XMU′s Simultaneous Translation System at NAACL 2021

Shuangtao Li, Jinming Hu, Boli Wang,
Xiaodong Shi, Yidong Chen

Xiamen University, China

Outline

- Tasks
- Data and preprocessing
- Text-to-text system
- Speech-to-text system
- Experiments
- Conclusion and future work

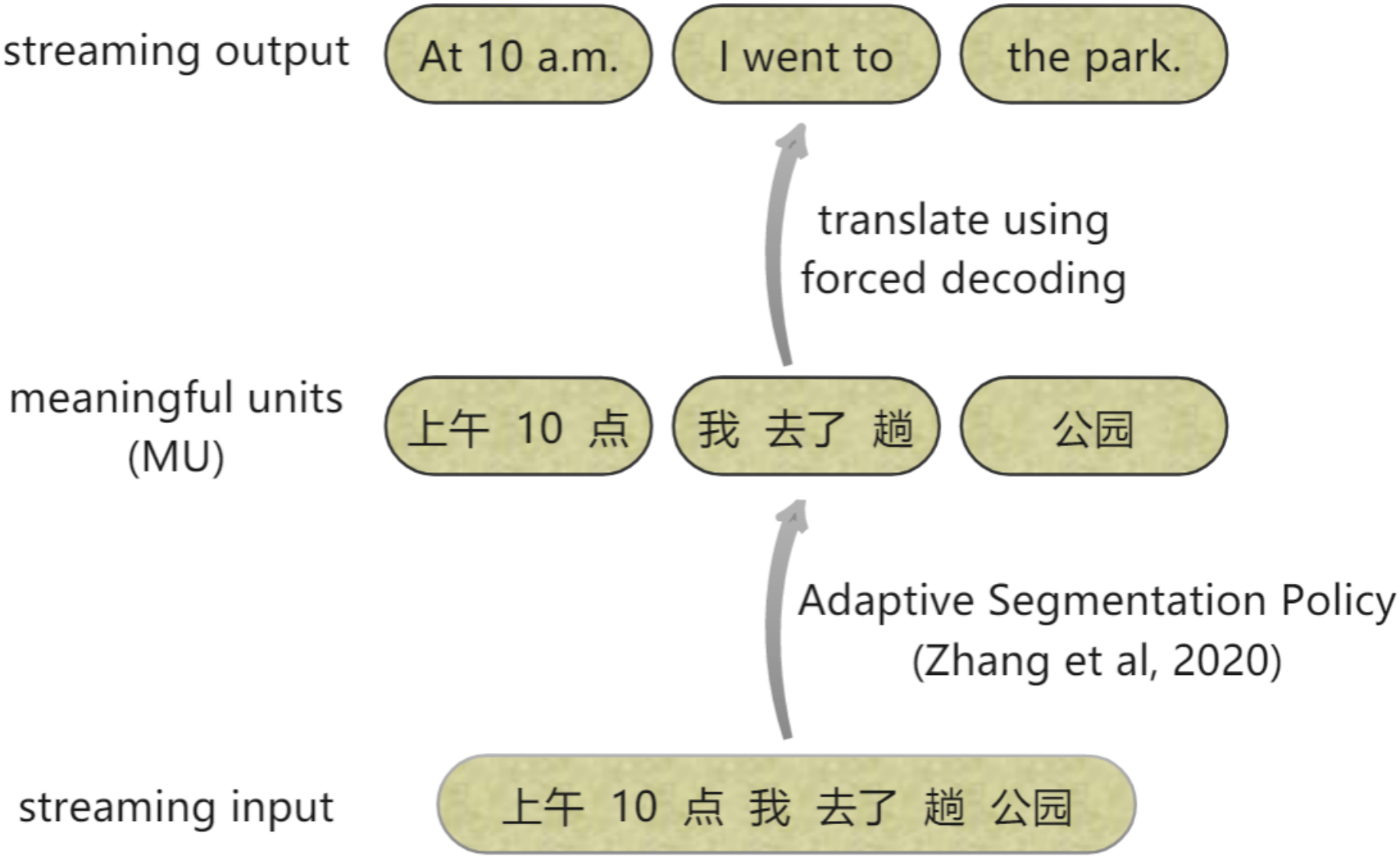
Tasks

- We participated in two tasks:
  - Zh->En Translation, input: streaming transcription. (text-to-text)
  - Zh->En Translation, input: audio file. (speech-to-text)

Data and preprocessing

- Datasets
  - Our MT model is pretrained on CWMT19 (9.1M parallel sentence pairs).
  - Our MT model is fine-tuned on Baidu Speech Translation Corpus (39K parallel sentence pairs).
- Preprocessing
  - Filter out long sentence pairs.
  - Convert full-width characters into half-width characters.
  - Segment Chinese text and tokenize English text.
  - Apply byte-pair-encoding to all sentences.

