を非侵襲的に観測するために非侵襲光電グロトグラフ装置を開発したりするなど、発話運動観測のための様々な新技術を開発した。さらに、得られたデータに基づいて、発話運動の解析や声道音響特性のシミュレーションを行った。

研究成果の概要（英文）： This research project studied measuring methods for articulatory movement of continuous speech and mechanisms of production of consonants. We developed new methods for measuring articulatory gesture and open/close of the glottis during production of consonants; for example, an ultra-high-speed MRI movie generation method and external lighting and sensing photoelectroglottography. We then analyzed articulatory movement and calculated the vocal-tract acoustic characteristics on the basis of the obtained data.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2009年度	2,800,000	840,000	3,640,000
2010年度	5,100,000	1,530,000	6,630,000
2011年度	4,600,000	1,380,000	5,980,000
2012年度	1,000,000	300,000	1,300,000
年度			
総計	13,500,000	4,050,000	17,550,000

研究分野：総合領域

科研費の分科・細目：情報学，知覚情報処理・知能ロボティクス

キーワード：音声情報処理，音声生成

1. 研究開始当初の背景

研究開始当初の時点では、時間分解能が高い 3 次元 MRI 動画撮像法が存在しなかった。MRI 同期撮像法という技術があったものの、フレームレートが 100 fps 未満で、子音の発話運動観測には不十分であった。また、声帯振動における声門開口面積を非侵襲的に計測する手法もなかった。そのため、複数の発

話器官が動的に協調することによって生み出される子音の生成メカニズムには不明な点が多かった。

2. 研究の目的

本研究では、高い時間分解能をもつ 3 次元 MRI 動画撮像法を開発し、得られたデータに基づいて連続音声の生成メカニズムの解明

を試みる。

具体的には、以下の検討を行った。(1) 高フレームレート MRI 動画撮像法の開発、(2) 非侵襲光電グロトグラフ装置の開発、(3) 発話時の皮膚振動解析、(4) 磁気センサシステムの改良、(5) 調音運動の計測・解析、(6) 声道音響解析。以下、それぞれの方法と得られた成果について概説する。

3. 研究の方法

(1) 高フレームレート MRI 動画撮像法の開発

従来法で得られる動画は、空間分解能は高いものの、そのフレームレートは破裂子音などの発話運動の観測には不十分であった。そこで、本研究では、Dynamic MRI によって得られた動画データと磁気センサシステム (EMA) によって得られた高時間分解能の発話運動データを融合して、250 Hz の高フレームレート MRI 動画を得る技術 (スパース法) を開発した。

(2) 非侵襲光電グロトグラフ装置の開発

光電グロトグラフ (PGG) は、内視鏡を用いて喉頭を照明し、声門からの光の通過量を測定することによって声帯振動を観測するシステムである。この方式は内視鏡を使うため侵襲的であり、医師による操作が必要であるなどの問題があった。そこで、体外から下咽頭の皮膚を通して声門に光を入れ、声門を通過した光を声門下の皮膚上で検出する、非侵襲光電グロトグラフ装置 (ePGG) を開発した (図 1)。

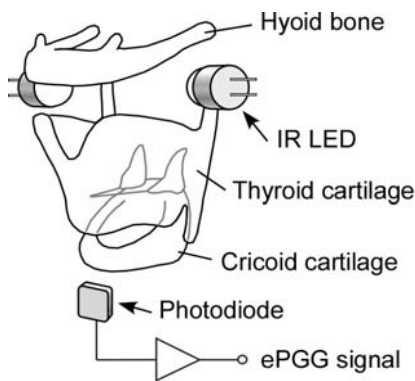


図 1 : ePGG の概念図

(3) 発話時の皮膚振動解析

音声は口や鼻孔のみならず体表面の振動を介しても空中に放射されている。そこで、スキャニング型レーザードップラ振動計を用いて発話時の顔の皮膚振動速度を計測した。この装置は、ドップラ効果を利用して対象物の振動速度を得ることができる。本研究では母音/a/と鼻子音/N/を対象にして計測を行った。



