

CJNLP 2023



# NLP/SLP research activity in NAIST AHC Lab.

Katsuhito Sudoh

Associate Professor

Nara Institute of Science and Technology (NAIST), Japan

Part of this work is supported by JSPS KAKENHI (Grant numbers JP21H05054, 21H03500, 17H06101)

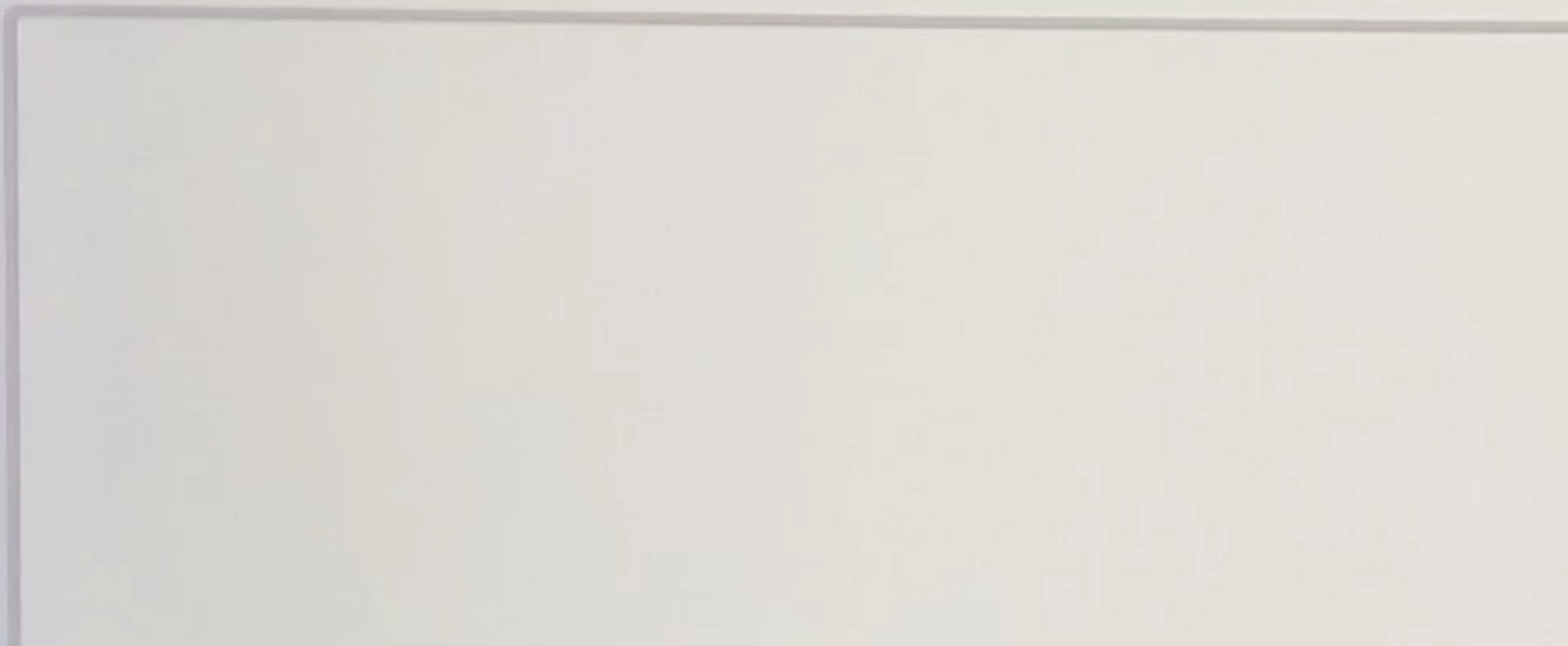
# Simultaneous speech translation and its evaluation

Y. Kano+, “Simultaneous Neural Machine Translation with Prefix Alignment,” IWSLT 2022

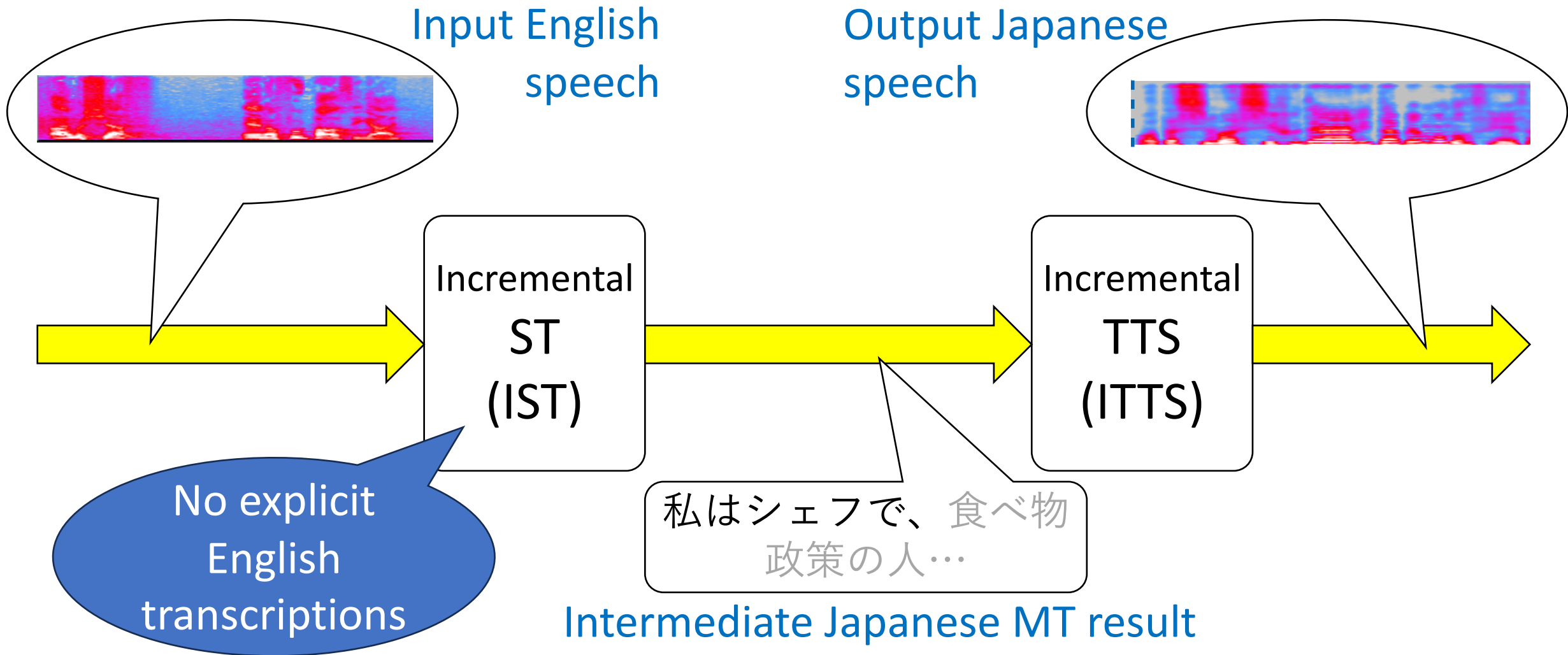
R. Fukuda+, “NAIST Simultaneous Speech-to-speech Translation System for IWSLT 2023,” IWSLT 2023

