

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

2021 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フェーズ後半として、主に機能回復応用を想定しつつ、機能拡張基盤技術の改善に取り組んだ。特に、質の高い即時的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。また、第2フェーズとして、多元動作信号による挙動制御機能を備えた機能拡張基盤技術に関する研究に着手した。

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、変換処理の高精度化・高速化に継続して取り組むとともに、身体動作を活用した音声表情制御機能の実現に取り組んだ。発声機能拡張プロトタイプ構築、系列ベース音声変換に基づく発声機能拡張技術の構築、教師なし学習に基づく低遅延リアルタイム深層音声変換技術の構築、音声変換基盤フリーソフトウェアの開発、合成音声品質評価の研究など、多くの技術的進展が得られた。

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、前年度に引き続き、マイクロホンアレイによる音源分離アルゴリズムの高速化・低遅延化に取り組んだ。また、拡張聴覚システムとしての音光変換センサノード「ブリンキー」の検討も継続した。新たな成果として、因果性ビームフォーマの設計法、音源方向を用いた音源選択、音源移動に柔軟なオンライン音源分離、ブリンキーによるシーン認識、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て音声を制御するため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 (日本電信電話株式会社 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) Yukoh Wakabayashi, Kouei Yamaoka, Nobutaka Ono, “Rotation-robust beamforming based on sound field interpolation with regularly circular microphone array”, Proc. IEEE ICASSP, pp. 771-775, June 2021.
- 2) Patrick Lumban Tobing, Tomoki Toda, “Low-latency real-time non-parallel voice conversion based on cyclic variational autoencoder and multiband WaveRNN with data-driven linear prediction”, Proc. 11th ISCA Speech Synthesis Workshop (SSW), pp. 142-147, Aug. 2021.
- 3) Kouei Yamaoka, Nobutaka Ono, Shoji Makino, “Time-frequency-bin-wise linear combination of beamformers for distortionless signal enhancement”, IEEE/ACM Transactions on Audio, Speech, and Language Processing, Vol. 29, pp. 3461-3475, Nov. 2021.
- 4) Ding Ma, Wen-Chin Huang, Tomoki Toda, “Investigation of text-to-speech-based synthetic parallel data for sequence-to-sequence non-parallel voice conversion”, Proc. APSIPA ASC, pp. 870-877, Dec. 2021. 【APSIPA ASC 2021 The Best Paper Award 受賞】
- 5) Yuma Kinoshita, Nobutaka Ono, “Analysis on roles of DNNs in end-to-end acoustic scene analysis framework with distributed sound-to-light conversion devices”, Proc. APSIPA ASC, pp. 1167-1172, Dec. 2021. 【APSIPA ASC 2021 The Best Paper Award 受賞】