様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

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：　基盤研究(B)（一般）
研究期間：　2015～2017
課題番号：　１５Ｈ０２７２６

研究課題名（和文）聴覚情報の静的表現に基づく高度音声処理基盤の構築

研究課題名（英文）Development of high-level speech processing infrastructure based on static representations of auditory information

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）　　12,400,000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 revolutionalized speech applications based on deep learning.

研究分野：　聴覚メディア処理

キーワード：　音声情報処理　音声分析変換合成　基本周波数　声帯音源　情報表現　深層学習　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 26 年の半ばでは、音声認識・合成技術は深層学習の成果を取入れることにより、機械の介在を想定した音声であれば実用の域に達しつつあった。しかし、感情音声や障害音声、歌唱音声など、日常的に少なく無い頻度で生ずる広範な音声を処理することは、方法論を含め、困難な状況にあった。

２．研究の目的

本課題では、当時の技術では十分に研究することが困難な、感情音声や障害音声、歌唱音声など、広範な音声を、高い品質を保持したまま柔軟に処理するための計算基盤を構築し、広く公開することで音声の研究および応用を促進することを目的とした。

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 とその情報表現を出発点とし、多様な音声の分析・処理と応用開発のサイクルを迅速に回す。その過程では、逐次、開発されたシステムを公開することにより、内外の知見および応用からのフィードバックを取入れて、音声処理基盤を進化させる。また、音声コミュニケーションの主体である人間の聴覚末梢系の数理モデルを背景として利用する。さらに、世界の第一線で研究を進める実力およびプレゼンスの高いメンバーで研究組織を構成することにより、効率的に研究を進めるとともに、成果の普及を促進する。

４．研究成果

本研究課題の最大の成果の一つは、STRAIGHT とは独立に開発された音声分析・変換・合成の基盤となる WORLD というアルゴリズムである。STRAIGHT での経験を踏まえ、この WORLD をオープンソースとして公開することとした。その結果、当初想定することのできなかった広範な分野での応用が学術のみならず、商用システムとしても多数出現するに至っている。STRAIGHT も WORLD も、音声の有する周期性の本質的理解に基づくアルゴリズムが中核となっている。これらの方法を有効に動作させるためには、音声の周期性および周期性からの逸脱を客観的に高い信頼性で求めることが必要になる。これらの音源情報の信頼できる抽出法を新たに発見できたことも、本課題の大きな成果である。

また、平成 27 年 9 月に Google と DeepMind の研究者が発表した深層学習にもとづく WaveNet の革新性と重要性をいち早く把握し、多くの応用技術を開発したことも特記したい。さらに、最終年度において、将来に大きな波及効果を有する発見があった。

このように、本課題は当初予想していなかった大きな発明・発見により、当初計画を上回る成果を上げた。以下、具体的に主な成果について説明する。

(1)音声分析・変換・合成システム WORLD：研究代表者が発明した STRAIGHT は、音声を音源情報とスペクトル包絡情報に分解し、変換と再合成するための方法である。周期性による干渉を組織的に排除することにより、従来の方法では不可能であった高い品質での柔軟な操作を可能とする方法である。この高い品質と柔軟性により、STRAIGHT は音声研究のデファクトスタンダードとなっている。(引用件数 1,927：Google Scholar による) しかし、前述の特許による制約と計算の複雑さにより、応用が限られる状況もあった。

WORLD では、スペクトルの分析に用いられる窓関数の特殊な性質を利用することにより、STRAIGHT を凌ぐ品質と柔軟性を遥かに軽い処理で実現することに成功した。([雑誌論文][1][4][学会発表][15]) また、STRAIGHT での経験を踏まえて、GitHub のレポジトリにオープンソースとして公開することにより、世界中の利用者や開発者の知見や提案を取り入れることを可能にした。([その他][1]) その結果、複数の国際的な音声処理プロジェクトの中核技術として採用されるとともに、多数の商用システムにも組み込まれるに至っている。また WORLD を最初に提案した論文は、2 年を経ずに Google Scholar による引用件数が 80 となっている。

