科学研究費助成事業　　研究成果報告書

平成 ２６ 年　 ６ 月　 ３ 日現在

機関番号：１４６０３
研究種目：若手研究(A)
研究期間：2010～2013
課題番号：２２６８００１６

研究課題名（和文）バリアフリー音声コミュニケーションのための次世代ボイスチェンジャー技術の構築

研究課題名（英文）Development of technologies for next-generation voice changer towards barrier-free speech communication

研究代表者
戸田　智基（Toda, Tomoki）

奈良先端科学技術大学院大学・情報科学研究科・准教授

研究者番号：９０４０３３２８

交付決定額（研究期間全体）：（直接経費）　　19,200,000円、（間接経費）　　5,760,000円

研究成果の概要（和文）：音声の生成過程における物理的制約は強く、例えば、発声器官の一部が正常に動作しなくなると深刻な発声障害を患うなど、時として音声コミュニケーションにおいて様々な障壁をもたらす。これに対し、統計的声質変換技術を用いて、発声された音声を加工することで、音声生成機能を仮想的に拡張するという解決策が考えられるが、音声コミュニケーションでの使用を可能とするには、変換音声の品質のみでなく、リアルタイム処理の実現が重要となる。本研究では、高い品質を保ちながらリアルタイムで動作する統計的声質変換技術に基づく次世代ボイスチェンジャー基盤技術を構築し、さらに、音声生成機能を拡張する様々な応用技術の開発に取り組んだ。

研究成果の概要（英文）：Physical constraints in our speech production mechanisms sometimes cause various barriers in speech communication, such as vocal disorder, limitation of singing expressions, and so on. Statistical voice conversion capable of flexibly converting speech sounds is a potential technology to break down these barriers by augmenting speech production beyond the physical constraints. To use this technology in speech communication, it is essential to develop a high-quality and real-time conversion algorithm. In this research project, we have developed basic technologies for a next-generation voice changer based on real-time statistical voice conversion. Moreover, we have also developed various applications using the next-generation voice changer to break down the existing barriers in speech communication.

研究分野：総合領域

科研費の分科・細目：知能情報処理・知能ロボティクス

キーワード：音声情報処理　音声信号処理　音声合成　声質変換　声質制御　リアルタイム処理　自動適応　バリアフリー

１．研究開始当初の背景

(1) 音声コミュニケーションでは、所望の言語情報およびパラ言語情報、さらには非言語情報も合わせて、同時に相手に伝達することができる。音声信号には多くの情報が埋め込まれるものの、その生成にかかる時間は短く、即時性は極めて高い、また、音声生成過程における物理的制約は、個人が生成できる声色の範囲を限定し、音声信号に個人性を与える一つの要因となる。

一方で、この制約の強さ故に、時として、コミュニケーションにおいて障壁が生じる。例えば、音源生成器官や調音器官の一部でも正常に動作しなくなると、深刻な発声障害を患い、音声コミュニケーションに支障をきたす。仮に、物理的制約を超えた音声生成を可能とする機能が実現されれば、このような障壁を無くすことができる。さらには，各個人が意図的に制御できる声質の範囲が広がり、より多様な歌唱表現や発声表現も生まれると予想される。

(2) ボイスチェンジャーは音声を変換する技術である。既存技術では、通常、極めて簡易な変換規則を採用しており、リアルタイム処理が可能である反面、変換先の声質は入力音声の声質に大きく依存する。結果、実現できる変換音声の声質は極めて限定され、例えば発声障害者の音声をより自然な音声へと変換するといった高度な処理は実現できない。

言語情報を保持しながら、所望の非言語情報やパラ言語情報を自在に変換する音声情報処理技術として、声質変換が古くから研究されている。統計的手法により高度な変換処理を行う枠組みが主流であり、確率モデルを用いた手法が目覚ましい発展を遂げている。近年では，音声信号の時系列データとしての特徴を最大限に活用する変換法が提案され、その性能は大幅に改善された。しかしながら、高い変換性能を得るためには、一発話単位で得られる統計量に基づく変換処理が必須となる。声質変換技術を音声コミュニケーションへと応用するためには、高品質なリアルタイム変換処理の実現が望まれる。

