

令和 6 年 6 月 15 日現在

機関番号：14301

研究種目：若手研究

研究期間：2020～2023

課題番号：20K20222

研究課題名（和文）聴覚末梢系数理モデルを用いた嚙声の音響学的特徴の解明

研究課題名（英文）Exploration the acoustic characteristics of voice disorders using mathematical auditory periphery model

研究代表者

藤村 真太郎（Fujimura, Shintaro）

京都大学・医学研究科・助教

研究者番号：50815751

交付決定額（研究期間全体）：（直接経費） 3,100,000円

研究成果の概要（和文）：ヒトが直感的に評価する嚙声や不快音の「質感」の特徴解明を目指し、聴覚末梢系の数理モデルを用いて時間領域の情報を高解像度で表現可能な新しい解析手法を開発、GPGPUを用いて解析プログラムを実装した。不快音の質感の特徴を求め、また得られた特徴表現を深層学習の入力として用いる方法などを検討したが、モデルの巨大化など新たな課題が見つかった。一方で提案手法のもたらす時間・周波数分解能を両立した新規表現は、例えば過渡現象を多く含む語音の解析において、子音や半母音などの周波数構造の過渡現象をこれまでにない形で視覚化し解析できることがわかった。

研究成果の学術的意義や社会的意義

本手法がもたらす時間・周波数分解能を両立した表現は、過渡現象を多く含む語音の解析において、既存手法にない特徴表現を可能としており、構音障害の診療において新たな治療・リハビリテーション法につながる可能性がある。

研究成果の概要（英文）：Aiming to clarify the characteristics of the "texture" of hoarseness and unpleasant sounds that humans intuitively evaluate, we developed a new analysis method that can express time domain information at high resolution using a mathematical model of the peripheral auditory system, and implemented the analysis program using a GPGPU. We investigated methods to obtain the characteristics of the texture of unpleasant sounds and use the obtained feature representation as input for deep learning, but new issues were found, such as the large size of the model. On the other hand, the new representation that combines time and frequency resolution brought about by the proposed method was found to be able to visualize and analyze transient phenomena in the frequency structure of consonants and semivowels in an unprecedented way, for example, when analyzing speech that contains many transient phenomena.

研究分野：音声障害、音声信号処理

キーワード：聴覚末梢系 数理モデル 構音障害

1. 研究開始当初の背景

音声障害の診療においては様々な質、程度の嘔声に直面する。ヒトは容易にこれらの「音の質感」を評価し、その特徴を疾患診断の一助として利用することができる。しかし既存の信号処理手法、例えば高調波雑音比(HNR)、周期変動 (Jitter, PPQ)、振幅変動 (Shimmer, APQ)もしくは時間周波数解析(短時間フーリエ変換(STFT)) では各々の特徴を十分に説明できない。

その理由として既存手法では時間情報の表現に限界があるためではないかと考えた。end-to-endな次元畳み込みニューラルネットワークを用いた音声解析法を研究において、さらに性能を高めるため音の「質感」を表現する本質的なモデル構造/入力特徴量を模索する中で、後述する独自設計の聴覚末梢系モデルを前処理として用い、これを入力特徴量として最適化することを考えた。

2. 研究の目的

嘔声や不快音の「質感」をヒトは直感的に評価できるが、従来の調波構造やその包絡を用いた信号処理法ではその特徴を十分に説明できない。本研究の目的は、1)聴覚末梢系の数理モデルを用いて時間領域の情報を高解像度で表現可能な新しい解析手法を開発し、2)嘔声や不快音の「質感」を特徴づける音響学的構造を解明することとした。

3. 研究の方法

1) 聴覚末梢系数理モデルによる音声解析プログラムの開発

モデルの設計と実験用プログラムの実装

蝸牛基底膜および内有毛細胞を減衰自由振動系とし、調和変位を入力とするモデルとしてフィルタアレイを設計した。本モデルは $O(n^3)$ の計算量が必要となるが、この計算を高速に行うため、線形部分を行列演算に変換し、NVIDIA 社の GPGPU および同コンピューティングプラットフォーム CUDA を用いて、プログラムを実装した。

モデル基本性能の評価と最適化

上記で作成したプログラムを利用して、音声波形を高密度の時間周波数表現に変換し、これを入力として深層学習判別モデルを作成する実験を計画した。このための基礎的実験として、雑音(嘔声において、喉頭における乱流を発生源として生成される雑音を想定した、広帯域ノイズ)が、深層学習モデルの判別性能にどの程度影響を与えるかを調べた。

2)嘔声や不快音の「質感」を特徴づける音響学的構造を解明

深層ニューラルネットワーク(深層 NN)の構築

1- で作成したプログラムを利用して、音声波形を高密度の時間周波数表現に変換し、これを入力として深層学習判別モデルを作成する実験を行った。持続母音「あ」の音声波形 945 サンプルを用いて、これを高密度の時間周波数表現に変換し、これを ResNet、Inception-v3 等の入力として嘔声の重症度を判別するモデルを作成した。

その他: 構音の評価

本提案手法のもたらす時間・周波数分解能を両立した表現は、語音の解析において、特に子音や半母音などの周波数構造の過渡現象をこれまでにない形で視覚化できることができた。これを活用して特に人工内耳装用者における構音のひずみに着目し、その異常を定性的・半定量的に評価する手法の研究を進めた。上記に関連する補助的調査として、人工内耳装

ユーザーにおけるコミュニケーションにおける発話によるコミュニケーションの役割と構音障害の与える影響を調べるため、21名の学生と32名の就労者の人工内耳装用者を対象に質問紙調査を行った。

