

18

声と話し方を好みのスタイルに変える

～系列変換モデルに基づく声質と韻律の同時変換～

どんな研究

話し声は抑揚や声質、リズムによって雰囲気が大きく変化します。系列変換モデルを用い、音声の様々な変動要素を柔軟かつ高品質に変換・生成する技術の研究です。例えば、**ユーザーが望んだ人の音声に自動変換したり、聞き取りやすい母語話者の発音とリズムに変換することが可能です。**

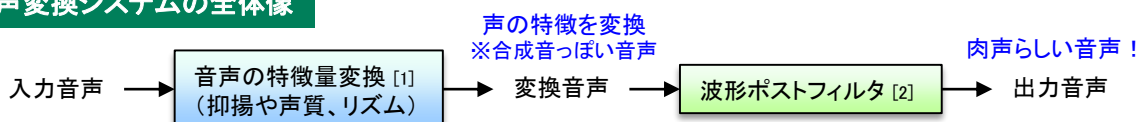
どこが凄い

従来の音声変換技術の研究では声質の変換のみに着目したものが多かったのですが、本技術では**声質だけでなく抑揚やリズムも自動変換することが可能です。**また、**機械的に感じる合成音声の波形を心地よく聞き取りやすい音声へと直接変換する**深層学習アプローチを世界で初めて実現しました。

めざす未来

本技術を用いることで、既存の音声変換で実現し得なかった**高品質な音声変換が可能になります。**今後は、感情音声変換や非母語話者の聞き取りにくい音声を母語話者風の音声に変換する研究、発声障がい者支援への応用などを行い、多様で表情豊かな音声変換技術の実現を目指しています。

音声変換システムの全体像

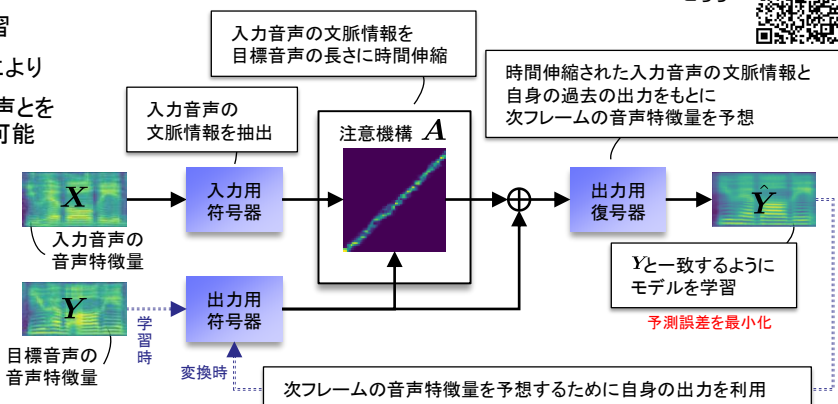


音声の特徴量変換 [1]

(例: 声を別の人の声に変換する、英語の発音をネイティブ風に変換する)

音声デモはこちら→

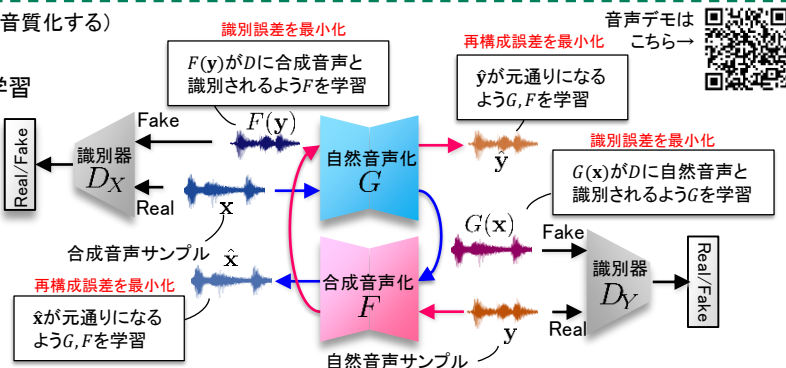
- 各符号器、復号器と注意機構を学習
- 符号器復号器モデルを用いることにより
 - 異なる長さの入力音声と目標音声とを一つのモデルで処理することが可能
 - その結果、声質だけでなくリズムも変換することが可能に!
- 注意機構を用いることにより
 - 入力音声のどの部分を考慮して音声特徴量を予測するべきかを判断することが可能
 - その結果、長期的な依存関係を考慮でき、韻律も変換可能に!



波形ポストフィルタ [2]

(例: 音声を高音質化する)

- 合成と自然音声の間の変換関数 G, F を学習
- 循環型モデルを用いることにより
 - 合成音声と自然音声の平行データが不要
 - 音声波形を再構成する際に位相情報まで考慮することが可能
- 敵対的生成モデルを用いることにより、クリアな音質の音声を生成可能



関連文献

- [1] K. Tanaka, H. Kameoka, T. Kaneko, N. Hojo, "AttS2S-VC: Sequence-to-Sequence Voice Conversion with Attention and Context Preservation Mechanisms," in *Proc. 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2019)*, May 2019.
- [2] K. Tanaka, H. Kameoka, T. Kaneko, N. Hojo, "WaveCycleGAN2: Time-domain Neural Post-filter for Speech Waveform Generation," *arXiv:1904.02892*, Apr. 2019, (submitted to *Interspeech2019*).

連絡先

田中 宏 (Kou Tanaka) メディア情報研究部 メディア認識研究グループ
Email: cs-liaison-ml at hco.ntt.co.jp



Innovative R&D by NTT

オープンハウス 2019

Copyright © 2019 NTT. All Rights Reserved.