Text-to-text system



How the system translates streaming text.

Text-to-text system

- The MU segmentor is a text classifier based on BERT.
- Once the probability of class 1 is larger than a threshold δ, the input text is segmented.

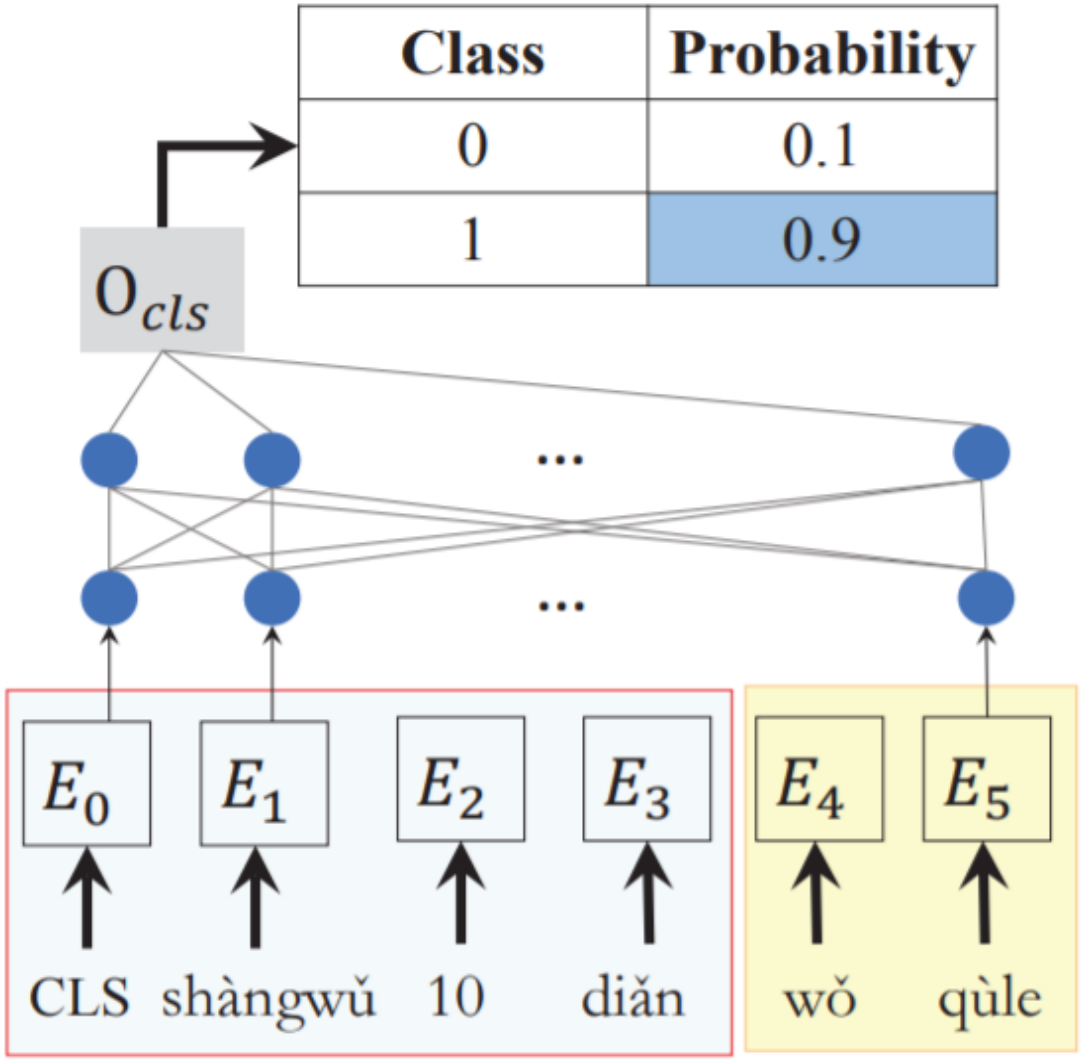
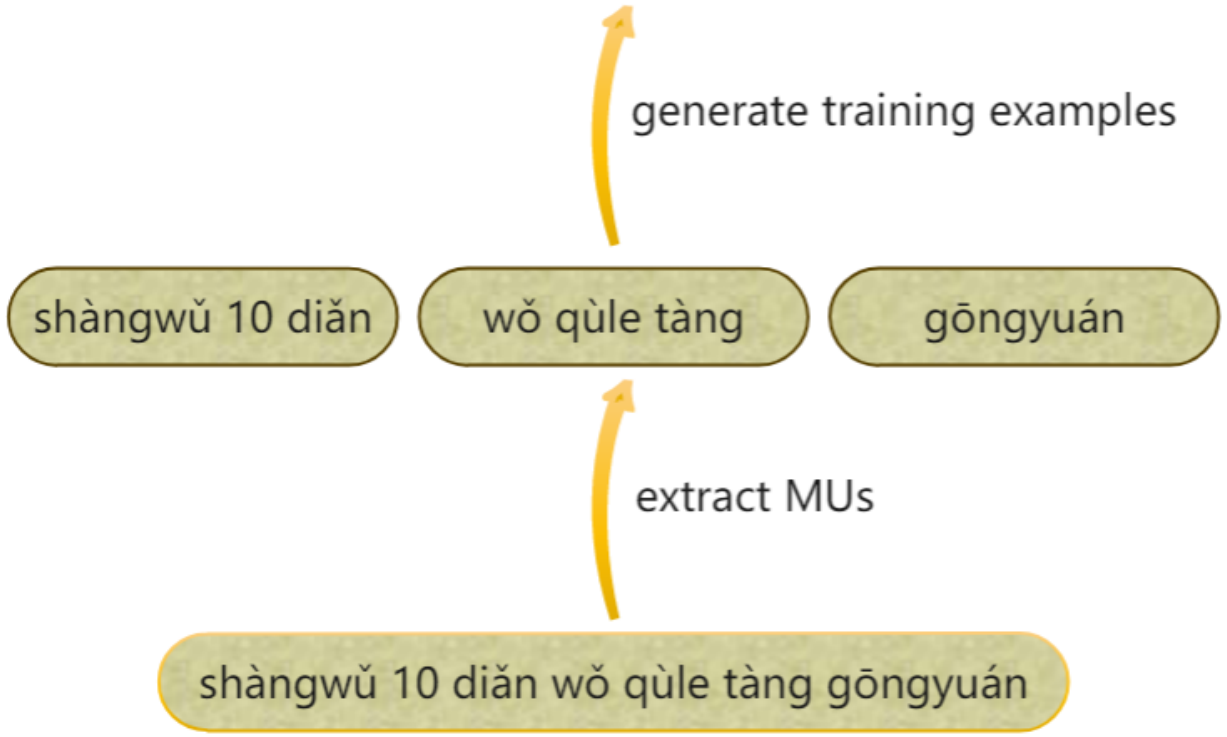