図 2 : NDI 社 Wave を用いた実験の様子

(4) 磁気センサシステムの改良

Northern Digital Inc. 社の磁気センサシステム Wave (図 2) は、磁気を用いて舌や顎に貼り付けたセンサの 3 次元位置を追跡するシステムである。観測点は限られるものの、空間分解能と時間分解能が共に高く、動きの速い子音発話運動の観測に適している。しかし、センサに接続されているツイストペアワイヤが太くて柔軟性に欠けるため発話を阻害するという問題があった。そこで、そのワイヤを細く柔軟性のあるものに交換するという改良を施した。

(5) 調音運動の計測・解析

① 子音調音運動の計測

閉鎖子音は、閉鎖の形成と開放、声帯振動開始などの様々な事象が動的に協調することによって生成される。そのため、その生成メカニズムには不明な点が多い。そこで、閉鎖子音の有声音と無声音の調音運動の違いに着目して、子音の調音運動の過渡的な特性を調べた。関西方言の成人話者 4 名を対象にして MRI 同期撮像法により /akise/, /agise/ 発話時の調音運動を計測した。

② 子音生成時の声道容積変化の計測

有声音と無声音の生成における差異は音源のみにとどまらず、調音動態にも及ぶ。その一例として、有声音生成時に声帯振動を維持するために咽頭腔の容積が拡大するという報告がある。しかし、従来は声道容積の変化を直接計測した研究がなかった。そこで、本研究では、成人男性 1 名を対象として MRI 同期撮像法を用いて有声音、無声音の破裂子音および摩擦子音の調音動態を観測した。そして、これらの子音が生成される際の声道容積変化を比較した。

③ 英語弱母音発話時の調音運動の解析

英語は、強勢を有する母音と有さない母音で音色が大きく異なる。強勢の間隔によるリズムとそれに伴って生じる弱音節の曖昧化は、英語発音の最も重要な特徴とされている。そこで、MRI 動画と X 線マイクロビームにより得られたデータを用いて、日英母語話者に

よる英語弱化石音の調音運動を解析した。

④ 声道長と音声の各種特徴量との関連

ホルマント周波数や基本周波数は声道長と相関が高いと言われてきた。しかし、音声と声道長が同時計測されたことはなく、この相関が実際に計算されたことはなかった。そこで、MRI を利用して成人男性 15 名の母音発話時の音声と声道長を同時計測し、ホルマント周波数、基本周波数、声道長の中の相関を求めた。

(6) 声道音響解析

① 声道断面積関数抽出ソフトウェアの開発

MRI は音声生成研究において重要なツールの 1 つである。しかし、それを用いた研究を支えるソフトウェアの整備は立ち遅れている。そこで、3 次元 MRI データから声道断面積関数を抽出するソフトウェアを開発した。本研究では、オープンソースでパブリックドメインの画像処理ソフトウェア ImageJ のプラグインとして声道断面積関数抽出プラグインを開発した。ImageJ は Java で開発されており、プラグインやマクロを追加できる仕様になっている。

② FDTD 法による声道音響解析

声道は複雑な形状をしており、複数の分岐管をもつ。従来は円筒管を連結した音響管で声道をモデル化していたが、声道内の音響現象を精密に理解するためには、3 次元声道形状に基づく音響解析が不可欠である。そこで、3 次元 MRI により得られた声道形状 (図 3) を対象にして時間領域差分法 (FDTD 法) を用いた解析を行った。さらに、得られた結果を声道の精密な 3 次元モデルの音響特性と比較した。

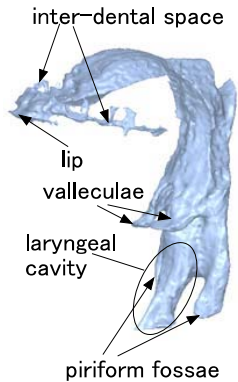


図 3 : 3 次元声道形状

③ FDTD 法の並列計算の試み

昨今、複数のコアが搭載された CPU や GPU を用いた並列演算プログラムの開発環境が整備されてきた。そこで、3 次元 FDTD 法を対象にして、CPU を用いた並列演算と GPU を用いた並列演算による計算時間短縮効果を比

