

科学研究費助成事業 研究成果報告書

平成 30 年 6 月 1 日現在

機関番号：14701

研究種目：挑戦的萌芽研究

研究期間：2016～2017

課題番号：16K12464

研究課題名(和文)対話的可視化可聴化に基づく音声コミュニケーション研究支援環境

研究課題名(英文)Research assisting environment of speech communication based on interactive visualization and sonification

研究代表者

河原 英紀(Kawahara, Hideki)

和歌山大学・学内共同利用施設等・名誉教授

研究者番号：40294300

交付決定額(研究期間全体):(直接経費) 2,600,000円

研究成果の概要(和文):本課題では、周期性に基づく干渉を排除した音声パラメタの表現に基づいて、音声の分析・処理・合成のための信号処理基盤を構築した。この過程で、我々が開発し、音声処理のデファクトスタンダードとなっているSTRAIGHTとは独立な新たなアルゴリズムを開発し、オープンソースとして公開した。また、STRAIGHTに関しては、学術的研究を支援するためのツール群を整備した。加えて、音声処理に革命を起こした深層学習に基づくWaveNetの応用基盤を確立した。

研究成果の概要(英文):We developed infrastructures of speech analysis, modification, and synthesis based on interference-free representations of speech parametric representations. In addition to our STRAIGHT-based infrastructure, which is a defacto standard in speech research, we developed a set of new independent algorithms. We made these algorithms as open-source. We elaborated on building supporting tools for promoting academic research using STRAIGHT systems. In addition to these planned accomplishments, we also established application infrastructure based on WaveNet, which revolutionized speech applications based on deep learning.

研究分野：聴覚メディア処理

キーワード：音声分析 音声合成 聴覚 感情音声 音声コミュニケーション 対話的研究環境 オープンソース

1. 研究開始当初の背景

本課題の申請書を作成していた平成 27 年の半ばの時点は、後に音声合成技術に革命をもたらす深層学習に基づく WaveNet が突然出現した時期にあっていた。しかし、これらの技術的進歩にもかかわらず、生物としての人間の音声コミュニケーションの本質を工学的に操作可能な形で理解することは、方法論を含め、困難な状況にあった。

2. 研究の目的

音声処理技術の進歩にも関わらず、生物としての人間と数学を基盤とするマシンの間には、出自に由来する埋めることのできない間隙が残されている。特に、悲鳴や慟哭の背景にある恐怖や深い悲しみ、優れた演劇や歌唱をもたらす感動を、機能的な意味で人間と同様に把握する方法論を確立することは、急速に技術の進歩する現在、喫緊の課題である。その研究の促進を狙い、対話的な可視化可聴化機能を有する研究支援環境を構築し公開する。

3. 研究の方法

研究代表者が発明し、内外の音声研究のデファクトスタンダードとなっている STRAIGHT とその情報表現に基づく操作環境を整備するとともに、近年の計算システムのリアルタイム処理能力を生かし、高度に対話的な可視化および可聴化の研究環境を用意する。また、それらの成果をオープンソースのシステムとして各段階で公開することにより、本課題の研究グループだけでは不可能な広範な分野をカバーするコミュニティーの形成を図る

4. 研究成果

本課題の最大の成果は、対話型のリアルタイム音声分析可視化と音声生成シミュレータを有する SparkNG という環境、STRAIGHT とは独立な原理に基づく音声分析・変換・合成システムである WORLD をオープンソースという形で公開し、それらを利用するコミュニティーを形成したことにある。以下、発表との関連を中心に、具体的な成果について説明する。

(1) 支援環境 SparkNG の公開と普及：平成 28 年度には、音声研究における最重要旗艦会議である Interspeech において、対話型のリアルタイム音声分析可視化と音声生成シミュレータを有する SparkNG という研究支援環境についての発表を行うとともに、日米を中心に、聴覚を始めとする広い分野の研究者が議論する ASA/ASJ Joint Meeting において、SparkNG の応用的側面について発表した。([学会発表][10][11])。

