CycleGAN-VC3: Examining and Improving CycleGAN-VCs for Mel-spectrogram Conversion

Takuhiro Kaneko
Hirokazu Kameoka
Kou Tanaka
Nobukatsu Hojo

NTT Communication Science Laboratories, NTT Corporation, Japan
Problem: Non-parallel Voice Conversion

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

Challenge: Linguistic Content Preservation

Required to convert only voice factors while retaining linguistic content factors
Proposal: Time-Frequency Adaptive Normalization

Extends Instance Normalization \cite{ulyanov2016instance} so that affine parameters become element-dependent and determined according to input mel-spectrogram.

Experiments: CycleGAN-VC3 > CycleGAN-VC2

MOS for naturalness

MOS for speaker similarity

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