Y. Kano+, “Average Token Delay: A Latency Metric for Simultaneous Translation,” Interspeech 2023

# SI Result



# Our SimulS2S System [Fukuda+ 2023 IWSLT]



# Prefix Alignment [Kano+ 2022 IWSLT]

- Induce *prefix-to-prefix* translation pairs from parallel corpora

Sentence pair { I bought a pen.  
僕はペンを買った。

Full-sentence

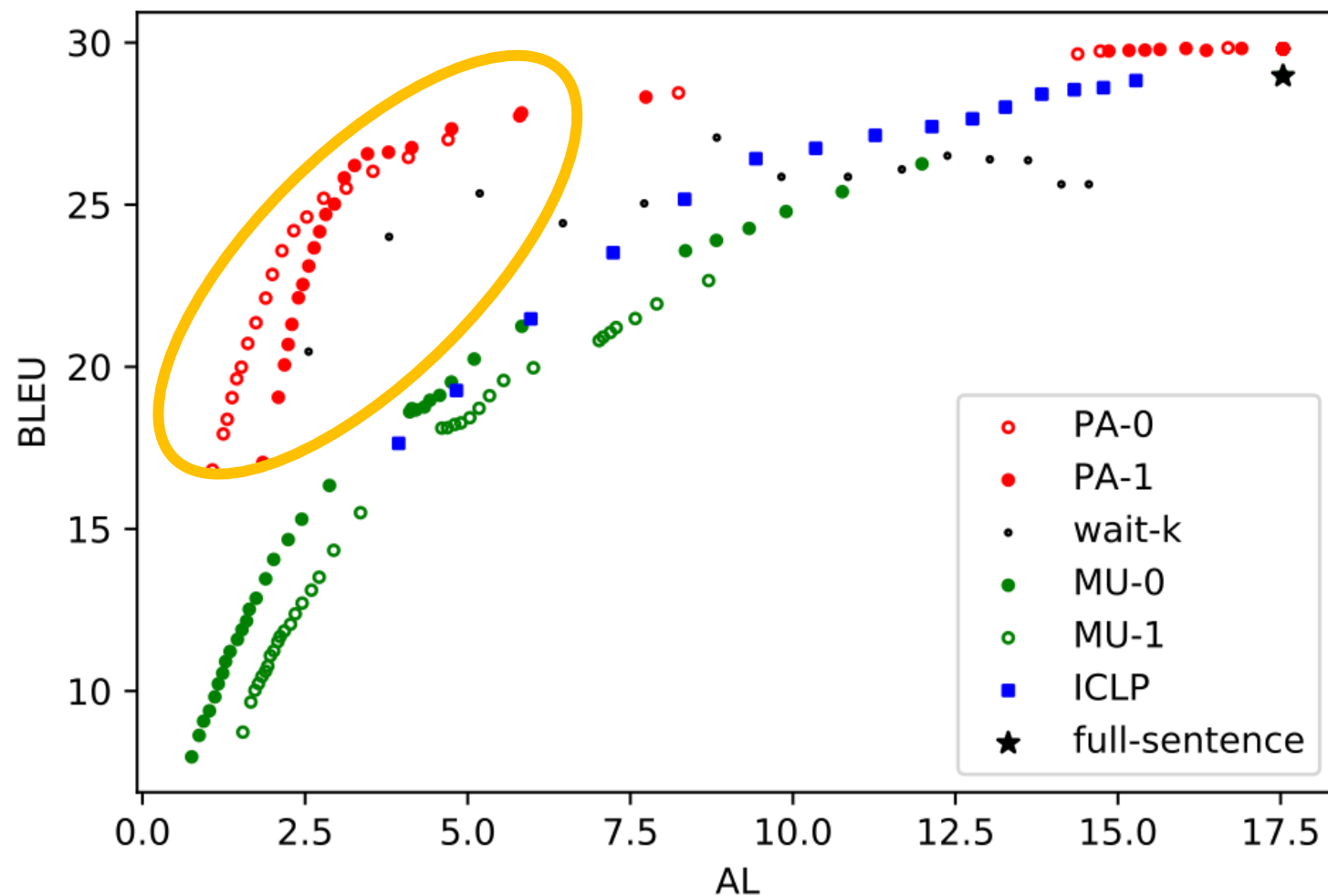
MT (pretrained)

Source prefix	Prefix translation	Full-sent. MT (from n-best)	Matched prefix
---------------	--------------------	-----------------------------	----------------

An induced target prefix is used as a *fixed (forced-decoded)* target prefix

# Results

- SimulMT model trained using the prefix pairs outperformed other methods (En-De)
- The advantage is smaller in En-Ja; PA failed to induce enough short prefix pairs that helps SimulMT



# IWSLT Evaluation Campaign – SimulST track

- History
  - 2020: text/speech-to-text, En-De
  - 2021-2022: text/speech-to-text, En-De/**Ja**/**Zh**
  - 2023: speech-to-text/**speech**, En-De/Ja/Zh
  - 2024: TBA
- Regulations
  - Use publicly-available speech/language resources
  - Configure SimulST systems to satisfy given latency limits
  - Submit systems in forms of Docker images

