



2023 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	クロスモーダルボイスクローニングのための深層学習に基づく音声合成技術の開発
研究の結果	<p>本研究では、顔画像からその人物の声質を予測し、その人物の声質を再現した音声合成システムを構築可能とする深層学習に基づくクロスモーダルボイスクローニング技術の開発に取り組んだ。</p> <p>本研究では、音声・テキスト・顔画像の関係をモデル化するために、音声から声質に関する潜在変数を獲得する speech encoder、顔画像から声質に関する潜在変数を獲得する face encoder を構成し、これらの潜在変数を対照学習によってモデル化する手法を提案した。提案法では獲得した潜在変数を用いて speech decoder を学習することで、潜在変数が表す人物の声質の音声を生成可能とする。対照学習によって同一人物の潜在変数の距離を小さくし、異なる人物の潜在変数の距離を大きくすることで、顔画像と音声の関係をより適切に表現可能な潜在変数を獲得することが可能となり、顔画像からその人物の声質を精度良く予測することを可能とした。また、speech encoder と face encoder の対照学習には発話内容を表すテキストを使用しないため、音声と顔画像のペアデータを用いて学習可能という利点もある。さらに、音声・テキストのペアデータを同時に用いることで、speech decoder の学習も同時に行うことが可能である。speech decoder を同時に学習することで、speech encoder と face encoder が獲得する潜在変数はその人物の声質を再現するために適したものとなり、声質の予測精度をさらに向上させた。主観評価実験による評価の結果、speech decoder との同時学習によって合成音声の自然性および顔画像に対する親和性を従来法から改善することが示された。また、speech encoder と face encoder の学習に二乗誤差を利用した場合では潜在変数が話者ごとに分類されておらず、音声と顔画像の関係を十分に表現できていなかったが、提案法では対照学習を導入することで潜在変数は話者ごとに分類され、音声と顔画像の関係を適切に学習していることが示された。これらに加えて、多様な声質の合成音声を高品質に生成するために深層学習に基づく音声波形モデルの研究に取り組んだ。</p>
研究発表 (実績)	<ol style="list-style-type: none"> 1. Yukiya Hono, Kei Hashimoto, Yoshihiko Nankaku, and Keiichi Tokuda, "PeriodGrad: Towards Pitch-Controllable Neural Vocoder Based on a Diffusion Probabilistic Model," ICASSP 2024, pp. 12782-12786, Korea, April 2024. 2. 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, pp.813-816, 2024 年 3 月. 3. 程塚海月, 橋本佳, 南角吉彦, 徳田恵一, "Noisy Student による半教師あり学習を用いた話者識別," 日本音響学会 2024 年春季研究発表会, pp.837-840, 2024 年 3 月. 4. 鈴木耀, 藤本崇人, 高木信二, 橋本佳, 南角吉彦, 徳田恵一, "出力遅延と時間伸縮変換を考慮したリアルタイム声質変換の検討," 日本音響学会 2023 年秋季研究発表会, pp. 1081-1084, 2023 年 9 月. 5. 福田至音, 法野行哉, 橋本佳, 南角吉彦, 徳田恵一, "基本周波数の制御性を考慮したピッチ抽出器を用いたニューラルボコーダ学習法の検討," 日本音響学会 2023 年秋季研究発表会, pp. 1065-1068, 2023 年 9 月. 6. 佐藤鈴夏, 藤本崇人, 法野行哉, 橋本佳, 南角吉彦, 徳田恵一, "基本周波数の制御性を考慮した特徴分離に基づくニューラルボコーダ構成法," 日本音響学会 2023 年秋季研究発表会, pp. 1061-1064, 2023 年 9 月. 7. 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, pp. 1051-1054, 2023 年 9 月. 8. 法野行哉, 橋本佳, 南角吉彦, 徳田恵一, "PeriodGrad: 基本周波数を制御可能な拡散確率モデル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, pp. 1045-1048, 2023 年 9 月.

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。