



Title of Project : A Study on Multi-modal Automatic Simultaneous Interpretation System and Evaluation Method

NAKAMURA Satoshi

(Nara Institute of Science and Technology, Data Science Center, Professor)

Research Project Number : 21H05054

Researcher Number : 30263429

Term of Project : FY2021-2025

Budget Allocation : 145,600 Thousand Yen

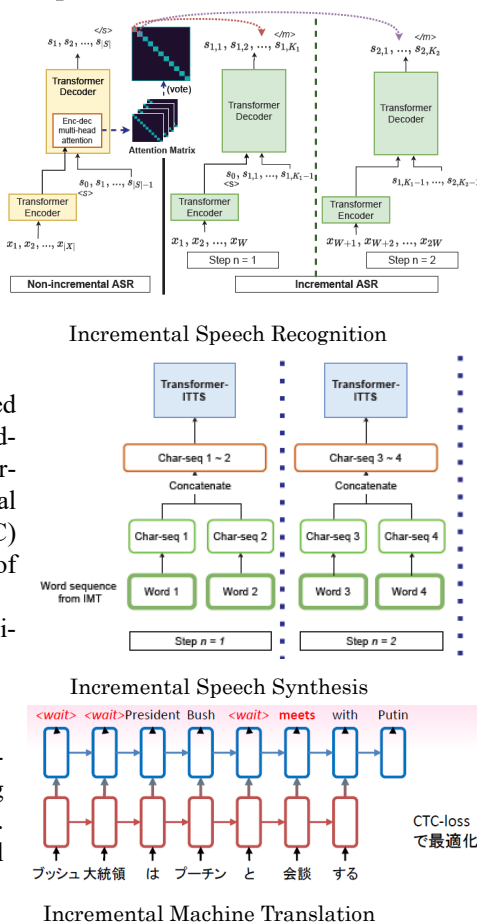
Keyword : Speech-to-speech translation

【Purpose and Background of the Research】

The conventional automatic speech translation, which translates a sentence after each utterance in a simple domain such as travel conversation, is advancing for practical use, but the human simultaneous interpretation is far difficult and its computational realization is understudied. We have collected around 400 hours of human interpreting corpus and developed incremental automatic speech recognition, synthesis, and machine translation as fundamental modules. However, further research and building an eco-system of collecting data and improving performance in real applications are necessary to approach human interpreter quality. This project focuses on fundamental interdisciplinary researches and building the eco-system with interpreters, computer scientists, and system users.

【Research Methods】

1. Multi-source, multi-modal interpretation. A) Multimodal translation of face, speech, emphasis, and emotion, B) Multi-source interpretation by a pre-trained model using video, prior information, external information, C) Improvement of incremental speech recognition/ synthesis. Incremental machine translation considering phrase structure. Latency control by speech synthesis and summarization of translation output.



2. Evaluation method of interpretation. A) Analysis and modeling of interpreting strategy, latency control, and summarization. Building interpreter supporting system. Implementation into an online interpretation service platform. B) Automatic evaluation of interpretation quality of human interpreter and the interpretation systems considering interpretation quality, latency, intention, content preservation, length of interpretation speech. C) Real-time analysis by brain-sensing for evaluation of mental load of translation

3. Corpus annotation and system development. A) Annotation of time alignment and interpretation quality to the simultaneous interpretation corpus. B) Collection of multimodal and multi-source interpretation corpus and annotation. C) Building eco-system of collecting data and improving performance in real online interpretation system.

【Expected Research Achievements and Scientific Significance】

- ① Simultaneous interpretation methods composed of incremental speech recognition, speech synthesis, and machine translation.
- ② Multi-source and multi-modal interpretation methods.
- ③ Automatic simultaneous interpretation system
- ④ Automatic evaluation system of interpretation quality of human interpreters and machines.
- ⑤ Automatic interpreter support system supporting human interpreters in real-time.
- ⑥ Annotations of time alignment and interpretation quality. The project will be enhanced by interdisciplinary and international collaboration.

【Publications Relevant to the Project】

- Automatic Translation of Spoken Language, Acoustic Science Series 18 CORONA Publishing 2018
- "Simultaneous Speech-to-speech Translation based on Neural Incremental ASR, MT, and TTS", Proc. Acoustical Society of Japan, 3-4-4 Spring 2020
- "NAIST English-to-Japanese Simultaneous Translation System for IWSLT 2021 Simultaneous Text-to-text Task", Proc. International Workshop of Spoken Language Translation 2021
- "Large-Scale English-Japanese Simultaneous Interpretation Corpus: Construction and Analyses with Sentence-Aligned Data", Proc. International Workshop of Spoken Language Translation 2021

【Homepage Address and Other Contact Information】

<https://ahcweb01.naist.jp/page-4794/kakenhi-ngst/>