# Automatic Evaluation in IWSLT 2023

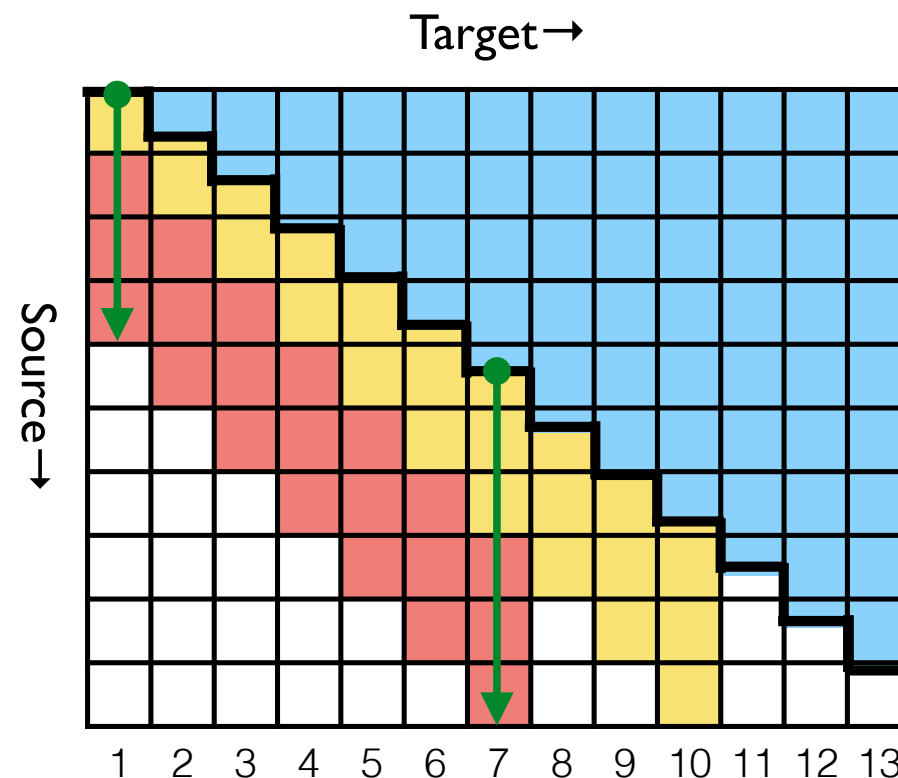
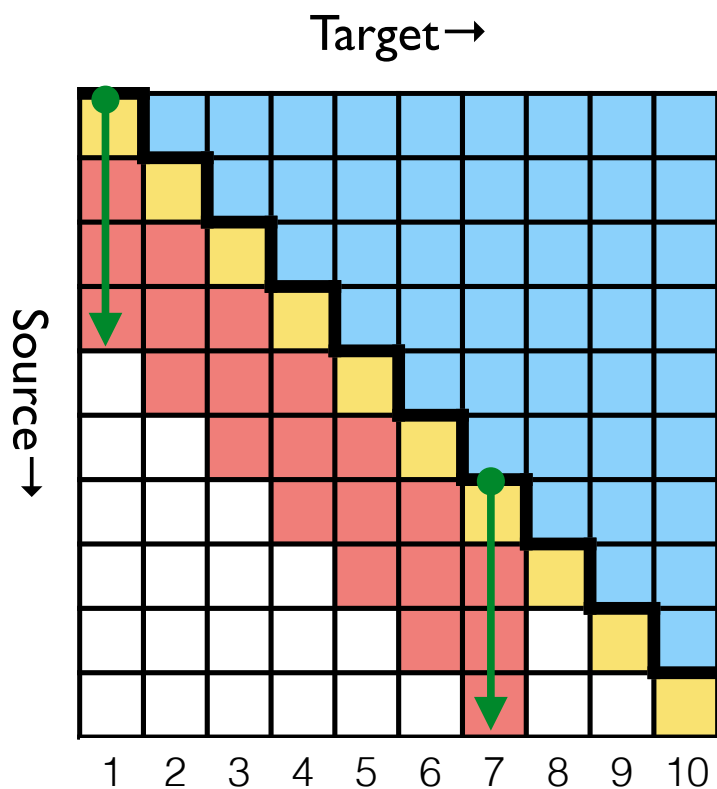
	Quality	Latency
Speech-to-text	BLEU	<b>Average Lagging (AL) and variants (LAAL, DAL)</b> Average Propotion (AP) <b>Average Token Delay (ATD)</b>
Speech-to-speech	ASR-BLEU BLASER	Start/End Offset <b>Average Token Delay (ATD)</b>

*\*Please find more details in the overview paper (archived in ACL Anthology)*



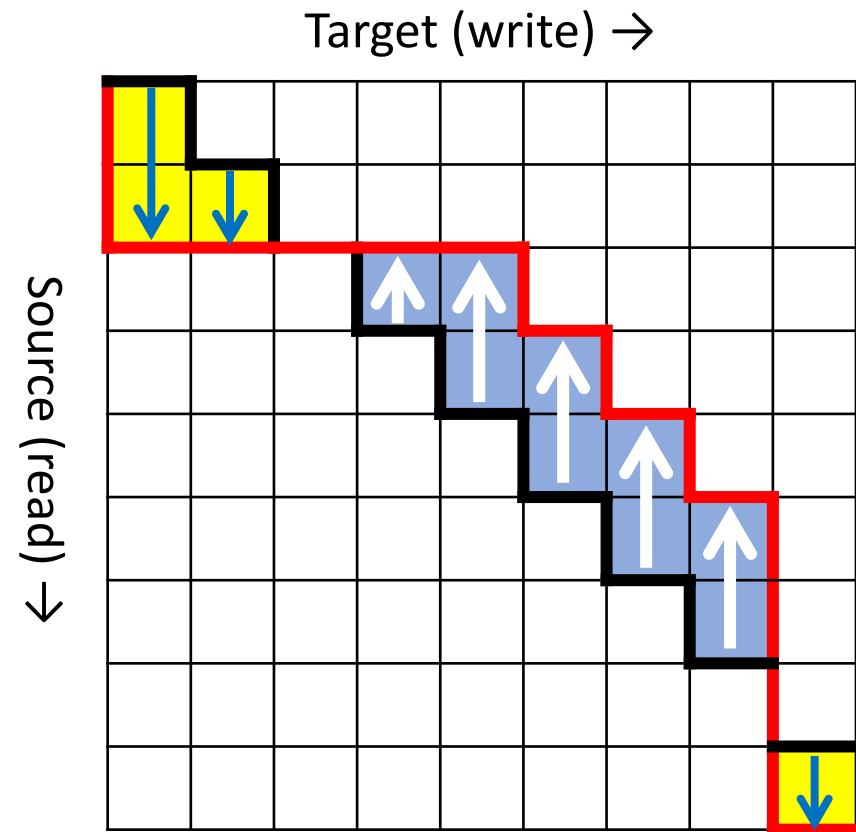
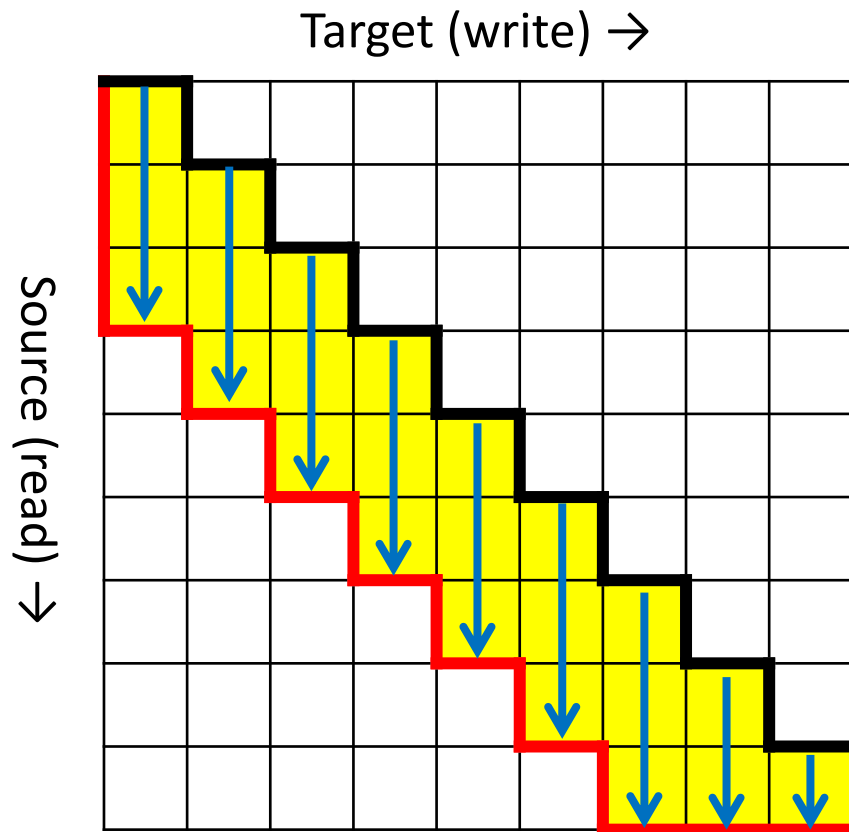
# Average Lagging (AL)

