Including Linguistic Knowledge in an Auxiliary Classifier CycleGAN for Corrective Feedback Generation in Korean Speech

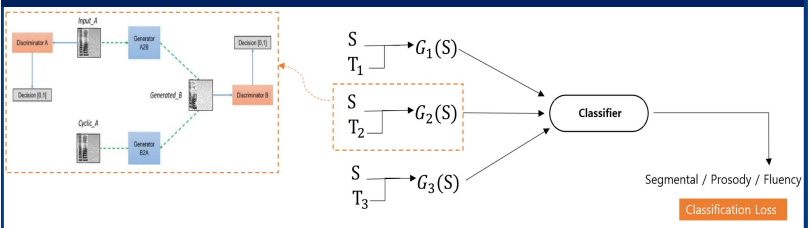
Seung Hee Yang [1], Minhwa Chung [1 2]
[1] Interdisciplinary Program in Cognitive Science, [2] Department of Linguistics, Seoul National University
sy2358@snu.ac.kr, mchung@snu.ac.kr

SLP Lab SNU Spoken Language Processing Lab
SEOUL NATIONAL UNIVERSITY
LT4ALL

1. Introduction

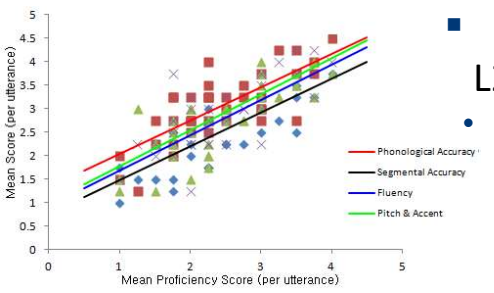
- Research Background
  - Corrective Feedback Generation for Computer-Assisted Pronunciation Training System Development for L2 Korean
- Self-imitating Feedback Generation Using CycleGAN (Yang, 2019))
  - Transplant native characteristics onto learners' voice
  - Interpret correction as a style transfer task
  - Utilise the generative ability for speech correction
- Problem Definition
  - Although the generator seems to have acquired what to correct with the adversarial training, this information is shared with the learner
  - Difficult to evaluate one's own pronunciations
  - Given non-native speech input, generate a corrected utterance with linguistic knowledge

2. Proposed Method



- AC-CycleGAN consists of 3 CycleGANs, corresponding to a linguistic class, and a domain classifier
- For each linguistic class, a CycleGAN with 2 discriminators and two 2 mapping functions as generators,
- synthetic sample is generated from the source
- The domain classifier learns to ensure the discriminability between the generated samples

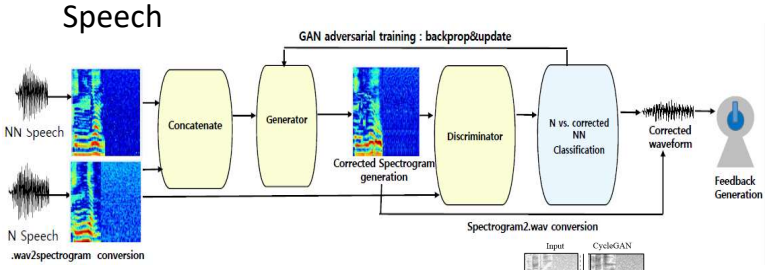
3. Linguistic Analysis



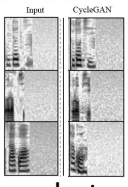
- Factors Affecting L2 Korean Speech
  - segmental accuracy (r= 0.81), fluency (r= 0.80), pitch and accent (r= 0.76), and phonological accuracy (r= 0.74)
- Correlation Analysis of Human Accented Ratings (Yang, 2017)

4. Experiment

- Corpus
  - 300 read speech from Korean as a Foreign Language Speech
  - 107 Native, 217 Learners with 27 L1 background
  - Training
    - 97,038 spectrogram images of 324 speakers
  - Testing
    - 162 spectrogram images
- Process
  - Auxiliary Classifier: Annotated Spectrograms →
  - CycleGAN: Speech → Spectrogram (Short-Time Fourier Transform using Griffin Lim Algorithm) → Adversarial Training → Spectrogram → Speech



- The Generator
  - Input: Learners' spectrogram
  - Output: Corrected spectrogram
  - Adversarial learning: Minimax game between G and D



$$\min_G \max_D V(D,G) = \mathbb{E}_{X\sim Pdata(X)}[\log D(X)] + \mathbb{E}_{Z\sim Pz(Z)}[\log(1-D(G(Z)))]$$

  - Add Cycle-consistency Loss
  - Cycle-consistency loss is built between real samples and their corresponding reconstructed samples.

$$L_{cyc}(G,F) = \mathbb{E}_{x\sim Pdata(x)}[\|F(G(x) - x\|_1] + \mathbb{E}_{y\sim Pdata(y)}[\|G(F(y) - y\|_1].$$

5. Conclusion

- New a machine learning method that retains domain knowledge
- AC-CycleGAN allows to work closely with linguistic analyses and machine learning
- Linguistic analysis results can be directly used in the automatic system

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2019-2016-0-00464) supervised by the IITP(Institute for Information & communications Technology Planning & Evaluation)