

人間と情報環境の共生インタラクション基盤技術の創出と展開  
2019 年度採択研究代表者

2020 年度 年次報告書
------------------

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

第1フェーズとなる2020年度は、発声機能拡張および聴覚機能拡張として、主に障害者用の機能回復応用を想定しつつ、基盤技術の改善に取り組んだ。特に、ユーザとシステム間の質の高い即時インタラクションの実現に向け、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。

発声機能拡張グループ(名大 戸田)では、動作信号を併用した発声機能拡張基盤技術の構築に向け、高精度化・高速化・制御性能改善や、喉頭摘出者向け発声・歌唱支援処理の実現に取り組んだ。国際音声変換技術評価会 Voice Conversion Challenge 2020 の開催ならびにベースラインシステムの提供、物理的な知見に基づく制御機構を備えた深層音声波形生成モデルの考案、マルチモーダルな多元動作信号を用いた歌唱支援システムの構築などの成果が得られた。

聴覚機能拡張グループ(都立大 小野)では、選択的聴取を可能とする低遅延リアルタイム音響信号処理基盤技術の構築を目指し、マイクロホンアレーによる音源分離アルゴリズムの高速化とその拡張に取り組んだ。また、音光変換センサノード「ブリンキー」を分散配置し、聴覚情報を視覚的にフィードバックする拡張聴覚システムの検討を開始し、リアルタイム音高可視化システムを構築した。その他、円状マイクロホンアレーの対称性を活用した音場補間法や、要素選択による乗算を用いない次元削減法などの成果が得られた。

機械学習基盤グループ(NTT 亀岡)では、物理パラメータを制御対象と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表現した深層生成モデルを考案するとともに、その生成過程パラメータを推定するための手法を構築した。また、多元動作信号を用いた音声制御の方法論の構築を目指し、画像と音声との間のクロスメディア変換の問題に取り組み、クロスモーダル音声・顔画像生成法を開発した。

## § 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 (東京都立大学システムデザイン研究科 教授)
- ② 研究項目
  - ・共創型聴覚機能拡張技術の構築

### (3) 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 上席特別研究員)
- ② 研究項目
  - ・挙動制御機能を備えた機械学習基盤の構築

### 【代表的な原著論文情報】

1. R. Scheibler, N. Ono, “Fast and stable blind source separation with rank-1 updates,” Proc. IEEE ICASSP, pp. 236-240, May 2020.
2. H. Kameoka, T. Kaneko, K. Tanaka, N. Hojo, “Nonparallel voice conversion with augmented classifier star generative adversarial networks,” IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 28, pp. 2982-2995, Nov. 2020.
3. L. Li, H. Kameoka, S. Inoue, S. Makino, “FastMVAE: a fast optimization algorithm for the multichannel variational autoencoder method,” IEEE Access, vol. 8, pp. 228740-228753, Dec. 2020.
4. W.-C. Huang, T. Hayashi, Y.-C. Wu, H. Kameoka, T. Toda, “Pretraining techniques for sequence-to-sequence voice conversion,” IEEE/ACM Transactions on Audio, Speech and Language Processing, Vol. 29, pp. 745-755, Jan. 2021.
5. Y.-C. Wu, T. Hayashi, T. Okamoto, H. Kawai, T. Toda, “Quasi-periodic parallel WaveGAN: a non-autoregressive raw waveform generative model with pitch-dependent dilated convolution neural network”, IEEE/ACM Transactions on Audio, Speech and Language Processing, Vol. 29, pp. 792-806, Jan. 2021.