INTERSPEECH 2020
OCTOBER 25-29/ SHANGHAI , CHINA
SHANGHAI INTERNATIONAL CONVENTION CENTER

Audio samples

CycleGAN-VC3:

Examining and Improving CycleGAN-VCs for Mel-spectrogram Conversion


Takuhiro Kaneko


Hirokazu Kameoka


Kou Tanaka


Nobukatsu Hojo

NTT NTT Communication Science Laboratories, NTT Corporation, Japan

Problem: Non-parallel Voice Conversion

Speaker A

This is a pen.

Unpaired utterances

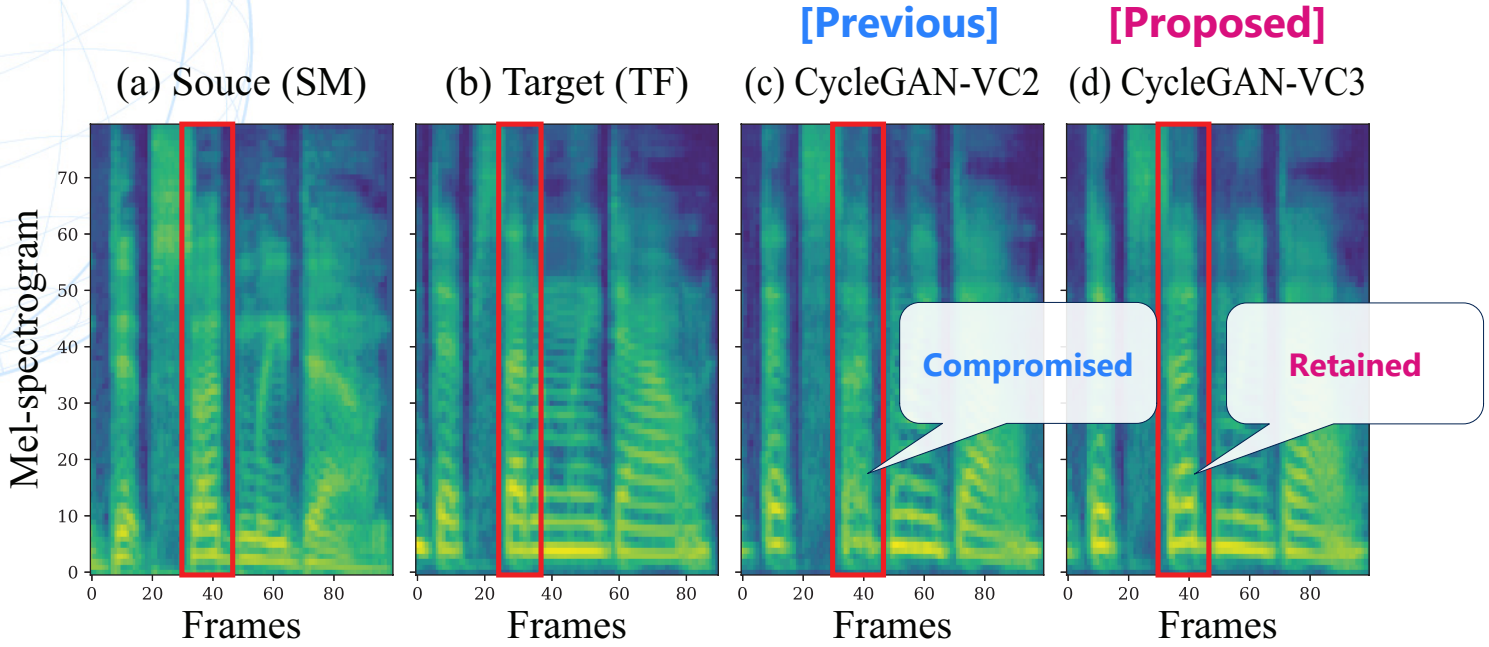
Speaker B

Hello world.

In particular, we focus on conversions
in **mel-spectrogram domain** based on **CycleGAN-VC** [1, 2]

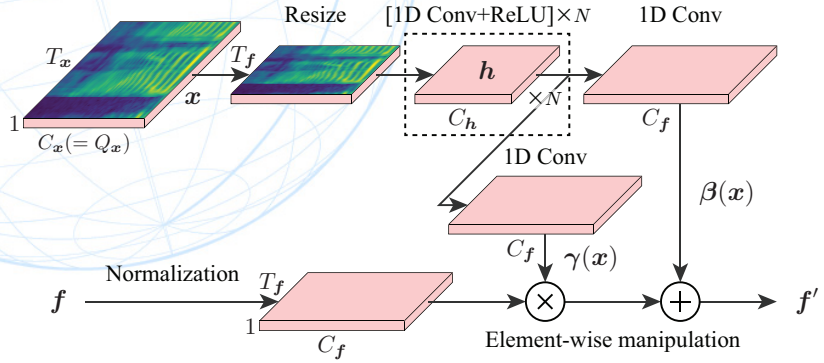
[1] Kaneko & Kameoka, "CycleGAN-VC: Non-parallel Voice Conversion Using Cycle-Consistent Adversarial Networks," EUSIPCO 2018.
[2] Kaneko et al., "CycleGAN-VC2: Improved CycleGAN-based Non-parallel Voice Conversion," ICASSP 2019.

