

**科学研究費助成事業 研究成果報告書**

平成 28 年 6 月 17 日現在

機関番号：14701

研究種目：挑戦的萌芽研究

研究期間：2014～2015

課題番号：26540087

研究課題名(和文)聴覚の情報表現に基づく機能性音声デザイン機構の研究

研究課題名(英文)Functional speech design framework based on auditory information representation

## 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) 2,800,000円

研究成果の概要(和文)：申請者の発明によるSTRAIGHTと時変多属性任意事例数モーフィングを利用することにより、様々な機能を有する音声による表現を探索的に研究するための基盤を実現した。また、それらの利用を促進するために、必要となる音声の物理特性、発声・調音器官、聴覚機構と機能に対話的に探求することのできる環境を構築し、オープンソースとして公開した。また、具体的な機能音声の実現例として、強い印象を与える歌唱で多用されるグロウル発声の特徴を、任意の歌唱に付加することのできるシステムを実現した。

研究成果の概要(英文)：We introduced a new foundation of interactive investigation of functional speech sounds, based on our inventions, STRAIGHT, a speech analysis, modification and resynthesis framework, and a generalised morphing framework, which can morph arbitrarily many voices with individual and temporal control of constituent attributes. We also introduced a set of realtime interactive tools for providing understanding of speech production physics, physiology of voicing and articulatory organs, auditory perception and engineering representations. We also made the latter tools open sourced. Finally, we implemented a system to add growl-like impression to any singing voices as an example of our proposed framework.

研究分野：音声情報処理

キーワード：音声情報処理 音声分析 音声変換 パラ言語情報 信号処理 音声生成 音声知覚

### 1. 研究開始当初の背景

申請者により開発された STRAIGHT と、それが可能にしたモーフィングは、画期的な音声処理技術として、学術だけではなく応用を通じて様々な分野に大きなインパクトを与えてきた。しかし、それにもかかわらず、例えば、内容を良く理解し記憶できる声、より深い感動を与える声、心に平安を与える声、警戒心を喚起する声など、様々な用途に適した機能を持つ声を自由に創り出すことは、未だに手の届かないところにあった。音声合成の分野では、当時も今も統計的手法に基づいて様々な声質を制御する研究が主流となっている。しかし、この方法論の根底にある平均化は、個別的で豊かな声の表現の追求とは対極的なものとならざるを得ない。

申請者らの手法であるモーフィングは、逆に個別性に立脚することにより統計的手法を補完するものと位置づけられる。このような状況の中で、モーフィングを時変かつ多属性で任意事例数の外挿を含む場合に拡張することに成功したことが、本課題の提案につながったのである。

### 2. 研究の目的

自然な文字である手書き文字よりも、適切にデザインされた活字は遥かに読み易い。音声言語も、適切にデザインすることにより、自然な肉声そのままのものよりも、明瞭に聴き取ることができ、内容を良く理解し記憶でき、より深い感動を得、心の平安を得、警戒心を喚起できるなど、用途に応じたものとすることができる可能性がある。本研究課題では、申請者らが進めている、聴覚と同型の情報表現に基づく最新の音声分析・変換・合成技術( STRAIGHT を分析・合成の基盤とし、時変多属性かつ任意事例数での外挿を可能とするように拡張されたモーフィング技術)を利用することにより、用途に応じた機能を有する音声のデザインを可能とするシステムのプロトタイプ構築を目的とする。

### 3. 研究の方法

まず、申請者らが発明した TANDEM-STRAIGHT を中核として、拡張された時変多属性任意事例数への適用と外挿が可能となったモーフィングおよび高速高精度高時間分解能の新しい F0 抽出法を加えて統合することにより、前意識的影響を定量化するための基盤となる対話的システムを構築する。このシステムを用いて、卓越した声を含む多様な音声資料の分析と探索的検討を進めることにより、目標とする機能を有する音声デザインするための基礎データを蓄積する。

これらの基礎データに基づき、機能性音声デザインのためのパラメタ操作の体系とインタフェースを構築するとともに、目的とする機能を有する音声を実現するデザインをテンプレートとして類型化する。なお、近赤外線を用いた脳機能の観測は、交付額を最も

