

科学研究費助成事業 研究成果報告書

平成 27 年 5 月 26 日現在

機関番号：17102

研究種目：基盤研究(B)

研究期間：2011～2014

課題番号：23300071

研究課題名(和文) 3次元非線形音声生成モデルの構築と病的音声の解析

研究課題名(英文) Construction of three-dimensional nonlinear speech production model and analysis of pathological voice

研究代表者

鈴木 時彦 (KABURAGI, Tokihiko)

九州大学・芸術工学研究科(研究院)・准教授

研究者番号：30325568

交付決定額(研究期間全体)：(直接経費) 13,700,000円

研究成果の概要(和文)：声帯音源の生成機構と声道フィルタの間の物理的な相互作用を考慮した、非線形音声生成モデルの構築と、病理発声において典型的に表れる不規則発声のメカニズム解明を目的として、研究を実施した。その結果、(1)病理発声における左右の声帯振動の非対称性を表す非線形ダイナミックモデルの構築、(2)左右の声帯形状の非対称性に対応した新しい呼気流解析法の開発、(3)不規則な声帯振動を生じる物理相互作用の実測、(4)二つの内視鏡カメラとステレオマッチング法による声帯振動の3次元動態の復元法の構築、(5)声道の3次元音響解析法の高度化とその声道形状との対応関係の解明を研究成果として得た。

研究成果の概要(英文)：This research was conducted to construct a nonlinear speech production model that takes into account the interaction between the voice production system and vocal-tract filter and also for the analysis of pathological voice. We produced the following results. (1) Asymmetrical behavior of the left and right vocal folds was modeled using a nonlinear dynamical system. (2) A novel method was developed to analyze airflow passing through an asymmetrical glottis. (3) Irregular vocal fold behaviors caused by the nonlinear interaction between the voice production system and vocal-tract filter were measured for human subjects. (4) Three-dimensional motion of the vocal folds was investigated using two endoscopy devices and a stereo-matching method. (5) We developed novel methods of analyzing the acoustic characteristics of the vocal tract. In addition, we investigated the relationship between the acoustic characteristics of the vocal tract and its fine morphological structure.

研究分野：音声情報処理

キーワード：音声生成 病的発声 声帯音源 声道音響解析 音源-フィルタ相互作用 非線形モデル

1. 研究開始当初の背景

(1) 音声生成のモデルは、音源機構と声道の音響フィルタの独立性を仮定した音源-フィルタ理論が G. Fant により提唱されて以来、着実な進展を遂げてきた。声道については、MRI (磁気共鳴画像法) による 3 次元形状観測、有限要素法による音響解析が大きな発展を遂げた。声帯音源についても、呼気流を粘性解析することで、声帯表面に生じる境界層や流れの剥離を精緻に調べることが可能となった。

(2) これらの結果、声帯の自励振動を生じる上で本質的な、呼気の圧力と声帯形状の依存性が明らかにされた。声帯音源と声道の依存性 (音源-フィルタ相互作用) にも注意が向けられ、声道の音響負荷によって、声帯に働く音響的駆動力や声門体積流の時間波形に大きな変化が生じることがわかった。このような物理事象間の相互依存性は、音声の生成に非線形性をもたらす。研究代表者の籙木は、呼気流の境界層解析のもとで音源-フィルタ相互作用を定式化し、通常発声における声帯音源のメカニズムを詳細に検討した。

(3) 一方、嘔声 (させい) は、病的音声の中でも特に声帯音源に関連したものであり、声帯疾患、神経障害のほか、職業的な声帯酷使や加齢でも生じる。嘔声では、左右の声帯が非対称に振動して声門の開閉が不規則になり、二つの異なる基本周波数を有した振動パターンを示して、二重音声を生じることもある。声帯の過緊張や低緊張による痙攣性発声障害が原因となっており、声帯音源を生成できず、日常会話が困難となるケースも多い。

2. 研究の目的

(1) 本課題では、音声生成モデルによる発声シミュレーションを通じ、嘔声における声帯の不規則振動、二重音声など、特異的な声帯現象が生じる要因を解明する。ここでは特に、音声の生成における呼気-声帯間、ならびに音源-フィルタ間の相互依存関係に着目し、それらが声帯駆動力の符号、強弱、時間パターンに与える影響を調べる。

