



Findings of the Second Workshop on Automatic Simultaneous Translation

Ruiqing Zhang, Chuanqiang Zhang, Zhongjun He, Hua Wu, Haifeng Wang

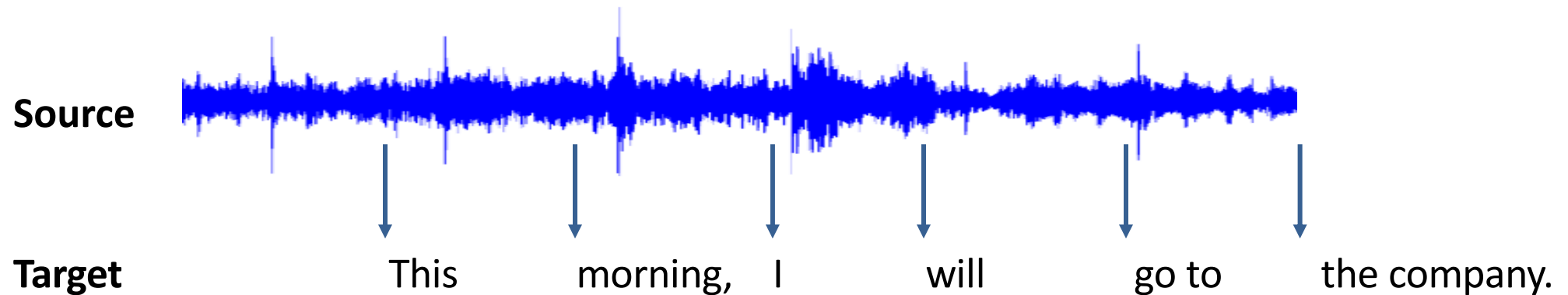
Baidu Inc.

Shared Task of Automatic Simultaneous Translation (AST)

1. Text-to-text Track

Source	今天	上午	我	要	去趟	公司。	
Target		This	morning,	I	will	go to	the company.

