

科学研究費助成事業 研究成果報告書

令和元年6月7日現在

機関番号：17102

研究種目：基盤研究(C) (一般)

研究期間：2016～2018

課題番号：16K00242

研究課題名(和文) 発話運動の磁気センシングに基づく調音・音響データベースの構築

研究課題名(英文) Construction of articulatory-acoustic database based on electro-magnetic sensing of articulatory motion

研究代表者

鍋木 時彦 (Kaburagi, Tokihiko)

九州大学・芸術工学研究院・教授

研究者番号：30325568

交付決定額(研究期間全体)：(直接経費) 3,300,000円

研究成果の概要(和文)：本研究では、音声発話時の舌、唇、下顎などの調音運動を音声とともに測定し、日本語の調音・音声データベースを構築することを目的として、調音器官の表面に装着した位置マーカ一用の受信コイルの3次元位置を高精度に測定する磁気センサシステムを設計・構築した。本システムは従来システムより多い8個の送信チャンネルを有し、さらにこれらの送信コイルの配置を最適化することで高い測定精度を得た。構築した測定システムと、独自に作成した発話文セットを用いて、1時間におよぶ日本語としては最大規模のデータ収録を行うことができた。得られた調音・音声データより調音運動と音声スペクトルの相互変換を行い、その有効性を検証した。

研究成果の学術的意義や社会的意義

音声の発話において、舌、唇、下顎などの調音器官はさまざまな役割を担っている。母音においては、調音運動は声道の形状を調整し、その音響特性や母音の識別に重要なホルマントを定める。子音では、摩擦性や破裂性の音源信号の生成に関わる。このように、人の音声コミュニケーションを理解し、その知見に立脚した音声情報技術を構築するには、調音運動の測定という困難な問題を解決することが不可欠である。本研究では、調音運動を3次元的に測定可能な磁気センサシステムを開発し、音声と調音運動の同時測定を可能にした。さらに、声帯振動を測定する電気式喉頭計を併用することで、音声発話の全体的なプロセスを調べることが可能になった。

研究成果の概要(英文)：This study aimed at observing movements of the articulatory organs, i.e., tongue, lips, and jaw, simultaneously with speech and at constructing a large set of articulatory and speech dataset of Japanese. To this end, we constructed an electro-magnetic position sensing system for tracking the three-dimensional position of multiple marker coils attached to the articulatory organs. Our system had eight transmitting channels from which alternating magnetic fields were generated. We obtained a fine measurement accuracy by optimizing the alignment of each transmitting coil. With this position sensing system, we collected a large amount of articulatory data simultaneously with speech from a speaker of Japanese. Here, we used a new utterance list and then the size of the dataset was about one hour. We also examined the relationship between articulatory movements and speech spectrum, revealing the effectiveness of the dataset for investigating articulatory-based speech information processing.

研究分野：音声情報処理

キーワード：音声発話 調音運動 磁気センサ 運動計測 音声情報処理

様式 C - 19、F - 19 - 1、Z - 19、CK - 19 (共通)

1. 研究開始当初の背景

(1) 近年、MOCHA EMA コーパスや USC-TIMIT コーパスなど、音声と音声器官の発話運動を同時計測した、マルチモーダルな英語音声データベースが整備されてきている。これらのコーパスは、調音・音響マッピングをはじめ、音声からの発話運動の逆推定、発話運動をパラメータに取り入れた声質変換処理、サイレント音声認識、バイオフィードバックを用いた第二言語の発話学習などに応用され、有効性が確かめられている。他方、日本語音声に関しては、調音・音響データベースは整備されておらず、研究を進める上での障壁となっている。その理由としては、発話運動の測定には、磁気センサや MRI (磁気共鳴画像法) などの特殊な装置が必要で、これらの装置を使いこなして高品質な発話データを取得するには、実験上のノウハウを積み重ねることが必要なためと考えられる。

(2) 発話運動の測定手段としては、MRI でも毎秒 10 フレーム程度のリアルタイム測定が可能となってきているが、過渡的な音声現象をとらえるには不十分である。それに対して、磁気センサは毎秒 100~200 フレームの測定が可能であり、破裂子音などの高速な運動に対しても、十分な時間分解能を有している。また、MRI では測定時にノイズが発生するため、クリーンな音声収録が困難だが、磁気センサでは、音声はもとより、電気式喉頭計 (EGG) による声帯振動の同時測定も可能であり、音声の生成過程を総合的に観測できる。