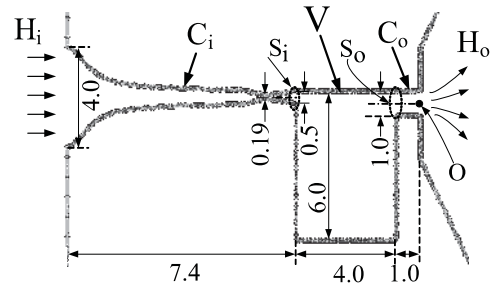


図 4 : 無声歯茎音 [s] の 2 次元領域モデル

較した。

CPU を用いた並列演算は、OpenMP を用いて実現した。一方、GPU を用いた並列演算は NVIDIA 社の CUDA ver. 3.1 を用いて実現した。利用した GPU は NVIDIA Tesla C1060 である。ATR 母音発話 MRI データベースから抽出した 3 次元声道形状を対象にして、FDTD 法による音響解析に要する時間を求めた。

④ 摩擦子音生成の音響数値解析

ナビエ-ストークス方程式の数値計算によって無声歯茎音 [s] の生成を試みた。X 線画像や MRI 画像に基づき、[s] 生成時の声道形状を単純化した 2 次元形状 (図 4) を計算領域とした。このモデルは、従来摩擦子音生成に不可欠と考えられてきた歯列を含まない。

4. 研究成果

(1) 高フレームレート MRI 動画撮像法の開発

本研究で開発したスパース法においては、まず Non-negative temporal decomposition (NTD) によって磁気センサシステムのデータ (250 Hz) を分析する。そして、その時間構造情報を用いて、Dynamic MRI の動画 (76 Hz) を 250 Hz にアップサンプリングする。

Dynamic MRI とスパース法により得られた画像を比較したところ、提案法により得られた画像の方が、母音 /u/ の口唇の突き出し、子音 /b/ の口唇の閉鎖の特徴を良く捉えられることが確認できた。

(2) 非侵襲光電グロトグラフ装置の開発

ePGG 装置により記録される信号の性質を確認するため、PGG 信号および内視鏡画像と比較した。男性話者が「それを軽々 (けーけー) という」と発話したデータを対象とした。

PGG 信号と ePGG 信号は、吸気時に大きな声門開大が記録されること、無声子音 /k/ 生成時の声門開大が単語中の第 2 子音で小さいことなどの点において共通の特徴を示した。

本研究で開発した ePGG システムは、その有用性が認められ、(有) 追坂電子から販売されるに至っている。

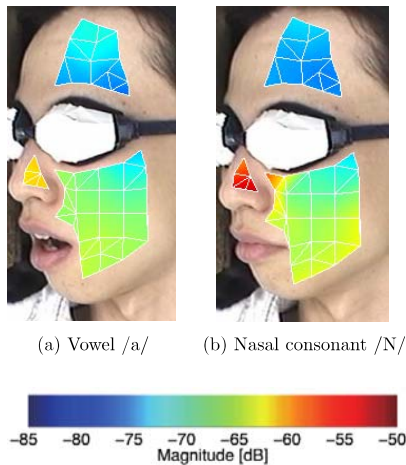


図 5: 母音/a/ (左), 鼻音/N/ (右) 発話時の皮膚振動速度パターン

(3) 発話時の皮膚振動解析

母音/a/, 鼻音/N/発話時の皮膚振動速度パターンを図 5に示す. このような計測結果を示したのは本研究が世界で初めてである.

母音と鼻音の皮膚振動速度パターンには明確な差異が見られる. 前者では口唇の周囲における値が他の部分と比較して大きい. 一方, 後者では鼻の周囲における値が他の部分と比較して大きい. これは, 鼻音生成時に鼻腔や副鼻腔が共鳴するためと考えられる.

一般に, 母音生成時には鼻腔内での共鳴は生じないと考えられてきた. しかし, 図 5 (左) から, 母音生成時にも鼻腔内共鳴が存在することが示された.

