

科学研究費助成事業 研究成果報告書

平成 26 年 6 月 13 日現在

機関番号：13501

研究種目：若手研究(B)

研究期間：2011～2013

課題番号：23700221

研究課題名(和文)高齢者・若年者の聴覚バリアフリーを実現する次世代音声明瞭化技術の研究

研究課題名(英文) Research on next-generation speech enhancement for auditory barrier-free between elderly and younger persons

研究代表者

森勢 将雅 (MORISE, Masanori)

山梨大学・医学工学総合研究部・助教

研究者番号：60510013

交付決定額(研究期間全体)：(直接経費) 3,300,000円、(間接経費) 990,000円

研究成果の概要(和文)：本研究課題では、聞き取りにくい発話を聞き取りやすい発話に変換し、誰でも聞き取りやすい発話を可能にする聴覚バリアフリーを目指した検討を実施した。研究期間の前半では、聴感的印象に対応するパラメータの策定を行い、最終年度で発話の聴感的印象を改善する具体的なアルゴリズムの開発と、提案したアルゴリズムの評価を実施した。その他にも、シャウトや掠れ声など、従来の音声分析法では処理が困難な音声进行分析のためのアルゴリズムや雑音除去法を提案し、有効性を示した。

研究成果の概要(英文)：A speech enhancement technology for auditory barrier-free between elderly and younger persons was introduced. In the first study, an acoustic feature associated with the auditory impression was investigated. Then, the enhancement technology based on the feature was proposed and evaluated. The results suggested that the proposed technology was effective to improve the auditory impression. Furthermore, speech analysis methods for atypical speech such as shout and hoarse speech and noise suppression methods for noisy speech were also proposed. Evaluations for these methods were carried out, and their effectiveness was demonstrated.

研究分野：総合領域

科研費の分科・細目：情報学・知覚情報処理・知能ロボティクス

キーワード：音声合成 声質変換 音声分析 聴覚

## 1. 研究開始当初の背景

加齢による視聴覚能力の低下は古くから知られており、衰えた能力を補うための技術が研究されてきた。視覚の低下に関しては眼鏡による補正が一般的であり、聴力の低下では補聴器による改善がなされてきた。しかしながら、補聴器は基本的にオーダーメイドで作ることが必要となるうえ、改善効果は眼鏡と比較すると限定的であった。加えて、高齢者の発話そのものが若年者と比較して聞き取りにくいことによる、若年者・高齢者間、高齢者・高齢者間での会話に対するストレスの蓄積が、未だに課題として残されている。本課題を解決するためには、根底にある「発話における聴感的印象（一般的に定義される明瞭度と区別するため、「聴感的印象」と呼称することとした）について基礎から見直す研究を実施し、その検討内容に基づいて聴感的印象を改善する声質変換技術を実現することが望ましいといえる。

補聴器は主に、聴取者側の能力低下を補うための技術である。一方、発話する側の原因で、聴力とは無関係に聞き取りにくい音声を発話したものに対する補助技術は未だ実現されていない。発話と聴感的印象に関する研究事例は複数存在しているが、声質変換による改善を意識した実用例は今のところ存在しない。このような研究を実施するためには、音声に含まれる聴感的印象に関連するパラメータを、声質変換に利用されるパラメータ（本研究ではVocoderに基づく声質変換を目指すため、基本周波数、スペクトル包絡、非周期性指標とする）として抽出する必要がある。

本研究では、これらの課題を解決し、高齢者に限らず、聞き取りにくい話者の発話内容を聞き取りやすくする聴感的印象の改善技術の開発に取り組む。本技術により、高齢者・若年者、高齢者・高齢者間のストレスのない「聴覚バリアフリー」を実現することは、介護施設におけるトラブル、および会話におけるストレスの低減など、未来の高齢化社会を支えるための重要な役割を担う。

