

どんな研究

従来の音声変換技術は、発話の終了を待ってから発話単位で変換を行うため、ライブ配信での使用に不向きでした。本技術は、複数の音声表現学習手法を組み合わせ、知識蒸留した深層生成モデルに基づき、**高音質で低遅延なリアルタイム音声変換およびライブ配信**を実現しました。

どこが凄い

話者変換に有用な音声表現を学習するために、**複数の音声表現学習手法を組み合わせ、知識蒸留**することにより、**発話の終わりを待たない高音質なリアルタイム音声変換**を実現しました。また、波形合成器の改善により演算量を大幅に削減し、スマートフォン上での動作も可能にしました。

めざす未来

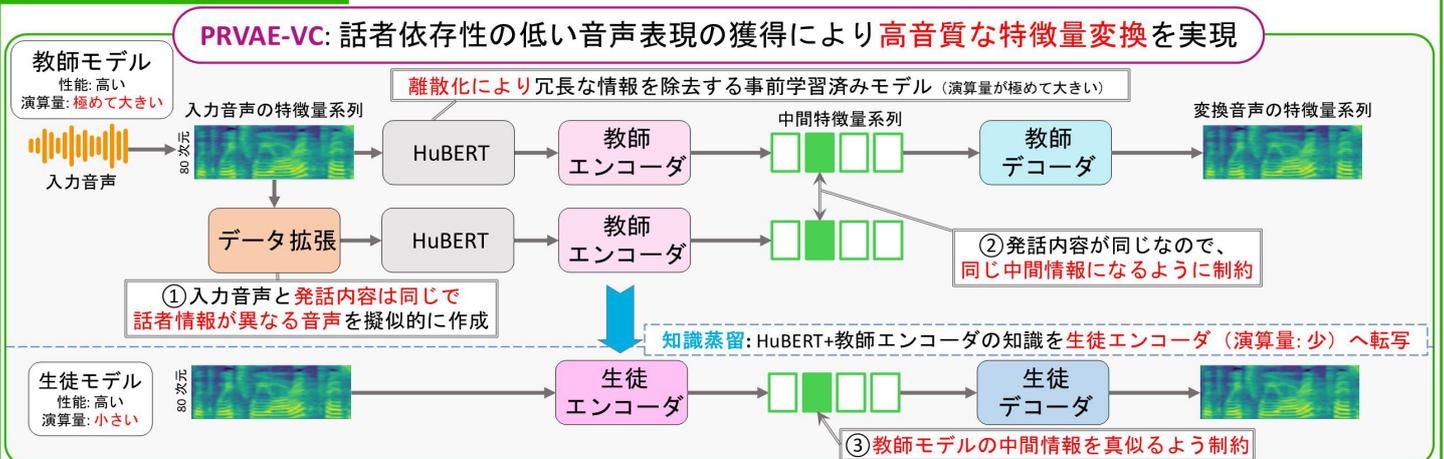
キャラクタになりきるライブ配信や声に悩みを抱える方のウェルビーイング向上など、**リアルタイム音声変換は様々な音声コミュニケーションへの応用を可能にします**。今後は声質以外の音声特徴を変換し、より聞きたい声・話したい声でコミュニケーションできる未来をめざします。

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



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)*, pp. 88-93, 2023.
- [2] K. Tanaka, H. Kameoka, T. Kaneko, Y. Kondo, "PRVAE-VC2: Non-parallel voice conversion by distillation of speech representations," in *Proc. INTERSPEECH*, pp. 4363-4367, 2024.
- [3] T. Kaneko, H. Kameoka, K. Tanaka, S. Seki, "iSTFTNet: Fast and lightweight mel-spectrogram vocoder incorporating inverse short-time Fourier transform," in *Proc. ICASSP*, pp. 6207-6211, 2022.
- [4] T. Kaneko, H. Kameoka, K. Tanaka, S. Seki, "iSTFTNet2: Faster and more lightweight iSTFT-based neural vocoder using 1D-2D CNN," in *Proc. INTERSPEECH*, pp. 4369-4373, 2023.

連絡先

田中 宏 (Kou Tanaka) メディア情報研究部 事象モデリング研究グループ