論 文 内 容 の 要 旨

博士論文題目

Towards Streaming Speech Translation for Real-world Scenarios
（実世界シナリオに向けたストリーミング音声翻訳）

氏　名　　福田 りょう

（論文内容の要旨）

Speech translation (ST) is a technology that automatically translates speech into other languages. With the increasing demand for communication support in a global society, developing advanced ST systems has become more crucial than ever. In particular, streaming ST for long-form audio has a wide range of potential applications. However, streaming ST is still an emerging field of research, and there are many challenges to its practical application. 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. A Machine Translation (MT) model trained only with human transcripts performs poorly on Automatic Speech Recognition (ASR) results. The MT model should be trained considering the presence of ASR errors during inference. We introduced knowledge distillation (KD), combined with fine-tuning, to train an MT model robust against ASR errors. Experiments demonstrate that our approach is consistently more robust to ASR errors in various error conditions than baselines. Secondly, we proposed three speech segmentation methods suitable for STs to overcome the difficulty of ST systems' inability to translate unsegmented long audio correctly. We proposed a speech segmentation method using a segmentation model trained using a segmented bilingual speech corpus. In addition, we used a pre-trained speech model to enhance the performance of the segmentation model. We also proposed an approach to perform translation while leaving multiple segmentation candidates and to decide on segmentation and translation simultaneously. We contributed to improving the performance of end-to-end and cascade STs for unsegmented speech through the

proposal of the segmentation methods. Finally, we addressed the development of streaming simultaneous ST systems, which achieve low latency translation of unsegmented speech. 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.

氏　名	福田 りょう

（論文審査結果の要旨）

This thesis work aims to overcome difficulties in speech translation (ST) for streaming speech inputs that arise in practical, real-world scenarios. It mainly focuses on two problems in ST: error propagation in a cascaded framework with automatic speech recognition (ASR) and machine translation (MT), and non-obvious segment boundaries in spontaneous speech. For the first problem, this work proposed a method to train an MT model working robustly against ASR errors through knowledge distillation and fine-tuning and demonstrated better results than baselines. For the second problem, this work proposed a series of three ST-oriented speech segmentation methods that aim to optimize speech segmentation to improve ST performance. These methods use speech segment boundaries in speech translation corpora as the references to train their segmentation models. In experiments, these methods improved translation accuracy and efficiency by finding relatively short segments compared to baseline methods. In addition, this work addressed the development of a streaming simultaneous ST system by integrating proposed techniques with a simultaneous ST framework.

This thesis work contributed to the progress of practical ST technologies by tackling typical problems in real-world scenarios using novel techniques and investigating the advantages of the proposed methods in detail. This research resulted in two peer-reviewed journal articles and three peer-reviewed international conference papers. As a result, this thesis sufficiently qualified as a Doctoral thesis of Engineering.