(3) 本課題の代表者は、科研費 (挑戦的萌芽研究、課題番号 25540071、H25~H27 年度) において、発話運動を測定するための 3 次元磁気センサの開発に取り組んだ。この磁気センサでは、送信コイルから生成される交流磁界を位置マーカー用のソレノイドコイルで受信することで、舌・口唇・下顎などの音声器官に装着したマーカーの 3 次元位置と傾きを推定する。検討の結果、3 次元磁気センサにおいては、『位置推定における解の不確定性』および『位置推定精度の角度依存性』の 2 点が根本的な問題であることがわかり、高精度な測定システムを構築する上で、これらの問題を解決することが課題となっている。

2. 研究の目的

(1) 磁気センサでは、位置マーカーの持つ 5 自由度のパラメータ (位置 3 個、傾き 2 個) の値を決定するため、磁界を生成する送信コイルを 5 個以上使用し、それらに対応する複数の磁界方程式を連立させる。『位置推定における解の不確定性』とは、受信信号からマーカーの位置を求める際に、パラメータの多自由度性が主要因となって、複数の解が同一の受信信号を生成し得る状況が生じ、解を一意に決定できなくなる問題である。一方、磁界から受信コイルに誘導される信号は、コイルの位置と傾きに依存するが、コイルの平行移動よりも回転の方が顕著に受信信号を変化させる。『位置推定精度の角度依存性』とは、受信信号に対する位置パラメータの感度が相対的に低下してしまい、受信コイルの角度変化が大きくなると、位置推定精度の劣化を生じる問題である。本課題では、3 次元磁気センサにおける『位置推定における解の不確定性』および『位置推定精度の角度依存性』の問題を解決し、高精度な発話運動の測定システムを構築することを目的とする。

3. 研究の方法

(1) これらの問題に対して、本課題では、まず、受信コイルの角度によらず、十分な強度を持った磁界成分を受信できるようにするため、従来よりも送信コイルの個数を増やしてシステムを設計する。磁界を生成する送信コイルの配置については、これまでの検討からも、測定精度に大きく影響することが明らかになっている。ここでは、送信コイルから生成される交流磁界を、近接場の磁気双極子を用いてモデル化し、この磁場シミュレーションから予測されるマーカー位置の推定結果を基として、推定誤差が最小となる、最適な送信コイルの配置を決定する。一方、送信コイルの個数を増やすことは、マーカー位置の推定における磁界方程式の個数をも増やすことになる。マーカーの位置や傾きは、これらの磁界方程式を連立させることで推定されるため、より多数の拘束式を用いることで解を安定化させ、受信コイルの角度変化に依存しない、高精度な測定システムとする。さらに、構築された磁気センサを用いて、舌・口唇・下顎の発話運動、音声、および電気式喉頭計による声帯振動の同時観測システムを実現し、

MOCHA EMA 英語音声コーパスに匹敵するような、日本語音声の調音・音響データベースを作成する。得られた調音・音響データを利用して、声道の形状を定める調音器官の運動と音声スペクトルの間の写像関係を検討し、データの有効性を検証する。

4. 研究成果

(1) 本研究では、まず、送信コイルの生成する磁界を磁気双極子モデルを用いて表現し、磁場シミュレーションに基づいて送信コイルの最適な配置方法について検討した。この磁場シミュレーションは、送信コイルを6個有する従来のシステムについて位置の推定精度が向上するなどの有効性が確認されている。ここでは、『位置推定における解の不確定性』および『位置推定精度の角度依存性』の根本的な問題を解決することを想定して、送信コイルの個数を8個に増やしてシミュレーションを行った。送信コイルを設置する位置については、現在と同様に被験者を取り囲む立方体の頂点に配置する方法や、この立方体を上下方向に縮小してコンパクトにする方法、さらに同一平面上に配置する方法などを比較した。この配置は、実際に測定装置を製作する際の作りやすさに関わるため重要である。このように送信コイルを設置する位置を複数のパターンで仮定した上で、さらに個々のコイルの向きを最適化を行った。最適化の基準としては、受信コイルの位置推定精度を用いた。なお、磁気双極子モデルや、受信信号からのマーカー位置推定法は、実際の測定システムにおいて使用されるものと同じである。8個のコイルの向きを最適化には、単純な組み合わせ最適化を用いた。結果的に、同一平面上にすべての送信コイルを配置しても、十分な精度が得られるとの知見を得た。この結果をもとにして、実際に稼働するハードウェアシステムを構築した。

