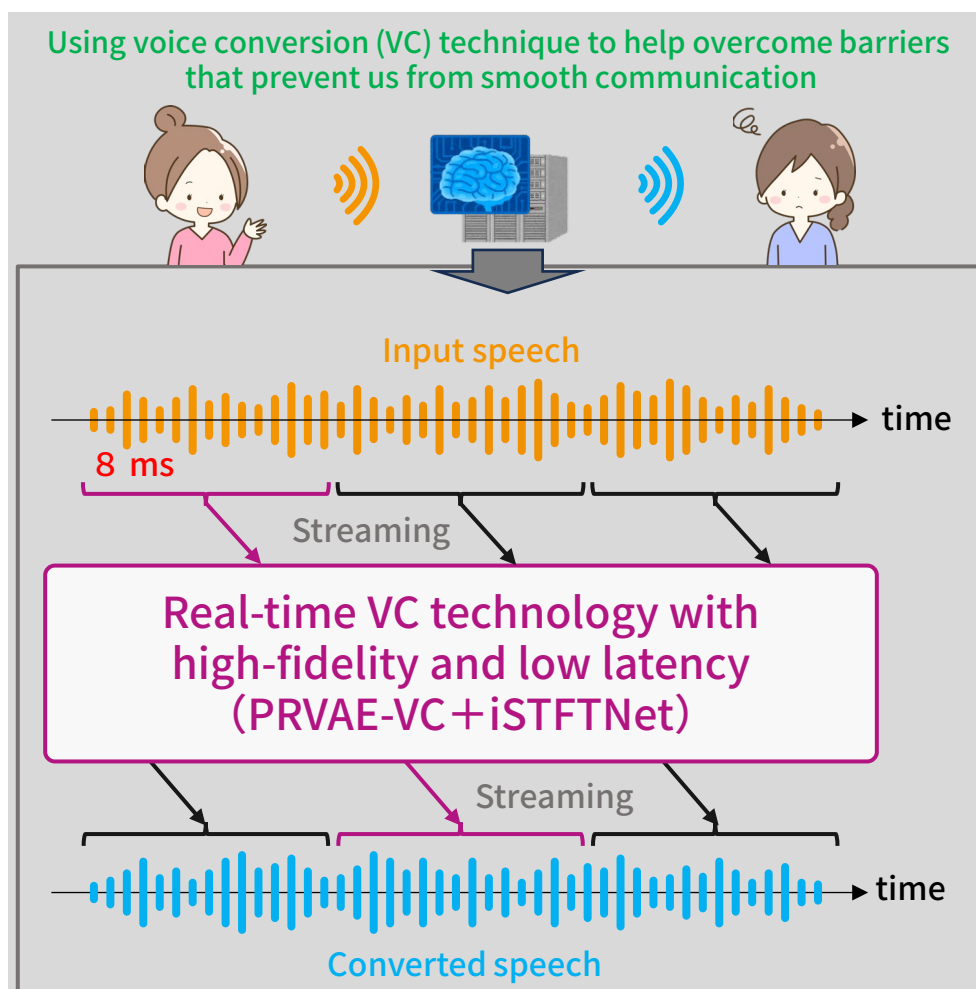# IOWN INTEGRAL
NTT R&D FORUM 2024

Real-time voice conversion changing your voice and speaking style
Rich speech communication where you can speak and hear in the voice you want to speak and hear

#Customer Experience Value Creation # Productivity Improvement
#Well-being, Human Capital Management

Using voice conversion (VC) technique to help overcome barriers that prevent us from smooth communication

Input speech

time

8 ms

Streaming

Real-time VC technology with high-fidelity and low latency
（PRVAE-VC＋iSTFTNet）

Streaming

time

Converted speech

///Technical Issue

The challenge was to obtain useful, speaker-independent, and generalized speech representation for speech conversion from only speech data.

///Research Goal

There are various forms of limitations in communication due to physical, ability, and psychological conditions. Our goal is to remove these restrictions and make comfortable communication.

---Technology

- Our deep generative model can obtain useful speech expressions for speech conversion by considering the differences between speaker identity.
- Improvement of the waveform synthesizer to achieve higher speed and lower power consumption.

---Novelty

- High sound quality by technology that takes speaker variation into account
- Low latency speech conversion with a logic delay of 35 ms

---Applicable Business

Making it possible to convert one speaker's voice into another speaker's voice for live streaming, or to convert the other party's voice into an easily audible voice for web conferencing or call centers [Customer Experience Value Creation]