なお、構成要素の各々の技術には、研究の進展に伴い新たに発明した手法を取込み、継続的に更新を続けている。具体的には、平成 29 年度には、声帯音源モデルとして表現能力の高い Fujisaki-Ljungqvist model を組込むためのアルゴリズムを開発し、旗艦国際会議である Interspeech で発表した。([学会発表][6])

また、アジアを中心とする音声研究の国際会議の基調講演においても、その応用の可能性を示した。(学会発表[4]) これらの成果と拡張を踏まえ、SparkNG のソフトをオープンソースの流通・更新のプラットフォームである GitHub に公開した。([その他][1])

なお、これらの開発の過程で遭遇したデジタル信号処理アルゴリズムの実装における基本的な知見をまとめ、学術雑誌で公表した。

([雑誌論文][1]) また併せて、具体的なコードも公開した。([その他][3]) このように論文だけではなく、動作するアルゴリズムの実装を公開することは重要である。実際、これらを実装する段階で、既に発表された複数の学術論文の中に誤植があることを発見し、修正したものを正しいアルゴリズムとして実装し公開した。なお、それらの既存の論文は、全体としては正しいことを補足しておく。

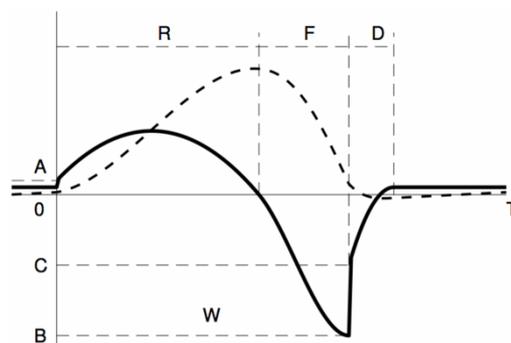


図 1 Fujisaki-Ljungqvist model。破線は声門を通過する呼気流、実線は音声の音源として寄与する成分。

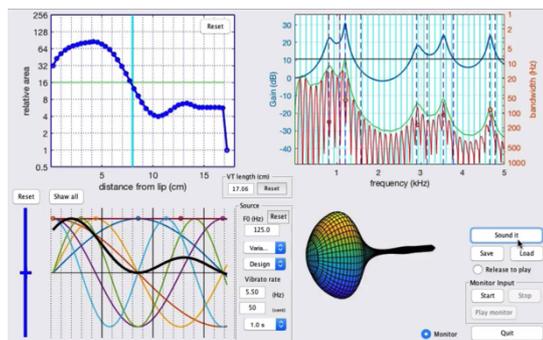


図 2 公開した対話的可視化可聴化環境 SparkNG のメイン GUI

(2) 基盤ソフトウェア WORLD の公開と普及：代表者が発明した STRAIGHT は、特許として登録していたことにより、オープンソースとして集合的知性を活用する道が閉ざされていた。この状況を変えるべく、分担者の森勢は、STRAIGHT とは独立なアルゴリズムに基づいて、新たな音声分析・変換・合成システム WORLD を発明した。STRAIGHT の経験を踏まえ、WORLD については、オープンソースとして公開することとした。具体的には、WORLD を前述の GitHub で公開するとともに([その他][2])、その様々

な応用可能性について機会を捉えて発表し普及に努めた。〔雑誌論文〕[1]、〔学会発表〕[8] その結果、WORLDの最初の論文は国内の雑誌にのみ掲載されたにも関わらず既に多数の研究に利用され引用されるに至っている。(Google Scholarで80件。2018.5.27) また、国際的なオープンソースプロジェクトや商用ソフトの重要な構成要素としても用いられるにいたっている。