## 2. 研究の目的

補聴器のように音圧の調整という受聴者に対するアプローチではなく、聞き取りにくい発話を聞き取りやすい声に変換するアプローチで本課題に取り組む。したがって、初めに、音声に含まれる聴感的印象に相当するパラメータの抽出について検討し、声質変換を実施するための基盤を構築する。また、本研究を通じて、音声分析、合成において従来研究では扱うことが困難であった「かすれ声」や「声帯異常を伴う音声」の検出などにも取り組む。これは、高齢者の発話が、若年者と比較して嚙れ声となりやすい傾向があるため、高齢者の音声分析法として重要であ

るという判断に基づく。

本取り組みは、外的要因により発話が困難となった話者との会話を支援することが可能となる。多数の音声を分析合成すること、また、将来的な目標となる実時間で変換するアプリケーションの実用化を視野に入れ、音声分析、合成技術の高速化、高精度化に加え、雑音除去などの信号処理技術についても取り組む。これらの研究を通じて、従来研究では扱うことが困難であった音声処理の重要性を示す。また、これらの分析方式については積極的に外部に公開することで、成果の還元を試みる。

## 3. 研究の方法

本研究では、主に、(1) 音声分析合成方式の高速・高精度化、および雑音除去などの信号処理技術の開発、(2) 音声に含まれる聴感的印象に対応する特徴量の策定、抽出法の検討、および(3) 求められた音声特徴量を制御することによる聴感的印象改善技術の開発に取り組んだ。以下では、それぞれの具体的な検討課題について概説する。

### (1) 音声信号処理に関する基礎研究

聴感的印象のように繊細な情報を扱うためには、高精度な音声分析方式が求められる。また、実用性の観点から実環境での利用に耐えうるシステムとするため、収録された音声に含まれる雑音を抑圧するための信号処理技術についても検討する。これらの検討に合わせ、聴感的印象に関連するパラメータに特化した分析技術についても検討する。

### (2) 音声の聴感的印象に対応するパラメータ策定

音声収録を実施し、様々な聴感的印象を持つ音声を収集する必要がある。任意の言葉とするのではなく、研究代表者が設定したテキストを読み上げさせることで、分析を容易にする。また、収録された音声の聴感的印象を主観評価でクラスタリングし、共通する特徴量についても検証する必要がある。また、音声収録時には聴感的印象に関する教示を行い、話者がイメージする聴感的印象が物理パラメータとして共通性を持つかについても検証する。

### (3) 聴感的印象改善手法の提案と評価

ここまでの検討により得られた音声特徴量を制御することで、聴感的印象を改善する方法を提案し、有効性について評価する。変換技術については、目標である会話支援への適用を容易にするため、入力された音声を加工するアプローチとする。(2)で収録された音声を分析し、聴感的印象に対応する音声パラメータを変換するルールを構築する。構築されたルールにより、あらゆる入力音声の聴感的印象を改善させる方法を実現する。

#### 4. 研究成果

##### (1) 音声信号処理に関する基礎研究

音声分析合成システム TANDEM-STRAIGHT の改良と GPGPU 版の公開, さらにアルゴリズムを改良した WORLD を提案し, 有効性を検証した. 本技術により, 音声に含まれる基本周波数, スペクトル包絡, 非周期性指標を, 元の音声を再現可能となるレベルで抽出することを可能にした. これらの分析法は, 聴感的印象を改善する技術の基盤を与える重要な技術となった.

高齢者の音声で特に多く観測される「しゃがれ声」を検出・解析するための信号処理技術を開発した. 提案法により, しゃがれ声の分析だけではなく, 歌声におけるシャウトや, 声帯にポリープなどの異常を含む音声にも同様の特性が生じることを確認した.

収録された音声に雑音が入っている場合, 提案法での分析性能が低下する. この問題に対して雑音除去を行う方法を提案した. 提案法は, 雑音抑圧手法において最もシンプルなスペクトルサブトラクション (SS) 法をベースにした方法である. SS 法は, 簡便な方法で雑音を抑圧可能である一方, 雑音抑圧後の音声にミュージカルノイズと呼ばれる不快な雑音を生じさせることが知られている. 提案法では, 人間のラウドネス特性に基づいて抑圧量を変動させることにより, ミュージカルノイズの発生を低減させることに成功した. また雑音抑圧音声の品質評価から, 音声の知覚的な劣化を従来法よりも低減できることを確認した.

