System Description on Automatic Simultaneous Translation Workshop

Linjie Chen, Jianzong Wang, Zhangcheng Huang, Xiongbin Ding, Jing Xiao
Ping An Technology (Shenzhen) Co., Ltd.

Tracks and Results

1. We participate in all the two directions of Chinese-to-English translation, Chinese audio to English text and Chinese text to English text.

2. We do data filtering and model training techniques to get the best BLEU score and reduce the average lagging.

3. We propose a two-stage simultaneous translation pipeline system which is composed of Quartznet and BPE-based transformer.

4. We propose a competitive simultaneous translation system and achieves a BLEU score of 24.39 in the audio input track.

Datasets

1.Audio Data

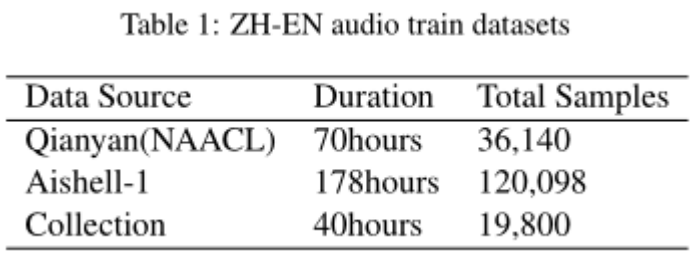


Table 1: ZH-EN audio train datasets

Data Source	Duration	Total Samples
Qianyan(NAACL)	70hours	36,140
Aishell-1	178hours	120,098
Collection	40hours	19,800

We use a collection of 19800 sentences as our training data (audio and transcription). Audio data were split into sentences by the sentence-level. We also use a pre-trained ASR model the similarity matching algorithm to filter audio and original transcription data of lower similarity.

2.Text Data

Dataset: CWMT 19 corpus

1) Data filtering

Human rules: Punctuation ratio, sentence length ratio, cross-entropy threshold and terminology substitution