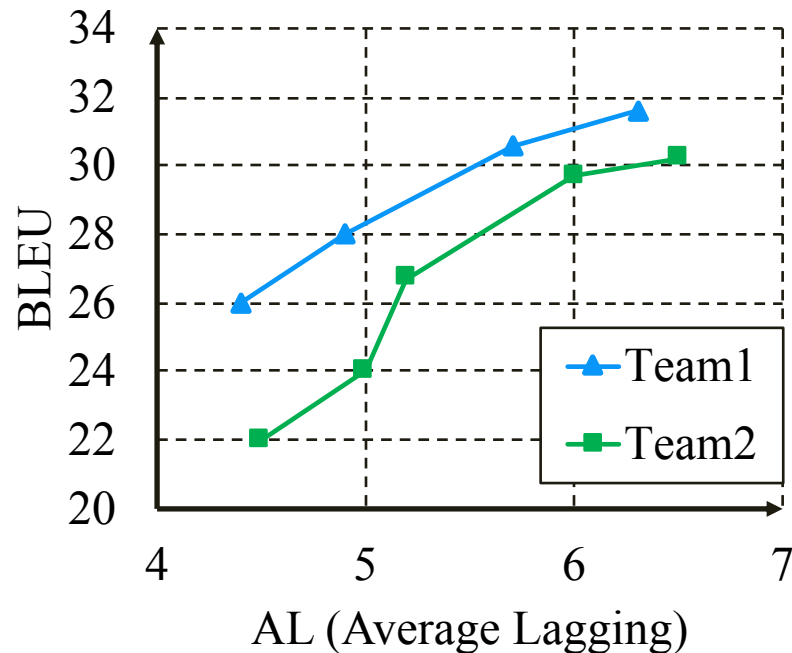
2. Speech-to-text Track



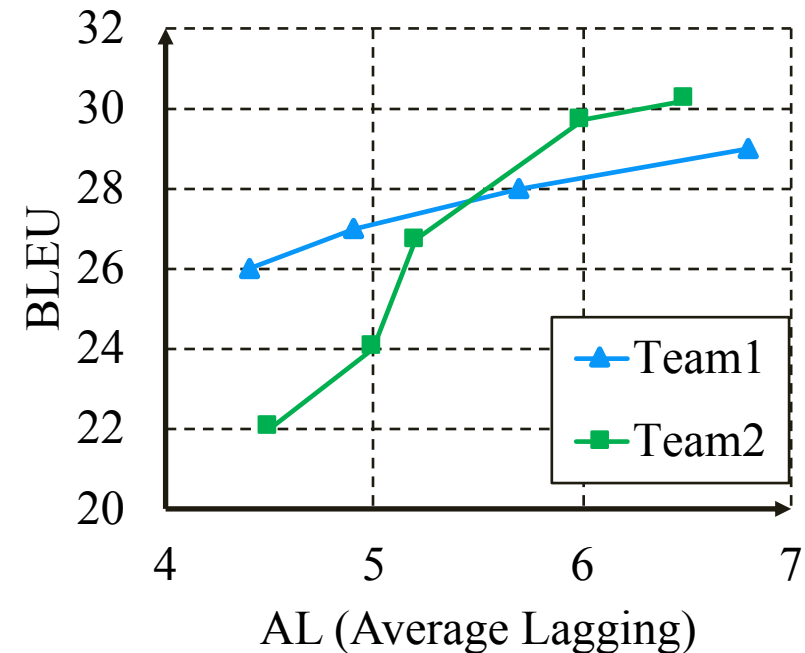
Evaluating AST Systems

Translation Quality: BLEU
Latency: Consecutive Wait (CW)
Average Lagging (AL)
...

Two examples of the results submitted by two teams



(a)



(b)

Evaluating AST Systems

—— Monotonic Optimal Sequence

Optimal Point:

One result is considered optimal if

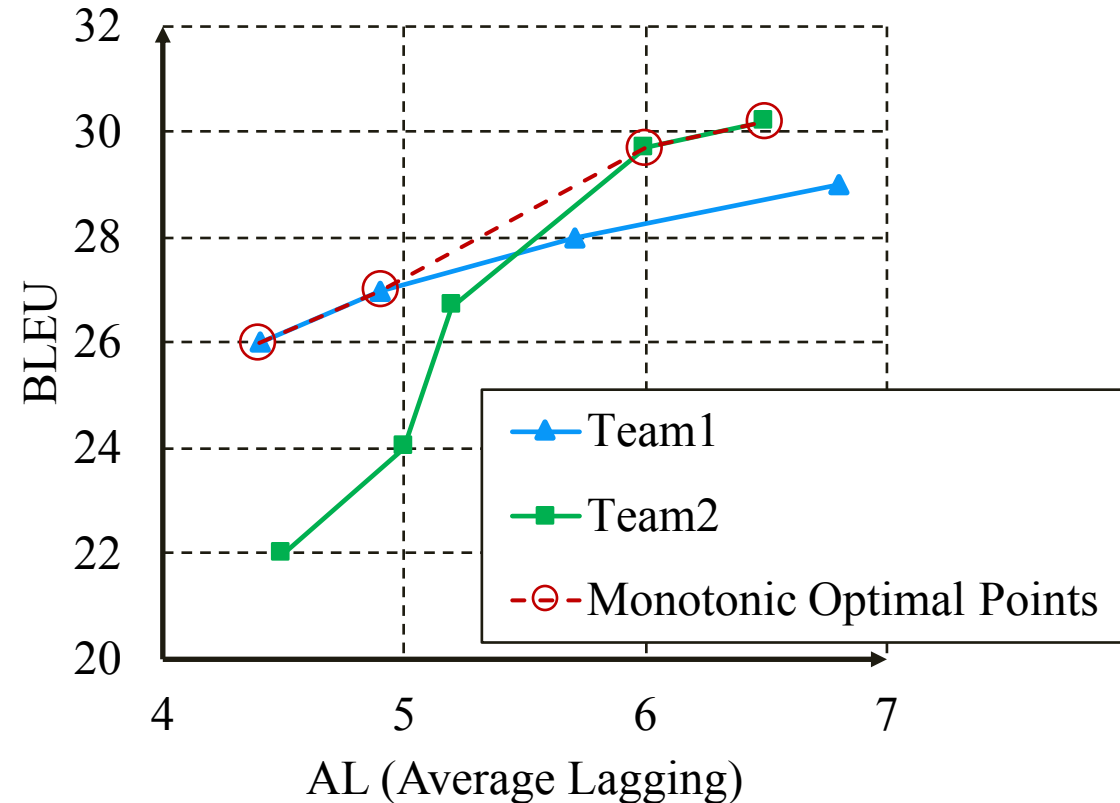
there is no other point or line

above it at an identical latency. In

this case, the result is of the highest

translation quality at that latency and

we define it as an Optimal Point.



