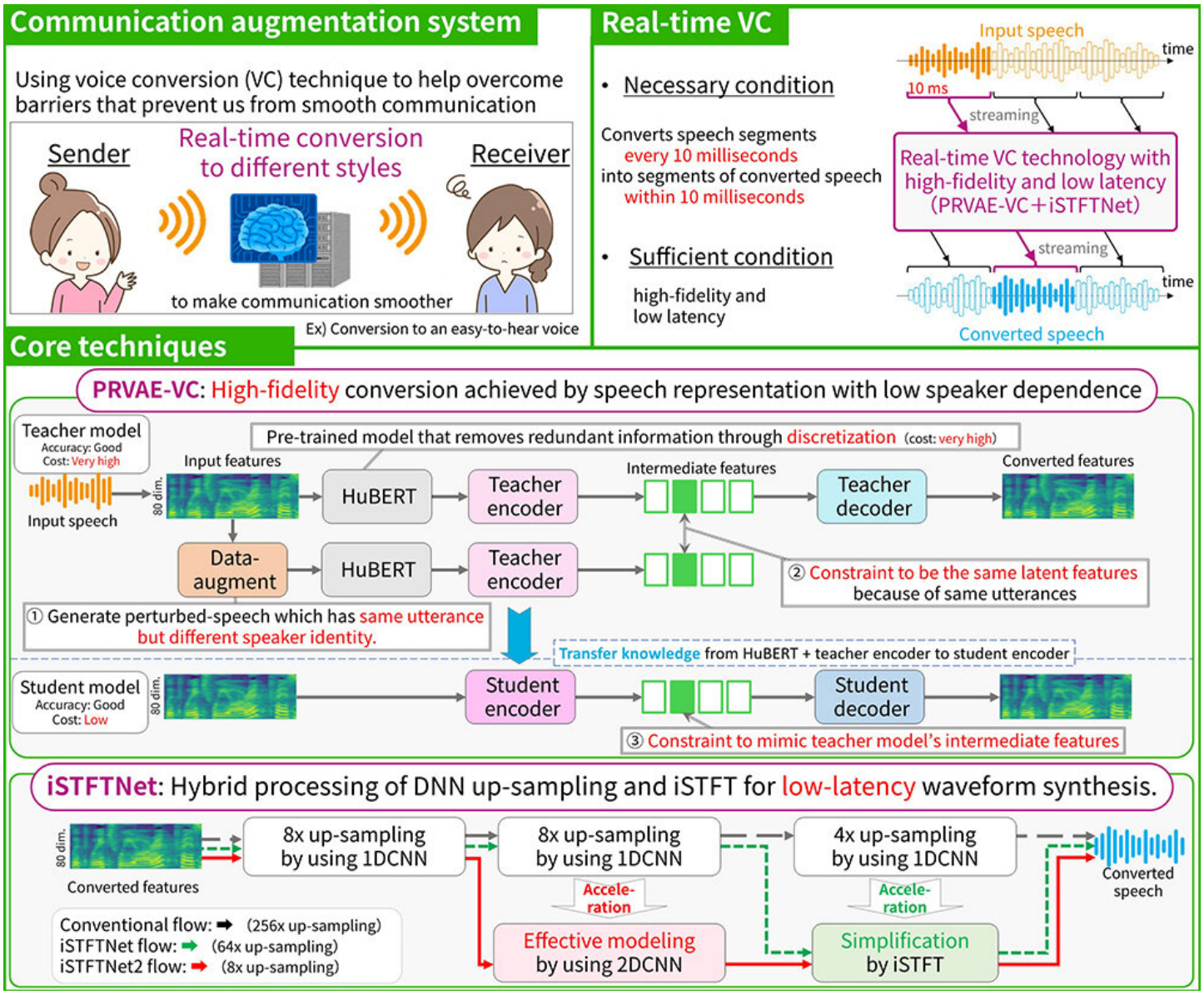
09

Real-time voice conversion with high quality and low latency

Live streaming with real-time voice conversion

Abstract

Conventional voice conversion technologies process speech after each utterance, making them unsuitable for live streaming. This research introduces a novel system that enables high-quality, low-latency real-time voice conversion and live streaming by combining multiple speech representation learning methods into a knowledge-distilled deep generative model. To improve reliability in voice conversion, various representation learning techniques are integrated and distilled, allowing real-time conversion without waiting for the end of speech. Additionally, waveform synthesis and inference costs are significantly reduced, making it feasible to run on smartphones. This technology aims to enhance well-being for individuals concerned about their voice and realize a live stream where you can pretend to be a certain character. It also opens up new possibilities for a variety of voice communication applications. Future developments will explore converting voice features other than voice timbre for more personalized and accessible communication.



References

[1] K. Tanaka, H. Kameoka, T. Kaneko, "PRVAE-VC: Non-parallel many- to-many voice conversion with perturbation-resistant variational autoencoder," in *Proc. the 12th Speech Synthesis Workshop (SSW)*, pp. 88-93, 2023.

[2] K. Tanaka, H. Kameoka, T. Kaneko, Y. Kondo, "PRVAE-VC2: Non-parallel voice conversion by distillation of speech representations," in *Proc. INTERSPEECH*, pp. 4363-4367, 2024.

[3] T. Kaneko, H. Kameoka, K. Tanaka, S. Seki, "iSTFTNet: Fast and lightweight mel-spectrogram vocoder incorporating inverse short-time Fourier transform," in *Proc. ICASSP*, pp. 6207-6211, 2022.

[4] T. Kaneko, H. Kameoka, K. Tanaka, S. Seki, "iSTFTNet2: Faster and more lightweight iSTFT-based neural vocoder using 1D-2D CNN," in *Proc. INTERSPEECH*, pp. 4369-4373, 2023.

Contact

Kou Tanaka, Computational Modeling Research Group, Media Information Laboratory