Illustration of the MU segmentor.

Text-to-text system

- To train the MU segmentor, we extract MUs and generate training data.

t	c_t	$f_t(m=2)$	l_t
1	shàngwǔ	10 diǎn	0
2	shàngwǔ 10	diǎn wǒ	0
3	shàngwǔ 10 diǎn	wǒ qùle	1
4	shàngwǔ 10 diǎn wǒ	qùle táng	0
5	shàngwǔ 10 diǎn wǒ qùle	táng gōngyuán	0
6	shàngwǔ 10 diǎn wǒ qùle táng	gōngyuán	1
...			



Generating examples for training the MU segmentor.

Text-to-text system

- Algorithm 1 extracts the MUs in an input sentence.
- We only use the *basic method* proposed by Zhang el al, 2020 for extracting MUs.

Algorithm 1: Extract MUs

Input: $\mathbf{x} = x_1, ..., x_T$ ▷ streaming input
Output: $\mathbf{S_{MU}}$ ▷ list of MU segmentation

$k = 0$ ▷ position of last MU boundary
$\widetilde{\mathbf{y}} = M_{nmt}(src = \mathbf{x}, tgt_{force} = None)$ ▷ full sentence decoding
while *Reading* x_t **do**
  $\mathbf{y}^t = M_{nmt}(src = \mathbf{x}_{\leq t}, tgt_{force} = \mathbf{y}^k)$
  if y^t *is a prefix of* $\widetilde{\mathbf{y}}$ **then**
    $\mathbf{S_{MU}} = \mathbf{S_{MU}} \cup \{x_{k+1}, ..., x_t\}$
    $k = t$
return $\mathbf{S_{MU}}$

The algorithm for extracting MUs.