##### (2) 聴感的印象に関するパラメータ策定

(1) の検討により, 聴感的印象を改善するための基盤が完成した. 次の検討として, 多くの話者を対象とした収録を行い, 聴感的印象 (本研究では「はきはき」した発話を聴感的印象が良い, 「もごもご」した発話を聴感的印象が悪いものとして定義する) に対応するパラメータの策定に向けた検討を実施した. 収録された音声について主観評価を実施し, 聴感的印象に関する知覚には被験者間のばらつきが少なく, 共通理解があることを確認した. この共通性は, 発話者が各々の解釈で発話した聴感的印象の良い (あるいは悪い) 音声について得られた結果であるため, 聴感的印象について話者間にも共通する表現方法があるといえる.

音声特徴量に関する予備的な検討を実施し, 聴感的印象には, 各音素における口の形 (本研究では声道断面積関数 (VTAF) を口の形として用いることとした) の変化速度が重要であることを突き止めた. 本仮説を検証するため, 聴感的印象の良い「はきはき」声,

悪い「もごもご」声を音声モーフィング技術により処理し, 聴感的印象が段階的に変化するかについて主観評価を実施した. 実験の結果, 口の形状の動きが重要であることが示されたが, 同時に, 音声のスペクトル包絡における高域成分を強調することが重要であることも突き止めた. 音声を構成する要素には, 基本周波数, スペクトル包絡, 非周期性指標があるが, 聴感的印象の制御にはスペクトル包絡だけで良いことも確認した.

##### (3) 聴感的印象改善手法の構築と評価

(2) で得られた結論に基づき, 音声の口の動きの時間変化, およびスペクトル包絡の高域を制御する声質変換法を実装し, 有効性について評価した. 声質変換については, 統計的手法ではなく, TANDEM-STRAIGHT で分析されたスペクトル包絡について, 変換ルールに基づいて制御する方法から構成される.

スペクトル包絡の制御については, 収録した話者全てのスペクトル包絡の変化を求めた結果, 低域は抑圧し高域を強調する形状のフィルタとして近似できることが確認された. また, 変化量は母音により異なることから, 制御フィルタは母音ごとに設計し, 音素間の境界は関数のパラメータをモーフィングすることで実装した. 各音素に対応するフィルタをシグモイド関数でフィッティングすることで, パラメータの変換によるモーフィングを行うこととした.

口の形状の時間的な変化については, 音声波形から VTAF を推定する方法に基づいて VTAF を制御する方法を採用した. VTAF についても母音ごとに特性が異なるが, すでに滑らかな特性が得られているため関数によるフィッティングは行わず, 平均的な VTAF をそのまま採用することとした. 音素と音素との中間については, VTAF をそのまま重み付き加算することで変換関数を設計した.

提案法の有効性を主観評価により検証した結果, 図 1 の通り, 提案法により聴感的印象が改善する結果を得た. なお, 変換元となる音声は元々どの程度の聴感的印象であったかに依存して最適な制御量が異なるため, 変換関数に重みを付けることで, あらゆる音声に対する理論的な最適値について検証することとした. 図 1 の Best が最適な結果を示すものであり, 変換前の音声と比較すると有意に聴感的印象を改善できることが明らかとなった.

今後は, 提案した聴感的印象改善法の実時間処理に対応可能となるようアルゴリズムを改善すること, さらに雑音抑圧アルゴリズムと統合したシステムを実装することで, さらに実用性の高い聴感的印象改善システムを模索する計画である.