2) Back translation

Monolingual data were translated by a pretrained English to Chinese model

3) Word segmentation

Chinese sentence: LAC toolkit

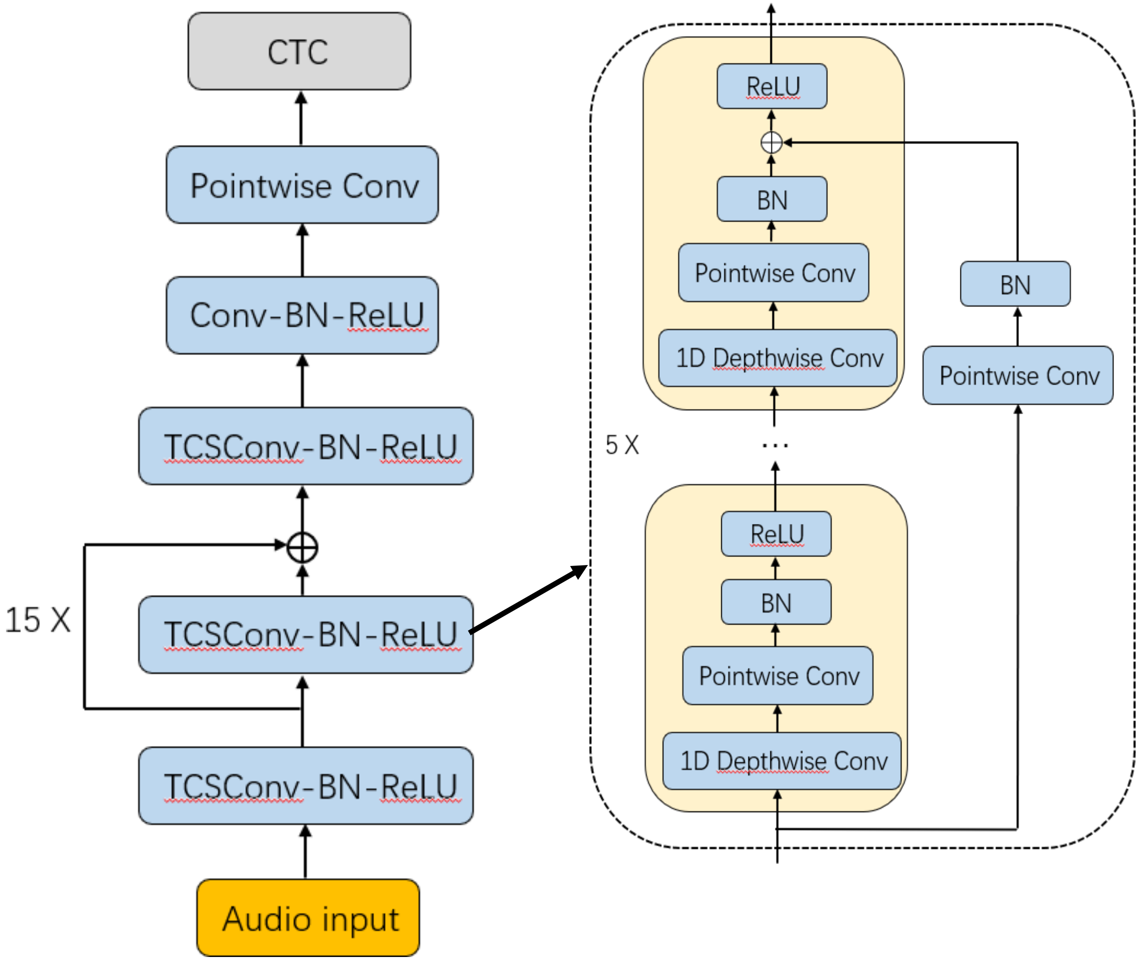
English sentence: Tokenizer and Truecaser (Moses scripts)

