

科学研究費助成事業 研究成果報告書

令和元年9月5日現在

機関番号：32689

研究種目：基盤研究(C) (一般)

研究期間：2015～2018

課題番号：15K01459

研究課題名(和文) 発話ロボットの音声学習における脳機能モデルの再現と発声障害のシミュレータの構築

研究課題名(英文) A talking robot having human-like brain functions for autonomous voice learning and its application to a vocal articulation simulator

研究代表者

澤田 秀之 (Sawada, Hideyuki)

早稲田大学・理工学術院・教授

研究者番号：00308206

交付決定額(研究期間全体)：(直接経費) 3,700,000円

研究成果の概要(和文)：申請者らがこれまでに研究を進めてきた、人間と同等の発声器官を全て機械的に構成した発話ロボットを再設計し、より柔軟に発話動作を獲得できる機構を実装した。聴覚フィードバック学習によって、ロボットが聴取音声を基に自律的に発話動作を獲得し、任意の音声を生成することができる学習機構を実装するため、人間の脳機能を再現した学習モデルを構築した。更にFPGAによって、音声獲得の高速オンライン学習を実現した。これら発話器官を再現した機械モデルと、脳の音声学習機能に着目したオンライン学習モデルを統合した新しい発話ロボットを構築し、脳内モデルの学習過程、獲得される発話動作および、ロボットの生成音声の解析を行った。

研究成果の学術的意義や社会的意義

本研究は、人間の音声生成機構を全て機械的に再現した発話ロボットを構築するばかりでなく、音声の自律獲得に拘わる脳機能を再現した学習機構を実装し、音声から口内の発話動作を獲得して音声生成を行うシステムを実現した。これにより、人間の様に学習によって発話動作及び音声を獲得することが出来る機械機構を再現でき、更に人間の構音動作との比較から、発話障害の訓練装置の実現に貢献できる。人間は、聴覚フィードバックによって各発話器官を統合的に制御して、安定した発話をおこなう手法を、後天的に獲得している。人間の音声学習機構との比較、解析を行っていくことにより、脳における情報の認知・生成メカニズムの解明につながる。

研究成果の概要(英文)：We have developed a talking robot, which is a mechanical vocalization system by modeling the human vocal articulatory system. The robot is constructed with mechanical parts that are made by referring to human vocal organs biologically and functionally. In this project, newly redesigned organs were developed for extending the speaking capability. In human speech, the timing function is important for determining its duration, stress and rhythm. The cerebellum plays a key role in the coordination, precision and timing of motor responses. We have developed a robotic brain, which generates human-like vocal sounds using a simplified cerebellum-like neural network model as the timing function. The brain model was designed using the System Generator software, and implemented in a hardware co-simulated with a FPGA. We verified that the learning capability of the cerebellar-like neural network was applicable to speaking for generating a human-like utterance with prosodic features.

研究分野：情報制御工学

キーワード： 知能ロボット ソフトロボット 人工知能 音声生成 聴覚フィードバック 脳内ネットワーク 音声認識 発声障害

1. 研究開始当初の背景

人間の音声は、肺、声帯、気道、舌、軟口蓋、顎、唇、鼻腔などから成る音声生成器官の複雑な運動によって生成され、特定の音響のパターンである「言葉」として聞き手に伝達される。話者が音声によって聞き手に情報を伝えるとき、まず音声生成のための発話動作指令が脳内で生成され、運動神経を介して音声生成器官へ送られる。この指令を基に、それぞれの音声生成器官が適応的に動作をおこない、声帯振動による音源生成から声道形状の形成、鼻腔共鳴の付加などによって音声生成される。発話された音声は同時に、骨導と気導によるフィードバックにより聴覚神経を介して発話者自身にも伝わり、常に音声生成器官が適応的にコントロールされて、安定した発話が実現されている。以上の音声生成過程と聴覚フィードバックは、発話時に話者にとって無意識のうちにおこなわれている。