(4) 磁気センサシステムの改良

センサの改良によって, センサを舌に貼り付けたときに被験者が感じる違和感と発話運動への影響が大幅に減少した. 改良前のセンサを用いた場合, 口の中に異物があるような音声になり, 自然に話せるようになるまでに時間を要した. しかし, 改良後のセンサを用いると, センサを貼り付けた直後から普段通りの発話が可能であった.

また, センサワイヤの変更による計測精度への影響がないことを確認した. このセンサは(株)ATR-Promotions より販売されるに至っている.

(5) 調音運動の計測・解析

① 子音調音運動の計測

MRI動画を解析して得られた舌の最高点の高さを図 6に示す. この結果から, 舌が口蓋に達する時間が/k/の調音において/g/よりも先行することがわかった. 時間差の平均は 32 msであった.

無声閉鎖子音では閉鎖時の口腔内圧力が有声音よりも高くなることが知られている. このことが, この実験において/k/における

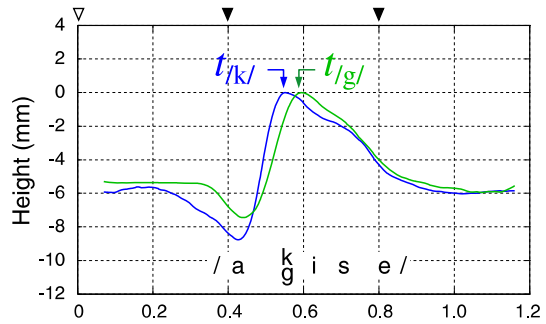


図 6: 舌の最高点の高さの時間変化 ($t_{/k/}$: /akise/, $t_{/g/}$: /agise/)

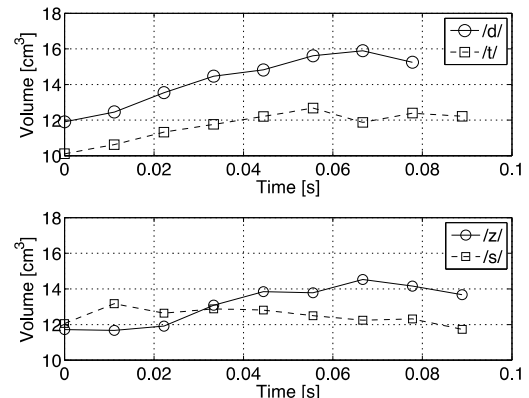


図 7: 有声, 無声の破裂子音および摩擦子音発話時の声道右半側容積の時間変化 (時刻の原点は舌が口蓋に接触もしくは最接近した時点)

舌運動の先行が観察された理由であると考えられる.

② 子音生成時の声道容積変化の計測

3次元MRIデータから抽出した声道右半側の容積変化を図 7に示す. 有声破裂子音/d/生成時の声道容積はほぼ単調に増加し, その値は常に無声破裂子音/t/よりも大きかった. また, 有声破裂子音/d/と有声摩擦子音/z/の生成時の声道容積を比較すると, 前者の方が最大値と増加幅が大きいという結果が得られた.

③ 英語弱母音発話時の調音運動

ここではX線マイクロビームデータの解析結果についてのみ記す. 日英母語話者による英語母音発話時の舌形状をそれぞれ, 図 8図 9に示す. 図の右側が前方で, T1~T4 が舌, ULおよびLLが上唇と下唇に貼り付けたマーカーの位置に対応している.

英語母語話者では T2 を頂点として舌全体が咽頭方向に傾いているのに対して, 日本語母語話者では T2 と T3 が盛り上がり, 全体が扁平である. このように2つの話者群間で舌形状が大きく異なることが明らかになった. この結果は, 語学教育分野にも音声生成の知見が有用であることを示したものである.