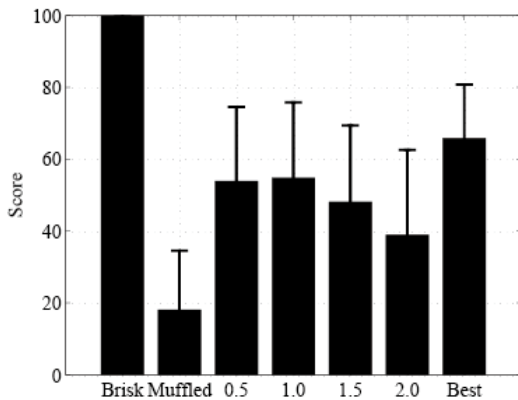


図 1：提案する手法の効果。提案法により Best のスコアまで聴感的印象が向上している。

## 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)

堀井圭祐, 福森隆寛, 森勢将雅, 中山雅人, 西浦敬信, 山下洋一: 雑音下音声受音における Weighted 反復スペクトル減算法を用いたミュージカルノイズの低減, 電子情報通信学会 論文誌 D, vol. J96-D, no. 3, pp. 664-674 (2013), 査読有

中野皓太, 森勢将雅, 西浦敬信, 山下洋一: 基本周波数の転写に基づく実時間歌唱制御システムの実現を目的とした高品質ボコーダ STRAIGHT の高速化, 電子情報通信学会 論文誌 A, vol. J95-A, no. 7, pp. 563-572 (2012), 査読有

M. Morise: PLATINUM: A method to extract excitation signals for voice synthesis system Acoust. Sci. & Tech, vol. 33, no. 2, pp. 123-125 (2012), 査読有

H. Kawahara and M. Morise: Technical foundations of TANDEM-STRAIGHT, a speech analysis, modification and synthesis framework SADHANA - Academy Proceedings in Engineering Sciences, vol. 36, no. 5, pp. 713-728 (2011), 査読有

赤桐隼人, 森勢将雅, 入野俊夫, 河原英紀: スペクトルピークを強調した F0 適応型スペクトル包絡抽出法の最適化と評価 電子情報通信学会論文誌, vol. J94-D, no. 8, pp. 557-567 (2011), 査読有

森勢将雅, 松原貴司, 中野皓太, 西浦敬信: 高品質音声合成を目的とした母音の高速スペクトル包絡推定法 電子情報通信学会論文, vol. J94-D, no. 7, pp. 1079-1087 (2011), 査読有

### [学会発表](計 20 件)

森勢将雅: オープンソース音声合成システム WORLD の現状と課題, 情報処理学会

音楽情報科学研究会, (東京都日本大学), 6-page, 2014 年 5 月 25 日, 査読無.

森勢将雅, 都築聡, 小澤賢司: 音声モーフィングによる聴感的印象制御に有効な特徴量の検討, 電子情報通信学会技術研究報告, vol. 113, no. 404, pp. 55-60, (愛知県名城大学), 2014 年 1 月 24 日. 査読無

都築聡, 森勢将雅, 小澤賢司: 声道断面積関数の時間変化に基づく聴感的印象評価の検討, 日本音響学会 2013 年秋季研究発表会, pp. 555-556, (愛知県豊橋技術科学大学), 2013 年 9 月 27 日. 査読無

M. Morise, H. Kawahara, K. Ozawa: Periodicity extraction for voiced sounds with multiple periodicity, Proc. INTERSPEECH2013, pp. 1921-1925, (Lyon, France), 2013 年 8 月 28 日. 査読有

M. Morise: An attempt to develop a singing synthesizer by collaborative creation, Proc. the Stockholm Music Acoustics Conference 2013 (SMAC2013), pp. 287-292, (Stockholm, Sweden), 2013 年 7 月 31 日. 査読有

H. Kawahara, M. Morise, R. Nisimura, and T. Irino: An interference-free representation of group delay for periodic signals, Proc. APSIPA2012, OS.17-SLA 8, (California, USA), 2012 年 12 月 4 日. 査読有

森勢将雅, 山下洋一: 調波間の干渉に基づく周期信号を対象とした位相推定法の提案, 第 27 回信号処理シンポジウム, pp.368-373, (沖縄県 ANA インターコンチネンタル石垣リゾート), 2012 年 11 月 28 日. 査読無

H. Kawahara, M. Morise: Simplified aperiodicity representation for high-quality speech manipulation systems, Proc. ICSP2012, pp. 579-584, (Beijing, China), 2012 年 10 月 22 日. 査読有.