２．研究の目的

高い品質を保ちながらリアルタイムで動作する統計的声質変換技術に基づく次世代ボイスチェンジャーの構築に取り組んだ。音声コミュニケーションにおいて欠かせない特徴である即時性を満たしながら、高度な変換処理を実現することで、仮想的に音声生成機能を拡張することができる技術の実現を目指した。また、環境変化に対応するための自動適応処理や、変換音声の声質制御処理など、利便性に優れた機能の実現にも取り組んだ。さらに、音声コミュニケーションにおける様々な障壁を取り除くために、音声生成機能拡張に基づく各種応用技術の構築に取り組んだ。

３．研究の方法

大きく分けて以下に示す４つの研究課題に取り組んだ。

(1) 高品質なリアルタイム声質変換技術の構築：発話全体から得られる統計量を考慮した高品質な声質変換処理に対して、近似処理を導入することで、短い処理遅延により動作する声質変換法の開発に取り組んだ。さらに、変換処理時の演算量削減に取り組み、高い品質を保持したリアルタイム声質変換処理の実現にも取り組んだ。

(2) 学習時と変換時の環境変化に対応するための自動モデル適応技術の構築：事前収録された多数話者の音声データを事前知識として活用することで、任意の入力音声に対する変換モデル適応（多対一変換）が可能となる。この技術を発展させて、ユーザーがボイスチェンジャーを使えば使うほど、変換音声の品質が改善される機能の実現を目指し、様々な変動要因（声質変動、発話様式、使用環境など）に対する変換モデルの自動適応技術の開発に取り組んだ。

(3) 操作性に優れた変換音声の声質制御技術の構築：予め多数事前収録話者の声質を主観的に評価した結果を用いることで、声質表現語（太い声－細い声など）に基づく直感的な声質制御（一対多変換）が可能となる。この技術をさらに発展させ、より操作性に優れた声質制御技術の開発に取り組んだ。

(4) 次世代ボイスチェンジャー応用技術の構築：上記要素技術を統合することで、次世代ボイスチェンジャー基礎技術の開発に取り組んだ。また、応用技術として、発声障害者である喉頭摘出者がより自然な音声を発声できるようにする無喉頭音声強調技術、所望の声色での歌唱を可能とする歌声声質変換技術、他人に迷惑を掛けないサイレント音声コミュニケーションを実現するための体内伝導音声強調技術などの開発に取り組んだ。基盤技術を個々の応用技術に特化させることで、より実用的な技術の構築を目指した。

４．研究成果

リアルタイムで声質を高精度に変換・制御できる次世代ボイスチェンジャー基盤技術とその応用技術を実現した。各研究課題に対する成果を以下に示す。

(1) 高品質なリアルタイム声質変換技術の構築：高品質なリアルタイム声質変換法として、時間フレーム間相関を考慮した短遅延変換処理と、高次統計量を考慮した変換音声強調処理を実現した。また、高い変換性能を保ったまま演算量の削減を行うために、全共分散混合正規分布モデルに対する同時対角化処理も実現し、リアルタイムボイスチェンジ

ャーソフトウェアを開発した。その結果、約50～80ms 程度の遅延時間で動作する変換処理の実装に成功した。また、実環境への応用を想定し、計算リソースが限られた状況下でも低演算量で動作するリアルタイムボイスチェンジャー技術を開発し、浮動小数点型ディジタルシグナルプロセッサ（DSP）上への実装にも成功した。

(2) 学習時と変換時の環境変化に対応するための自動モデル適応技術の構築：自動モデル適応法として、最尤線形回帰に基づく変換モデルの教師無し適応法を提案した。また、適応データが少量しか得られない際に、より頑健な変換モデル適応処理を実現するために、最大事後確率推定処理を導入した。これらの適応法を様々な入力話者に対する変換処理（多対一変換）に適用し、その有効性を示した。また、多対一変換技術の一つである固有声変換技術に対しては、ベイズ的アプローチに基づく適応パラメータおよびモデルパラメータのモデリング法を提案した。これにより、自動オンライン適応処理を、数理的により見通しの優れた枠組みで定式化した。

(3) 操作性に優れた変換音声の声質制御技術の確立：固有声変換技術に基づく声質制御法を拡張することで、声質制御機能を保ちつつ、声質適応機能を改善する手法を提案した。また、非線形処理に基づく声質制御法を提案し、声質制御性能を改善した。一対多変換である声質制御処理に対しても、ベイズ的アプローチに基づくモデリング法を導入し、声質制御パラメータおよびモデルパラメータに対する事前分布の使用を可能とした。これにより、目標とする参照音声データが少量の場合においても、高い声質制御性能が得られることを確認した。