(2) また一方、病的音声を定量的に解析するには、通常発声に比べて声帯の挙動が複雑になるため、声帯をはじめ、呼気流、声道の音響解析を 3 次元的に行った上でそれらを統合し、物理的な相互作用の影響を組み入れた上で、精緻な音声生成モデルを構築することが課題となる。声帯や声道については、従来法である有限要素法の解析精度を維持しつつ、時間領域の発声シミュレーションや並列計算による高速化が可能な、より実効性の高い解析モデルを構築する。声帯音源の実質を担う呼気流解析は、境界層解析では 2 次元が限度であり、3 次元化のためのパラダイムシフトが必要となっている。

(3) 本課題では、音声生成の最新の知見に基づいて、最先端の解析アルゴリズムや超並列計算を導入し、病的音声解析するため、音声科学、音響工学、計算機工学、病的音声の分野で顕著な業績を有する研究者の連携のもとに、この問題解決にあたる。

3. 研究の方法

(1) 病的音声で特徴的に生じる特異な声帯振動に対応するため、呼気の流体力によって変形運動する声帯の弾性体モデルと、声門流路において噴流ジェットを形成する呼気の粘性流解析を実現し、声帯音源モデルを構築する。声道の音響負荷の特性を考慮し、10 kHz 程度までの広帯域に亘る声道の 3 次元音響解析を行い、種々の相互作用を組み入れた非線形音声生成モデルを構築する。音声の生成に及ぼす各解析の効果や相互作用の影響を、音声器官の調音的特徴との関係も含めて詳細に検討し、音声生成モデルの有効性を定量的に明らかにする。

(2) 音声生成過程のシミュレーションにより、不規則発声における生理物理的な音声生成メカニズムを解明する。病的音声の生理物理的要因を特定するため、声帯の形状、質量、スティフネスなどの生理的特徴と、呼気-声帯間、音源-フィルタ間の相互作用に着目し、それらが声帯振動の機序に与える影響を調べ、正常発声と病的音声における差異を明らかにする。

4. 研究成果

(1) 病理発声における声帯の不規則振動の解析のため、ハイスピードカメラで撮影されたビデオデータ (図1) よりキモグラムや高速グロットグラムなどを作成し (図2)、振動様態の可視化を行った。その結果、左右の声帯が同位相、逆位相の振動を繰り返す、一種のモードスイッチング現象が観察された。この特異的な振動現象を理解するため、非線形振動子の結合系における「2つの周期解のスイッチング現象」のもとに、不規則発声における声帯のダイナミクスを検討した。

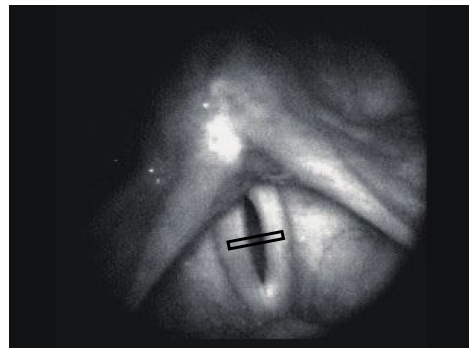


図1 ハイスピードカメラによる声帯の画像データ

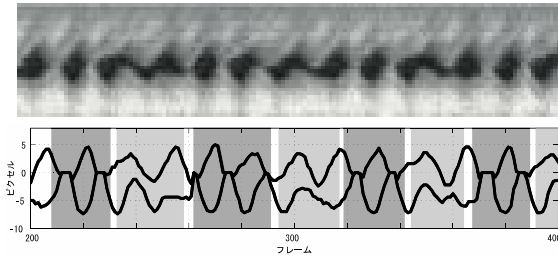


図2 キモグラフ（上）と声帯振動の時間パターン（下）

(2) 声帯振動の物理的機構に即したメカニズム解明のため、声帯間を通過する呼気流の振る舞いを2次元的に数値解析した。病的発声においては、左右の声帯振動が同期しないために、流路の形状が左右で非対称になる。そのような非対称性を有する流路における呼気流の解析は、いまだ十分に行われていない。我々は、2次元の境界層解析とポテンシャル流解析のハイブリッド法を開発し、様々な声門形状における流れの様相を詳細に明らかにすることに成功した。この解析法によって、声門内の各部分における流速、声帯表面にできる境界層の緒特性や流れの剥離位置、声帯を駆動する流体力などの計算を可能とした。

