

戸田 智基

名古屋大学情報基盤センター  
教授

## 音メディアコミュニケーションにおける共創型機能拡張技術の創出

### § 1. 研究成果の概要

音声による他人との意思疎通や、音楽による芸術・娯楽、環境音による周囲の状況把握など、音メディアを用いた情報伝達は、我々の社会生活の一部であり必要不可欠なものである。一方で、そのために必要となる発声機能や聴覚機能といった身体機能は、年齢とともに衰弱を迎え、また、障害を患うことでその機能自体が失われることもある。この問題に対し、本研究では、音メディアコミュニケーションにおいて身体的制約を超えて我々の発声機能および聴覚機能を拡張する基盤技術を創出することを目的とし、機械学習が持つ高い性能を保持しつつ、低遅延でリアルタイムに動作し、かつ、その挙動の意識的な制御を可能とするデータ駆動型システムの実現に取り組む。システムとのインタラクションを通して、意図したシステム挙動が得られるような入力動作をユーザが自ずと習得できるように、ユーザとシステムの共創的な働きかけを活用する身体機能拡張基盤の構築を目指す。

研究開始時点にあたる本年度においては、発声機能拡張および聴覚機能拡張として、主に、システムの使用状況が限定される障害者用の機能回復応用を想定しつつ、基盤となる技術の構築に取り組んだ。特に、ユーザとシステム間の質の高い即時的インタラクションの実現を目標に定め、低遅延かつ高精度なリアルタイム処理の実現に向けた研究に取り組んだ。各研究グループにおいて、次の研究成果が得られた。

1) 発声機能拡張グループ(名大 戸田)では、低遅延リアルタイム音声変換処理において、動作信号を併用した音声表情制御を可能とする基盤技術の構築を目指し、機械学習に基づく変換処理の高精度化および高速化、ならびに、多様な学習データの取り扱いを可能とする学習処理の実現に取り組んだ。高品質な音声変換処理を達成する上で重要となる深層波形生成モデルに対して、安定性を改善する手法や、高速化を実現する手法を考案するとともに、喉頭摘出者を対象とした

歌唱支援技術を構築し、低遅延リアルタイム動作が可能なテストベッドシステムを構築した。

2) 聴覚機能拡張グループ(首都大 小野)では、低遅延リアルタイム音響信号処理において、動作信号を併用した選択的聴取を可能とする基盤技術の構築を目指し、マイクロフォンアレイによる音源分離アルゴリズムのさらなる理論的高速化に取り組んだ。音源分離におけるこれまでの最速アルゴリズムを越える高速アルゴリズムの導出や、補聴器のような限られた計算リソースにおいて機械学習に基づく処理を動作させるための次元削減法など、多くの新しい理論的な成果が得られた。

3) 機械学習基盤グループ(NTT 亀岡)では、発声・聴覚機能拡張において、物理的制約を加味した音響信号制御方法の確立を目的として、音声生成過程や音響伝播過程を加味した深層生成モデルの構築に取り組むとともに、複数モダリティの多元動作信号を用いて音声を制御するための方法論に関する研究を進めた。音声生成過程や音響伝播過程を高精度にモデル化する深層メディア生成技術の構築や、クロスモーダル音声・顔画像生成技術の構築など、機械学習基盤において数多くの技術的進捗が得られた。

#### 【代表的な原著論文】

1. Yi-Chiao Wu, Patrick Lumban Tobing, Tomoki Hayashi, Kazuhiro Kobayashi, and Tomoki Toda, “Non-parallel voice conversion system with WaveNet vocoder and collapsed speech suppression,” IEEE Access, Vol. 8, No. 1, pp. 62094–62106, 2020.
2. Robin Scheibler and Nobutaka Ono, “Fast and stable blind source separation with rank-1 updates,” Proc. IEEE ICASSP, pp. 236–240, 2020.
3. Patrick Lumban Tobing, Yi-Chiao Wu, Tomoki Hayashi, Kazuhiro Kobayashi, and Tomoki Toda, “Efficient shallow WaveNet vocoder using multiple samples output based on Laplacian distribution and linear prediction,” Proc. IEEE ICASSP, pp. 7204–7208, 2020.

## § 2. 研究実施体制

### (1) 発声機能拡張グループ

- ① 研究代表者: 戸田 智基 (名古屋大学情報基盤センター 教授)
- ② 研究項目
  - ・共創型音声生成機能拡張技術の構築

### (2) 聴覚機能拡張グループ

- ① 主たる共同研究者: 小野 順貴 (首都大学東京システムデザイン学部 教授)
- ② 研究項目
  - ・共創型聴覚機能拡張技術の構築

### (2) 機械学習基盤グループ

- ① 主たる共同研究者: 亀岡 弘和 (日本電信電話(株)NTT コミュニケーション科学基礎研究所メディア情報研究部 特別研究員)
- ② 研究項目
  - ・挙動制御機能を備えた機械学習基盤の構築