2023 年度年次報告書
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

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

主たる共同研究者:
小野 順貴（東京都立大学 システムデザイン学部 教授）
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 特別研究員）

研究成果の概要

ユーザとシステム間のインタラクションを活用した共創型発声・聴覚機能拡張の実現に向け、高精度な実時間処理ならびに低遅延リアルタイム処理を可能とする基盤技術の研究を継続しつつ、発声・聴覚機能拡張システムの改善に取り組んだ。

発声機能拡張グループ（名大　戸田）では、音声・歌声・障害音声に対する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、インタラクション活用によるシステム挙動制御に関する研究を実施した。発声機能拡張におけるインタラクション効果の調査、音声・歌唱表情制御機能を備えた高速深層音声波形生成・加工、汎用深層音声・歌声変換、障碍者用発声機能拡張システムの改良・評価、合成音声品質評価・推定など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リアルタイム音響信号処理の低遅延化、低演算量化、並びに頭部装着時の音源分離制御手法に継続して取り組み、低遅延制約下でのビームフォーマーの最適化、低演算量化のための要素選択による次元削減の一般化、頭部回転への頑健性を有するブラインド音源分離、移動音源に対するオンライン音源分離の計算量の理論限界を達成する効率的アルゴリズムなどを開発した。

機械学習基盤グループ（NTT　亀岡）では、発声・聴覚機能拡張の基盤となる技術の構築に向け、遅延時間の短さと変換性能の高さを両立するリアルタイム音声変換、高品質リアルタイム変換処理の実現に向けた高効率ニューラルボコーダ、音声の声質や表情を柔軟に変換可能な拡散確率モデルベース音声変換、ニューラルボコーダの敵対学習において学習効率と生成音声品質を改善する汎用波形識別器およびその学習法、音声の抑揚と声質を個別操作可能な音声変換、音声と映像を併用したマルチモーダル音声表情変換を開発した。

CEATEC2023 にて、発声・聴覚機能拡張の各種プロトタイプの体験型デモ展示を行った。

【代表的な原著論文情報】

1) Yukoh Wakabayashi, Kouei Yamaoka, Nobutaka Ono, “Sound field interpolation for rotation-invariant multichannel array signal processing,” IEEE/ACM Transactions on Audio, Speech, and Language Processing, Vol. 31, pp. 2286—2298, June 1, 2023.
2) Reo Yoneyama, Yi-Chiao Wu, Tomoki Toda, “High-fidelity and pitch-controllable neural vocoder based on unified source-filter networks,” IEEE/ACM Transactions on Audio, Speech and Language Processing, Vol. 31, pp. 3717—3729, Sep. 11, 2023.
3) Chao Xie, Tomoki Toda, “Noisy-to-noisy voice conversion under variations of noisy condition,” IEEE/ACM Transactions on Audio, Speech and Language Processing, Vol. 31, pp. 3871—33882, Sep. 20, 2023.
4) Yoshiki Masuyama, Kouei Yamaoka, Yuma Kinoshita, Taishi Nakashima, Nobutaka Ono, “Causal and relaxed-distortionless response beamforming for online target source extraction,” IEEE/ACM Transactions on Audio, Speech, and Language Processing, Vol. 32, pp. 310—324, Nov. 1, 2023.
5) Hirokazu Kameoka, Takuhiro Kaneko, Kou Tanaka, Nobukatsu Hojo, Shogo Seki, “VoiceGrad: non-parallel any-to-many voice conversion with annealed Langevin dynamics,” IEEE/ACM Transactions on Audio, Speech and Language Processing, Vol. 32, pp. 2213—2226, Mar. 20, 2024.