

公益財団法人 立松財団 御中
様式 2021A1,A2,B

2024年10月25日

所属:名古屋工業大学大学院工学研究科

氏名: 南角 吉彦



2022年度助成 研究終了 報告書

※ゴシック文字で記入下さい。

研究テーマ	深層学習と統計的生成モデルの融合に基づく次世代音声合成技術の開発
研究の結果	<p>近年、深層学習に基づく音声合成が目覚ましい成果を上げており、特に入力から出力まですべてを単一のニューラルネットワークで構成する End-to-End 音声合成の研究が注目を集めている。しかし、End-to-End 音声合成には、次のような3つの本質的な問題点があった。1. 従来手法に比べ学習に大量のデータが必要(従来:1 時間程度⇒10 時間以上)。2. 直感的に理解しやすい音声特徴(声の高さ、音色、長さ)に基づいた合成音声のコントロールが困難。3. 学習に有用な入力と出力を繋ぐ中間的な特徴表現や不完全なデータを利用する枠組みが未確立。これらの問題に対し、本研究では従来のデファクトスタンダードであった HMM 音声合成と深層学習に基づく End-to-End 音声合成を融合によって、理論的に整理された形でこれらの問題の解決を目指して研究を行った。提案法の核となる具体的なアイデアは大きく2つある。</p> <p>アイデア1:統計的生成モデルに基づくニューラルネットワークの構造化 アイデア2:音声合成のための中間表現を利用した半教師有り学習の枠組みの確立</p> <p>本研究では、上記のアイデアを理論的に整理し、計算機による実装および評価実験を行った。アイデア 1 に関して研究成果を取りまとめ、トップカンファレンスにおいて発表を行った(研究発表[1,2])。また、アイデア2に関する研究成果を(研究発表[1,4])として発表した。さらには、アイデア 2 を発展させるための音声波形モデルについて、先行して成果を上げることができ、国内外の学会において発表を行った(研究発表[3,5,6])。</p>
研究発表 (実績)	<ol style="list-style-type: none"> 1. Takato Fujimoto, Kei Hashimoto, Yoshihiko Nankaku, and Keiichi Tokuda, "Autoregressive variational autoencoder with a hidden semi-Markov model-based structured attention for speech synthesis," Proceedings of 2022 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2022), pp. 7462-7466, Singapore, May 7-13, 2022. 2. Hono, Kei Hashimoto, Yoshihiko Nankaku, and Keiichi Tokuda, "Singing voice synthesis based on a musical note position-aware attention mechanism," Proceedings of 2023 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2023), Greece, June 4-10, 2023. 3. Takenori Yoshimura, Shinji Takaki, Kazuhiro Nakamura, Keiichi Oura, Yukiya Hono, Kei Hashimoto, Yoshihiko Nankaku, and Keiichi Tokuda, "Embedding a differentiable mel-cepstral synthesis filter to a neural speech synthesis system," Proceedings of 2023 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2023), Greece, June 4-10, 2023. Yukiya Hono, Kei Hashimoto, Yoshihiko Nankaku, and Keiichi Tokuda, "Embedding a differentiable mel-cepstral synthesis filter to a neural speech synthesis system," Proceedings of 2023 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2023), Greece, June 4-10, 2023. Yukiya Hono, Kei Hashimoto, Yoshihiko Nankaku, and Keiichi Tokuda, "Embedding a differentiable mel-cepstral synthesis filter to a neural speech synthesis system," Proceedings of 2023 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2023), Greece, June 4-10, 2023. Yukiya Hono, Kei Hashimoto, Yoshihiko Nankaku, and Keiichi Tokuda, "Embedding a differentiable mel-cepstral synthesis filter to a neural speech synthesis system," Proceedings of 2023 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2023), Greece, June 4-10, 2023. 4. 藤本崇人, 橋本佳, 南角吉彦, 徳田恵一, "半教師あり学習を用いた階層化生成モデルに基づく日本語 end-to-end 音声合成," 日本音響学会 2022 年秋季研究発表会講演論文集, pp. 1579-1582, September 14-16, 2022. (第7回 IEEE Signal Processing Society Tokyo Joint Chapter Student Award) 5. 青原光, 法野行哉, 橋本佳, 南角吉彦, 徳田恵一, "ソース・フィルタ型ニューラルボコーダにおける周期励振信号に関する考察," 日本音響学会 2024 年春季研究発表会講演論文集, pp. 813-816, March 6-8, 2024. (日本音響学会 学生優秀発表賞) 6. 藤本崇人, 橋本佳, 南角吉彦, 徳田恵一, "V2Coder: 階層型 VAE に基づくニューラルボコーダ," 日本音響学会 2023 年秋季研究発表会講演論文集, pp. 1051-1054, September 26-28, 2023.

提出期限: 研究期間終了後、すみやかに助成金の「必要経費使途明細書」「領収書」と合わせて提出下さい。
年度をまたぐ場合は毎年3月末日までに、途中経過をご記入の上、報告願います。