(3) 病的発声に代表される声帯の不規則振動に関して、喉頭における声帯音源の生成機構と声道の音響フィルタの間の物理的な相互作用について解明を進めた。発声のピッチが声道のフォルマント周波数に近接した場合、両機構の間に強い相互作用が生じて声帯振動が不規則になり、声区の転換が生じる。この特異的な現象を、計算機シミュレーションだけでなく、実際の人の発声においても観測するため、声道を音響的に雑音駆動して口唇から外部に放出される応答信号を測定し、高い発声ピッチにおいても安定にフォルマント周波数を求めることを可能とした。同時に、声帯振動を電気式喉頭計（EGG）で測定することで、発声ピッチとフォルマント周波数の関係を詳細に検討した。その結果、人の実際の発声でも、地声と裏声の声区転換において、声帯音源と声道フィルタの相互作用が関わっていることが強く示唆される実験結果を得た。また、発声ピッチの倍音周波数がフォルマント周波数に近接した場合でも、同様の不安定性が生じ得ることが明らかになった。一方、従来から考えられてきた声道フィルタが声帯音源に与える影響に加えて、声門は声道音響管の境界条件を与えるため、声帯の振動様態が声道音響特性に逆の影響を与える可能性が考えられる。我々はすでに、地声と裏声の発声においてフォルマント周波数が変化するとの実験結果を得ている。声帯音源と声道フィルタの相互作用に関しては、今後、より精緻な検討を継続する必要がある。

(4) 発声中の声帯振動の観測法に関して、二つの内視鏡カメラを用いて異なる角度から声帯振動の同時撮影を行い、ステレオマッチング法による声帯振動の3次元動態の復元を試みた。一名の被験者の通常の発声条件における声帯の3次元ダイナミクスを復元し、上唇下唇の位相差などの動的な特徴を捉えることができることを示した。さらに、このステレオマッチング法を改良するため、適応窓の設定およびストロボスコープの原理を取り入れることによって、性能を大幅に改善した。これにより、裏声および吸気発声では、通常の声帯発声よりも声帯の上下動が小さいことを観測することができた。本観測法の定量的な精度評価に関し、声帯レプリカモデルでは、前後・左右に0.2mm程度、上下方向で0.8mm程度の計測誤差がみられた。声帯長や声門面積などの測定には、ステレオ内視鏡は誤差10%以内であることを示した。

(5) 発話時における声道内の音響解析の高精度化に関しては、音場の時間的な推移を時間領域差分法（FDTD法）により可視化した。平面波及び高次モードを含む音場が概ね数ミリ秒で定常になり、音源付近の大きな振幅変化が局所的に存在して空間を伝搬しないエバネッセントモードにより表現される様子を示すことができた。さらに、声道壁を伝わるなどして口唇以外から放射される音声の特徴についても検討した。小型マイクロホンと直線アレイ状に配置し、受音した音圧傾度から頸部皮膚近傍の粒子速度を求めた。

(6) 数値シミュレーションや物理モデル作成のための形状データとして、MRI（磁気共鳴画像法）装置により、定常母音および連続母音発話時のMRIデータを収集した。さらに、母音発話時のMRI画像データから半自動的に取得した声道形状の音響特性を、3次元のGPGPU-FDTD法で計算するシミュレータを開発した。本シミュレータは計算しながら即時に計算結果を可視化できるだけでなく、インタラクティブに参照点を変更することが可能であることなどの特徴を有している。

(7) 声道の音響解析について、声道モデルの高度化のために壁インピーダンスを非対称に設定する解析的な手法を提案し、数値計算により妥当性を検討した。これにより、例えば、硬口蓋と舌面など壁特性が異なる部位の境界条件をより正確に表現することが可能となった。また、時間領域差分法（FDTD法）による声道内の音響解析において、数値計算をGPUで、計算結果の可視化をCPUで行う方法を確立し、声道形状のデータとして使用する磁気共鳴映像法により撮像された断面画像を時間領域差分法で計算できるように、自動的に画像処理する方法の検討を行った。

(8) 母音/a/発話時の3次元声道形状モデルの声道伝達特性に見られる極や零が、声道形状の微細構造のどのような部位に起因して生じているのかを明らかにした。喉頭腔の狭めが声道伝達特性上に極を生じさせ、約3kHzよりも高域の振幅を低下させるなど、微細構造が音響特性に大きく影響を与えていることが示された。また、非対称壁インピーダンスを導入した矩形声道モデルの音響特徴を検討した。高次モードの伝搬定数は遮断周波数付近で壁インピーダンスの影響が比較的大きく現れる。体組織として妥当な範囲で壁インピーダンスの値を変化させた場合には、高次モードにより生じる高域のピークは剛壁条件によるピークとほぼ一致することが示された。