- Average delay from the *ideal* policy through the diagonal line in a read-write chart



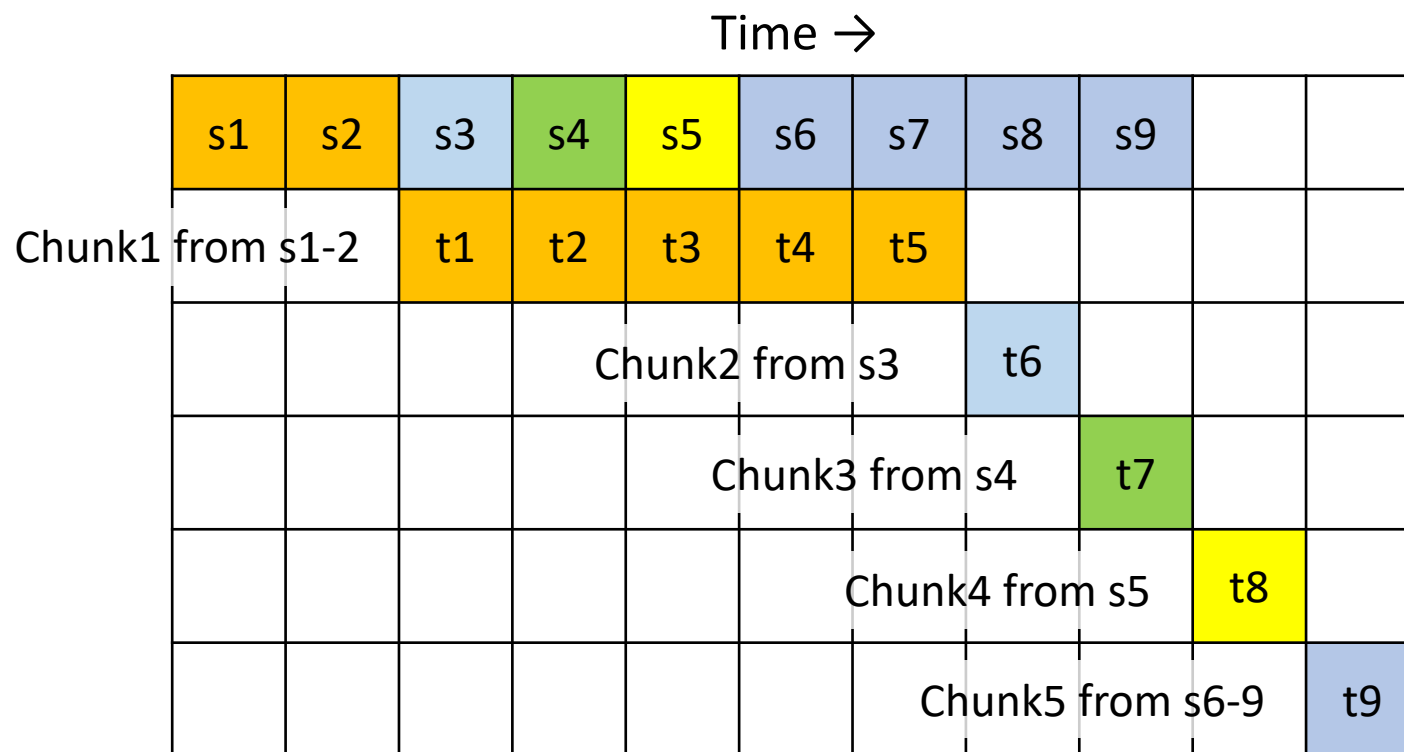
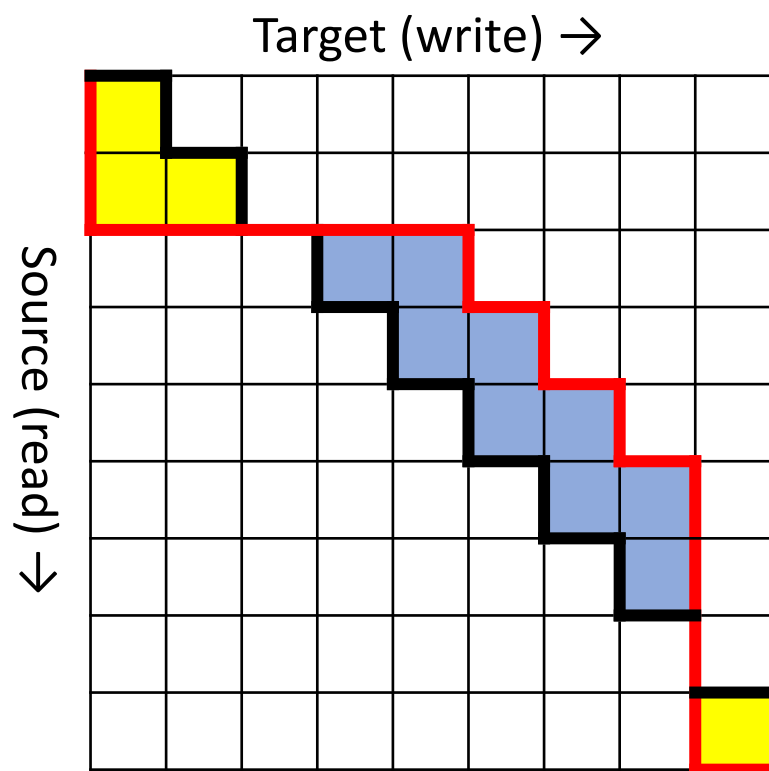
# Unintuitive Latency Measurement by AL