(4) 次世代ボイスチェンジャー応用技術の構築：開発した各種要素技術を統合し、次世代ボイスチェンジャー基盤技術を構築した。また、発声障害者補助のための音声強調、任意の歌手の声質による歌唱を実現する歌声用ボイスチェンジャー、周囲に迷惑をかけないサイレント音声強調、音声翻訳システムの出力音声声質制御といった応用技術を開発した。以下で各応用技術について述べる。

① 発声障害者補助
喉頭摘出者のための代替発声法として、電気式人工喉頭を用いた発声に着目し、生成される電気音声の自然性を改善するための音声強調技術を開発した。統計的声質変換処理により生じる誤差により、明瞭性が若干劣化する傾向が見られたため、変換対象とする音声特徴量を限定し、信号処理技術と組み合わせたハイブリッド変換技術を提案した。実験的評価結果から、明瞭性を保持しつつ大幅な自然性改善が達成されることを示した。

② 歌声用ボイスチェンジャー
声質制御技術および自動適応技術を導入することで、任意の歌手間において、容易に声質を変換することができる歌声用ボイスチェンジャーソフトウェアを開発した。また、変換システムを構築する際に、ある特定歌手による大量の歌声データを要するという問題点を解決するために、学習データを効率的に生成する技術についても構築した。さらに、声質制御技術を発展させて、歌声から知覚される年齢（知覚年齢）を自由に制御できる機能も実現した。個々の歌手の個人性を保持したまま、知覚年齢を自由に制御するための変換ソフトウェアを開発し、さらなる品質改善処理も導入した。実験的評価結果から、提案技術により、高精度な知覚年齢操作が可能であることを示した。

③ サイレント音声強調
体表密着型マイクを用いることで、周囲に聞こえないぐらい小さなささやき声や、外部雑音の影響が大幅に低減された通常音声を収録することができる。しかしながら、体内伝導収録の影響により、その品質は大幅に劣化するため、次世代ボイスチェンジャー技術を応用し、より自然な音声へと変換する体内音声強調技術を開発した。サイレント音声強調においては、従来法であるささやき声への変換処理のみでなく、聞き手が雑音環境下にいる状況も想定し、より聞き取りやすい目標音声に関する調査を行った。実験的評価結果から、聞き手側の雑音レベルが大きい際には、ささやき声を有声音化する変換処理が有効であることが分かった。

④ 音声翻訳システムの出力声質制御
音声翻訳では、通常、入力言語を発声した話者とは異なる話者による他言語合成音声が出力される。そこで、次世代ボイスチェンジャー技術を応用し、システムに入力される一発話分の音声のみを用いて、合成音声の声質を入力話者のものへと瞬時に変換するシステムを構築した。実験的評価結果から、提案法の有効性を確認した。

これらの研究成果は高く評価され、2010 年度音声研究会研究奨励賞、国際会議 APSIPA ASC 2012 The Best Paper Award (Short Paper in Regular Session Category)、第 96 回音楽情報科学研究会ベストプレゼンテーション賞、日本音響学会から第 35 回日本音響学会粟屋潔学術奨励賞・第 6 回日本音響学会学生優秀発表賞・第 16 回関西支部若手研究者交流研究発表会最優秀奨励賞を受賞した。また、国際会議 ISCSLP2012 における招待チュートリアル講演や、数百人規模の参加者を誇る Winter School on Speech and Audio Processing (WiSSAP 2013)での招待講義を初めとして、国内外において数多くの招待講演を行い、本研究成果を大いにアピールした。

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２４件）

① A Kazuhiro Kobayashi, Tomoki Toda, 他6名, Voice timbre control based on perceived age in singing voice conversion, IEICE Transactions on Information and Systems, 査読有, Vol. E97-D, No. 6, 2014, pp. 1419-1428. DOI: 10.1587/transinf.E97.D.1419

② Kou Tanaka, Tomoki Toda, 他3名, A hybrid approach to electrolaryngeal speech enhancement based on noise reduction and statistical excitation generation, IEICE Transactions on Information and Systems, 査読有, Vol. E97-D, No. 6, 2014, pp. 1429-1437. DOI: 10.1587/transinf.E97.D.1429

③ Hironori Doi, Tomoki Toda, 他3名, Alaryngeal speech enhancement based on one-to-many eigenvoice conversion, IEEE/ACM Transactions on Audio, Speech and Language Processing, 査読有, Vol. 22, No. 1, 2014, pp. 172-183, DOI: 10.1109/TASLP.2013.2286917

