研　究　報　告　書

「ユーザの適応能力を活用する共創型音声生成機能拡張技術の構築」

研究タイプ：通常型
研究期間：　2016 年 12 月～2020 年 3 月
研　究　者：　戸田　智基

1.　研究のねらい

音声は、言語情報（何を話しているかという情報）のみでなく、パラ言語情報（どのように話しているかという情報）や非言語情報（誰が話しているかという情報）といった様々な情報も同時に伝達可能であり、円滑な意思伝達を実現する上で欠かすことのできない媒体である。我々は、発声器官を巧みに動作させることで、所望の言語情報とパラ言語情報を持つ音声を即時に発声することができるが、同時に、身体的制約により個人に応じた声質が決定され、意図とは無関係に非言語情報が音声に埋め込まれる。この音声生成過程における物理的制約は、時として様々な障壁をもたらす要因となり、例えば、喉頭摘出による発声障害のように、発声器官の一部が動作しなくなるだけで、自然な音声の発声は困難となる。

物理的制約を超えて音声を自由に発声する機能の実現を目指す技術として、音声生成機能を拡張する情報通信技術が挙げられる。統計的音声変換処理を活用することで、発声された音声に対して、言語情報を保持したまま所望の情報のみを即時に変換する処理を施すものであり、非言語情報さえも意図的に制御可能となる。この音声生成機能拡張技術を発展させることで、発声障碍者や高齢者が自身の失われた声を取り戻すことや、健常者が個人の能力的制約を超越した音声の表情付けを可能とする新たな発声法・歌唱法を獲得することが可能となり、社会のさらなるバリアフリー化および活性化を推し進めることができると期待される。

これまでの音声生成機能拡張技術に関する研究は、音声変換システム側の性能改善に主眼を置いており、信号処理と機械学習という要素技術を改善することで、変換音声の品質改善や適用可能な応用範囲の拡大に取り組んできたが、未だ変換音声の品質は不十分であり、ユーザの意思を適切に伝達できる音声を生成できる水準には至っていない。本研究では、実用化に耐えうる変換音声の実現を目指し、従来とは異なるアプローチとして、人が持つ高い適応能力に着目し、システム側の性能改善のみではなく、ユーザ側の協力も最大限に活用することを考える。特に、音声生成機能拡張においては、最終的に生成される変換音声の品質は、入力であるユーザの音声に大きく依存する点に着目し、ユーザ側がシステム側に適応する仕組みを新たに導入することで、ユーザの意図を適切に反映した音声を高品質に生成する共創型音声生成機能拡張技術の構築に取り組む。

2.　研究成果

（1）概要

従来の信号処理と機械学習に基づく統計的音声変換処理に対して、新たにユーザによる協力的な行動を利用する仕組みを導入することで、従来のシステムからユーザへの歩み寄りに加えて、ユーザからシステムへの歩み寄りも活用する共創型音声生成機能拡張技術の構築に取り組んだ。共創型音声生成機能拡張技術の概要を図1に示す。

まず、従来のアプローチに基づくベースラインとして、A)リアルタイムに動作する統計的音声変換基盤技術を用いて、各種応用例を想定した音声生成機能拡張プロトタイプシステムを構築し、さらには、実環境下への適用に向けた技術改善に取り組んだ。次に、共創型音声生成機能拡張技術の実現に向けて、B)ユーザが発声した音声のみでなく、発声時に生じる動作信号も入力として併用することで、より表情豊かな音声へと変換する基盤技術の構築に取り組んだ。これらの研究成果をベースとして、C)ユーザとシステムの歩み寄りを利用する技術として、ユーザの協力的動作を活用する共創型音声生成機能拡張技術の構築に取り組んだ。具体的には、喉頭摘出者を対象とした応用技術として、電気式人工喉頭を用いた発声に加え、演奏動作なども協力的動作として併用することで、自由度の高い即興性に優れた歌唱を可能とする歌唱支援システムを構築した。また、健常者を対象とした応用技術として、特定のキャラクターの声色による発声・歌唱を可能とするために、声の高さを真似た特殊な発声・歌唱を協力的動作として活用することで、耐雑音性に優れた変換処理を実現するボイスチェンジャー・ボーカルエフェクターを構築した。これらの結果から、共創型音声生成機能拡張技術により、従来のシステムからユーザへ歩み寄るアプローチのみでは実現困難であった機能を実現できることを示した。

