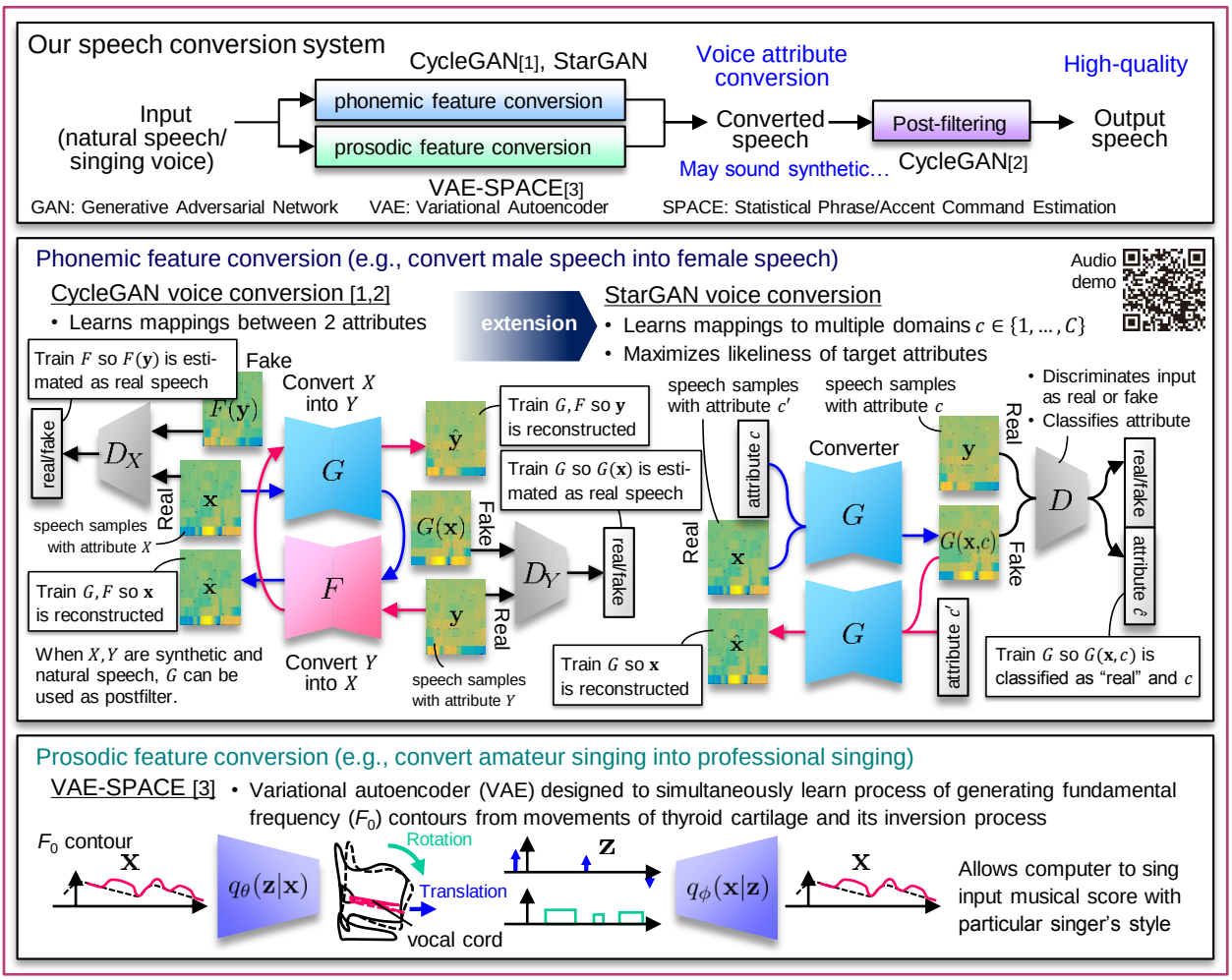
20 Converting impression and intelligibility of speech

- Speech attribute conversion using deep generative models -

Abstract

We are interested in developing a speech conversion system that can convert impressions of speech such as speaker identity, gender, and accents at will. Speech contains a number of variation factors in addition to linguistic information. These factors are highly entangled in speech waveforms and cannot be separated with simple transformations. We use deep generative models including generative adversarial networks (GANs), variational autoencoders (VAEs) and their extensions to develop methods that can effectively disentangle these factors and allow us to convert them individually. With our techniques, we can convert, for example, a male speech into a female speech, English-accented speech into American-accented speech, and amateur singing into professional singing. We hope to develop a real-time system using these techniques to overcome many kinds of barriers to our daily communication.



References

[1] T. Kaneko, H. Kameoka, "Non-Parallel voice conversion using cycle-consistent adversarial networks," submitted to *EUSIPCO2018*.
[2] K. Tanaka, T. Kaneko, N. Hojo, H. Kameoka, "Synthetic-to-natural speech conversion using cycle-consistent adversarial networks," submitted to *EUSIPCO2018*.
[3] K. Tanaka, H. Kameoka, Kazuho Morikawa, "VAE-SPACE: Deep generative model of voice fundamental frequency contours," in *Proc. ICASSP2018,* pp. 5779-5783, Apr. 2018.

Contact

Hirokazu Kameoka Recognition Research Group, Media Information Science Laboratory