④ Kazuhiro Kobayashi, Hironori Doi, Tomoki Toda, 他5名, An investigation of acoustic features for singing voice conversion based on perceptual age, Proceedings of INTERSPEECH, 査読有, Lyon, France, Aug. 2013, pp. 1057-1061, http://www.isca-speech.org/archive/interspeech_2013/i13_1057.html

⑤ Hironori Doi, Tomoki Toda, 他3名, Evaluation of a singing voice conversion method based on many-to-many eigenvoice conversion, Proceedings of INTERSPEECH, 査読有, Lyon, France, Aug. 2013, pp. 1067-1071, http://www.isca-speech.org/archive/interspeech_2013/i13_1067.html

⑥ Kou Tanaka, Tomoki Toda, 他3名, Hybrid approach to electrolaryngeal speech enhancement based on spectral subtraction and statistical voice conversion, Proceedings of INTERSPEECH, 査読有, Lyon, France, Aug. 2013, pp. 3067-3071, http://www.isca-speech.org/archive/interspeech_2013/i13_3067.html

⑦ Takuto Moriguchi, Tomoki Toda, 他5名, A digital signal processor implementation of silent/electrolaryngeal speech enhancement based on real-time statistical voice conversion, Proceedings of INTERSPEECH, 査読有, Lyon, France, Aug. 2013, pp. 3072-3076, http://www.isca-speech.org/archive/interspeech_2013/i13_3072.html

⑧ Hironori Doi, Tomoki Toda, 他3名, Singing voice conversion method based on many-to-many eigenvoice conversion and training data generation using a singing-to-singing synthesis system, Proceedings of APSIPA ASC, 査読有, Hollywood, USA, Nov. 2012, http://www.apsipa.org/proceedings_2012/papers/163.pdf
APSIPA ASC 2012 The Best Paper Award (Short Paper in Regular Session Category).

⑨ Mayumi Kishimoto, Tomoki Toda, 他3名, Model training using parallel data with mismatched pause positions in statistical esophageal speech enhancement, Proceedings of ICSP, 査読有, Beijing, China, Oct. 2012, pp. 590-594, DOI: 10.1109/ICoSP.2012.6491557

⑩ Tomoki Toda, 他2名, Statistical voice conversion techniques for body-conducted unvoiced speech enhancement, IEEE Transactions on Audio, Speech and Language Processing, 査読有, Vol. 20, No. 9, 2012, pp. 2505-2517, DOI: 10.1109/TASL.2012.2205241

⑪ Tomoki Toda, 他2名, Implementation of computationally efficient real-time voice conversion, Proceedings of INTERSPEECH, 査読有, Portland, USA, Sep. 2012, http://www.isca-speech.org/archive/interspeech_2012/i12_0094.html

⑫ Kenzo Yamamoto, Tomoki Toda, 他3名, Statistical approach to voice quality control in esophageal speech enhancement, Proceedings of ICASSP, 査読有, Kyoto, Japan, Mar. 2012, pp. 4497-4500, DOI: 10.1109/ICASSP.2012.6287949

⑬ Daisuke Deguchi, Tomoki Toda, 他3名, Computationally efficient body-conducted voice conversion with original excitation signals, Proceedings of APSIPA ASC, 査読有, Xi'an, China, Oct. 2011, http://www.apsipa.org/proceedings_2011/pdf/APSIPA112.pdf

⑭ Nobuaki Hattori, Tomoki Toda, 他3名, Speaker-adaptive speech synthesis based on eigenvoice conversion and language-dependent prosodic conversion in speech-to-speech

translation, Proceedings of INTERSPEECH, 査読有, Florence, Italy, Aug. 2011, pp. 2769-2772, http://www.isca-speech.org/archive/interspeech_2011/i11_2769.html

⑮ Kumi Ohta, Tomoki Toda, 他 3 名, Adaptive voice-quality control based on one-to-many eigenvoice conversion, Proceedings of INTERSPEECH, 査読有, Chiba, Japan, Sep. 2010, pp. 2158-2161, http://www.isca-speech.org/archive/interspeech_2010/i10_2158.html

⑯ Chie Hayashida, Tomoki Toda, 他 3 名, Linear transformation approaches to many-to-one voice conversion, Proceedings of 7th ISCA Speech Synthesis Workshop, 査読有, Kyoto, Japan, Sep. 2010, pp. 74-79, http://www.isca-speech.org/archive/ssw7/ssw7_074.html