効果的に用いる観点から、今回の計画から削除した。

### 4. 研究成果

まず、機能音声デザイン研究を支える基盤となるシステムの構築を進めた。ここでは、高い時間分解能を有する高精度な基本周波数分析アルゴリズムを TANDEM-STRAIGHT と併用することにより、グロウル系の歌唱音声などの印象の操作を可能とした。また、時変多属性任意事例数モーフィングを容易にするためのツール群の整備と、グラフィカルユーザインタフェース (GUI) の整備を進めた。このシステムは、高度で柔軟な操作を可能にするために自由度を高めているため、適切な操作には、音声生成・知覚・音声信号処理に関する総合的な理解が必要となる。そのために、音声生成・知覚・音声パラメタ間の理解を支援するための環境を構築し、オープンソースとして公開した([ホームページ等])。このツールは、科学技術計算環境である MATLAB の上に構築することにより、高度な可視化機能を実現している。その中でも、話されている音声の声道の形状を実時間で可視化するツールは、英国 Edinburgh 大学でも利用されるなど、大きな反響を呼んでいる。

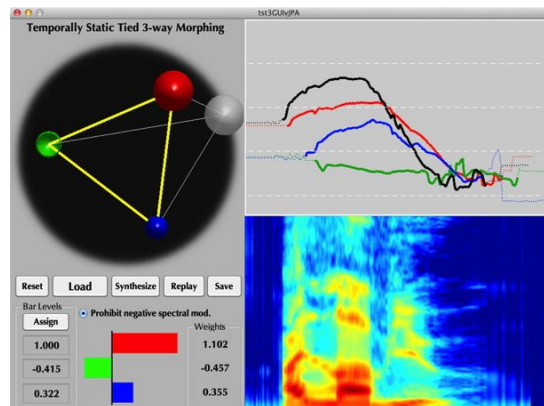


図 1 モーフィング操作 GUI

図 1 に、三種類の音声資料のモーフィングを対話的に操作する GUI の例を示す。ここでは、「怒り」「喜び」「悲しみ」の三種類の表現で発声された音声から、それらを混合した音声を合成している。画面の左上の半透明の灰色の球をマウスなどでドラッグ操作することにより、三種類の表現の混合比率を自由に変更し、ボタンの解放により即座にモーフィング音声合成される。この GUI では、音声を表現している五種類のパラメタの操作量を同一にし、また時間方向では定数とするという制限を加えており、すべての自由度を利用してはいない。

図 2 に、音声生成・知覚・音声パラメタ間の理解を支援するための環境に含まれる、声道形状可視化ツールの例を示す。ツールには、その他に、音声の時間周波数表現の基本であるフーリエ変換を理解するための対話的可視化可聴化ツール、聴覚での周波数分析を実時間で可視化するツールなどが含まれてい

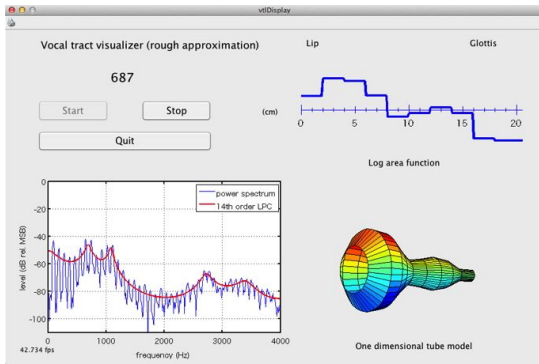


図 3 声道形状の実時間可視化ツール

る。図 2 の左下は、音声のスペクトルを青線で、そのスペクトルを最適に近似するモデルの特性を赤線で示している。右上は、モデルから求められた声道の形状を、喉から唇に至る音の伝搬経路である声道の断面積として示している。右下は、その形状を三次元表示することで、直感的に把握できるようにしたものである。

こうした基盤システムの構築と並行して、実用上の効用が期待される「高感度の高い音声のデザイン」の研究と、「歌唱音声へのグロウル系の印象の付与」の研究を、具体的なターゲットとして研究を進め、システムとして実現した。またそれらの効果の評価のために、被験者を用いた受聴試験を行い、基礎データを蓄積した。

これら当初計画に挙げた成果に加え、本課題を推進する過程で、想定を超える大きな成果が得られた。その一つは、TANDEM-STRAIGHT よりも遥かに少ない計算量で同等の分析合成を可能にする方法の発明である。この方法全体を WORLD という名前のシステムとして実現し、オープンソースとして公開した〔ホームページ等〕。また、その中核となるスペクトル推定方法である CheapTrick については、オープンアクセスの論文として刊行した〔雑誌論文〕。

もう一つの当初計画を超える成果は、信号のデジタル化に伴う不要信号（エアリング）を含まない音声の音源モデル（L-F モデル）の発明である。L-F モデルは 1985 年に提案されて以来、音声合成および知覚研究に広く用いられているにも関わらず、デジタル化の際の基本問題である不要信号の発生を避ける方法が明らかにされないままであった。上記の研究支援環境を構築する際に、この問題に遭遇し、明示的な数式として根本的な解を得ることに成功したのである。この結果を国際会議で発表〔学会発表〕するとともに、支援環境に組み込んでオープンソースとして公開した〔ホームページ等〕。なお、これらの当初計画を超える成果は、音声生成・知覚研究を大きく促進する可能性を有する基本的なものであり、その普及を促すことを狙う新たな挑戦的萌芽研究の提案につながったことを付記する。