Evaluating AST Systems

—— Optimal Points

Optimal Point:

One result is considered optimal if

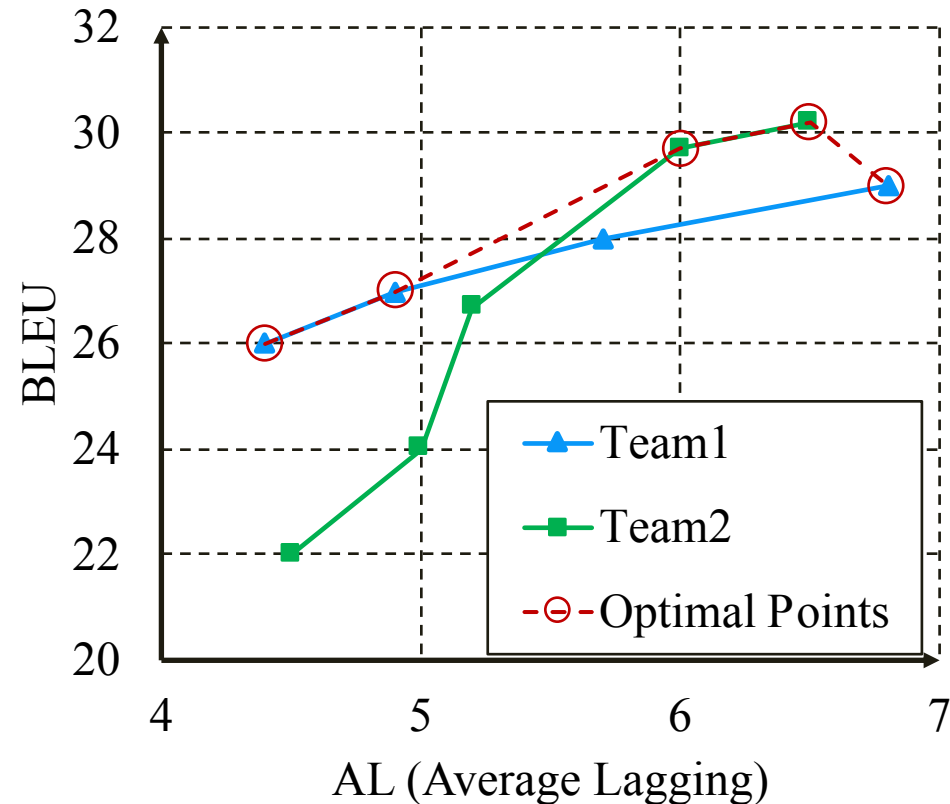
there is no other point or line

above it at an identical latency. In

this case, the result is of the highest

translation quality at that latency and

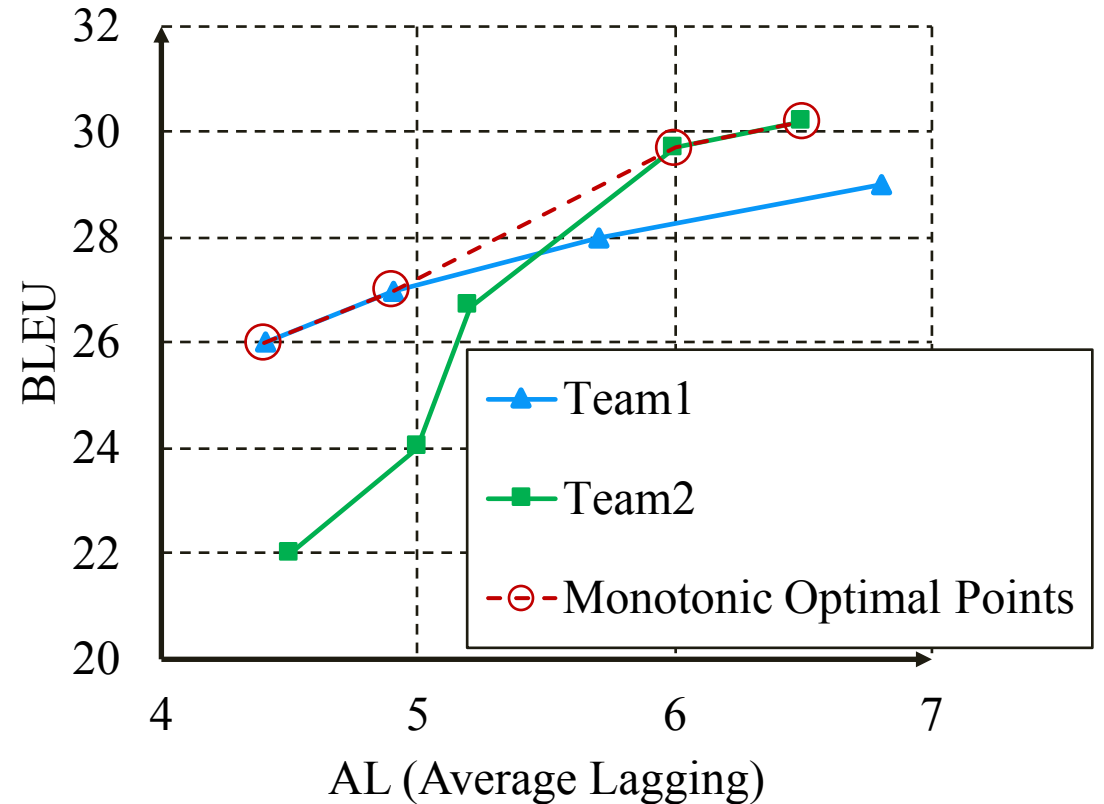
we define it as an Optimal Point.



