Speech conversion using vocal tract model and deep generative models

17 英語の発音をネイティブのように綺麗に変換

～声道モデルと深層生成モデルを用いた統計的音声変換～

どんな研究

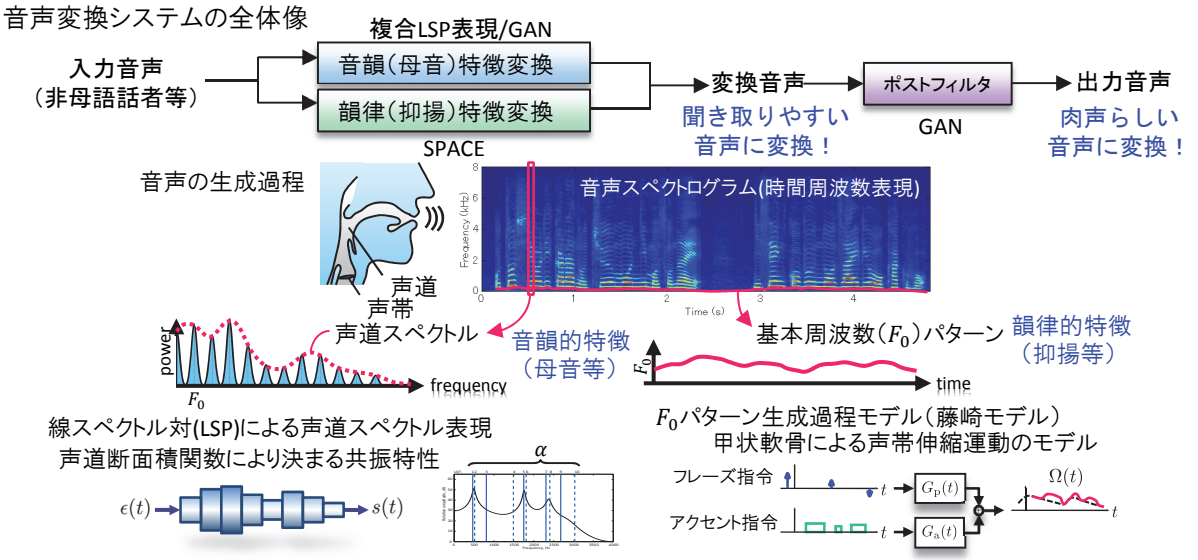
非母語話者の聞き取りにくい音声を母語話者風の発音の音声に自動変換する研究です。本技術では、複合線スペクトル対（LSP）表現と呼ぶ独自の声道モデルを用いた信号処理アプローチと、敵対的生成ネットワーク(GAN)を用いた深層学習アプローチをとっています。

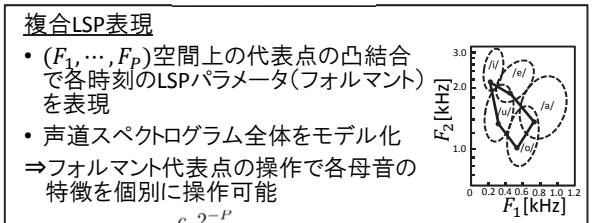
どこが凄い

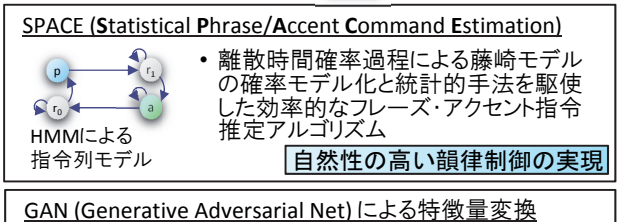
本技術は変換音声の自然さを多面的に追求しています。複合LSP表現では音声の生成過程のモデルを用いており、発声器官から生成され得る範囲の自然さを保ったまま各母音の特徴を個別に調整できます。GANポストフィルタでは合成音声をより肉声らしくなるように変換・合成します。

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の実現を目指しています。







関連文献

[1] H. Kameoka, K. Yoshizato, T. Ishihara, K. Kadowaki, Y. Ohishi, K. Kashino, "Generative modeling of voice fundamental frequency contours," *IEEE/ACM Transactions on Audio, Speech and Language Processing*, Vol. 23, No. 6, pp. 1042-1053, Jun. 2015.

[2] T. Kaneko, H. Kameoka, N. Hojo, Y. Ijima, K. Hiramatsu, K. Kashino, "Generative adversarial network-based postfilter for statistical parametric speech synthesis," in *Proc. 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2017)*, pp. 4910-4914, Mar. 2017.

連絡先

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

E-mail: kameoka.hirokazu(at)lab.ntt.co.jp