これらの研究に加え、近年の深層学習の発展に伴う基盤技術の改善により、統計的音声変換においても大きな技術の進展が得られる可能性が高まったため、D)深層学習に基づく統計的音声変換基盤技術の改善にも取り組んだ。特に、深層波形生成モデルに基づく統計的声質変換技術を世界に先駆けて提案するとともに、国際的音声変換技術評価会 Voice Conversion Challenge 2018 においては、23 グループ中、世界 2 位および3位相当の性能を実現するシステムを構築するまでに至った。

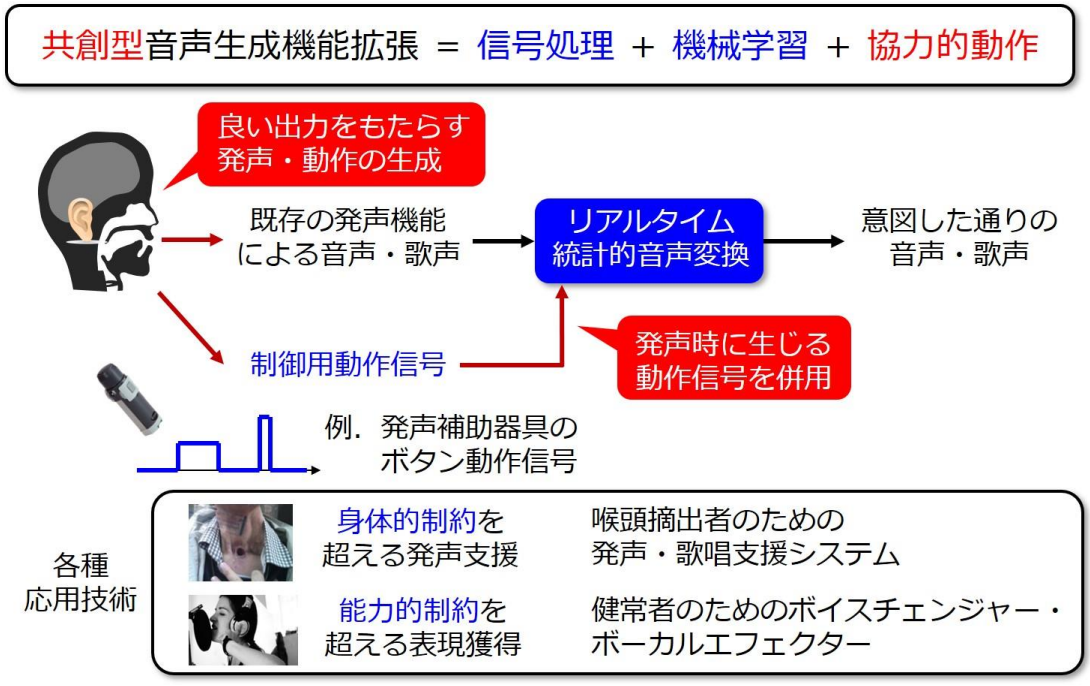

図1．共創型音声生成機能拡張技術の概要

（2）詳細

研究テーマ A「音声生成機能拡張プロトタイプシステムの構築」

リアルタイム統計的音声変換基盤技術を利用した音声生成機能拡張プロトタイプシステムとして、1）喉頭摘出者が電気式人工喉頭を用いて発声した電気音声を自然な音声へと変換することで、身体的制約を超えた発声機能を獲得する発声支援システムと、2）健常者が発声した音声を特定のキャラクターの声質へと変換することで、身体的制約を超えた音声表現を獲得するボイスチェンジャーシステムを構築した。また、各システムを実環境下で頑健に動作させるために、電気式人工喉頭の出力信号を直接制御する技術や、音声波形の直接加工処理に基づく音声変換技術【5. 主な研究成果リスト（1）1】など、変換モデル学習時と使用時の音響的不一致に対する頑健性を向上させる技術を構築した。

研究テーマ B「動作信号を併用した変換音声制御技術の構築」

喉頭摘出者のための発声支援システムにおいて、より表情豊かな音声の生成を可能とするシステムの実現を目指し、電気音声に加えて動作信号も併用する音声変換基盤技術の構築に取り組んだ。特に、多様な音声表情を含む歌声の生成を対象として、動作信号としてキーボードの演奏動作に着目し、MIDI（Musical Instrument Digital Interface）の音高情報から自然な歌声が持つ声の高さの変化パターンへと変換する技術、ならびに、電気音声の機械的な声質を緩和する技術を考案し、その有効性を示した【5. 主な研究成果リスト（3）1】。また、統計的音声変換技術も導入し、声質に関しても自然な歌声へと変換することで、より自然性の高い歌声による歌唱を可能とする歌唱支援システムを構築した。さらに、本技術の仕組みを応用し、喉頭摘出者の発声支援システムにおいて、電気式人工喉頭のボタン動作信号と変換音声の発話様式を統計的音声変換技術により結び付けることで、ボタン動作信号による発話様式制御機能を備えた発声支援システムを構築した。