堀井圭祐, 福森隆寛, 森勢将雅, 中山雅人, 西浦敬信, 山下洋一: 実騒音環境下における Weighted 反復スペクトル減算法のパラメータ最適化に関する実験的検討, 日本音響学会 2012 年秋季研究発表会, pp. 9-10, (長野県信州大学), 2012 年 9 月 19 日. 査読無.

池田健二, 森勢将雅, 中山雅人, 西浦敬信: 音高変化における歌声のスペクトル包絡の傾斜分析の検討, 日本音響学会 2012 年秋季研究発表会, pp. 373-374, (長野県信州大学), 2012 年 9 月 21 日. 査読無.

H. Kawahara, M. Morise, R. Nisimura, and T. Irino: Deviation measure of waveform symmetry and its application

to high-speed and temporally-fine F0 extraction for vocal sound texture manipulation, Proc. INTERSPEECH 2012, pp. 386-389, (Portland, USA), 2012年9月10日. 査読有.

河原英紀, 森勢将雅, 西村竜一, 入野俊夫: 周期信号の群遅延の安定な表現について, 電子情報通信学会応用音響研究会, vol. 112, no. 125, pp. 1-6, (北海道札幌医療大学), 2012年7月19日. 査読無.

和田芳佳, 森勢将雅, 西村竜一, 入野俊夫, 河原英紀: 障害音声および歌唱音声における音声の周期構造分析について, 日本音響学会 2012年春季研究発表会, pp. 375-376, (神奈川県神奈川大学), 2012年3月15日. 査読無.

河原英紀, 森勢将雅, 西村竜一, 入野俊夫: 異なった原理に基づく周期性検出器のアンサンブルによる音源情報の分析について, 日本音響学会 2012年春季研究発表会, pp. 385-388, (神奈川県神奈川大学), 2012年3月15日. 査読無.

森勢将雅, 杉林裕太郎, 上田麻理: パワースペクトルの非線形変換による騒音の遮蔽レベル評価, 日本音響学会 2012年春季研究発表会, pp. 651-652, (神奈川県神奈川大学), 2012年3月15日. 査読無.

河原英紀, 森勢将雅, 西村竜一, 入野俊夫: 基本波のFMとAM成分に基づく高速な基本周波数推定法について, 日本音響学会聴覚研究会, vol. 41, no. 9, pp. 679-684, (熊本県熊本県立大学), 2011年12月10日. 査読無.

河原英紀, 和田芳佳, 西大輝, 森勢将雅, 西村竜一, 入野俊夫: 音声分析変換合成系における時変フィルタの実装と駆動情報の表現について, 日本音響学会聴覚研究会, vol. 41, no. 7, pp. 561-566, (富山県牛岳温泉リゾート), 2011年10月1日. 査読無.

森勢将雅, 西浦敬信, 河原英紀: 高品質音声分析変換合成システム WORLD の提案と基礎的評価 ~ 基本周波数・スペクトル包絡制御が品質の知覚に与える影響 ~, 日本音響学会聴覚研究会, vol. 41, no. 7, pp. 555-560, (富山県牛岳温泉リゾート), 2011年10月1日. 査読無.

和田芳佳, 森勢将雅, 西村竜一, 入野俊夫, 河原英紀: 複数の周期成分を持つ音声のための周期構造抽出法と障害音声分析への応用について, 日本音響学会聴覚研究会(電子情報通信学会 応用音響研究会共催), vol. 41, no. 6, pp. 457-462, (宮城県東北大学), 2011年8月10日. 査読無.

M. Morise, T. Nishiura, and T. Matsubara: Performance estimation for high-quality voice synthesis based on

signal-to-noise ratio in the recording, Proc. HAID2011, Electrical proceedings, (Shiga, Japan), 2011年8月25, 26日. 査読有.

〔その他〕

ホームページ等

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

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

## 6. 研究組織

### (1) 研究代表者

森勢 将雅 (MORISE Masanori)

山梨大学・医学工学総合研究部・助教

研究者番号: 60510013