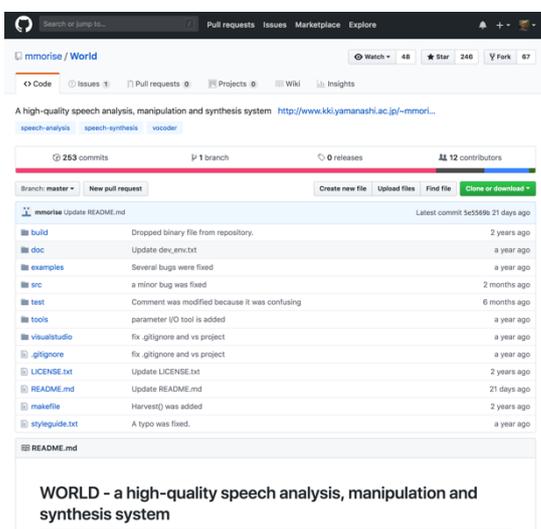


図 3 GitHub に公開した音声分析・変換・合成システム WORLD のページ。既に数多く利用されていることが分かる。



図 4 WORLD を応用した実時間基本周波数操作インターフェイス SOUND STONE。〔学会発表〕[8]

(3) その他の対話的環境：分担者の入野は聴覚末梢系における信号処理をモデル化した動的圧縮型 gammachirp filter bank (dCGCFB) の発明者であり、聴覚障害における音声知覚の劣化が、dCGCFB の適切なパラメタ設定により模擬できることを明らかにしてきた。本課題では、その成果を言語聴覚士の教育に還元すべく、対話的環境を構築し実際の教育の場における評価を進めている。〔学会発表〕[2][9]

(4) 本課題で発見／発明された個別技術：本課

題の研究を進める過程で、現実の課題解決に有用な技術が開発された。その一つは、SparkNG の声帯音源モデルの拡張の際に発見された、 \cos 級数で表される anti-alias 用の関数である。高い局在性と干渉の急速な減少により、瞬時周波数の精密な推定を可能にすることが明らかにされた。〔雑誌論文〕[2][学会発表][6]

また、対話的環境への導入には至っていないが、非周期性分を高い操作可能性と再現性を有する形で表現できる新しい信号を明らかにした。この信号は、聴覚研究用の試験信号としても多くの可能性を有するものであり、問題提起するとともに実装例を公開した。〔学会発表][1]

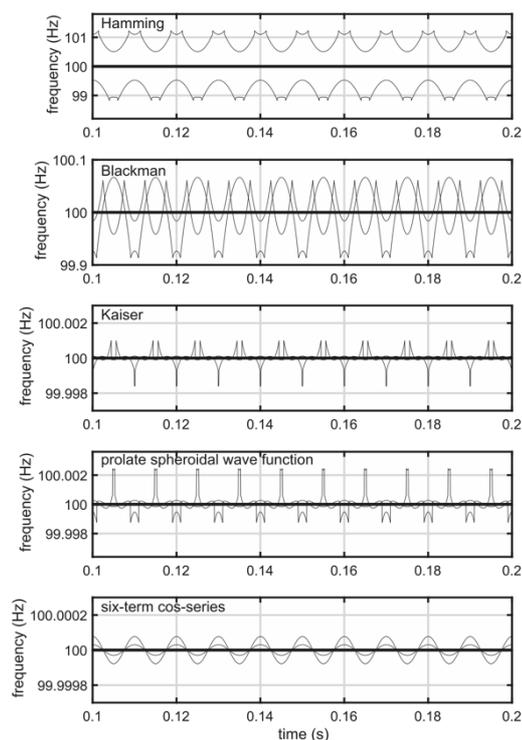


図 5 瞬時周波数の計算における新しい \cos 級数の効果。最下段が新しい関数による結果。

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 2 件)

(以下は、招待論文を含め査読有)

[1] Masanori Morise, Yusuke Watanabe, “Sound quality comparison among high-quality vocoders by using re-synthesized speech,” *Acoustical Science and Technology*, vol. 39, pp. 263–265, DOI: 10.1250/ast.39.263, オープンアクセス

[2] 河原英紀, “デジタル信号処理の落と

し穴,” 日本音響学会誌, vol. 73, pp. 592-599, DOI: 10.20697/jasj.73.9_592, (招待論文), オープンアクセス

[学会発表] (計 21 件)
(発表は筆頭著者)

[1] 河原 英紀、津崎 実、坂野秀樹、森勢将雅、松井淑恵、入野俊夫, “velvet noise とその変種の聴覚心理・生理研究への応用可能性について,” 日本音響学会聴覚研究会, 2018. 3.