- Latency can be negative for outputs with long chunks



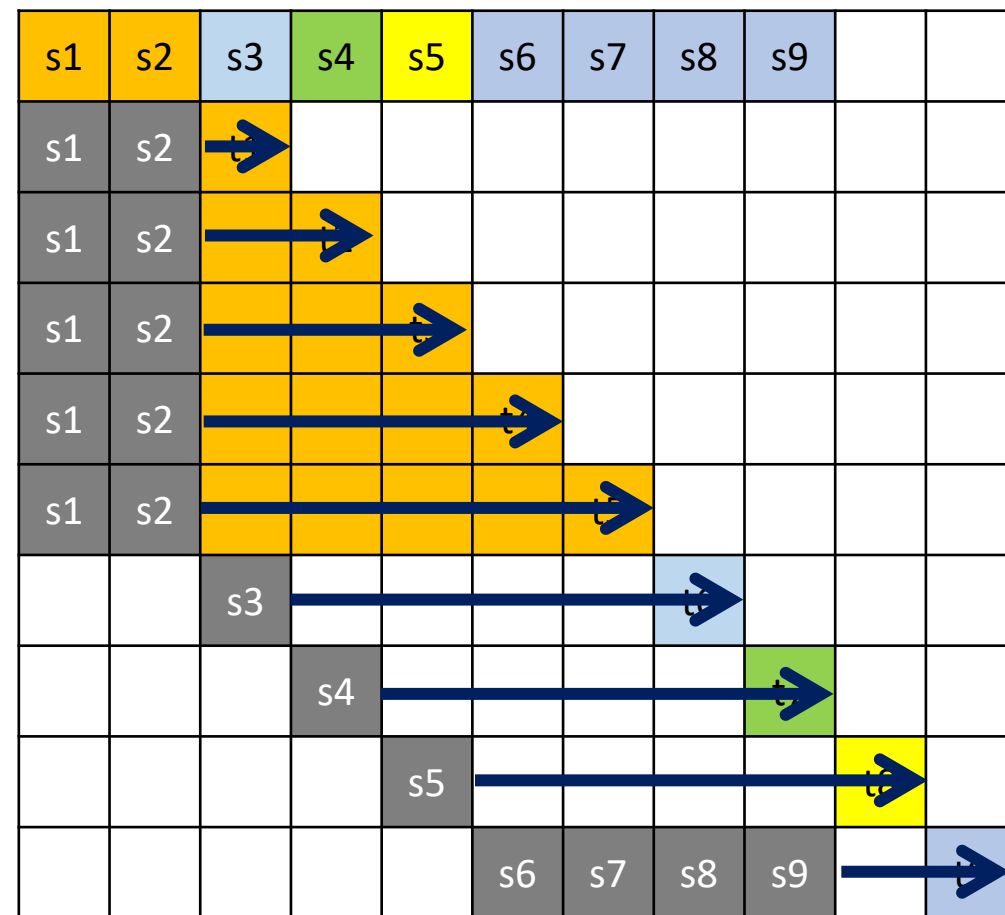
# Average Token Delay (ATD) [Kano+ 2023 Interspeech]

- Inspired by *Ear-Voice Span (EVS)* in interpretation studies



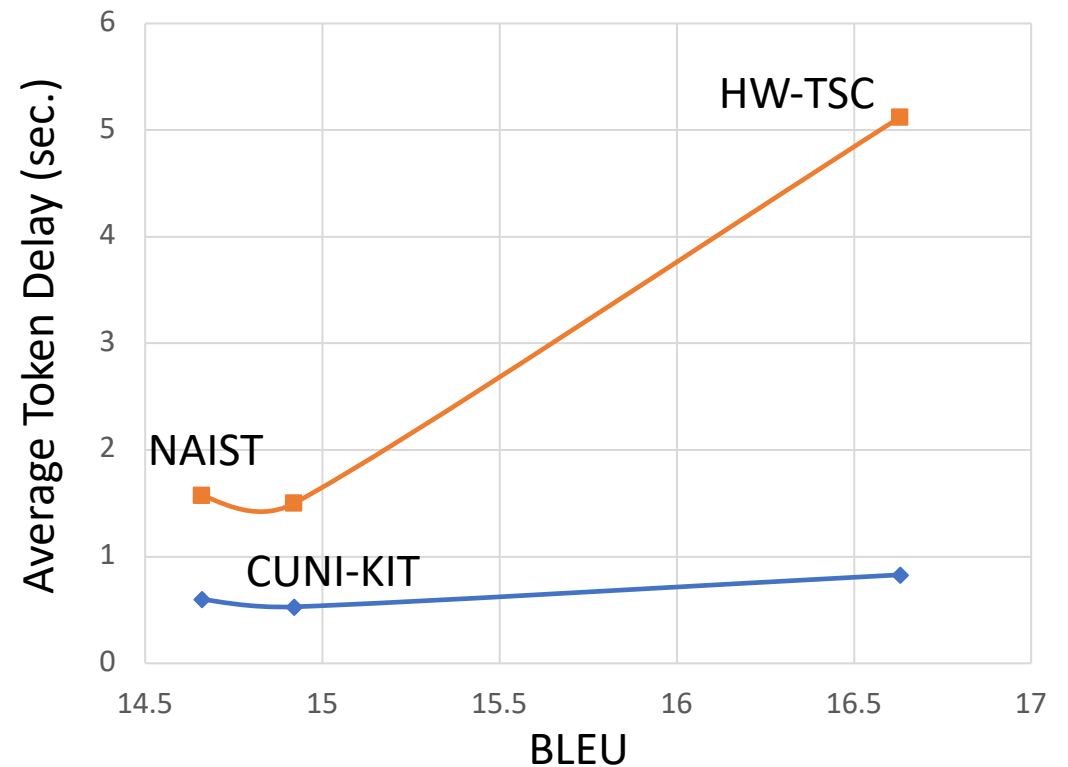
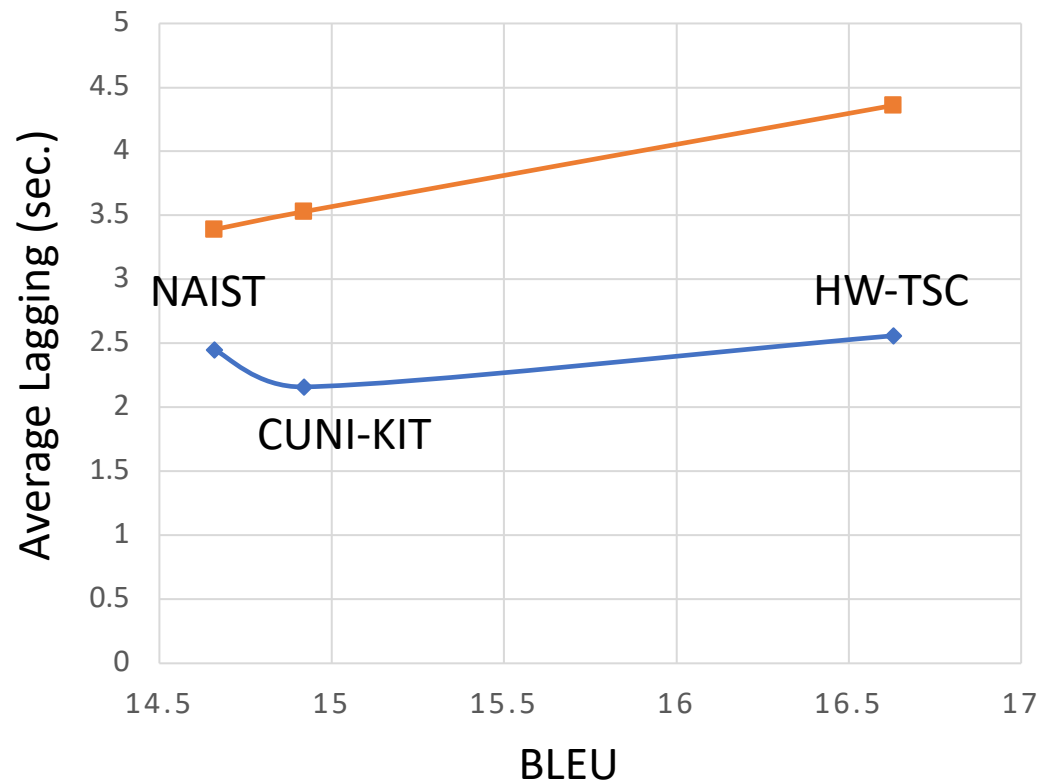
# Average Token Delay (ATD) [cont'd]