Challenge: Linguistic Content Preservation

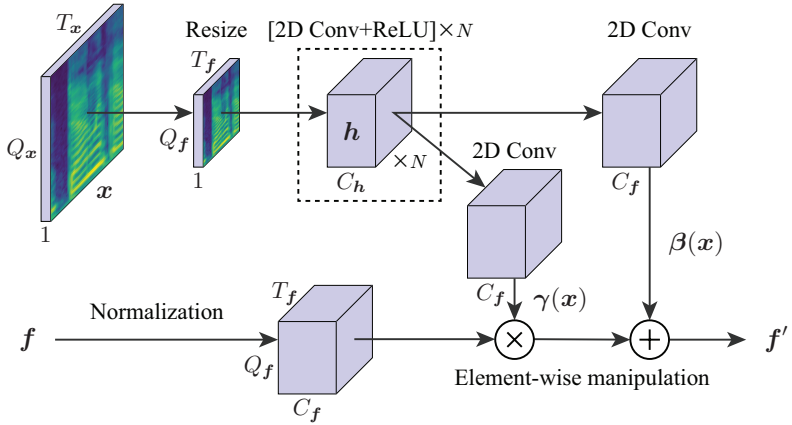


Required to **convert only voice factors**
while **retaining linguistic content factors**

Proposal: Time-Frequency Adaptive Normalization



1D TFAN

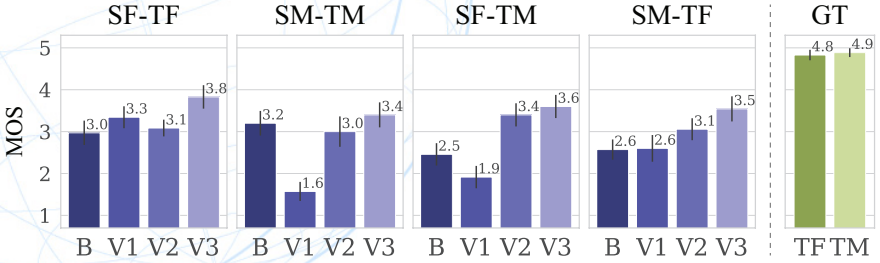


2D TFAN

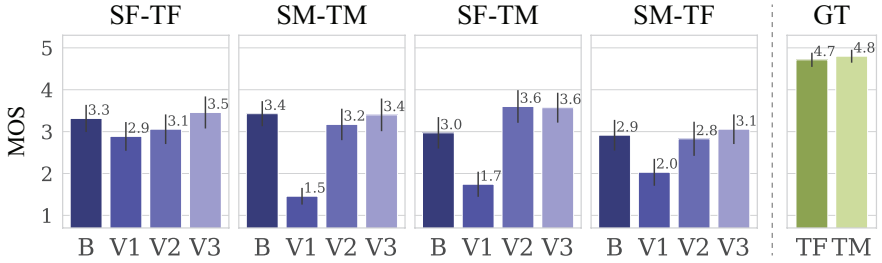
Extends **Instance Normalization** [3] so that affine parameters become element-dependent and determined **according to input mel-spectrogram**

[3] Ulyanov et al., "Instance Normalization: The Missing Ingredient for Fast Stylization," arXiv 2016.

Experiments: CycleGAN-VC3 > CycleGAN-VC2



MOS for naturalness



MOS for speaker similarity

B: Mel-cepstrum conversion by CycleGAN-VC2 [2] + WORLD vocoder [4]
V1: Mel-spectrogram conversion by CycleGAN-VC [1] + MelGAN vocoder [5]
V2: Mel-spectrogram conversion by CycleGAN-VC2 [2] + MelGAN vocoder [5]
V3: Mel-spectrogram conversion by CycleGAN-VC3 + MelGAN vocoder [5]

CycleGAN-VC3 showed its potential as **new benchmark!**

Audio samples are available at http://www.kecl.ntt.co.jp/people/kaneko.takuhiro/projects/cyclegan-vc3/index.html

[1] Kaneko & Kameoka, "CycleGAN-VC: Non-parallel Voice Conversion Using Cycle-Consistent Adversarial Networks," EUSIPCO 2018.
[2] Kaneko et al., "CycleGAN-VC2: Improved CycleGAN-based Non-parallel Voice Conversion," ICASSP 2019.
[4] Morise et al., "WORLD: A Vocoder-Based High-Quality Speech Synthesis System for Real-Time Applications," IEICE Trans. Inf. Syst., 2016.
[5] Kumar et al., "MelGAN: Generative Adversarial Networks for Conditional Waveform Synthesis," NeurIPS 2019.