[2] Toshio Irino, Kenji Yokota, Toshie Matsui, and Roy D. Patterson, “Incorporating absolute threshold and a cochlear noise floor into the GammaChirp model of masking,” ARO 41st midwinter meeting, 2018. 2. (国際学会)

[3] Hideki Kawahara, Eri Haneishi, Kaori Hagiwara, “Realtime feedback of singing voice information for assisting students learning music therapy,” 2017 International Conference on Orange Technologies, 2017. 12. (国際学会)

[4] Hideki Kawahara, “Making speech tangible for better understanding of human speech communication,” The 21th International Conference on Asian Language Processing, 2017. 12. (招待講演)(国際学会)

[5] Hideki Kawahara, Ken-Ichi Sakakibara, “Characterization of subharmonic voices using phase derivatives,” Pan-European Voice Conference, 2017. 9. (国際学会)

[6] Hideki Kawahara, K. Sakakibara, H. Banno, M. Morise, T. Toda, T. Irino, “A new cosine series antialiasing function and its application to aliasing-free glottal source models for speech and singing synthesis,” Interspeech 2017, 2017. 8-9. (国際学会)

[7] 河原 英紀, “VOCODER 再訪 ~ なぜ元の音声の位相を保存したくないのか ~,” 電子情報通信学会応用音響研究会, 2017. 7. (招待講演)

[8] 渡邊 優介, 森勢 将雅, 小澤 賢司, “高品質音声分析合成を用いた基本周波数の実時間操作インタフェースの実装,” 情報処理学会音楽情報科学研究会, 2017. 6.

[9] 米満麻弥, 入野俊夫, 松井淑恵, 西村竜一, 吐師道子, 長谷川 純, “模擬難聴システムを用いた言語聴覚士養成課程での演習と Web アプリ化の検討,” 電子情報通信学会ヒ

ューマン情報処理研究会, 2017. 5.

[10] Hideki Kawahara, Masanori Morise, Ryuichi Nisimura, Toshio Irino, “Realtime and interactive tools for speech and hearing science education,” ASA/ASJ Joint meeting, 2016. 11-12. (国際学会)

[11] Hideki Kawahara, “SparkNG: Interactive MATLAB tools for introduction to speech production, perception and processing fundamentals and application of the aliasing-free LF model component,” Interspeech 2016, 2016. 9. (国際学会)

[その他]

ホームページ等

[1] SparkNG: MATLAB realtime/interactive tools for speech science research and education, GitHub,
<https://github.com/HidekiKawahara/SparkNG>

[2] WORLD: a high-quality speech analysis, manipulation and synthesis system, GitHub,
<https://github.com/mmorise/World>

[3] 研究代表者のポータルサイト:
<http://www.wakayama-u.ac.jp/~kawahara/>

6. 研究組織

(1) 研究代表者

河原 英紀 (KAWAHARA, Hideki)
和歌山大学・学内共同利用施設等
・名誉教授
研究者番号: 40294300

(2) 研究分担者

入野 俊夫 (IRINO, Toshio)
和歌山大学・システム工学部・教授
研究者番号: 20346331

森勢 将雅 (MORISE, Masanori)
山梨大学・大学院総合研究部・准教授
研究者番号: 60510013

(3) 連携研究者

戸田 智基 (TODA, Tomoki)
名古屋大学・情報基盤センター・教授
研究者番号: 90403328

榊原 健一 (SAKAKIBARA, Ken-Ichi)
北海道医療大学・リハビリテーション科学部・准教授
研究者番号: 80396168

羽石 英里 (HANEISHI, Eri)

昭和音楽大学・音楽学部・教授
研究者番号： 70350684

坂野 秀樹 (BANNO, Hideki)
名城大学・理工学部・准教授
研究者番号： 20335003

(4)研究協力者

パターソン ロイ (Roy D. Patterson)
Cambridge 大学 (英国)

シュバインベルガー ステファン (Stefan
Schweinberger)
Jena 大学 (独)

エリス ダン (Dan Ellis)
Columbia 大学、Google (米)

マクダーモット ジョッシュ (Josh
McDermott)
MIT (米)