(2) 下図は、磁気センサシステムの構成を表している。まず、計算機で作成した駆動用の正弦信号を D/A 変換器から出力し、増幅した後に送信用ソレノイドコイルに送り、話者の頭部周辺に交流磁界を生成する。マーカー位置推定におけるパラメータの自由度の観点から、送信コイルは5個以上が必要になる。個々の送信用コイルに対しては、異なる周波数の正弦信号を用いる。受信側では、小型・軽量の受信コイル（直径、幅ともに2~3 mm程度）をマーカーとして音声器官に装着し、磁界から誘導される受信信号を増幅した後に、A/D変換器で計算機に取り込む。主要な音声器官の運動をカバーするため、12個の受信コイルを同時に使用可能とする。次に、受信信号をそれぞれの送信コイルに割り当てられた周波数成分に分離して、受信コイルの位置を推定する。このマーカー位置の推定では、送信コイルが生成する交流磁界を近接場の磁気双極子によりモデル化し、磁界モデルからの予測値と実際に測定される信号値の間の自乗誤差が最小となるように、パラメータ（受信コイルの3次元位置と傾き）の値を推定する。このとき、送信コイルの個数分の非線形な磁界方程式を連立させることで、位置や傾きが計算される。

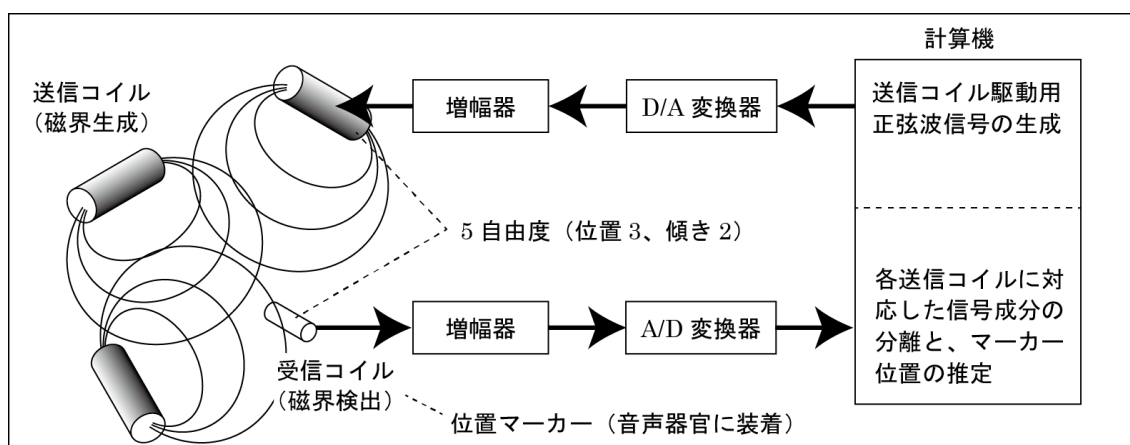


図1：磁気センサシステムの基本的な構成

(3) 構築された磁気センサシステムを用いて、調音器官の発話運動、音声、および電気式喉頭計による声帯振動の同時観測実験を行った。位置マーカー用の受信コイルは、上下の唇に各1個、下顎に1個、舌面に3個、さらに、頭部位置のリファレンスとして3個を被験者に装着し