研究テーマ C「発声者の協力的動作を活用した共創型音声生成機能拡張技術の構築」

ユーザである発声者・歌唱者とシステムがお互いに歩み寄り、より良い音声生成を可能とする音声生成機能拡張技術の一つとして、発声者の協力的動作を活用する共創型音声生成機能拡張技術の構築に取り組んだ。喉頭摘出者向けの応用技術として、キーボードの演奏動作を協力的動作として活用する技術を核とし、自由度の高い音高（メロディー）制御に基づく歌唱を可能とする歌唱支援システムを実現した（図2左図参照）。また、演奏を不得手とする喉頭摘出者向けには、協力的動作として、システムが提示する伴奏に合わせた電気式人工喉頭による同期発声を用いることで、事前に埋め込まれた音高パターンによる歌唱を可能とする音高パターン埋め込み型（カラオケ型）システムを構築した（図2右図参照）。一方で、健常者向けの応用技術としては、発声者の協力的動作として、同一音高歌唱・発声として、裏声などの特殊な歌唱・発声様式を活用することで、特定のキャラクターへの高精度な変換処理を実現するボーカルエフェクタ・ボイスチェンジャーを構築した。これにより、声の高さの変換処理を回避して、音声波形直接加工処理のみで統計的音声変換を実施することが可能となり、変換歌声、変換音声の耐雑音性の向上に成功した【5. 主な研究成果リスト（1）3】。一方で、裏声などの特殊な歌唱・発声様式の使用は、変換精度を若干劣化させる原因となり得ることも分かった。

研究テーマ D「統計的音声変換基盤の改善」

深層学習の発展に伴い、音声波形を直接モデル化する深層波形生成モデルが考案され、

音声合成基盤技術の性能が大幅に改善された。これにより、統計的音声変換においても大きな技術の進展が得られる可能性が高まったため、4）深層学習に基づく統計的音声変換基盤技術の改善にも取り組んだ。音声特徴量から音声波形を合成するボコーダ処理に深層波形生成モデルを適用し、少量の音声データからでも高精度な音声波形を合成できることを世界に先駆けて明らかにし、深層波形生成モデルを用いた統計的音声変換技術を考案した【5．主な研究成果リスト（3）4】。国際的音声変換技術評価会 Voice Conversion Challenge 2018 においては、従来型の音声変換基盤技術に基づき、国際的ベースラインシステム【5．主な研究成果リスト（1）2】を構築するとともに、深層波形生成モデルを用いた統計的音声変換システムについても構築した【5．主な研究成果リスト（1）5】。その結果、図3に示す通り、両システムは 23 グループ中、世界 2 位および3位相当の性能を達成するに至った。その後、統計的音声変換処理を考慮した深層波形生成モデル学習法を提案し、さらなる性能改善に成功した【5．主な研究成果リスト（1）4】。

以上の通り、システムからユーザへの歩み寄りに相当する基盤技術の改善と、ユーザからシステムへの歩み寄りに相当するユーザの協力的動作を活用した統計的音声変換技術の構築を通して、共創型音声生成機能拡張技術を実現した。そのため、研究目的は概ね達成することができたといえる。国際講習会【5．主な研究成果リスト（3）2】、国際会議チュートリアル【5．主な研究成果リスト（3）3】、招待講演【5．主な研究成果リスト（3）5】、Web ニュースでの技術紹介などのアウトリーチ活動を通して、本研究成果を広く周知することにも尽力した。