この他に８件

〔学会発表〕（計３５件）

① 小林 和弘、差分スペクトル補正に基づく統計的歌声声質変換、日本音響学会春季研究発表会、2014 年 3 月 12 日、日本大学（東京都千代田区）

② 鶴田 さくら、雑音環境下での非可聴つぶやき強調システムにおける目標音声の評価、日本音響学会春季研究発表会、2014 年 3 月 12 日、日本大学（東京都千代田区）

③ 田中 宏、統計的音源予測に基づく電気式人工喉頭制御法、日本音響学会春季研究発表会、2014 年 3 月 12 日、日本大学（東京都千代田区）

④ 小林 和弘、統計的歌声声質変換における知覚年齢に基づく声質制御、電子情報通信学会音声研究会、2013 年 11 月 21 日、奈良先端科学技術大学院大学（奈良県生駒市）

⑤ 田中 宏、ハイブリッド式電気音声強調法における音源特徴量予測の評価、電子情報通信学会音声研究会、2013 年 11 月 21 日、奈良先端科学技術大学院大学（奈良県生駒市）

⑥ 戸田 智基、統計的手法に基づくリアルタイム声質変換による音声生成機能拡張、日本音響学会秋季研究発表会、2013 年 9 月 27 日、豊橋技術科学大学（愛知県豊橋市）、**招待講演**

⑦ 高道 慎之介、変調スペクトルを考慮した HMM 音声合成、日本音響学会秋季研究発表会、2013 年 9 月 26 日、豊橋技術科学大学（愛知県豊橋市）、**第 35 回粟屋潔学術奨励賞**

⑧ 田中 宏、スペクトル補正及び統計的音源生成に基づくハイブリッド電気音声強調、電子情報通信学会音声研究会、2013 年 6 月 13 日、新潟大学（新潟県新潟市）

⑨ 小林 和弘、知覚年齢に沿った歌声声質制御のための音響特徴量の調査、情報処理学会音楽情報科学研究会、2013 年 5 月 12 日、お茶の水女子大学（東京都文京区）

⑩ 森口 拓人、統計的手法に基づくリアルタイム声質変換処理の DSP 上への実装、電子情報通信学会音声研究会、2012 年 11 月 8 日、東北工業大学（宮城県）

⑪ 土井 啓成、多対多固有声変換に基づく歌声声質変換及び歌声合成を用いた学習データ生成、日本音響学会秋季研究発表会、2012 年 9 月 19 日、信州大学（長野県）、**第 6 回日本音響学会学生優秀発表賞**

⑫ 土井 啓成、VocaListener による学習データ生成を利用した多対多固有声変換に基く歌声声質変換、情報処理学会音楽情報科学研究会、2012 年 8 月 9 日、近江町交流プラザ（石川県）、**第 96 回音楽情報科学研究会ベストプレゼンテーション賞**

⑬ 岸本 真由美、統計的無喉頭音声強調における学習データのポーズ位置不一致への対応、電子情報通信学会音声研究会、2011 年 11 月 29 日、九州大学（福岡県）

⑭ 戸田 智基、統計的手法に基づく声質分析・変換・制御技術とその応用、日本音響学会秋季研究発表会、2011 年 9 月 20 日、島根大学（島根県）、**招待講演**

⑮ 出口 大祐、残差波形の使用による肉伝導音声変換処理の演算量削減、日本音響学会春季研究発表会、2011 年 3 月 11 日、早稲田大学（東京都新宿区）

⑯ 服部 信彦、音声翻訳システムのための声質変換の性能評価、日本音響学会春季研究発表会、2011 年 3 月 9 日、早稲田大学（東京都新宿区）

⑰ 山本 憲三、食道音声強調における声質制御技術の検討、日本音響学会春季研究発表会、2011 年 3 月 9 日、早稲田大学（東京都新宿区）

この他に１８件

〔産業財産権〕

○出願状況（計１件）

名称：電気式人工喉頭装置
発明者：戸田 智基、他 4 名
権利者：国立大学法人 奈良先端科学技術大学院大学
種類：特許
番号：特願 2013-165087
出願年月日：2013 年 8 月 8 日
国内外の別：国内

６．研究組織

(1)研究代表者

戸田　智基（TODA, Tomoki）
奈良先端科学技術大学院大学・情報科学研究科・准教授
研究者番号：９０４０３３２８