

# Towards Streaming Speech Translation for Real-world Scenarios

Name Ryo Fukuda

Laboratory's name Augmented Human Communication Laboratory

Supervisor's name Professor Satoshi Nakamura

Abstract [\(should be within 1st page\)](#)

In a global society, speech translation (ST) systems supporting communication are helpful. In particular, streaming ST for long-form audio has a wide range of potential applications. This thesis focuses on the difficulties that arise when assuming situations similar to real-world scenarios to realize a practical streaming ST. Firstly, we tackled an error propagation problem in cascade ST systems. Knowledge distillation (KD), combined with fine-tuning, is introduced to train a Machine Translation (MT) model robust against Automatic Speech Recognition (ASR) errors. Experiments demonstrate that our approach is consistently more robust to ASR errors in various error conditions than baselines. Secondly, we proposed speech segmentation methods suitable for STs to overcome the difficulty of ST systems' inability to translate unsegmented long audio correctly. We contributed to improving the performance of end-to-end and cascade STs for unsegmented speech through the proposal of three segmentation methods. Finally, we addressed the construction of streaming simultaneous ST systems, which achieve low latency translation of unsegmented speech. Technical contributions in building the systems include developing and integrating online-adapted segmentation methods and simultaneous ST models with competitive performance based on recent research. Experiments under conditions using segmented and unsegmented speech demonstrate the effectiveness of our simultaneous ST models and online segmentation approaches. Overall, this thesis contributes to the advancement of streaming ST for real-world scenarios by improving ST models and introducing segmentation methods.