Evaluating AST Systems

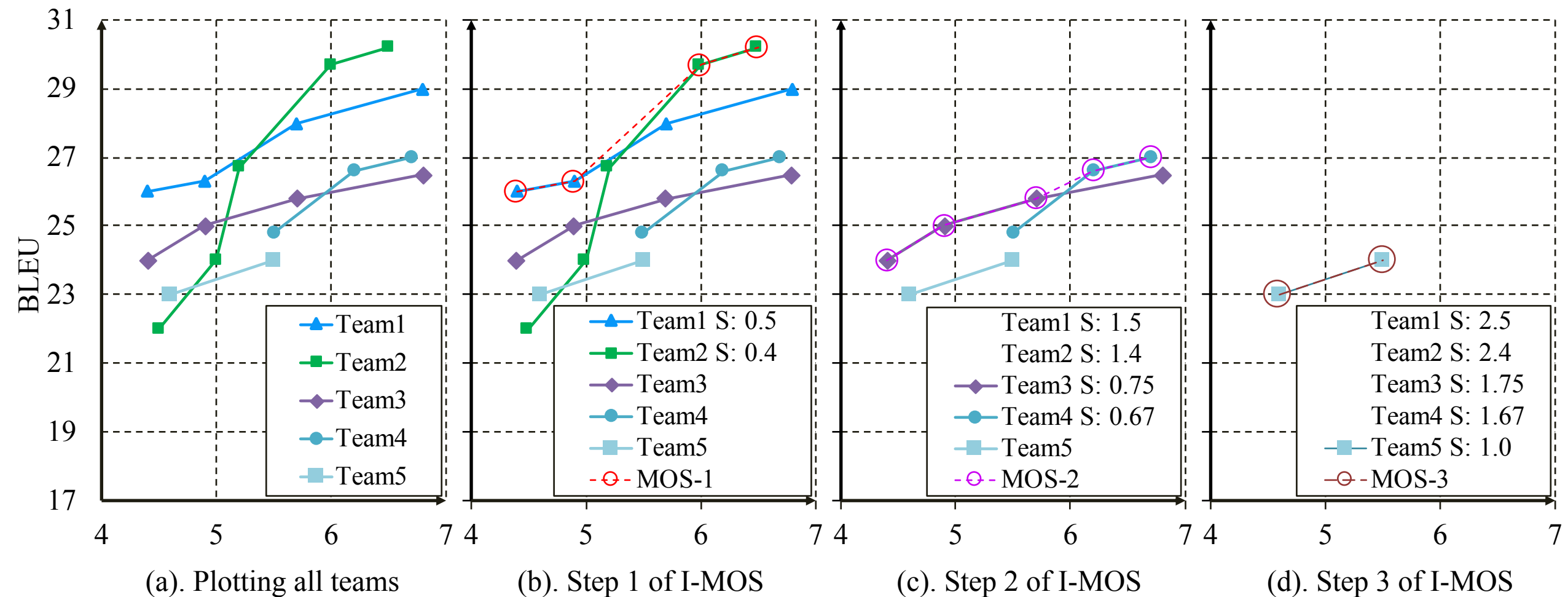
—— Monotonic Optimal Sequence

$$S_{T_i} = N(p_{t_i}^*) / N(p_{t_i})$$



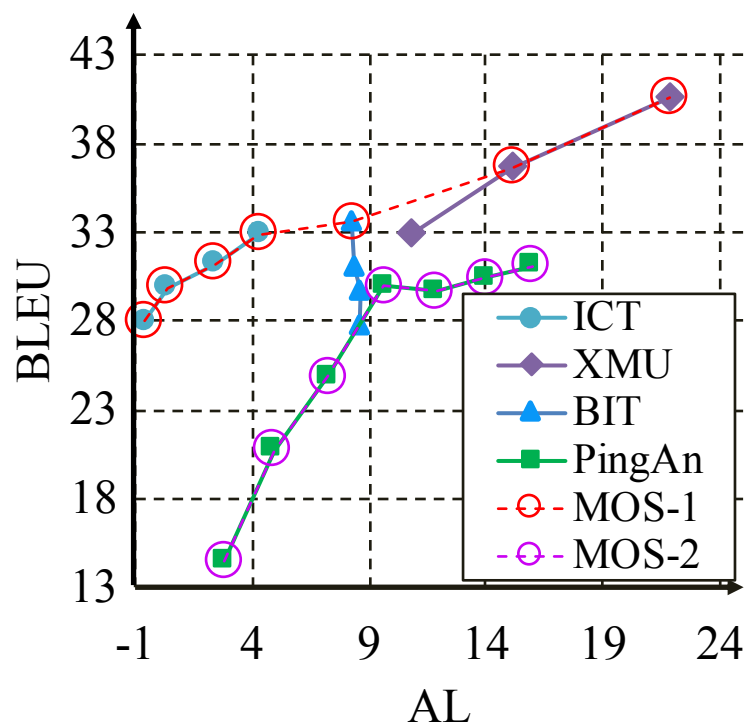
Evaluating AST Systems

— Iterative Monotonic Optimal Sequence

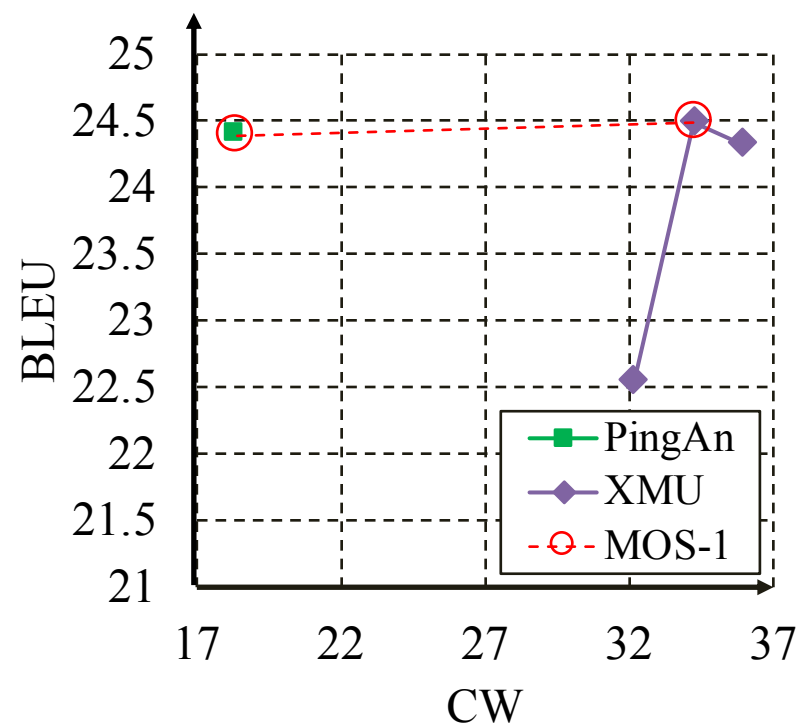


Results of the shared task

— 6 systems evaluated on the two tracks



(a). Track1. Text-to-Text AST



(b). Track2. Speech-to-Text AST

Discussion

——important challenges for AST

- **Data Scarcity**
 - BSTC is still insufficient to the data-hungry E2E speech translation models.
 - **Evaluation Dilemma**
 - It remains open to question whether it is reasonable to compare two systems with no intersection in latency.
 - **Applications**
 - Robustness and controllability
-