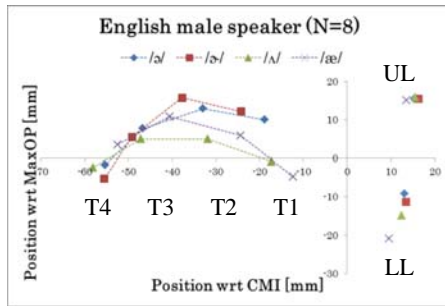


図 8：英語母語話者（男性）の英語母音発話時の舌形状。T1～T4 は舌上のマーカー（T1 が前方），UL と LL はそれぞれ上唇と下唇のマーカーを表している。

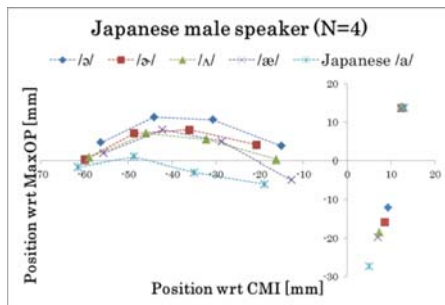


図 9：日本語母語話者（男性）の英語母音発話時の舌形状。マーカー位置は図 8 と同様である。

④ 声道長と音声の各種特徴量との関連

分析の結果，第 1，第 2，第 3，第 4 ホルマント周波数と声道長の中に相関がみられないことが明らかになった。また，基本周波数との相関もみられなかった。これらの結果は，従来の定説とは異なるもので，引き続き検討が必要である。

(6) 声道音響解析

① 声道断面積関数抽出ソフトウェアの開発
声道断面積関数抽出法は様々なものが提案されているが，我々は Takemoto ら (2006) が提案した手法を採用した。

この方法では，まず，3次元 MRI データから正中矢状断面を決定し，その断面上で声門の位置を指定する。次に，声道領域を対象にして，声門からの距離を表すコンター・マップ（等高線図）を作成する。求めたコンター・マップにおいて同じ値を持つコンターごとに重心を計算し，それらの点からスプライン曲線を求めることによって声道中心線を決定する。その声道中心線に直交する線を声道中心線に沿って等間隔に求める。最後に，その線を含み矢状断面に直交する断面を抽出し，声道領域の面積を求める。

以上の手順を ImageJ のプラグインとして実装し，我々の Web ページ上で公開した。

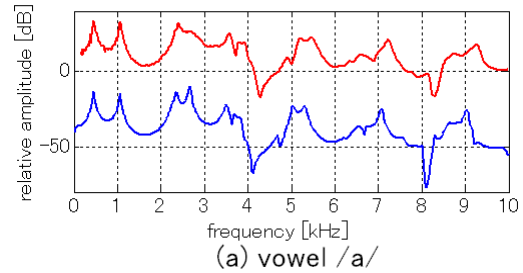


図 10：実測（赤）および FDTD 法（青）により求めた母音/a/の声道伝達関数

② FDTD 法による声道音響解析

FDTD法により求めた母音/a/の声道伝達関数と対応する声道模型の音響特性の実測値を 図 10 に示す。計算値と実測値における極と零の周波数が良く対応している。特に，9 kHz までの対応は良好である。この結果から，本研究で開発した音響シミュレータは，声道内部で生じている音響現象を忠実にシミュレート可能であるといえる。このシミュレータを用いれば，極と零の成因を明らかにすることができる。

③ GPGPU による FDTD 法の並列計算

総ボクセル数 1,859,000 の計算領域を逐次計算と 2 つの並列化手法 OpenMP, CUDA により計算した。その結果，OpenMP は逐次計算の 2.6 倍，CUDA は逐次計算の 35.0 倍の速度向上を達成した。CUDA 実行時には高いメモリバンド幅を引き出すことができた。本研究によって，コストパフォーマンスの高い計算時間短縮法を示すことができた。

④ 摩擦子音生成の音響数値解析

4 重極音源に 5 kHz のハイパスフィルタ処理を施した変量の強度分布を分析したところ，長方形の領域（図 4 における V）の流出開口付近で強度が著しく増加していた。また，数値解析により得られたスペクトルは実際の音声の特性を有していた。この結果は，摩擦子音生成に歯列が必須ではないことを示している。

5. 主な発表論文等

〔雑誌論文〕（計 8 件）

- ① **Tatsuya Kitamura**, Measurement of vibration velocity pattern of facial surface during phonation using scanning vibrometer, *Acoustical Science and Technology*, 33, 2, 126-128 (2012). (査読有)
DOI:
<http://dx.doi.org/10.1250/ast.33.126>