図 3 に、L-F モデルを様々な方法でデジタ

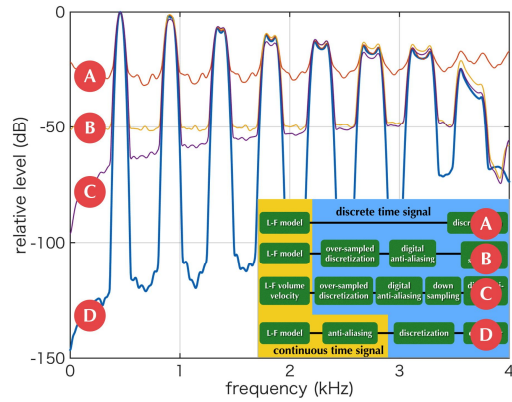


図 2 不要信号を含まない音源信号

ル化する際に生ずる不要信号の例を示す。串歯状に見える調波成分の間にあるのが、不要信号である。A は、元の数式から直接デジタル化したもの、B, C は、目的とする標本化周波数の 6 倍の周波数でデジタル化し、間引きを行った場合、D が今回解いた数式を用いてデジタル化した場合である。

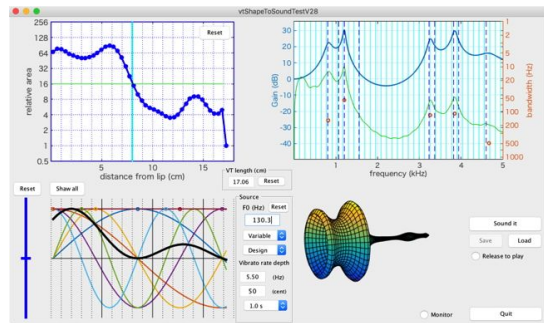


図 4 音声生成過程のシミュレータ GUI

図 4 は、このモデルを組み込んだ、音声生成過程のシミュレータの GUI を示す。このモデルでは、声道形状、声道のサイズ（長さ）、L-F モデルのパラメタに加え、声道伝達特性の極周波数を対話的に操作して、音源を合成することができる。

## 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)

溝淵 翔平、西村 竜一、松井 淑恵、入野 俊夫、河原 英紀: 声道形状と声帯音源特性の操作に基づいたグロウル系歌唱の印象付与法、電子情報通信学会 論文誌 D ,Vol. J99-D, No. 3, pp.283-292, Mar. 2016. (査読有)

Masanori Morise and Kenzi Ozawa: Speaker identification framework by peripheral and central auditory models, Acoust. Sci. & Tech., vol. 36, no. 4, pp. 340-343, July 2015. (査読有)

Masanori Morise: Error evaluation of an F0-adaptive spectral envelope

estimator in robustness against the additive noise and F0 error, IEICE transactions on information and systems, vol. E98-D, no. 7, pp. 1405-1408, July 2015. (査読有)

Masanori Morise: CheapTrick, a spectral envelope estimator for high-quality speech synthesis, Speech Communication, vol.67, pp.1-7, 2015. (査読有)

河原 英紀、音声分析変換合成基盤ソフトウェア STRAIGHT とその応用、コンピュータソフトウェア、日本ソフトウェア科学会、Vol.32, No.3, pp.3\_23-3\_28, 2015. (査読有)

河原 英紀：音声の実時間表示とモーフィングで探る声の多様性、音声研究、18巻、pp.578-588、2014。(査読無、招待)

[学会発表](計 30 件)

横森 文哉、二宮 大和、森勢 将雅、田中 章浩、小澤 賢司：好感度の男女差に着目した女性発話のスペクトル分析、日本音響学会 2016 年春季研究発表会、pp. 281-284, Kanagawa, March 9-11, 2016. H. Kawahara, K-I. Sakakibara, H. Banno, M. Morise, T. Toda, T. Irino: Aliasing-free implementation of discrete-time glottal source models and their applications to speech synthesis and F0 extractor evaluation, Proc. APSIPA2015 ASC, Hong Kong, Dec. 16-19, 2015.

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、vol. 45, no. 8, pp. 697-702, Yamanashi, Nov. 13-14, 2015. (招待)

横森 文哉、二宮 大和、森勢 将雅、田中 章浩、小澤 賢司：女性発話を対象とした好感度評価の男女差の検討、日本音響学会聴覚研究会、vol. 45, no. 8, pp. 631-636, Yamanashi, Nov. 13-14, 2015.