Speech-to-text system

- The speech-to-text system is a pipeline of three components:
  - Baidu’s real time ASR service.
  - A repunctuation model, which is a BERT-based sequence labeling model.
  - A text-to-text translation subsystem.

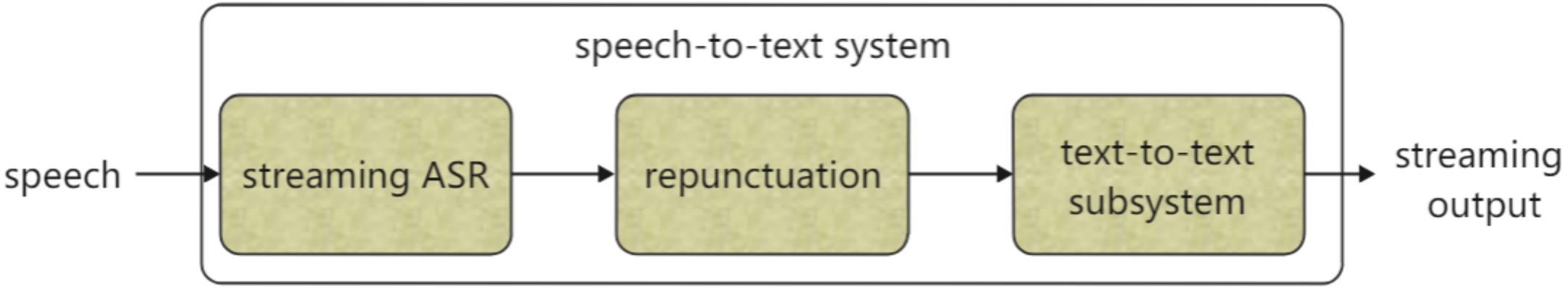
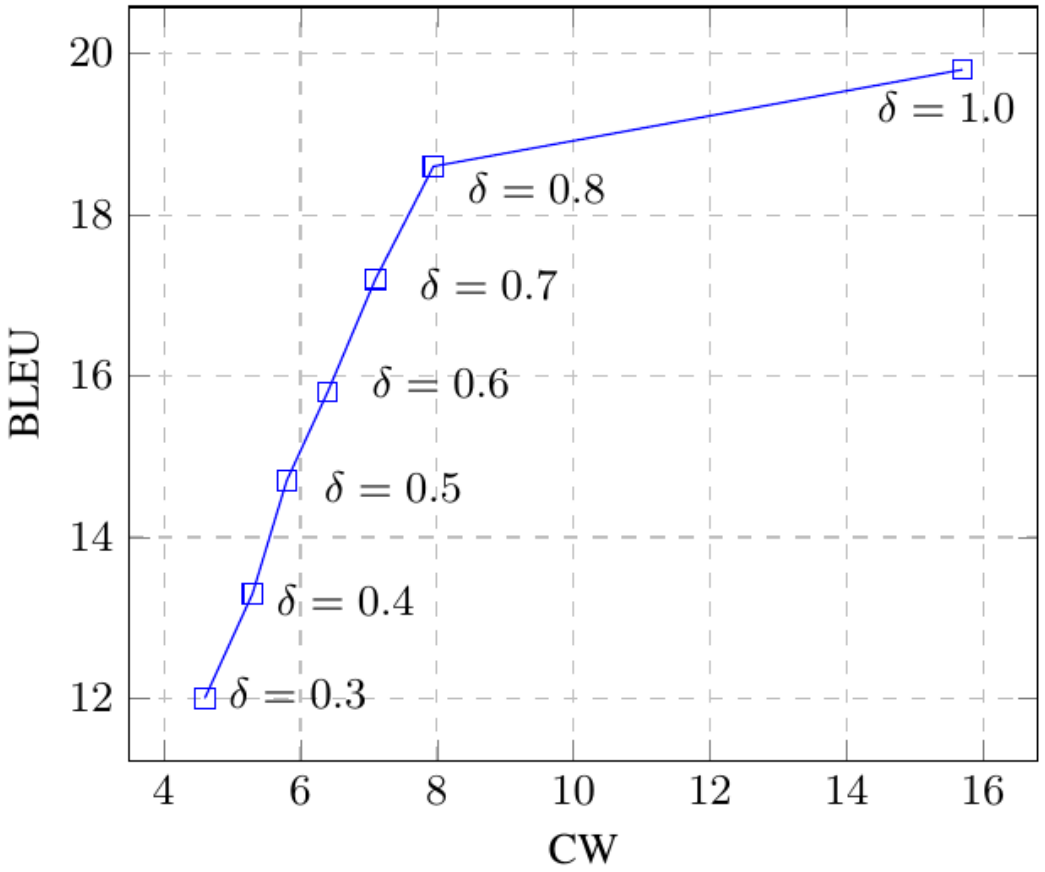