このような能力は、乳幼児が生まれてから数年のあいだに言語を習得していく過程において、発声と聞き取りの試行錯誤を繰り返すことによって獲得されるものである。また音声獲得後であっても、人間は練習によって声真似をすることができる。これは、発声のために必要な器官が生まれながらに備わっている一方で、発話技術は学習によって後天的に獲得されるものであるためと考えられる。例えば、聴覚障害や、脳性麻痺などの動作器官の障害によりフィードバック学習が正常に行なわれない場合には、音声の獲得、生成が困難となる。また高齢化に伴い、脳内出血や脳梗塞の後遺症として、発話障害や構音障害の報告も増えているが、これは脳機能の低下や障害により、構音動作が制限されることによって引き起こされる症状である。

申請者は研究開始当初までの研究で、人間と同等の発話機構を全て機械的に再現した発話ロボットを構築し、聴覚フィードバック学習により、自律的に発話動作を獲得できることを示してきた。特に日本語音声の発声に着目し、日本人男性の平均値を基に各発話器官を製作して、日本語五母音、破裂音、鼻音の発話と、声帯および空気流の制御と適応学習によって、抑揚の表現や歌声の生成が可能であることを示した。ここでの自律学習には、音声の音響特徴とモータ制御パラメータを適応的に対応づけることを目的として、自己組織化ニューラルネットワークを提案し、見本となる人間の音声を与えれば、学習によって同様の音声生成するための発話動作が獲得できることが解った。これらの成果を国内外で発表し、これまでに New Technology Foundation Award Finalist、電子情報通信学会 ヒューマンコミュニケーション賞、IEEE/RSJ IROS Hyper Human Tech Award、各種 ベストプレゼンテーション賞など、多くの賞を受けている。また、国内ばかりでなく、海外からも招待講演、基調講演、ロボットのデモンストレーションの依頼を受け、研究成果を広く公表してきた。

音声生成を機械的に再現する研究は、日本を中心にいくつかの報告があった。早稲田大学の WT-7RII (Waseda Talker No.7 Refined II)、大阪大学の Lingua、Indiana 大の VoxHead などに代表される研究成果があるが、これらの多くは人間の頭部 MR 画像を基に、発話器官をメカニカルモデルによって再現することに重点が置かれたものであり、機械系が人間の様に対話学習によって自律的に発話手法を獲得したり、脳内の学習過程を再現するような成果には繋がっていなかった。

2. 研究の目的

本研究では、申請者らがこれまでに研究を進めてきた、人間と同等の発声器官を全て機械的に構成した発話ロボットを再設計し、より柔軟に発話動作を獲得できる機構を実装する。聴覚フィードバック学習によって、ロボットが聴取音声を基に自律的に発話動作を獲得し、任意の音声生成することができる学習機構を実装するため、人間の脳機能を再現した学習モデルを構築する。これを FPGA (Field- Programmable Gate Array) によって、音声獲得の高速オンライン学習を実現する。これら発話器官を再現した機械モデルと、脳の音声学習機能に着目したオンライン学習モデルを統合した新しい発話ロボットを構築し、脳内モデルの学習過程、獲得される発話動作および、ロボットの生成音声を詳細に解析することにより、人間の発話機構の科学的解明に繋げていく。

3. 研究の方法

人間の音声生成器官と等価な発話機構を持つロボットの構築

人間の発声は大きく分けて、声帯振動による音源の生成と、共鳴によるホルマントの付加という、2つの働きによって構成されている。肺からの呼気流は気道を通して声帯の振動を引き起こし、音源を生成する。更にこの声帯音源波に対して声道および鼻腔が音響フィルタの役割を果たすことによって、音素が構成される。これら人間の発声器官に対応する機能を、それぞれエアポンプ、人工声帯、調音用共鳴部、マイク、音響解析部によって、全てメカニカルに再現する。