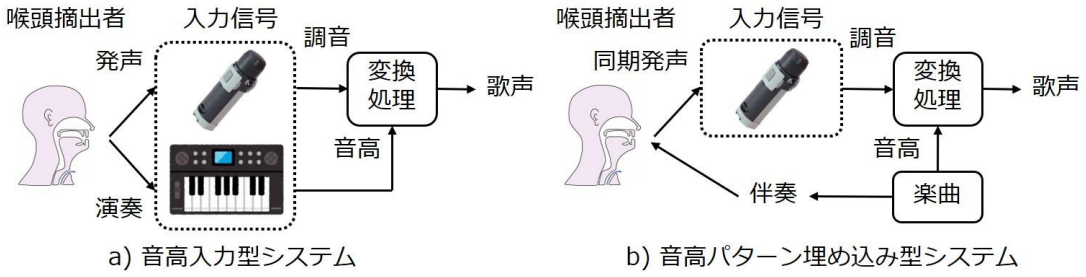

図2．共創型音声生成機能拡張に基づく歌唱支援システム

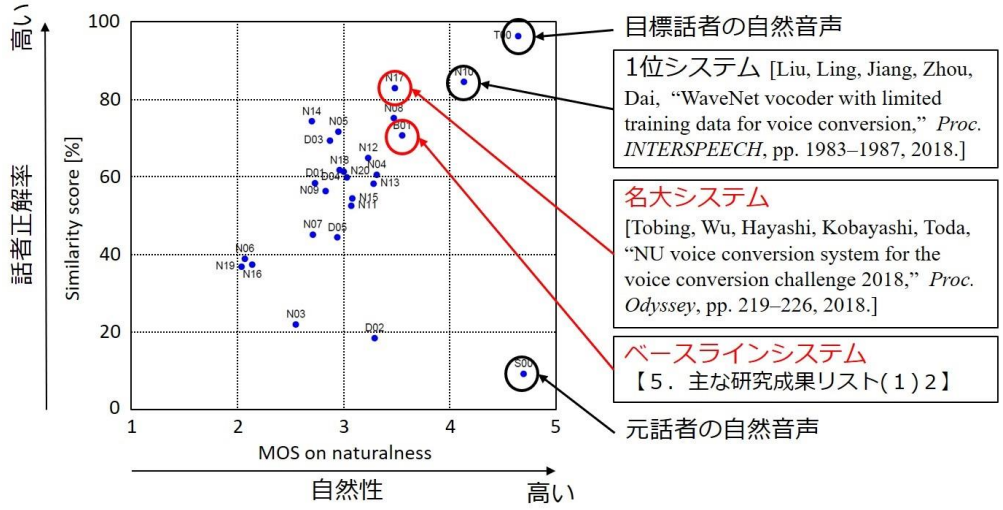

図3．国際的音声変換技術評価会 Voice Conversion Challenge 2018 の結果

3.　今後の展開

本研究課題を通して、音声生成機能拡張技術の実用化に向けて、解決すべき技術的課題が明確となった。まず、研究テーマ C の結果から、協力的動作として裏声などの特殊な発話様式を用いる場合、音高変換処理の回避による耐雑音性改善効果は得られるものの、若干の変換精度の低下を招く場合もあるといったトレードオフの存在が明らかとなった。また、特殊な発話様式は使用できる条件が限られることからも、通常の発話様式についても利用可能な枠組みを検討する必要があるといえる。これに対し、研究テーマ D の結果から、深層波形生成モデルの利用により、統計的音声変換において、高精度な音高変換処理の実現が可能であることが示された。そのため、システムからユーザへの歩み寄りとして、深層波形生成モデルを用いたリアルタイム統計的音声変換処理を実現することは、重要な研究課題といえる。

本研究課題では、様々な環境下で共創型音声生成機能拡張システムのデモを実施したが、その経験を通して、ユーザからシステムへの歩み寄りをさらに有効に利用するためには、データ駆動方式であるリアルタイム音声変換システムの挙動を、ユーザ自身が把握できる仕組みが必要であるとの考えに至った。そのための一手法として、インタラクションを通じてユーザがシステム挙動を把握する枠組みの構築が考えられ、1）発声・歌唱時に伴う多様なマルチモーダル動作信号の併用により、変換システム出力の意図的制御を高精度化し、2）機械学習への物理的制約の導入により、起こり得ないシステム出力を抑圧する不随意的制御を実現し、3）高精度かつ即時的なシステム出力フィードバックを達成する、といった技術的課題の解決が重要となる。これにより、即時的インタラクションの有効性と、システムを長期間使い続けることで生じる長期的インタラクションの有効性を調査することも可能となり、ユーザが自ずと協力的動作を習得していく過程をモデル化できると期待される。

以上の研究課題に取り組むプロジェクトを、JST CREST「共生インタラクション」領域の研究課題「音メディア機能拡張」（代表：戸田）として、2019 年 10 月から開始している。