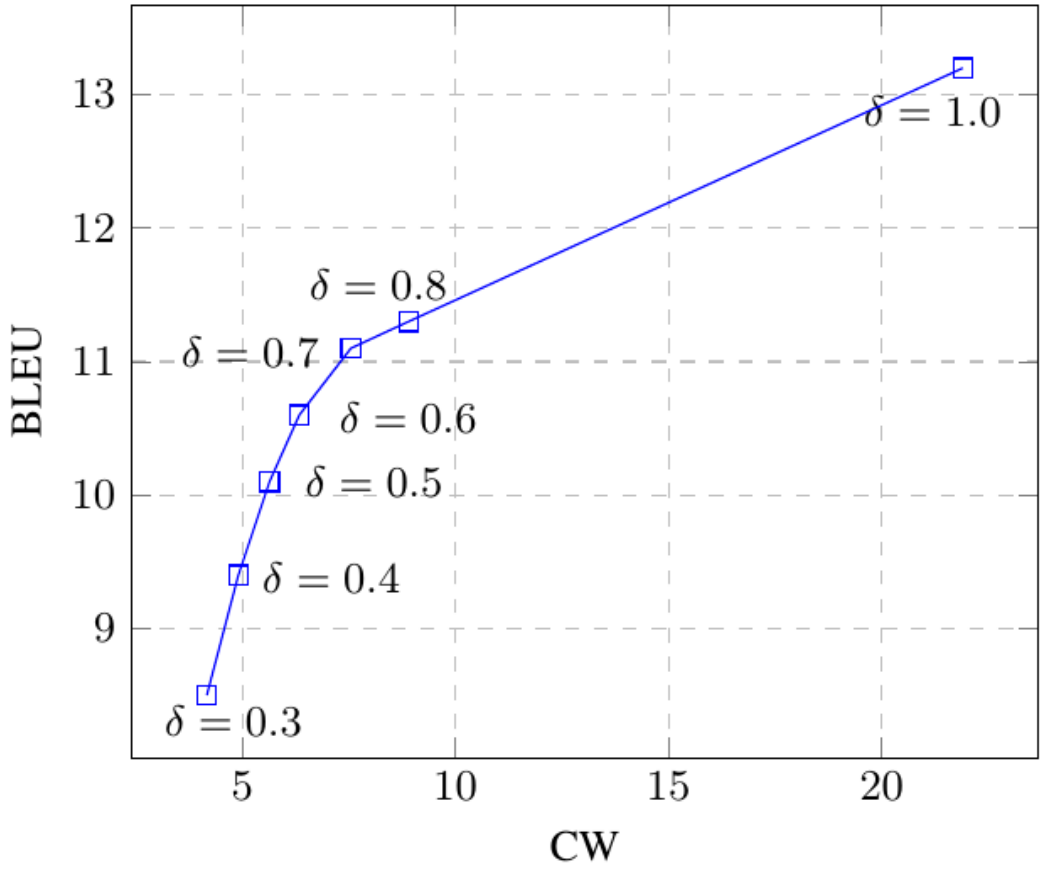
Illustration of speech-to-text system.

Experiments

- The two systems are evaluated on the development set of the Baidu Speech Translation Corpus.



The text-to-text system.



The speech-to-text system.

Conclusion and future work

- Adaptive Segmentation Policy (Zhang et al., 2021) is effective.

- Our systems use a conventional MT model that is not designed for simultaneous translation. We will study how to train the MT model that is more suitable for simultaneous translation.