本申請研究では、未だ実現が困難であった、柔軟な舌の動作を再現してより明瞭な音素の発話を可能とするメカニズムと、摩擦音や破裂音を生成するための歯および口唇の開発をおこない、生成音声の明瞭度を上げる。特に舌の変形自由度ならびに変形速度は、音声の明瞭度の向上に大きく貢献することが解っている。しかし、人間の音声と同等の空洞共鳴効果を得るためには、声道部の体積を人間のそれと同程度にする必要があり、サーボモータによるアクチュエーションでは実現が困難である。そこで新しいソフトアクチュエータを舌内に配

置し、人間の舌と類似した変形をさせる機構を実現する。更に、摩擦音の生成のために、口唇の後方部に歯を組み込み、舌の動作と同期させて発話をおこなう制御手法を実現する。同時に、声質を決定する声帯と声道部の材質、形状および制御点数について詳細に検討をおこない、より人間に近い音声を生成するメカニカルシステムを構築する。

聴覚フィードバック学習による自律音声獲得を可能とする脳機能モデルの構築

発話ロボットが、聴覚フィードバックによる模倣学習によって音声を獲得する課程について、人間の脳機能を忠実に再現した学習モデルを構築する。ここでは、小脳、皮質視床、大脳基底核から成る、音声の聴覚フィードバック学習の脳内ネットワークに着目し、これを計算機モデルとして実装する。これまで申請者が構築してきた発話ロボットにおいては、音響特徴とモータ制御量のパターンの対応付けに着目して、自己組織化ニューラルネットワークを応用した学習モデルを用いていた。これは厳密な意味で、人間の乳幼児の様に、音声獲得過程の時系列を再現したもとはなっていない。脳機能の障害と発話不明瞭要因との関連に基づく、発話障害シミュレータの開発には、より忠実に脳機能の学習モデルを再現したモデルが必要となる。

本研究では、小脳(Cerebellum system)、皮質視床(Cortico-thalamic system)、大脳基底核(Basal ganglia and limbic system)から成る、音声の聴覚フィードバック学習の脳内ネットワークに着目し、これを計算機モデルとして実装する。小脳は、脳内において知覚と運動機能の統合を行う機能を司り、音声の生成においては聴覚から得られる音声知覚情報を元に、発話器官の動作生成をおこなうための制御信号を作り出している。皮質視床は間脳の一部を占め、聴覚からの入力情報を大脳新皮質へ中継する役割を担っている。この信号を元に、大脳は音声を解釈し、言葉や声の特徴の理解をおこなっている。また大脳基底核は、大脳皮質と視床、脳幹を結びつけている神経核の集まりであり、音声学習における運動調節や認知機能を担っている。

発話ロボットが模倣学習によって音声を獲得する課程について、これら3つの脳機能を忠実に再現した学習モデルを構築する。脳内学習においては、各受容野の神経活動電位に基づく神経伝達パルス密度のモデル化が必要となる。これをFPGAを使ってロボットの音声学習制御部の実装し、音声獲得の高速オンライン学習および実時間発話制御システムを実現する。

発話障がい患者の不明瞭音声の再現

上記の研究成果を基に、発話障がい者音声の不明瞭要因を発話ロボットによって再現させる。特に、脳機能モデルにおいて、神経伝達パルスの密度の変化、神経ネットワークの閉塞・解放を元に、発話障害の動作、生成音声を再現させることにより、その原因を、構音動作と脳障害の両面から特定するためのロボットシミュレータを開発する。

音声インタフェースとしてのロボットの構築

ロボティクス、ヒューマンインタフェース開発の視点から、人間と同等の発話器官を用いて音声対話をおこなうことが可能な、自律発話ロボットの実現を目指す。本ロボットは、一連の発話動作を、音源生成部と共鳴特性付加部に分離して考えることにより、この二つの制御部を独立に操作することが可能である。これは、音声生成の物理シミュレータとも考えることができる。人間の発話原理を、目に見ながら誰でもが解りやすく理解するための教育教材としても、意味のあるシステムとなる。