4.　自己評価

研究目的の達成状況：システムからユーザへの歩み寄りとユーザからシステムへの歩み寄りを活用した枠組みとして、ユーザの協力的動作を活用した統計的音声変換技術を構築し、共創型音声生成機能拡張の応用技術をいくつか具体化することができた。また、当初の計画にはなかった深層波形生成モデルに基づく統計的音声変換技術に関する研究についても、世界に先駆けて取り組むことができ、統計的音声変換基盤技術を大きく発展させることに貢献した。聴覚フィードバック制御による発話誘導に関しては、十分な成果は得られていないものの、本研究課題の主たる研究目的は概ね達成することができ、新たな技術課題を明らかにすることができた。

研究実施体制及び研究費執行状況：：研究補助者を雇用することで、複数の研究課題を効率的に進めることができた。結果として、当初計画していなかった深層学習に基づく音声変換基盤技術の改善といった研究課題にも取り組むことができた。また、様々な国際会議にて論文発表を行うことで、音声変換分野以外の研究者に対しても、本研究成果を広く周知することができた。国際的音声変換評価会 Voice Conversion Challenge も盛り上がりを見せるなど、音声変換研究の活性化に大いに役立てることができた。

研究成果の科学技術及び社会・経済への波及効果：発声支援技術に関しては、富士通クラ

イアントコンピューティング株式会社と実用化に向けた共同研究を実施している。未だ、実用化には至っていないものの、実際の喉頭摘出者による評価実験を通して、多くの有益な知見を得ることができた。実用化に向けて、引き続き、JST CREST「音メディア機能拡張」プロジェクトで継続的に取り組む予定である。また、アウトリーチ活動の一環として、招待講演や各種メディアの取材に積極的に応じた。それに関連して、共創型音声生成機能拡張に基づくボイスチェンジャー・ボーカルエフェクタのデモを実施している様子が、動画として SNS 上で公開され、50 万回以上視聴されるなど、本研究課題の研究成果を社会的に知ってもらう機会を提供することができた。

5. 主な研究成果リスト

(1)論文(原著論文)発表

1. Patrick Lumban Tobing, Kazuhiro Kobayashi, Tomoki Toda, "Articulatory controllable speech modification based on statistical inversion and production mappings," IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2017, Vol. 25, No. 12, pp. 2337-2350.
2. 戸田 智基, 小林 和弘, "統計的声質変換ソフトウェア入門," システム／制御／情報, システム制御情報学会, 2018, Vol. 62, No. 2, pp. 69-75.
3. Kazuhiro Kobayashi, Tomoki Toda, Satoshi Nakamura, "Intra-gender statistical singing voice conversion with direct waveform modification using log-spectral differential," Speech Communication, 2018, Vol. 99, pp. 211-220.
4. Patrick Lumban Tobing, Yi-Chiao, Wu, Tomoki Hayashi, Kazuhiro Kobayashi, Tomoki Toda, "Voice conversion with CycleRNN-based spectral mapping and finely tuned WaveNet vocoder," IEEE Access, 2019, Vol. 7, No. 1, pp. 171114-171125.
5. Yi-Chiao, Wu, Patrick Lumban Tobing, Tomoki Hayashi, Kazuhiro Kobayashi, Tomoki Toda, "Non-parallel voice conversion system with WaveNet vocoder and collapsed speech suppression," IEEE Access, 2020, Vol. 8, No. 1, pp. 62094-62106.

(2)特許出願

研究期間累積件数: 0 件(公開前の出願件名については件数のみ記載)

(3)その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

1. 国際会議論文: Kazuho Morikawa, Tomoki Toda, "Electrolaryngeal speech modification towards singing aid system for laryngectomees," Proc. APSIPA ASC, 4 pages, 2017.
2. 国際講習会: Tomoki Toda, "Advanced voice conversion" and "Hands on voice conversion," Speech Processing Courses in Crete (SPCC), 2018 and 2019, Greece.
3. 国際会議チュートリアル: Tomoki Toda, Kazuhiro Kobayashi, Tomoki Hayashi, "Statistical voice conversion with direct waveform modeling," Tutorial, INTERSPEECH, 2019.
4. 解説論文: 戸田 智基, "機械学習と音声生成:音声波形モデリングの進展," 計測と制御, 2019, Vol. 58, No. 12, pp. 951-954.

5. 招待講演: 戸田 智基, “音声変換技術と音声生成機能拡張への応用,” 電子情報通信学会 2020 年総合大会 ソサイエティ合同企画「情報通信技術と人間相互理解の未来」, 2020, 2 pages.