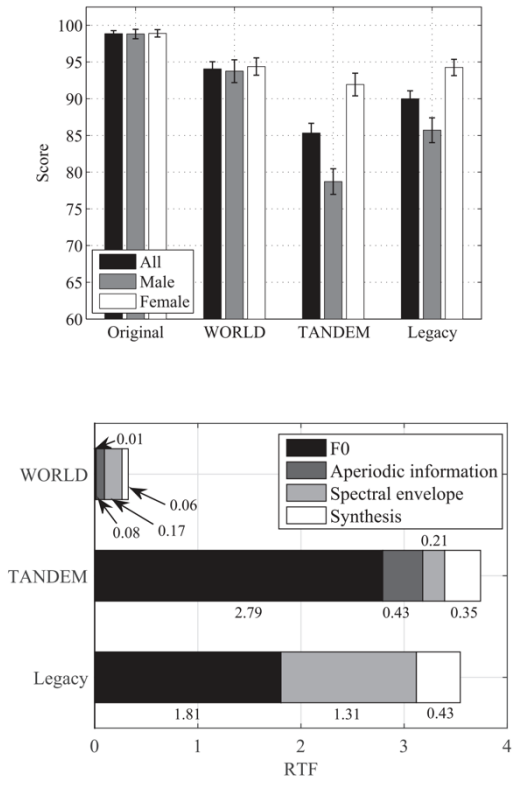

図 1 WORLD の品質（上）と処理時間(下) TANDEM と legacy は STRAIGHT の種類。下図の RTF は音声の時間長に対する処理時間の比。

図 2 WORLD の情報サイト。新たな技術を取り入れて継続的に更新している。

(2) 精密な音源情報の分析法：研究代表者が Google に滞在した際に開発した方法を発展させて、音源情報である基本周波数と非周期性分を同時に正確に抽出する方法を発明した。また、この方法をそれまで精密な研究が困難であった音声の分析に応用した。([雑誌論文][2][学会発表][2][4][8])
また、これに先立って、WORLD における音源情報の分析方法として音声の位相情報のより理解しやすい表現である群遅延の特殊な性質を利用した方法を開発するとともに、それら音源情報の分析方法の客観的評価法を明らかにした。([雑誌論文][3][学会発表][12])

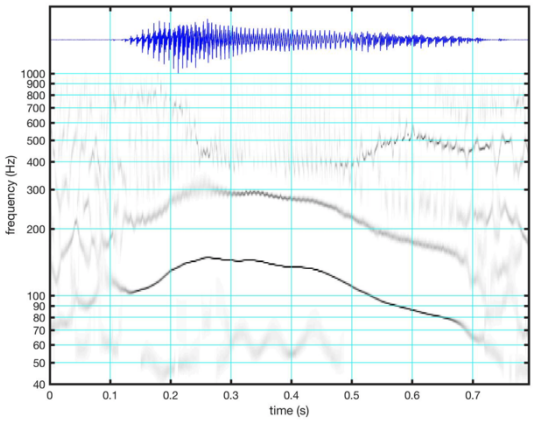

図 3 精密な音源情報の分析例。連続母音「あいうえお」の波形と基本周波数の存在確率が表示されている。

(3) 深層学習に基づく WaveNet の技術基盤：それまでの音声処理技術を根本的に革新する WaveNet の Google と DeepMind による発表は、旗艦国際会議である Interspeech 2016 の直前になされた。引き続くワークショプでの WaveNet 議論に、本課題の分担者である戸田は積極的に関与し、その後の我が国での研究を主導するに至っている。戸田は、これまで自らが主導して進めてきた統計的音声合成法も、統計的手法に基づく音声変換も、品質が高い方法として導入したフィルタ処理による音声変換方式も、全て WaveNet に基づいて開発した技術により置換えることにより、画期的に品質が改善されることを具体的に実証した。([学会発表][6][10][13]) この WaveNet の出現と成功は、本課題を提案した時には予想することができなかった革命とも呼べる変化である。この技術の重要性をその出現から直ちに把握し、音声処理の様々な分野における応用を明らかにして技術基盤を整備したことは、本課題の大きな成果である。

