Optimizing Speech Translation for Low Latency and High Robustness

Name: Yuka Ko
Laboratory’ｓ name: Human-AI Interaction Laboratory
Supervisor’ｓ name: Sakriani Sakti

Abstract

Speech translation (ST), which automates the translation of spoken language, is a crucial technology that bridges linguistic barriers by converting spoken language in real-time. This dissertation addresses for achieving low latency and high robustness, especially in scenarios involving disfluencies, and spontaneous speech.

Firstly, we focus on a novel multi-task end-to-end ST incorporating automatic speech recognition (ASR) posterior distribution-based loss for improving robustness against ASR ambiguities.

Experiments demonstrate the improvements over baseline methods, showing robustness of the disfluent inputs. Secondly, we propose a effective multi-stage fine-tuning methods for training disfluent-to-fluent speech translation models by integrating augmented disfluency-tagged data.

Experimental results show that the approach effectively identifies and removes disfluencies, leading to more fluent and accurate translations in spontaneous speech. Thirdly, we propose a fine-tuning approach using both offline and simultaneous speech translation data to tackle the problem of small amount of simultaneous interpretation (SI) data. This method achieves a balance between latency and translation quality, providing practical solutions for real-time applications. Lastly, we construct a robust simultaneous ST system, which takes disfluency into account. Experiments show the accuracy improvement in disfluency-aware simultaneous ST.

These contributions provide practical solutions to address the key challenges techniques of robust simultaneous ST in real-world applications. The methods proposed in this thesis represent a significant step forward in delivering accurate and high-quality real-time speech translation ensuring both robustness and low latency.