- ② Hironori Takemoto, Parham Mokhtari, Tatsuya Kitamura, Acoustic analysis of the vocal tract during vowel production by finite-difference time-domain method, *Journal of the acoustical society of America*, 128, 6, 3724-3738 (2010). (査読有)
DOI:
<http://dx.doi.org/10.1121/1.3502470>
- ③ Kiyoshi Honda, Tatsuya Kitamura, Hironori Takemoto, Seiji Adachi, Parham Mokhtari, Sayoko Takano, Yukiko Nota, Hiroyuki Hirata, Yasuhiro Shimada, Ichiro Fujimoto, Shinobu Masaki, Satoru Fujita, Jinwau Dang, Visualisation of hypopharyngeal cavities and vocal-tract acoustic modelling, *Computer Methods in Biomechanics & Biomedical Engineering*, 13(4), 443-453 (2010). (査読有)
<http://www.tandfonline.com/doi/abs/10.1080/10255842.2010.490528>

[学会発表] (計 36 件)

- ① 波多野博顕, 北村達也, 林良子, ドナ・エリクソン, 調音と音響の観測に基づく英語曖昧母音の分析, 音声学会第 326 回研究例会, 国立情報学研究所 (2013/3/23)
- ② 北村達也, スキャニング型レーザー Doppler 振動計による発話時の皮膚振動計測, 日本音響学会研究発表会講演論文集, 2012, 1, 329-33, 神奈川大学 (2012/3/13).
- ③ Sadao Hiroya, Tatsuya Kitamura, Generation of a vocal-tract MRI movie based on sparse sampling, *Proceedings of International Seminar of Speech Production 2011, Montreal, Canada* (2011/6/20).

[図書] (計 1 件)

- ① 鏑木時彦, 正木信夫, 元木邦俊, 松崎博季, 北村達也, 音声生成の計算モデルと可視化, コロナ社, 203-252 (2010).

[その他]

- ① ホームページ等
<http://basil.is.konan-u.ac.jp/>
- ② アウトリーチ活動
(1) 本研究課題に関連した「ひらめき☆ときめきサイエンス」を実施予定 (2013 年 10 月 5 日)
(2) 発話観測システム NDI Wave のセミ

ナーを開催予定 (2013 年 7 月 6 日)
(3) 神戸市立青少年科学館にて「おもしろ科学館クラブ」を担当 (2012 年 6 月 23 日)

- ③ 研究成果の公開
(1) (有)追坂電子が ePGG システムを販売
(2) (株)ATR-Promotions が ATR 声道 MRI データを公開
(3) 声道断面積関数抽出用 ImageJ プラグインを公開
- ④ 報道関連
(1) 京阪神エルマガジン社「Meets Regional」にて研究紹介 (2013 年 3 月号)
(2) ABC テレビ「キャスト」にて解説 (2013 年 1 月 9 日)
(3) 読売新聞にて解説 (2012 年 12 月 11 日)
(4) 読売テレビ「ZIP!」にて解説 (2012 年 11 月 7 日)
(5) TBS ラジオ「久米宏 ラジオなんですけど」にて解説 (2011 年 7 月 23 日)

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

なし

(3) 連携研究者

本多 清志 (HONDA KIYOSHI)
天津大学・計算科学技術学院・教授
研究者番号: 90395088
正木 信夫 (MASAKI SHINOBU)
(株)ATR-Promotions・代表取締役社長
研究者番号: 40181630
竹本 浩典 (TAKEMOTO HIRONORI)
(独)情報通信研究機構・ユニバーサルコミュニケーション研究所・研究員
研究者番号: 40374102
モクタリ パーハム (MOKHTARI PARHAM)
(独)情報通信研究機構・ユニバーサルコミュニケーション研究所・研究員
研究者番号: 00395089

(4) 研究協力者

田口 友康 (TAGUCHI TOMOYASU)
田口計算解析研究所
橋 亮輔 (TACHIBANA O. RYOSUKE)
現東京大学・研究員
波多野 博顕 (HATANO HIROAKI)
現神戸大学・研究員