

どんな研究	従来の音声変換技術の研究では発話の終了を待ってから発話単位で変換しています。そのため、話をしている最中に音声を変換することができません。抑揚や声質における話者ごとの違いを考慮した深層生成モデルを用い、 高音質かつ低遅延でリアルタイム変換する技術 を紹介します。
どこが凄い	話者間変動に対する頑健性を考慮することで、話者依存性の低い音声表現を学習 することが可能になり、発話の終わりを待たない高音質なリアルタイム変換を実現しました。また、波形合成器の改善にも取り組むことで、更なる高速化・低消費電力化も同時に実現しています。
めざす未来	対面・遠隔とを問わず、様々な音声コミュニケーションでの応用が可能 になります。例えば、ライブ配信やコールセンターに用いると、発話者を秘匿できます。今後はテキスト音声合成技術や音声認識技術を組みこみ、より安定して好きな声でコミュニケーションできる未来をめざします。

音声変換によるコミュニケーション拡張

音声変換技術を通じて多様なシーンにおいて人が不自由なくコミュニケーションできる手段を創出



提案する要素技術

リアルタイム音声変換

動作する必須条件

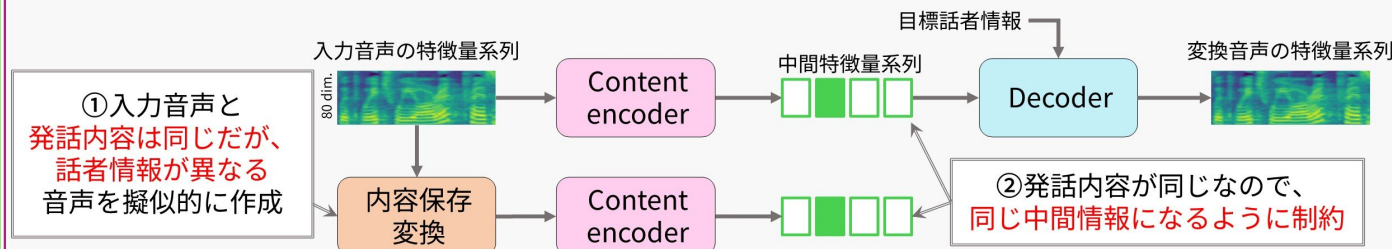
8ミリ秒毎に
入力される音声のセグメントを
8ミリ秒以内に
変換音声のセグメントへ変換

望まれる条件

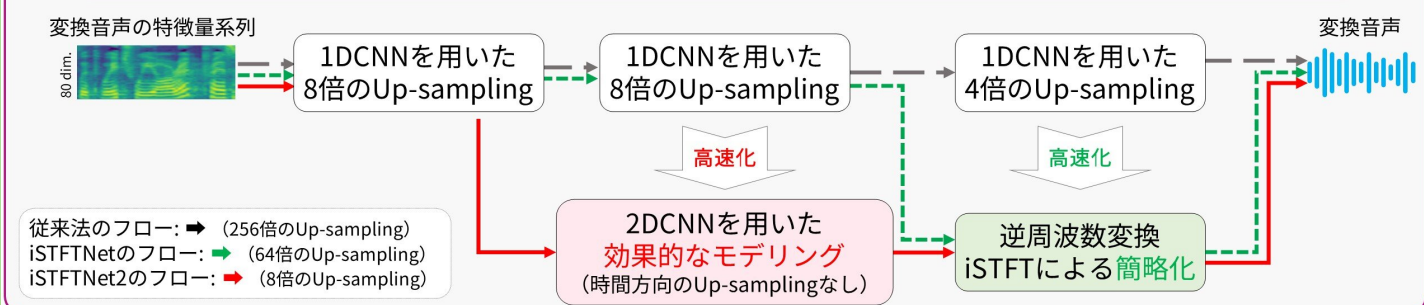
高音質かつ低遅延



PRVAE-VC: 話者依存性の低い音声表現の獲得により高音質な特徴量変換を実現



iSTFTNet: DNNアップサンプリング+iSTFTのハイブリッド処理により低遅延な波形合成を実現



関連文献

- [1] K. Tanaka, H. Kameoka, T. Kaneko, "PRVAE-VC: Non-Parallel Many-to-Many Voice Conversion with Perturbation-Resistant Variational Autoencoder," in *Proc. the 12th Speech Synthesis Workshop (SSW)*, 2023.
- [2] T. Kaneko, H. Kameoka, K. Tanaka, S. Seki, "iSTFTNet2: Faster and More Lightweight iSTFT-Based Neural Vocoder Using 1D-2D CNN," in *Proc. INTERSPEECH*, 2023.

連絡先

田中 宏 (Kou Tanaka) メディア情報研究部 基礎事象研究グループ