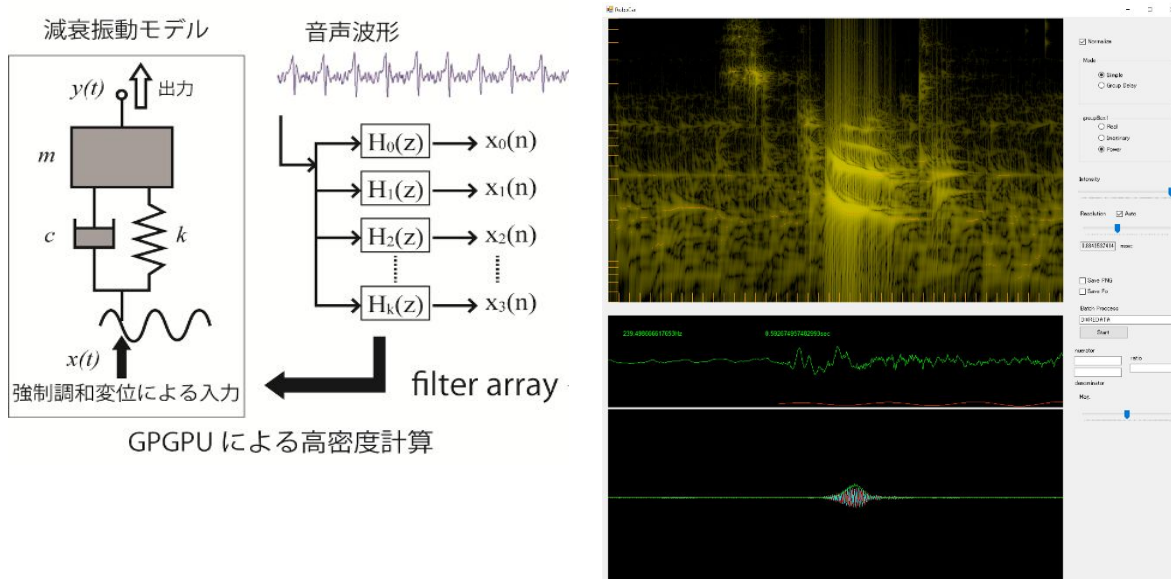
高度難聴・人工内耳装用者において、発話明瞭度を評価するための13種の単語について、約20例から音声サンプルを収集し、提案手法により高密度の時間周波数表現に変換、視覚化し、発話の各段階におけるエネルギー分布の傾向などを解析中である。

4. 研究成果

1) 聴覚末梢数値モデルによる音声解析プログラムの開発

モデルの設計と実験用プログラムの実装

3-1) で記載の方法により、本計算を General-purpose computing on graphics processing units(GPGPU)を用いて並列実行することにより、任意の時間・周波数解像度で高速度にこれを計算するプログラムを開発した。



モデル基本性能の評価と最適化

上記で作成したプログラムを利用して、音声波形を高密度の時間周波数表現に変換し、これを入力として深層学習判別モデルを作成する実験を計画した。このための基礎的実験として、雑音(嘔声において、喉頭における乱流を発生源として生成される雑音を想定した、広帯域ノイズ)が、深層学習モデルの判別性能にどの程度影響を与えるかを調べた。

2) 嘔声や不快音の「質感」を特徴づける音響学的構造を解明

深層ニューラルネットワーク(深層 NN)の構築

提案手法を入力とした 2d-CNN モデルの性能は Accuracy 0.6 ~ 0.7 程度にとどまり、過去に我々が作成した 1d-CNN モデルの性能 (Accuracy: ~0.771) を超えることはできなかった。高密度の時間周波数表現は、深層学習モデルの入力として非常に大きなデータとなるため、モデルの入りに近い部分でパラメータ数が多くなりモデルが巨大化するため、有効な学習が行えなかったものと考えられる。この部分は提案手法のパラメータ調整により、今後改善が見込まれる。

その他: 構音の評価

本提案手法のもたらす時間・周波数分解能を両立した表現は、語音の解析において、特に子音や半母音などの周波数構造の過渡現象をこれまでにない形で視覚化できることができることがわかった。これを活用して特に人工内耳装用者における構音のひずみに着目し、その異常を定性的・半定量的に評価する手法の研究を進めている。

高度難聴・人工内耳装用者において、発話明瞭度を評価するための 13 種の単語について、約 20 例から音声サンプルを収集し、提案手法により高密度の時間周波数表現に変換、視覚化し、発話の各段階におけるエネルギー分布の傾向などを解析中である。

5. 主な発表論文等

〔雑誌論文〕 計1件（うち査読付論文 1件 / うち国際共著 0件 / うちオープンアクセス 0件）

1. 著者名 Fujimura Shintaro, Kojima Tsuyoshi, Okanoue Yusuke, Shoji Kazuhiko, Inoue Masato, Omori Koichi, Hori Ryusuke	4. 巻 36
2. 論文標題 Classification of Voice Disorders Using a One-Dimensional Convolutional Neural Network	5. 発行年 2022年
3. 雑誌名 Journal of Voice	6. 最初と最後の頁 15 ~ 20
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.jvoice.2020.02.009	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計1件（うち招待講演 0件 / うち国際学会 0件）

1. 発表者名 藤村 真太郎、児嶋 剛、大森 孝一
2. 発表標題 聴覚モデルとしてみた病的音声の深層学習
3. 学会等名 第122回日本耳鼻咽喉科学会・学術講演会
4. 発表年 2021年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------