

2023 年度年次報告書

次世代 AI を築く数理・情報科学の革新

2023 年度採択研究代表者

齋藤 佑樹

東京大学 大学院情報理工学系研究科

助教

インタラクティブ強化学習に基づく人間参加型適応的テキスト音声合成

研究成果の概要

2023 年度は, (1)本研究の応用先として着目している方言音声合成モデルのベースライン構築と(2)音声発話スタイルの自動キャプション技術に取り組んだ。

(1) 方言音声はアクセントのラベル付けが困難という課題に対し, データ駆動でアクセントの潜在変数を獲得する音声合成モデル学習法を提案した。評価結果から, 提案技術は対象方言のネイティブ話者・非ネイティブ話者を両方含むクラウドソーシングの評価者のみならず, ネイティブにとっても自然と知覚される方言アクセントを再現できることを示した。本研究成果は国内研究会で発表し, 筆頭著者の学生が IEICE SP 研究会の学生ポスター賞を受賞した。2024 年度はこの提案技術をより多くの方言に対して適用できるよう拡張し, 原著論文としてまとめる予定である。

(2) 方言に限らず, より広範な発話スタイルを自然言語で記述する技術は, 本研究のみならず, 音声合成モデルを大規模言語モデルと統合するための基盤技術となりうる, この技術の初期検討として, 入力された音声の音高, 音量, 話速, 話者の性別に関するスタイル記述を生成する “StyleCap” を提案した。評価結果から, 高精度なスタイル記述生成には大規模音声データで事前学習された自己教師あり学習モデルと, よりパラメータ数が多いリッチな大規模言語モデルの活用が不可欠であることを示した。本研究成果は音声研究分野におけるトップカンファレンスである ICASSP2024 にてポスターで発表した。2024 年度はこの提案技術をより多くの発話スタイルに対して適用できるよう拡張し, 原著論文としてまとめる予定である。

以上に加え, 2024 年度は方言音声合成のための Web インタフェースや, 発話スタイル予測モデルを報酬として考慮したインタラクティブ強化学習アルゴリズムの開発について取り組む予定である。