- Measure the diff. btw. the end of input chunk and the corresponding output tokens
  - Delays are always positive
- The largest difference from AL
  - AL ignores the output duration
  - ATD takes it into account;
    - A long outputs can cause further delay in later outputs
- Differences are measured by # tokens (text) or actual time (speech)



# Results in Computation-aware Latency

- In real situations, SimulST systems pose some delays due to their computations



# NAIST-SIC (Simultaneous Interpretation Corpus)

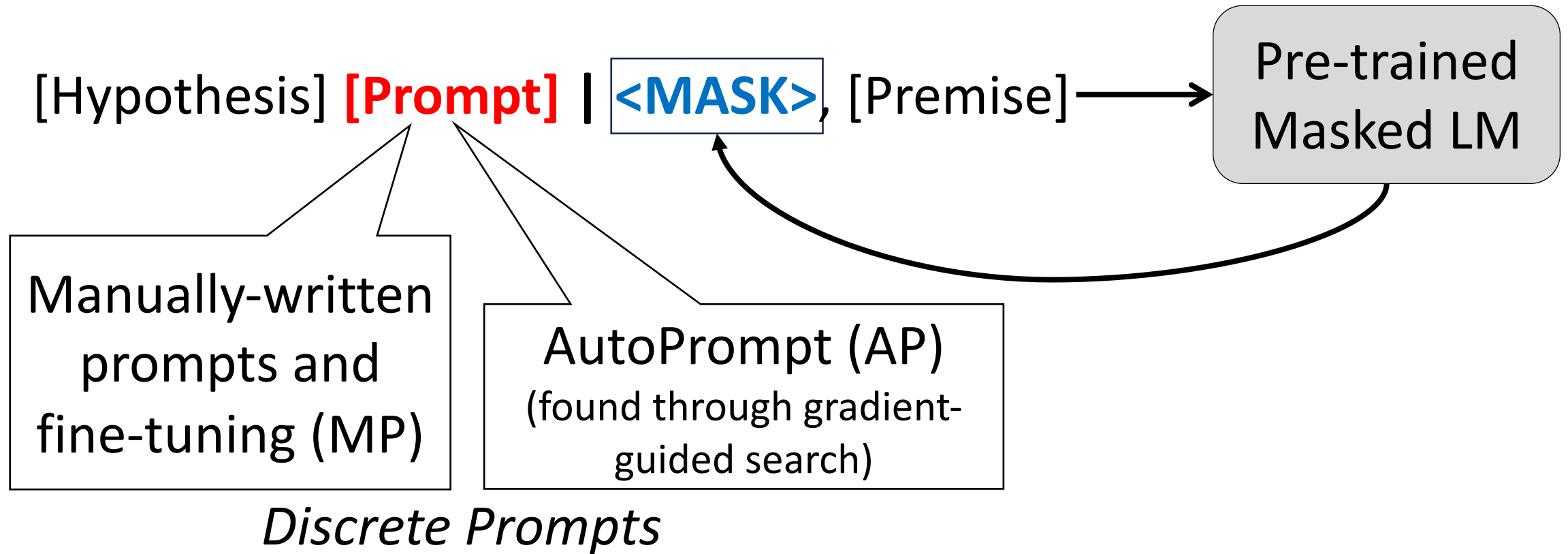
- A collection of Simultaneous Interpretation
  - <https://dsc-nlp.naist.jp/data/NAIST-SIC/>
  - You can (easily) find by searching “NAIST-SIC”
  - A part of this corpus (NAIST-SIC 2021) was used for IWSLT Simultaneous Translation shared task
    - It was also presented at IWSLT 2021
      - Doi et al., Large-Scale English-Japanese Simultaneous Interpretation Corpus: Construction and Analyses with Sentence-Aligned Data, Proc. IWSLT 2021.
  - An additional release (NAIST-SIC 2022) includes automatic source-target sentence alignment

# Robustness of discrete prompts for pre-trained models

Y. Ishibashi+, “Evaluating the Robustness of Discrete Prompts,” EACL 2023

# Prompt-based Problem Solving

- E.g., Natural Language Inference (NLI)



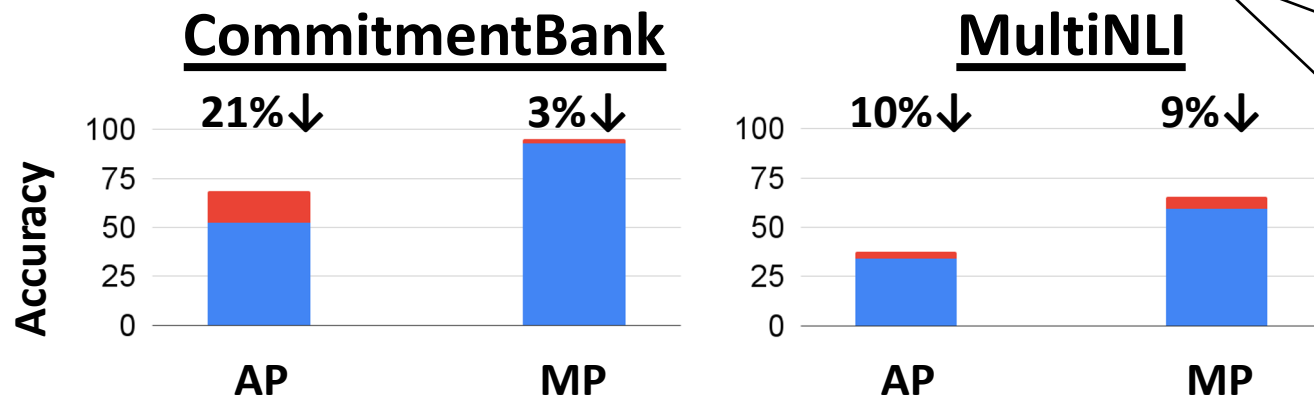


# Robustness of Discrete Prompts

- In this work, we evaluated the following on the NLI task:
  - Robustness against perturbations on discrete prompts
    - Token reordering (shuffling)
    - Token deletion
  - Robustness against different datasets
    - Out-of-domain data
    - Perturbed data