横森 文哉、森勢 将雅、小澤 賢司：好感度に着目した発話音声の音響的特徴分析の検討、日本音響学会 2015 年秋季研究発表会、pp. 251-252, Fukushima, Sept. 16-18, 2015.

河原 英紀、榊原 健一、坂野 秀樹、森勢 将雅、戸田 智基、入野 俊夫：音声科学教育用対話的ツールのための エリアシングの無い L-F モデルの実装について、電子情報通信学会技術研究報告、EA2015-08, , 東北大学、仙台、3-4 Aug. 2015.

森勢 将雅：帯域毎の非周期性指標推定法とその誤差評価、電子情報通信学会技術研究報告、vol. 115, no. 99, pp. 13-18, Niigata, June 18-19, 2015.

横森 文哉、大柴 まりや、森勢 将雅、

小澤 賢司：スペクトル包絡情報を入力とした Deep Neural Network に基づく歌声のための声質評価、情報処理学会音楽情報科学研究会、vol. 2015-MUS-107, no. 61, pp. 1-6, Tokyo, May 23-24, 2015.

Hideki Kawahara: Speech analysis modification and synthesis tool STRAIGHT and extended voice morphing, ARO midwinter meeting, Baltimore USA, Feb. 21-25, 2015. (招待)

森勢 将雅：基本波検出に基づく F0 推定法の耐雑音性向上、情報処理学会音声言語情報処理研究会、vol.2016-SLP-110, no.5, pp.1-6, Toyama, Feb. 5-6, 2015.

Hideki Kawahara, Masanori Morise, Ken-Ichi Sakakibara, Tomoki Toda, Hideki Banno Ryuichi Nisimura, Toshio Irino: Excitation source design for high-quality speech manipulation systems based on a temporally static group delay representation of periodic signals, APSIPA ASC 2014, Siem Reap, Cambodia, 9-12 Dec., 2014. Hideki Kawahara: STRAIGHT speech analysis, Tutorial of APSIPA ASC 2014, Siem Reap, Cambodia, 9-12 Dec., 2014. (招待)

Hideki Kawahara, Tatsuya Kitamura, Hironori Takemoto, Ryuichi Nisimura, Toshio Irino: Vocal tract length estimation based on vowels using a database consisting of 385 speakers and a database with MRI-based vocal tract shape information, Proc. Interspeech2014, Singapore, 14-18 Sept., 2014.

溝淵 翔平、伊佐 衣代、西村 竜一、入野 俊夫、河原 英紀：グロウル系歌唱から求められる声道断面積関数の特徴について、日本音響学会聴覚研究会、Vol.44, No.7, H-2014-80, pp.431-432, 南紀白浜温泉ホテルシーモア、和歌山、23-24 Oct. 2014.

河原 英紀：周期信号の静的表現に基づく音声処理、日本音響学会聴覚研究会、Vol.44, No.7, H-2014-80, pp.467-472, 南紀白浜温泉ホテルシーモア、和歌山、23-24 Oct. 2014. (招待)

吉元 照貴、伊佐 衣代、溝淵 翔平、西村 竜一、入野 俊夫、河原 英紀：声道形状と音源情報に注目した音声の好感度改善システムの検討について、日本音響学会 2014 年秋季研究発表会、2-Q-46, pp. 373-375, 北海道学園大、札幌、3-5 Sept. 2014.

河原 英紀：Matlab のリアルタイム処理を利用した音声聴覚教育用ツールについて、日本音響学会 2014 年秋季研究発

表会、北海道学園大、札幌、3-5 Sept. 2014.

河原 英紀: 音声・聴覚の理解と信号処理、音学シンポジウム、日本大学、東京、24-15 May 2014. (招待)

〔図書〕(計 1 件)

Hideki Kawahara: Temporally variable multi attribute morphing of arbitrarily many voices for exploratory research of speech prosody, eds. Keikichi Hirose and Jianhua Tao, Speech Prosody, pp.109-120, 2015.

〔その他〕

ホームページ等

<http://www.wakayama-u.ac.jp/~kawahara/MatlabRealtimeSpeechTools/>

[http://www.wakayama-u.ac.jp/~kawahara/STRAIGHTadv/index\\_j.html](http://www.wakayama-u.ac.jp/~kawahara/STRAIGHTadv/index_j.html)

<http://ml.cs.yamanashi.ac.jp/world/>

<https://github.com/mmorise/World>

## 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

松井 淑恵 (MATSUI, Toshie)

和歌山大学・システム工学部・助教

研究者番号：10510034

森勢 将雅 (MORISE, Masanori)

山梨大学・総合研究部・助教

研究者番号：60510013