4) Bytes pair encoding

Trained a BPE model and applied it for both Chinese and English sentences

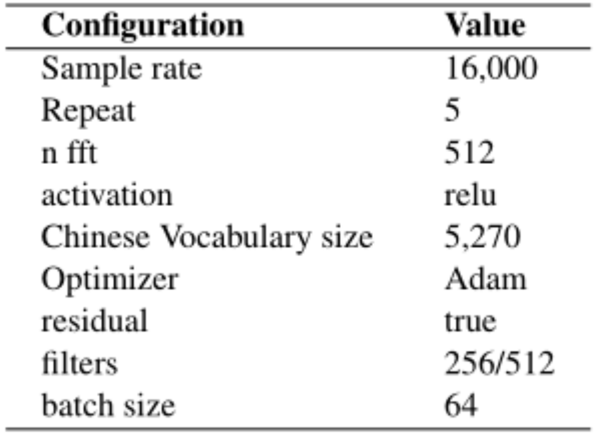
Automatic Speech Recognition

1.Model



2.Setups

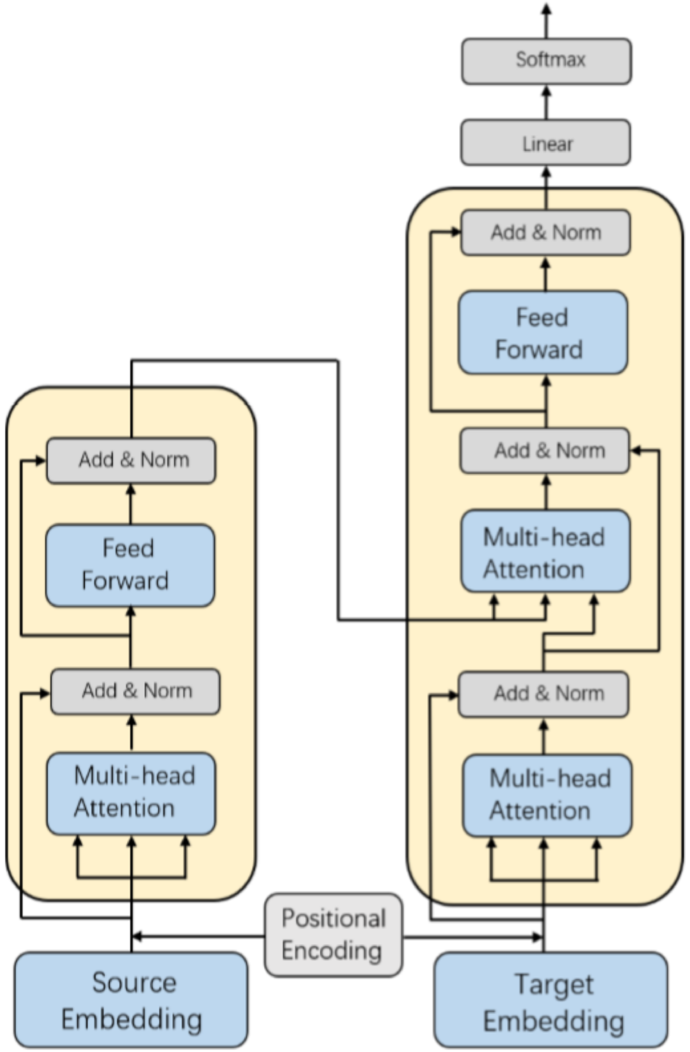
Table 2: Model Configuration



Configuration	**Value**
Sample rate	16,000
Repeat	5
n fft	512
activation	relu
Chinese Vocabulary size	5,270
Optimizer	Adam
residual	true
filters	256/512
batch size	64

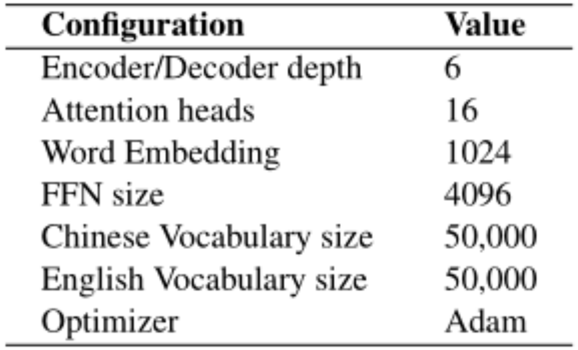
Machine Translation

1.Model



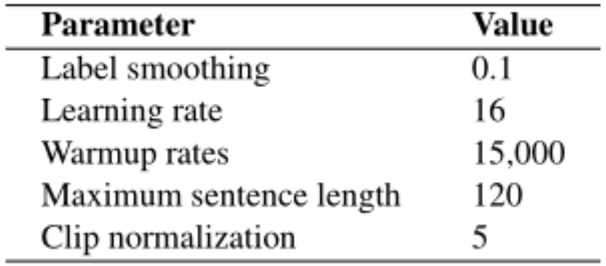
2.Setups

Table 3: Model Configuration



Configuration	Value
Encoder/Decoder depth	6
Attention heads	16
Word Embedding	1024
FFN size	4096
Chinese Vocabulary size	50,000
English Vocabulary size	50,000
Optimizer	Adam

Table 4: Training Parameters



Parameter	Value
Label smoothing	0.1
Learning rate	16
Warmup rates	15,000
Maximum sentence length	120
Clip normalization	5

3.Fine tuning

We implement fine-tuning on our model using the development set of qianyan audio datasets （956 sentence pairs) to improve the translation quality on auotomatic simultaneous translation task.

Conclusion

1. This paper describes a pipeline automatic simultaneous translation system and details the process of data filtering and model training.

2. The consecutive wait of the best point reached to 18.4 while this simultaneous translation system achieves a BLEU score of 24.39 in the audio input track.

3. We will continue to research on developing an end-to-end speech translation model from Chinese speech input to English text output.

THANK YOU!