(9) 喉頭腔を狭めた時の声道内の音圧分布より、狭めより先の部分が一つの開端音響管として見え、声道全体の特性として現れることを示した。また、声門音源の境界条件を、声門下部を見込む高次モードを含むアドミタンス行列として電気回路的に表現し、声道の入力インピーダンスと駆動音源境界の相互作用について検討した。声道の3次元解析で現れる帯域幅の狭い極の発生は、音源境界条件として事前に設定した振動速度分布が、声道の入力インピーダンスとの関係により、実際には実現困難な分布であることに起因していることを示した。

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16件)

- (1) M. Kob, N. Henrich, H. Herzel, D. Howard, I. Tokuda, J. Wolfe, Analysing and understanding the singing voice: recent progress and open questions, *Current Bioinformatics*, 査読有, vol. 6, 2011, 362-374.
- (2) F. Alipour, C. Brucker, D. D. Cook, A. Gommel, M. Kaltenbacher, W. Mattheus, L. Mongeau, E. Nauman, R. Schwarze, I. Tokuda, S. Zorner, Mathematical Models and numerical schemes for the simulation of human phonation, *Current Bioinformatics*, 査読有, vol. 6, 2011, 323-343.
- (3) T. Riede, I. Tokuda, C. G. Farmer, Subglottal pressure and fundamental frequency control in contact calls of juvenile Alligator mississippiensis, *Journal of Experimental Biology*, 査読有, vol. 214, 2011, 3082-3095.
- (4) T. Kaburagi, Effect of channel asymmetry on the behavior of flow passing through the glottis, *Acoustical Science and Technology*, 査読有, vol. 33, 2012, 348-358.

DOI:10.1250/ast.33.348

- (5) K. Motoki, A parametric method of computing acoustic characteristics of simplified three-dimensional vocal-tract model with wall impedance, *Acoustical Science and Technology*, 査読有, vol. 34, 2013, 113-122.
DOI:10.1250/ast.34.113
- (6) 松崎博季, 紺野健幸, 真田博文, Graphics Processing Unitを用いた時間領域差分法による声道の音響数値解析と解析結果の即時可視化, *日本音響学会誌*, 査読有, vol. 69, 2013, 58-62.
- (7) I. Tokuda, A. Kuwahara, Synchronization analysis of choir singing, *Nonlinear Theory and Its Applications*, IEICE, 査読有, vol. 3, 2012, 215-221.
DOI 10.1588/nolta.3.215
- (8) T. Kaburagi, T. Takano, Y. Sakamoto, Estimating area function of the vocal tract from formants using a sensitivity function and least-squares, *Acoustical Science and Technology*, 査読有, vol. 34, 2013, 301-310.
DOI: 10.1250/ast.34.301
- (9) 元木邦俊, 矩形声道モデルにおける非対称壁インピーダンスを考慮した伝搬定数の表現法, *北海学園大学工学部研究報告*, 査読無, vol. 41, 2014, 57-74.
- (10) H. Matsuzaki, A. Serrurier, P. Badin, K. Motoki, Acoustic simulation of Japanese and French vowel /a/ with nasal coupling by time-domain finite-element method, *Acoustical Science and Technology*, 査読有, vol. 34, 2013, 284-286.
DOI: 10.1250/ast.34.284
- (11) H. Matsuzaki, A. Serrurier, P. Badin, K. Motoki, One-dimensional and three-dimensional propagation analyses of acoustic characteristics of Japanese and French vowel /a/ with nasal coupling, *Acoustical Science and Technology*, 査読有, vol. 35, 2014, 35-41.
DOI: 10.1250/ast.35.35
- (12) I. Tokuda, M. Iwawaki, K. Sakakibara, H. Imagawa, T. Nito, T. Yamasoba, N. Tayama, Reconstructing three-dimensional vocal fold movement via stereo matching, *Acoustical Science and Technology*, 査読有, vol. 34, 2013, 374-377.
DOI: 10.1250/ast.34.374
- (13) H. Matsuzaki, K. Motoki, H. Sanada, Finite element analysis of fine-structural effect on

acoustic characteristics of vocal tract, Acoustical Science and Technology, 査読有, vol. 36, 2015, 39-41.
DOI: 10.1250/ast.36.39

(14) 徳田功, 歌声の非線形科学, 日本音響学会誌, 査読有, 9巻, 2014, 512-518.