4. 研究成果

前章の4つの研究項目について、全て良好な成果を得た。

まず発話動作をおこなう器官を声道物理モデルに基づいて全て機械的に構成した、発話ロボットを構築した。ロボットの外観を図1に示すが、声道部と舌、鼻腔部をあわせて、全12個のサーボモータによって声道共鳴管の変形ならびに空気流量の適応制御をおこなっている。聴覚フィードバック制御によって、ロボットが自律的に発話動作を獲得し、目標となる人間の音声と同様な音響を生成する。本ロボットシステムは、エアポンプ、圧力・流量制御弁、人工声帯、声道共鳴管、鼻腔共鳴部、マイクロフォン、音響解析部から成り、それぞれ人間の肺、気道、声帯、声道、鼻腔、聴覚部に対応する。発声手法として、まずエアークンプレッサから圧縮空気を送り出す。送り出された空気は減圧弁、流量制御弁を経由し、人工声帯を振動させて音源を生成する。このとき流量制御弁により、生成される音声の大小と高低を変化させることができる。音源は、共鳴管としての役割を果たす声道に導かれ、声道形状を変化させることにより、様々な音響特性を音源に付加することが可能となっている。人間は軟口蓋により口腔と鼻腔を遮断することが可能である。本ロボットは、声道と鼻腔接続部に回転式弁を持ち、鼻腔共鳴の有無を操作することにより鼻音/mの生成が可能である。また、声道内部にシリコンゴムで成形された舌を持ち、これをモータにより上げ下げすること

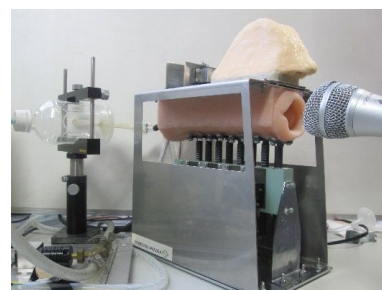


図1 発話ロボットの外観

で側音/l/、破裂音/t/、鼻音/n/などを生成することができる。音響分析器は、人間の聴覚部と脳に対応しており発話ロボットが生成する音声をリアルタイムで解析し、発話動作の獲得に必要な特徴量を抽出する。システムコントローラは生成された音声に対応するモータ制御量を計算し、聴覚フィードバック学習により抽出された音響特徴量とモータ制御量の対応関係を学習していく。

人間の発声においては、母音は定常的な声道形状を形成することにより生成され、子音は顎や舌、軟口蓋などの非定常的な動作から引き起こされる変化により生成される。よって、任意の音を発声するための声道形状を決定することができれば、ロボットがその発声器官を使用して口を動かしながら発話動作を学習することが可能となる。加えて似たような音響特徴を持つ音声は、似たような声道形状により生成されることは、我々の発話中の口内形状からもわかる。発話ロボットの生成する音響特徴量と声道形状の対応関係を構築するために、自己組織化マップ(SOM)を3次元構造に拡張し、3次元構造内に音響特徴量ならびにモータ制御量を配置する3D-SOMを提案した(図2)。SOMはデータの近隣関係を自己組織的に獲得することができるため、特徴マップ上の各セルには、音響特徴量とモータ制御量が対応づいた状態で配置されることになる。発話ロボットの自律学習により、発話動作と生成音響の特徴量の対応が獲得される。

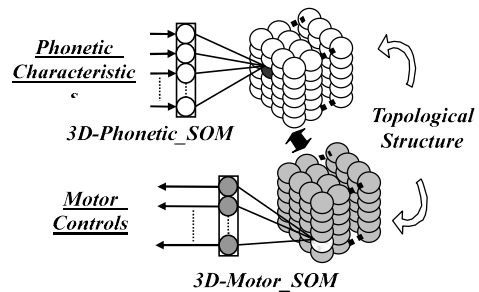


図2 3D-SOMの構造

一方、ロボットの発話においては、声の音響的特徴を生成するだけでなく、抑揚や会話の速度が重要となる。本研究では、リズムの認知が小脳において成されていることに着目し、小脳、皮質視床、大脳基底核から成る音声の聴覚フィードバック学習の脳内ネットワークを、計算機モデルとして実装した。神経パルスの生成と音声学習を高速に実行するため、FPGA(Field-Programmable Gate Array)を導入した。脳神経モデルの実装結果を図3に示す。