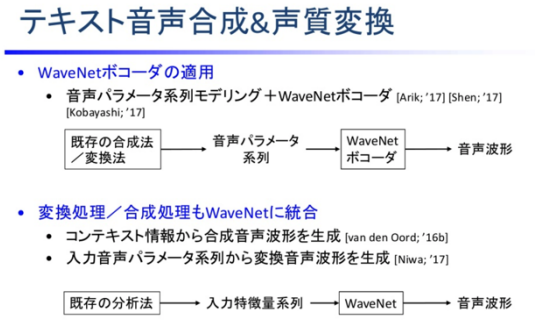

図 4 WaveNet による音声合成と変換の置換え。

(4) 支援技術の開発：音声の教育研究を支援するための対話的可視化可聴化環境 SparkNG の開発を進めている。この環境の拡張の過程で、音源情報抽出法の客観評価の精度を向上させるだけではなく、抽出法自体の基盤を強化することのできる cos 級数で表される関数を発見し実装した。([学会発表][2][8][9])
また、人間の聴覚末梢系の数理モデルである動的圧縮型 gammachirp filter bank を用いることで、これまで困難であった音声処理系での信号の（広義の）歪みによる明瞭度の劣化を予測する方法を提案した。([学会発表][5])

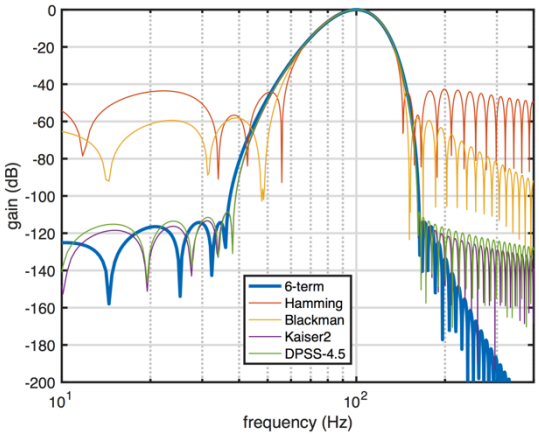

図 5 新たな関数（cos 級数：図中では 6-term）の選択性。細線で示した既存の関数と比較し、細かな変動を示す部分が急速に減衰している。

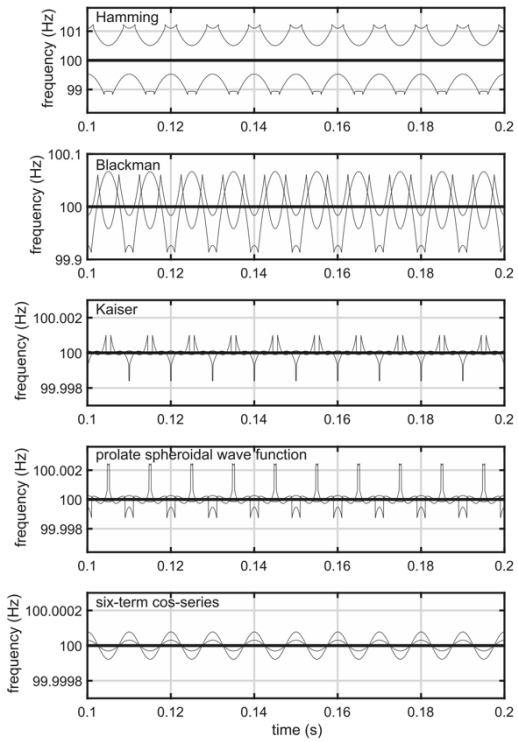

図 6　新たな関数を利用することによる瞬時周波数計算の精密化。細線が誤差を示す。縦軸の単位に注目。

(5)技術基盤の広報・普及：メンバーの内外での高いプレゼンスを利用し、招待講演などの機会をとらえて本課題の成果の広報と普及に努めた。([学会発表][3][13][15]) また、これらの招待講演とは別に、研究代表者がアジア太平洋・信号情報処理学会(APSIPA)の Distinguished Lecturer として選ばれたことを利用し、大学を含む様々な研究機関において講演し、その中で本課題の成果を紹介した。([その他][4]) この際には、オープンソースの公開用の GitHub と、研究代表者が更新しているポータルサイトを有効に活用した。([その他][1][2][3])

