Crossmodal voice conversion with deep generative models

19 顔に合わせて声を作り、声に合わせて顔を作る

～深層生成モデルによるクロスモーダル音声変換～

どんな研究

私たちは、声の印象からその人がどんな顔か、また、顔の印象からその人がどんな声かをある程度想像できます。これは、声と顔には何らかの相関があるからだと考えられます。本研究では、与えられた顔画像の印象に合った声を作り出すクロスモーダル音声合成の問題に初めて取り組みました。

どこが凄い

音声変換器を深層生成モデルで表し、出力音声と入力顔画像との相互情報量を規準として音声変換器を学習する情報論的アプローチを考案しました。これにより、入力顔画像に合った声質に入力音声を変換するクロスモーダル声質変換技術を実現することに初めて成功しました。

めざす未来

私たち人間は、異なる感覚器官から得られる情報（視覚情報や聴覚情報など）を無矛盾に関連付けてモノや出来事を認識しています。本研究では、人間のこの知的な認識機能を実現することを究極の目標としています。

クロスモーダル声質変換問題の提案

声のみから顔を想像したり顔のみから声の雰囲気を想像したりできる⇒声質と容貌には相関があることを示唆

→
- 音声特徴量系列と顔画像との間の相関を捉え、入力音声を入力顔画像に適合した声質に変換する
- さらに入力音声に適合した容貌の顔画像を生成する

情報論的アプローチによる問題の定式化

変換器G：音声$\mathbf{x}$と顔画像$\mathbf{y}$を入力として変換音声$\hat{\mathbf{x}} = G(\mathbf{x},\mathbf{y})$を出力するニューラルネットワーク(NN)

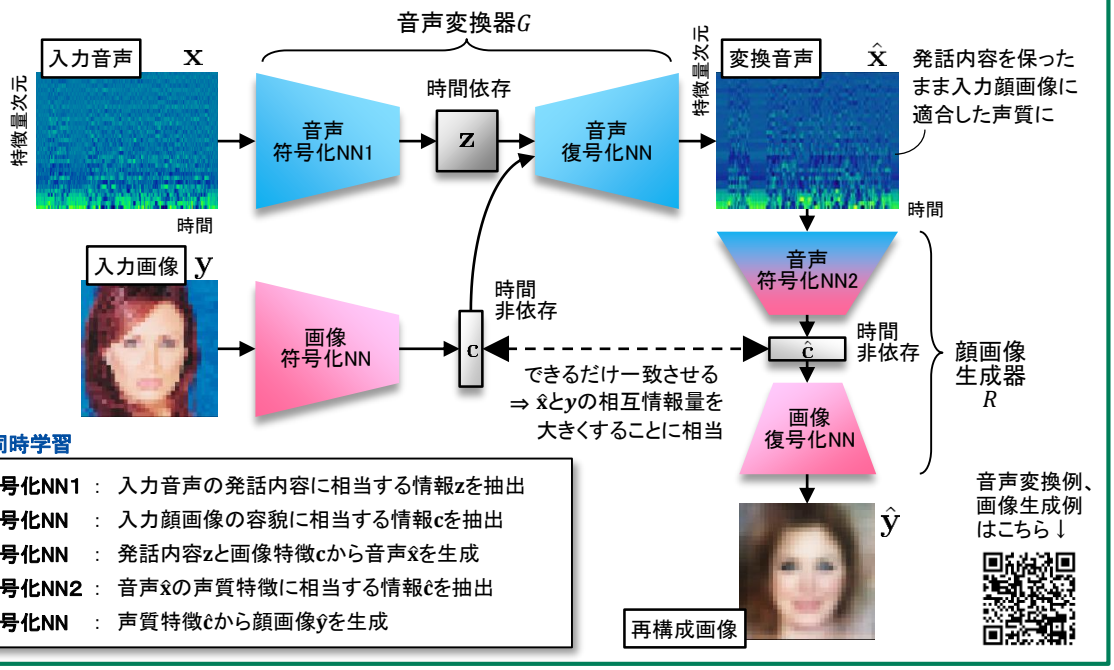
変換器Gの学習：$\hat{\mathbf{x}} = G(\mathbf{x},\mathbf{y})$と$\mathbf{y}$の相互情報量を規準として$G$を学習する問題として定式化

$$I[G(\mathbf{x},\mathbf{y})\|\mathbf{y}] \geq \mathbb{E}_{(\mathbf{x},\mathbf{y})\sim p(\mathbf{x},\mathbf{y})}[\log R(\mathbf{y}|G(\mathbf{x},\mathbf{y}))]$$

→ 下界をGとRに関して最大化

音声と顔画像のペアデータ

音声$G(\mathbf{x},\mathbf{y})$から予測される顔画像$\mathbf{y}$の条件付分布を近似するNN



5つのNNを同時学習

音声符号化NN1	：	入力音声の発話内容に相当する情報$\mathbf{z}$を抽出
画像符号化NN	：	入力顔画像の容貌に相当する情報$\mathbf{c}$を抽出
音声復号化NN	：	発話内容$\mathbf{z}$と画像特徴$\mathbf{c}$から音声$\hat{\mathbf{x}}$を生成
音声符号化NN2	：	音声$\hat{\mathbf{x}}$の声質特徴に相当する情報$\hat{\mathbf{c}}$を抽出
画像復号化NN	：	声質特徴$\hat{\mathbf{c}}$から顔画像$\hat{\mathbf{y}}$を生成

関連文献

[1] H. Kameoka, T. Kaneko, K. Tanaka, N. Hojo, "StarGAN-VC: Non-parallel many-to-many voice conversion using star generative adversarial networks," in *Proc. 2018 IEEE Workshop on Spoken Language Technology (SLT 2018)*, pp. 266-273, 2018.

[2] H. Kameoka, T. Kaneko, K. Tanaka, N. Hojo, "ACVAE-VC: Non-parallel voice conversion with auxiliary classifier variational autoencoder," *arXiv:1808.05092 [stat.ML]*, 2018.

[3] H. Kameoka, K. Tanaka, A. Valero Puche, Y. Ohishi, T. Kaneko, "Crossmodal Voice Conversion," *arXiv:1904.04540 [cs.SD]*, 2019.

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

Email: cs-liaison-ml at hco.ntt.co.jp

Innovative R&D by NTT

オープンハウス 2019