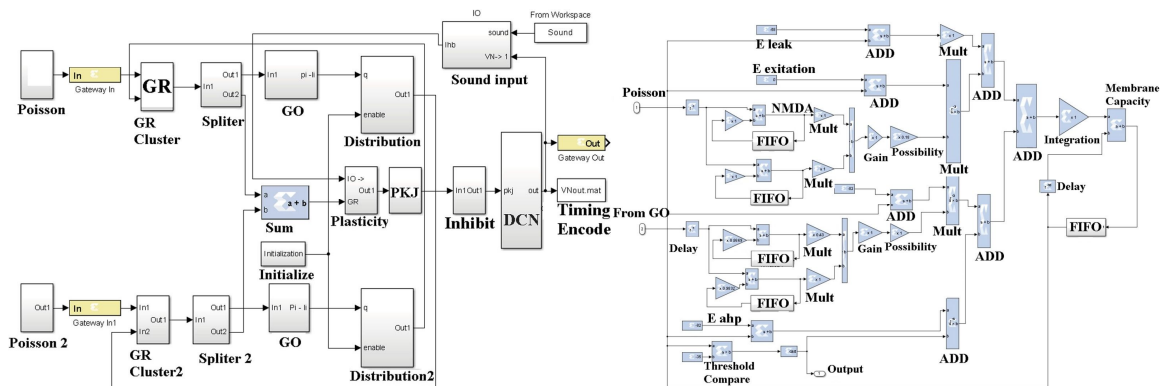


図3 脳神経モデルの構築とFPGAへの実装

以上で述べた学習機構の実装により、ロボットの自律的音声獲得とならびに柔軟な発話が可能となった。人間の音声の声真似実験を行った結果、良好に音声の再現が可能であり、また人間の抑揚や発話速度も再現できることが示された。

次に、聴覚障がい者のための対話型ロボット発話訓練システムの構築を行った。出生時から言語獲得期までに聴覚に障害を持つことが発見される患者は、出生人口の0.5~1%程度にのぼると言われる。発話障害、聴覚障がい患者の発声訓練は現在、言語聴覚士(Speech Therapist: ST)が一人一人の患者に対して、症状に応じた手法でおこなっている。口唇や顎を使う破裂子音は、口先部の動作を観察することにより、比較的容易にその動作を模倣することができる。しかし、口内の舌や軟口蓋を使った複雑な構音動作を必要とする母音や摩擦子音の獲得においては、その動きが外部から観察できないため、訓練手法は未だ確立されていない。本ロボットは、音声の特徴から、その音声を生成するのに必要な発話動作を再現できる。本研究では、聴覚障がい者や発話障がい者が、見本となるロボットの構音動作や口内の動きを見ながら、対話的に発話訓練をすることができるシステムの構築を行った。聴覚障害のある被験者らによる発話訓練実験をおこない、本ロボットシステムの有効性を見いだした。また、ロボットの動作によって口の動きが見られるのは面白いという意見も得られた。

更に、音声インタフェース構築の観点から本ロボットの評価を行い、「発話楽器」としての可能性を見出した。現在、マサチューセッツ工科大学(MIT)のDiemut Strebe氏をはじめとする芸術家、サイエンティストらと協同し、AIを実装した発話ロボットを核とする芸術作品の製

作を進めている。2020年夏より、フランス・パリの Centre Pompidou 美術館において展示が決まり、現在、鋭意、研究開発と製作を進めている。

本研究機期間の成果は、学術論文 10 編と、国際学会での Best Paper Award および Best Presentation Award の受賞を得た。

5 . 主な発表論文等

[雑誌論文] (計 10 件)