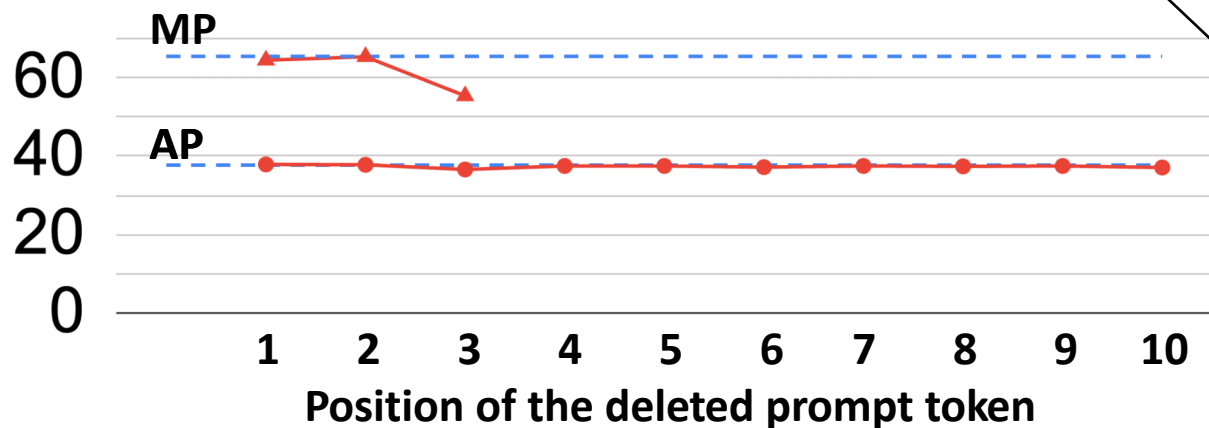
# Experimental Results (1/2)

## - Token reordering



Larger drop in AutoPrompt;  
“AP relies on token (word)  
order more than MP”

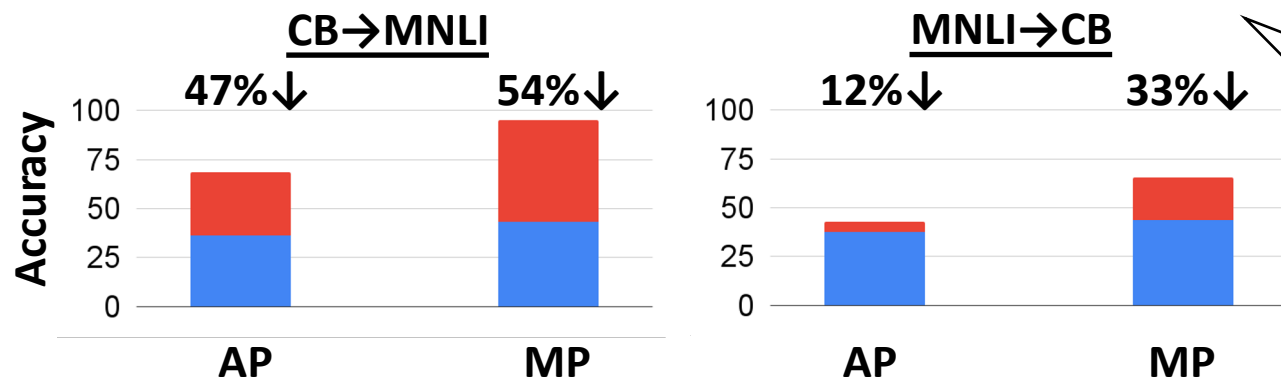
## - Token deletion



Even a single word deletion  
may hurt manually-written  
prompts seriously

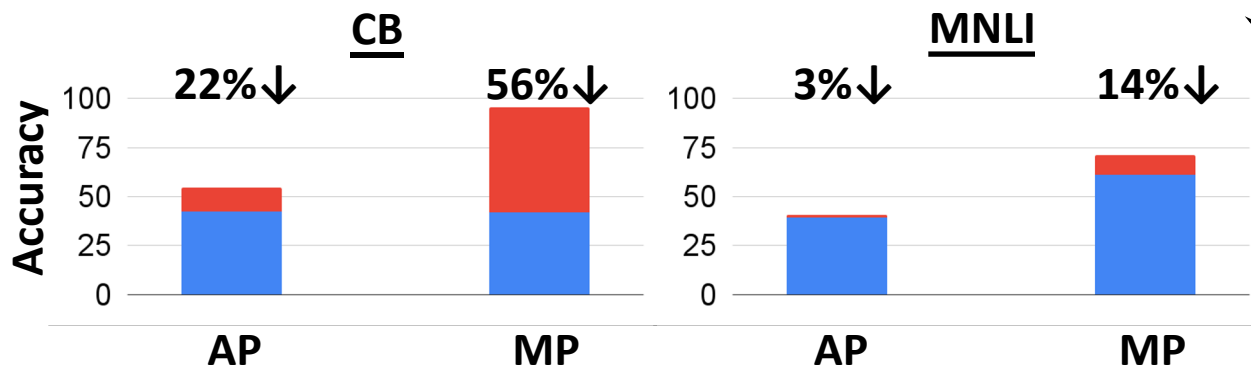
# Experimental Results (2/2)

- Out-of-domain data (cross-dataset evaluation)



Larger drop in MP;  
“MP does not generalize well”

- Evaluation data perturbation (rewrite hypotheses and labels)



Larger drop in MP;  
“MP may overfit with specific data distribution”

