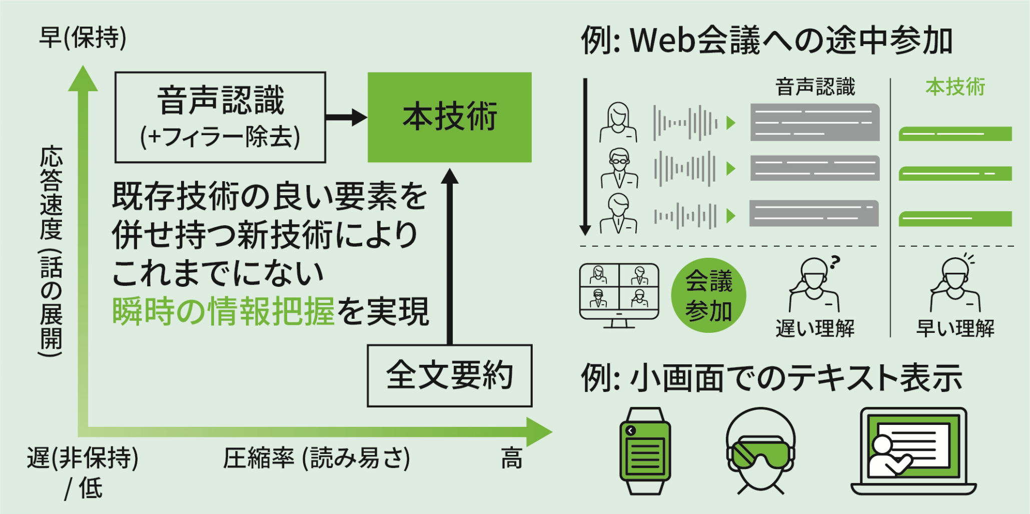
IOWN
INTEGRAL
NTT R&D FORUM 2024

DEVELOPMENT
δ 01-03

Real-time speech summarization for developing human-like intent understanding

Next-generation speech recognition technology for instant information grasp and decision-making support

#Productivity Improvement, #Customer Experience Value Creation



早(保持)

応答速度(話の展開)

音声認識
(+フィラー除去)

本技術

既存技術の良い要素を
併せ持つ新技術により
これまでにない
瞬時の情報把握を実現

全文要約

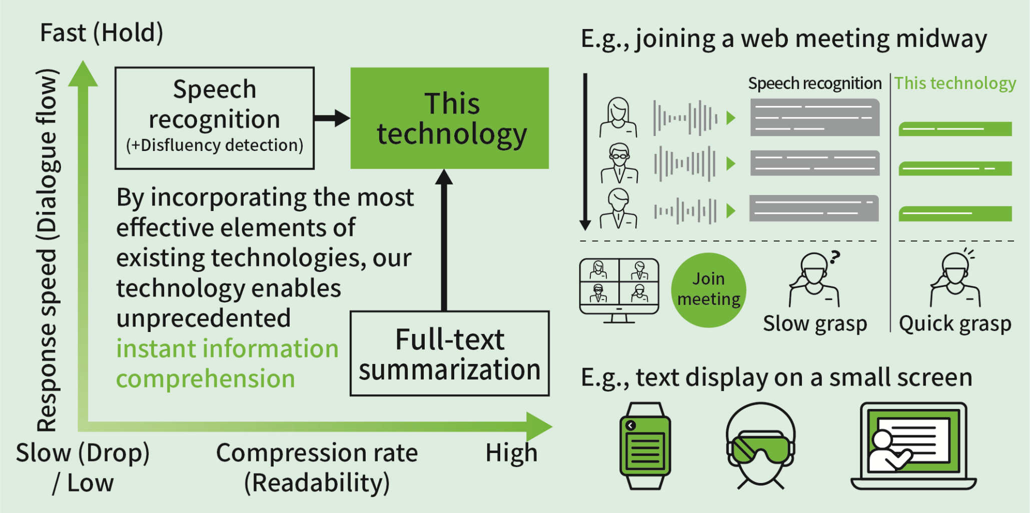
遅(非保持)
/ 低

圧縮率 (読み易さ)

高

例: Web会議への途中参加

例: 小画面でのテキスト表示



Fast (Hold)

Response speed (Dialogue flow)

Speech recognition
(+Disfluency detection)

This technology

By incorporating the most effective elements of existing technologies, our technology enables unprecedented instant information comprehension

Full-text summarization

Slow (Drop)
/ Low

Compression rate
(Readability)

High

E.g., joining a web meeting midway

E.g., text display on a small screen

///Technical Issue

When capturing the content of long meetings or similar audio in a short time, speech recognition provides real-time processing, but the output tends to be verbose. In contrast, full-text summarization excels in readability but struggles to maintain real-time processing.

///Research Goal

This technology allows users to grasp the content of long audio recordings, such as meetings or presentations, in real-time without having to listen to the entire recording. It enables improved work efficiency and better accessibility to information.

---Technology

We are developing a new technology that combines the strengths of speech recognition and full-text summarization to capture key points accurately, much like a human would, without the need to process all the audio. Using a compact model, it summarizes audio segments effectively, achieving real-time processing while ensuring efficient content understanding.

---Novelty

Existing speech recognition and full-text summarization technologies each have their own strengths and weaknesses in terms of real-time processing and content simplicity (readability). By combining the best features of these two technologies, our new approach offers a faster and more efficient way to deliver critical information, surpassing the limitations of previous methods in speech processing.

---Applicable Business

Information and communication technology sector:
Our technology can be applied to speech-based applications, such as real-time meeting minutes in web conferences and text display for people with hearing impairments on devices with small screens, such as smartwatches.
(expected to be established by Q1 2026)

Exhibitors=NIPPON TELEGRAPH AND TELEPHONE CORPORATION

Contact URL