た。音声信号、電気式喉頭計の信号は、受信信号と一緒にAD変換したため、サンプリング周波数は50 kHzとなった。受信信号からの位置推定により、各マーカーの位置情報は250 Hzのサンプリング周波数で計算された。発話用の文章セットとしては、当初、日本語音素の出現頻度が考慮された文章セットであるATR503文を用いた。しかしながら、ATR503文で収録されるデータの量は約20分となり、MOCHA EMAコーパスのような英語コーパスには量的に及ばないことがわかった。そこで、日本語wikiペディアから3組音素の組み合わせで出現頻度を考慮する形でテキストの抽出を行い、発話用の文章セットを1298個追加した。なお、この中には文章のほかに文節の形の発話文が多く含まれている。磁気センサによる長時間の測定実験は被験者への負担が非常に大きく、また測定の途中で装着した受信コイルが脱落することもある。複数回の測定実験を重ね、男性の成人話者1名について、約1時間と日本語音声としては最大規模の調音・音声データの取得に成功した。

(4) 最後に、取得した調音・音声データを用いて、声道形状を制御する調音運動と音声のスペクトル特性の間の写像関係を調査し、データセットの有効性を検証した。これは、調音運動から音声スペクトルを予測する順問題と、音声スペクトルから調音運動を予測する逆問題とに分けられる。音声スペクトルは声道の音響特性を反映したものであり、この特性は声道の形状によって定まる。一方、磁気センサで得られるデータは複数のマーカー位置のみであり、声道全体の形状を復元することは困難である。そこで、ここでは、調音・音声データを学習データとして用いて、ディープニューラルネットワークのフレームワークによって順問題、逆問題それぞれの写像関数を求め、十分な写像性能が得られることを確認した。

5. 主な発表論文等

[雑誌論文](計2件)

- 1 Tokihiko Kaburagi, Momoyo Ando, and Yasufumi Uezu, Source-filter interaction in phonation: A study using vocal-tract data of a soprano singer, *Acoustical Science and Technology* (Accepted for publication)
- 2 Hidetsugu Uchida, Kohei Wakamiya, and Tokihiko Kaburagi, Improvement of measurement accuracy for the three-dimensional electromagnetic articulograph by optimizing the alignment of the transmitter coils, *Acoustical Science and Technology*, Vol. 37, No. 3, 2016, pp. 106-114
[doi:10.1250/ast.37.106]

[学会発表](計10件)

- 1 田口史朗、鎗木時彦、End-to-End 調音-音声変換に関する検討、日本音響学会音声研究会、2019
- 2 Tokihiko Kaburagi, Momoyo Ando, and Yasufumi Uezu, Source-filter interaction: A study using vocal-tract data of a soprano singer, 175th Meeting of the Acoustical Society of America, 2018
- 3 Fumiaki Taguchi and Tokihiko Kaburagi, Articulatory-to-speech conversion using bi-directional long short-term memory, *Interspeech2018*, 2018
- 4 田口史朗、鎗木時彦、畳み込みニューラルネットワークを用いた音声-調音逆マッピングの検討、日本音響学会秋季研究発表会、2018
- 5 田口史朗、鎗木時彦、深層学習による調音-音声変換に関する検討、日本音響学会音声研究会、2018
- 6 Yasufumi Uezu and Tokihiko Kaburagi, A simulation study on the effect of glottal boundary conditions on vocal tract formants, *Interspeech2017*, 2017
- 7 田口史朗、鎗木時彦、DNNを用いた調音-音響マッピングの構築に関する検討、日本音響学会音声研究会、2017
- 8 Tokihiko Kaburagi, Synergistic Interactions Underlying the Production of Human Voice (invited), Youngnam-Kyushu Joint Conference on Acoustics 2017, 2017
- 9 Kohei Wakamiya, Hidetsugu Uchida, and Tokihiko Kaburagi, An investigation of the alignment of the transmitter coils in the three-dimensional electromagnetic articulography having eight transmission channels, Youngnam-Kyushu Joint Conference on Acoustics 2017, 2017

10 若宮幸平、内田秀継、鎗木時彦、8個の送信チャネルを持つ磁気センサシステムの検討、日本音響学会秋季研究発表会, 2016

〔図書〕(計0件)

〔産業財産権〕

出願状況(計0件)

名称:

発明者:

権利者:

種類:

番号:

出願年:

国内外の別:

取得状況(計0件)

名称:

発明者:

権利者:

種類:

番号:

取得年:

国内外の別:

〔その他〕

ホームページ等

6. 研究組織

(1)研究分担者

なし。

(2)研究協力者

なし。

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。