

20

声の雰囲気や聞き取りやすさを変換する

～深層生成モデルを用いた音声属性変換～



どんな研究

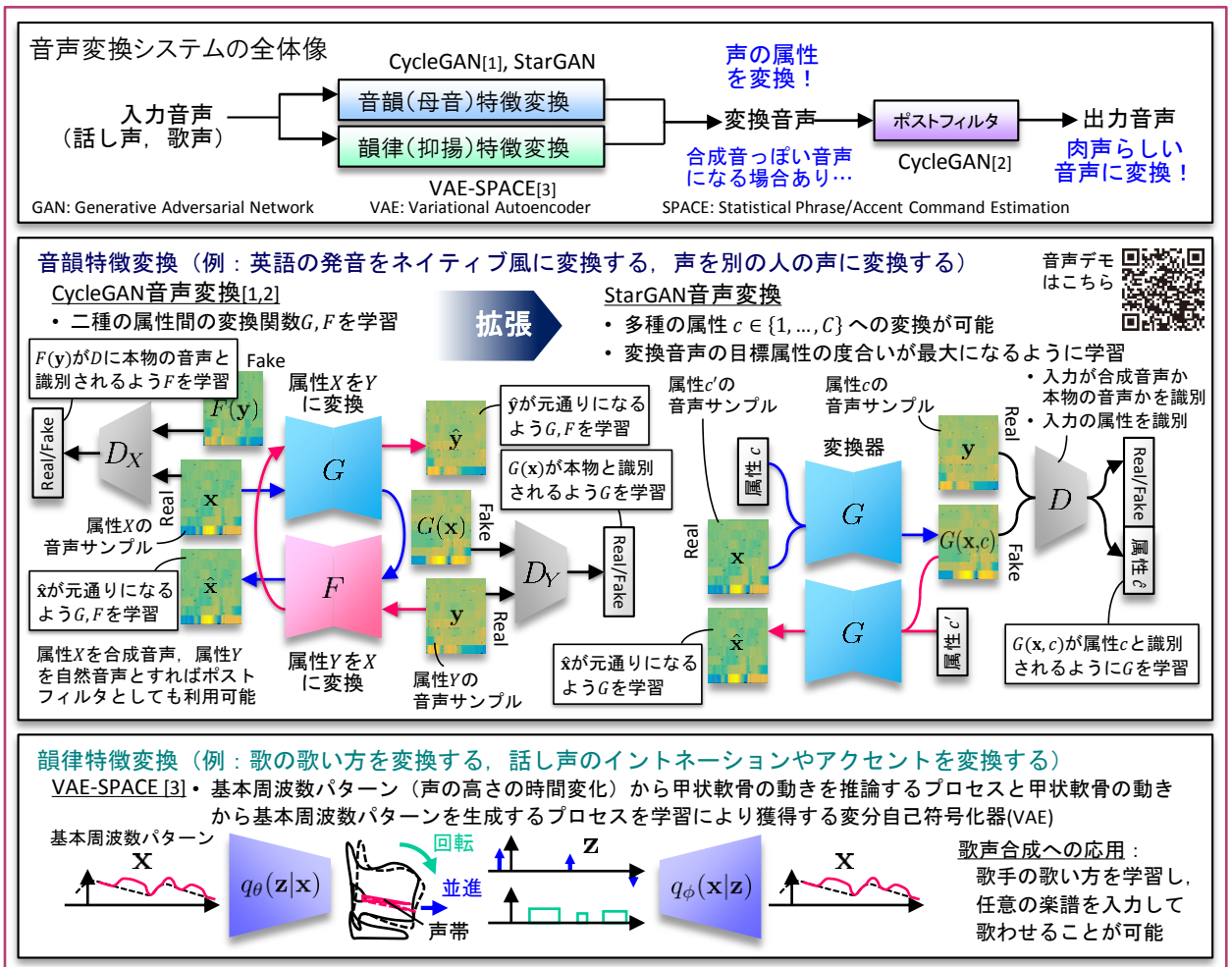
深層生成モデルを用い、音声の属性や非言語情報を柔軟かつ高品質に変換・生成する技術の研究です。例えば非母語話者の音声を**母語話者風の発音の音声に自動変換**したり、好きな歌手の歌い方で**任意の楽譜を歌わせたり**することが可能です。

どこが凄い

音声という現象には言語情報以外に様々な変動要因が複雑にもつれ合って混在しています。本研究では、敵対的生成ネットワーク (GAN) や変分自己符号化器 (VAE) などの深層生成モデルを用いて**これら変動要因のもつれを解くことで属性の個別操作が可能な音声変換法を検討**しています。

めざす未来

聞き取りにくい音声を聞き取りやすい音声に変換することでコミュニケーションを円滑化することができます。本研究では、音声変換技術を通して**非母語話者音声や喉頭摘出者などの音声を聞き取りやすい発音や抑揚の音声に変換するシステム**などの実現を目指しています。



関連文献

- [1] T. Kaneko, H. Kameoka, "Non-Parallel voice conversion using cycle-consistent adversarial networks," accepted for publication in *Proc. 26th European Signal Processing Conference (EUSIPCO 2018)*.
- [2] 田中宏, 金子卓弘, 北条伸克, 亀岡弘和, "CycleGANを用いた合成音声から自然音声への波形変換," 日本音響学会2018年春季研究発表会講演論文集, 1-Q-32, pp. 285-286, Mar. 2018.
- [3] K. Tanaka, H. Kameoka, Kazuho Morikawa, "VAE-SPACE: Deep generative model of voice fundamental frequency contours," in *Proc. 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2018)*, pp. 5779-5783, Apr. 2018.

担当者

亀岡 弘和 (Hirokazu Kameoka) メディア情報研究部 メディア認識研究グループ