- [1] Vo Nhu Thanh and Hideyuki Sawada, "Simplified cerebellum-like spiking neural network as short-range timing function for the talking robot", Connection Science, Vol.30, Issue 4, DOI:10.1080/09540091.2018.1510901, August, 2018.
- [2] **Best Paper Award 受賞**, Thanh Vo Nhu and Hideyuki Sawada, "Intoning Speech Performance of the Talking Robot for Vietnamese Language Case," International Symposium on Micro-NanoMechatronics and Human Science, pp. 160-163, 2018.
- [3] Thanh Vo Nhu and Hideyuki Sawada, "Development of Vietnamese Voice Chatbot with Emotion Expression," International Symposium on Micro-NanoMechatronics and Human Science, pp. 22-27, 2018.
- [4] **Best Presentation Award 受賞**, Thanh Vo Nhu and Hideyuki Sawada, "Cerebellum-like Neural Network for Short-range Timing Function of a Robotic Speaking System," Proceedings of International Conference on Control, Automation and Robotics, pp.184-187, 2017.
- [5] Vo Nhu Thanh and Hideyuki Sawada, "Text-to-Speech of a Talking Robot for Interactive Speech Training of Hearing Impaired," Proceedings of IEEE International Conference on Human System Interaction, pp. 166-171, 2017.
- [6] Vo Nhu Thanh and Hideyuki Sawada, "A Talking Robot and Its Real-time Interactive Modification for Speech Clarification", SICE Journal of Control, Measurement, and System Integration, 計測自動制御学会, pp. 251-256, Vol.9, No.6, November 2016.
- [7] Vo Nhu Thanh and Hideyuki Sawada, "Automatic Vowel Sequence Reproduction for a Talking Robot Based on PARCOR Coefficient Template Matching", IEIE Transactions on Smart Processing and Computing, Vol. 5, No. 3, pp.215-221, 2016.
- [8] Vo Nhu Thanh and Hideyuki Sawada, "Comparison of Several Acoustic Features for the Vowel Sequence Reproduction of a Talking Robot", 2016 IEEE International Conference on Mechatronics and Automation, Harbin, China, pp. 1137-1142, 2016.
- [9] Hideyuki Sawada, "A Talking Robot and the Expressive Speech Communication with Human", International Journal of Affective Engineering, Vol.14, No.2, pp.95-102, 2015.
- [10] Vo Nhu Thanh and Hideyuki Sawada, "Autonomous Vowels Sequence Reproduction of a Talking Robot Using PARCOR Coefficients", Proceedings of International Conference on Electronics, Information and Communication, CD-ROM TM1-4, Danang, Vietnam, 2016.

[学会発表] (計 3 件)

- [1] Vo Nhu Thanh and Hideyuki Sawada, "Vietnamese Language Speech Performance by the Talking Robot", 電気関係学会 四国支部連合大会 講演論文集, 2016年9月17日(土), 徳島大学, p.82, 2016.
- [2] **招待講演**, Hideyuki Sawada, "A Talking Robot and the Autonomous Learning of Speech Articulation for the Communication with Humans", International Workshop on Speech Robotics, Dresden, Germany, 2015.
- [3] Vo Nhu Thanh and Hideyuki Sawada, "A Real Time Visualization System for Articulatory Analysis of the Talking Robot", 電気関係学会 四国支部連合大会 講演論文集, 2015年9月26日(土), p. 86, 2015.

[図書] (計 2 件)

- [1] 前田真吾, 澤田秀之, 重宗宏毅, 三輪貴信, 「運動パターンを創り出すマテリアルと知能の設計」, 日本機械学会誌 Vol.122, No.1205, April, 2019.
- [2] 鈴森康一, 新山龍馬, 前田真吾, 「新学術領域「ソフトロボット学」の概要」, pp. 10-12, Soft Matter, October, 2018.

[産業財産権]

出願状況 (計 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 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 <http://www.sawada.phys.waseda.ac.jp/>

6. 研究組織

(1) 研究分担者

研究分担者氏名：

ローマ字氏名：

所属研究機関名：

部局名：

職名：

研究者番号（8桁）：

(2) 研究協力者

研究協力者氏名：Vo Nhu Thanh, University of Science and Technology, The University of Danang,
Vietnam

ローマ字氏名：

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。