(15) D. E. Sommer, I. Tokuda, S. D. Peterson, K. Sakakibara, H. Imagawa, A. Yamauchi, T. Nito, T. Yamasoba, N. Tayama, Estimation of inferior-superior vocal fold kinematics from high-speed stereo endoscopy data in vivo, Journal of the Acoustical Society of America, vol. 136, 2014, 3290-3300.
DOI : 10.1121/1.4900572

(16) T. Kaburagi, Determining the length and cross-sectional area of the vocal tract jointly from formants using acoustic sensitivity function, Acoustical Science and Technology, 査読有, vol. 35, 2014, 290-299.
DOI: 10.1250/ast.35.290

[学会発表](計12件)

(1) T. Kaburagi, Frequency-domain representation of source-filter coupling and its effect in the production of voice, Proc. Interspeech2011, 2011年8月30日, Florence (Italy).

(2) T. Kaburagi, Interactive processes underlying the production of voice, Proc. APSIPA ASC 2011, 2011年10月20日, Xian (China).

(3) H. Matsuzaki, K. Motoki, Numerical simulation of acoustic characteristics of vocal-tract model with 3-D radiation and wall impedance, Proc. APSIPA ASC 2011, 2011年10月20日, Xian (China).

(4) I. Tokuda, M. Iwawaki, K.-I. Sakakibara, H. Imagawa, T. Nito, T. Yamasoba, N. Tayama, Reconstruction of three-dimensional laryngeal dynamics via stereo-matching technique, Proc. Nonlinear Theory and Its Applications, 2011年9月5日, KIPC (兵庫県神戸市).

(5) T. Kaburagi, T. Takano, Y. Sakamoto, Estimating the vocal-tract area function from formants using a sensitivity function and least square, Proc. Interspeech2012, 2012年9月13日, Portland (USA).

(6) I. Tokuda, M. Iwawaki, K. Sakakibara, H. Imagawa, T. Nito, T. Yamasoba, N. Tayama, Three-dimensional reconstruction of vocal fold vibration by stereo matching technique, Proc. International Conference on Voice

Physiology and Biomechanics2012, 2012年7月6日, Erlangen (Germany).

(7) I. Tokuda, H. Koda, T. Nishimura, C. Oyakawa, T. Nihonmatsu, N. Masataka, Formant tuning in gibbon songs, Proc. International Conference on Voice Physiology and Biomechanics2012, 2012年7月6日, Erlangen (Germany).

(8) T. Kaburagi, Synergistic interactions underlying the production of voice, Proc. ICA2013, 2013年5月5日, Montreal (Canada).

(9) T. Kaburagi, N. Yamada, T. Fukui, E. Minamiya, A morphological and acoustic study on the effect of a trumpet player's vocal tract, Proc. ICA2013, 2013年5月3日, Montreal (Canada).

(10) K. Motoki, Three-dimensional rectangular vocal-tract model with asymmetric wall impedances, Proc. Interspeech2013, 2013年8月26日, Lyon (France).

(11) T. Kaburagi, Estimation of vocal-tract shape from speech spectrum and speech resynthesis based on a generative model, Proc. Interspeech2014, 2014年9月16日, Singapore.

(12) H. Uchida, K. Wakamiya, T. Kaburagi, A study on the improvement of measurement accuracy of the three-dimensional electromagnetic articulography, Proc. Interspeech2014, 2014年9月16日, Singapore.

6. 研究組織

(1) 研究代表者

鏑木 時彦 (KABURAGI, Tokihiko)
九州大学・芸術工学研究院・准教授
研究者番号: 30325568

(2) 研究分担者

元木 邦俊 (MOTOKI, Kunitoshi)
北海学園大学・工学部・教授
研究者番号: 80219980

松崎 博季 (MATSUZAKI, Hiroki)
北海道科学大学・工学部・教授
研究者番号: 80219980

徳田功 (TOKUDA, Isao)
立命館大学・理工学部・教授
研究者番号: 00261389

(3) 連携研究者

三木信弘 (MIKI, Nobuhiro)
公立はこだて未来大学・システム情報科学部・名誉教授