(6)新たな信号の発見：本課題の成果として構築している技術基盤に組込むことには間に合わなかったが、音声合成・変換ならびに人間の聴覚機構の研究に大きく貢献すると考えられる信号を発見した。これは当初の計画では想定されていなかった発見であり、研究代表者が成果発表のために参加した学会([学会発表][2]) での議論で知った信号から派生したものである。この信号（正規乱数による雑音よりも滑らかな雑音として知覚されるために velvet noise と名付けられている）の生成過程を周波数領域で再構築し変形することによって、新たな有用な信号を構成できることを発見した。([学会発表][1]) 本課題の終了後になるが、その重要性に鑑み、音声処理・研究基盤として公開する準備を進めている。

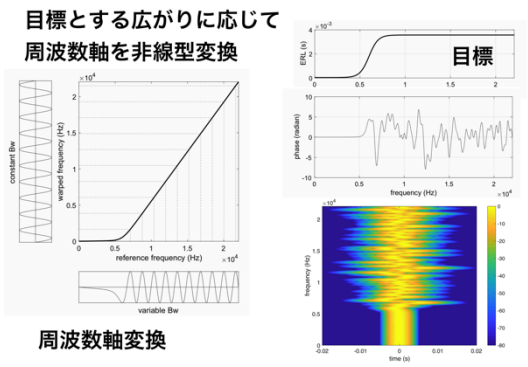

図 7 Velvet noise から派生した新たな信号。周波数帯域ごとのランダム性の広がりを自由に設定できる。

(6)応用事例：本課題の研究期間の終了直後に、本課題で開発している高度音声処理基盤の応用に関わる事例が、学術的に重要な論文が掲載される Nature Communications においてオンラインで公刊された。([その他][5]) この論文につながる研究は 2012 年に開始されたものであり、本課題の成果である音声処理基盤がどのように応用できるかの具体例となっている。この論文で用いられた拡張機能（任意の駆動信号を合成音源として用いることができる）は、本課題の成果である処理基盤に組み込まれており、利用者がそれぞれの研究に応用することができる。

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　6 件）
（以下は全て査読有）

[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, 2018, DOI: 10.1250/ast.39.263, オープンアクセス

[2] Hideki Kawahara, “Application of time-frequency representations of aperiodicity and instantaneous frequency for detailed analysis of filled pauses,” Journal of the Phonetic Society of Japan, vol.21, pp.63-73, 2017, DOI: 10.24467/onseikenkyu.21.3_63, オープンアクセス

[3] Masanori Morise, “D4C, a band-aperiodicity estimator for high-quality speech synthesis,” Speech Communication, vol.84, pp.57-65, 2016, DOI: 10.1016/j.specom.2016.09.001, オープンアクセス

[4] Masanori Morise, Fumiya Yokomori,

Kenji Ozawa, “WORLD: a vocoder-based high-quality speech synthesis system for real-time applications,” IEICE transactions on information and systems, vol.E99-D, pp.1877-1884, 2016, DOI: 10.1587/transinf.2015EDP7457, オープンアクセス

〔学会発表〕(計 66 件)
(以下、発表者は筆頭著者)
[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] Hideki Kawahara, Ken-Ichi Sakakibara, Masanori Morise, Hideki Banno, Tomoki Toda, “Accurate estimation of fo and aperiodicity based on periodicity detector residuals and deviations of phase derivatives,” APSIPA ASC 2017, 2017.12. (国際学会)

[3] 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)

[4] Hideki Kawahara, Ken-Ichi Sakakibara, “Characterization of subharmonic voices using phase derivatives,” Pan-European Voice Conference, 2017.9. (国際学会)

[5] Katsuhiko Yamamoto, Toshio Irino, Toshie Matsui, Shoko Araki, Keisuke Kinoshita, and Tomohiro Nakatani, “Predicting speech intelligibility using a gammachirp envelope distortion index based on the signal-to-distortion ratio,” Interspeech 2017, 2017.9. (国際学会)

[6] A. Tamamori, T. Hayashi, K. Kobayashi, K. Takeda, T. Toda, “Speaker-dependent WaveNet vocoder,” Interspeech 2017, 2017.9. (国際学会)

[7] Toshie Matsui, Toshio Irino, Kodai Yamamoto, Hideki Kawahara, Roy D. Patterson, “The effect of spectral tilt on size discrimination of voiced speech sounds,” Interspeech 2017, 2017.9. (国際学会)

[8] Hideki Kawahara, Ken-Ichi Sakakibara, Masanori Morise, Hideki Banno and Tomoki Toda, “A Modulation Property of Time-Frequency Derivatives of Filtered Phase and its Application to Aperiodicity and fo Estimation,” Interspeech 2017, 2017.9. (国際学会)

[9] 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, 2017.9. (国際学会)

[10] 玉森 聡, 林 知樹, 戸田 智基, 武田一哉, “音声生成過程を考慮した WaveNet に基づく音声波形合成法,” 電子情報通信学会/日本音響学会 音声研究会, 2017.3.

[11] Hideki Kawahara, “Realtime and interactive tools for speech and hearing science education,” ASA/ASJ Joint meeting, 2016.11-12. (国際学会)

[12] Masanori Morise, Hideki Kawahara, “TUSK: A framework for overviewing the performance of F0 estimators,” Interspeech 2016, 2016.9. (国際学会)

[13] 戸田智基, “音情報処理における特徴表現,” 音学シンポジウム 2016, 2016.5. (招待講演)

[14] H. Kawahara, K. Sakakibara, H. Banno, M. Morise, T. Toda and T. Irino, “Aliasing-free implementation of discrete-time glottal source models and their applications to speech synthesis and F0 extractor evaluation,” APSIPA ASC 2015, 2015.12. (国際学会)

[15] 森勢将雅, “目指せ音声分析合成マスター!「よくわからない」から「ちょっとわかる」へのチュートリアル,” 日本音響学会聴覚研究会, 2015.11. (招待講演)

〔図書〕(計 0 件)

〔産業財産権〕

○出願状況 (計 0 件)

〔その他〕
ホームページ等
[1] WORLD: a high-quality speech analysis, manipulation and synthesis system, GitHub, https://github.com/mmorise/World

[2] SparkNG: MATLAB realtime/interactive tools for speech science research and education, GitHub,

https://github.com/HidekiKawahara/SparkNG

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

講演活動
［4］ APSIPA 2015-2016 Distinguished Lecturer

応用事例［雑誌論文］
［5］Sara Popham, Dana Boebinger, Dan P. W. Ellis, Hideki Kawahara, Josh H. McDermott, “Inharmonic speech reveals the role of harmonicity in the cocktail party problem,” Nature Communications, vol.9(1), Article number.2122, 2018.5.29 公開, DOI: 10.1038/s41467-018-04551-8, オープンアクセス、（査読有）
https://www.nature.com/articles/s41467-018-04551-8

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

西村　竜一（NISHIMURA, Ryuichi）
和歌山大学・システム工学部・助教
研究者番号： 00379611

松井　淑惠（MATSUI, Toshie）
豊橋技術科学大学・工学研究科・准教授
研究者番号： 10510034

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

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

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

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

(3)連携研究者

(4)研究協力者
シュバインベルガー　ステファン（Stefan Schweinberger）
Jena 大学（独）