# Adaptive and efficient speech segmentation for speech translation

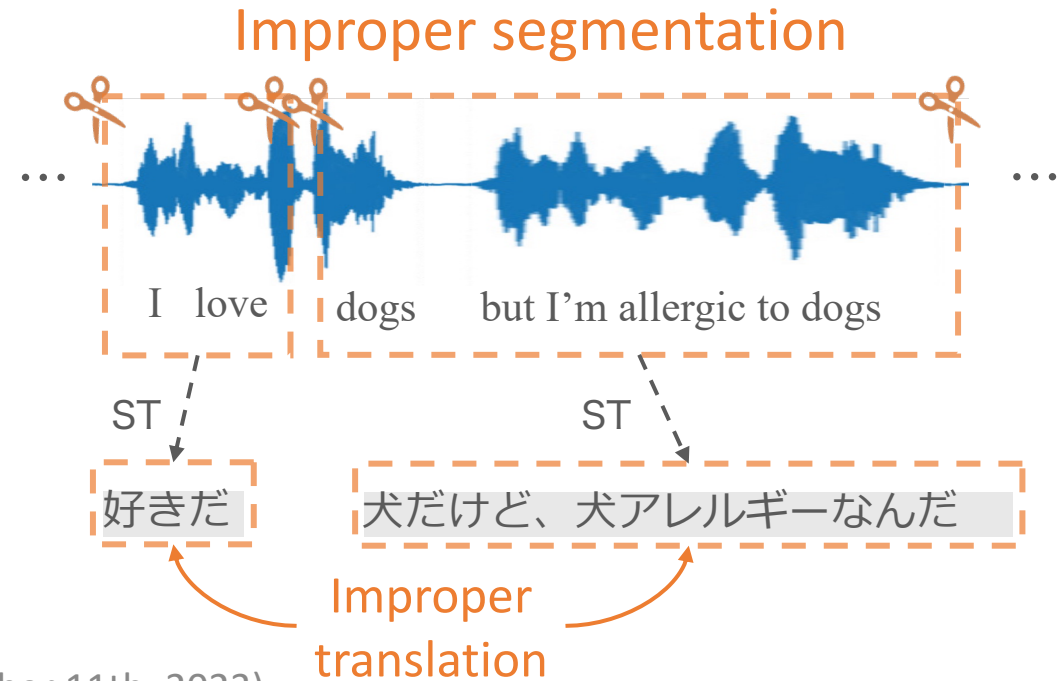
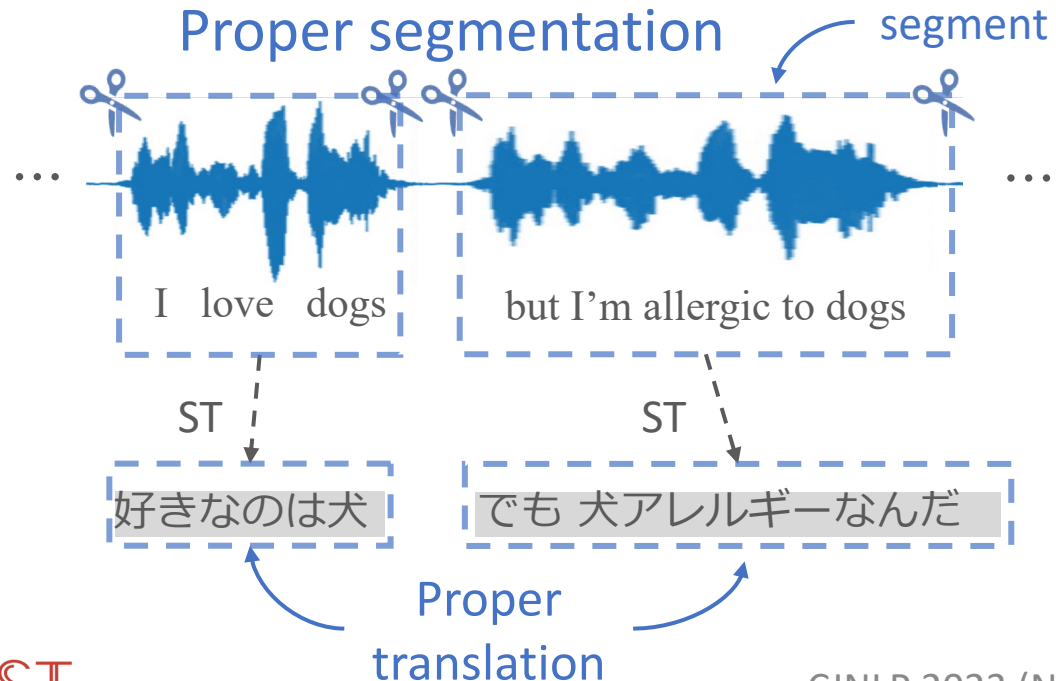
R. Fukuda+, “Speech Segmentation Optimization using Segmented Bilingual Speech Corpus for End-to-end Speech Translation,” Interspeech 2022

*+ Recent progress (not published yet)*

# Background: Speech Segmentation

**Segmentation** is a fundamental process required for Speech Translation (ST).

- Splitting continuous speech into translation units (segments).
- It's difficult because explicit boundaries such as punctuation marks are not available.
- It's important because it greatly affects translation results.

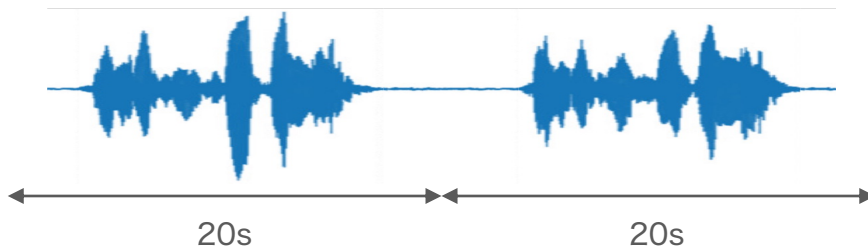


# Previous Approaches to Speech Segmentation

- Pause-based
  - Voice Activity Detection (VAD)

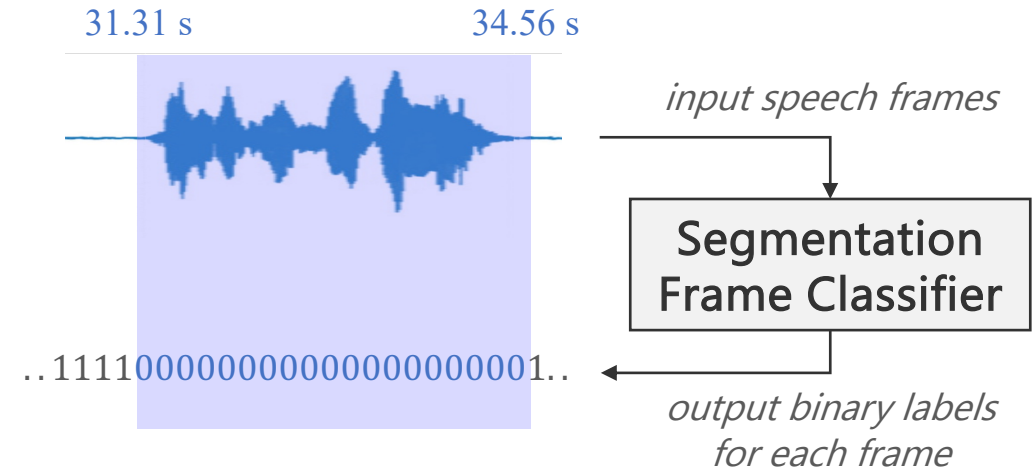


- Length-based



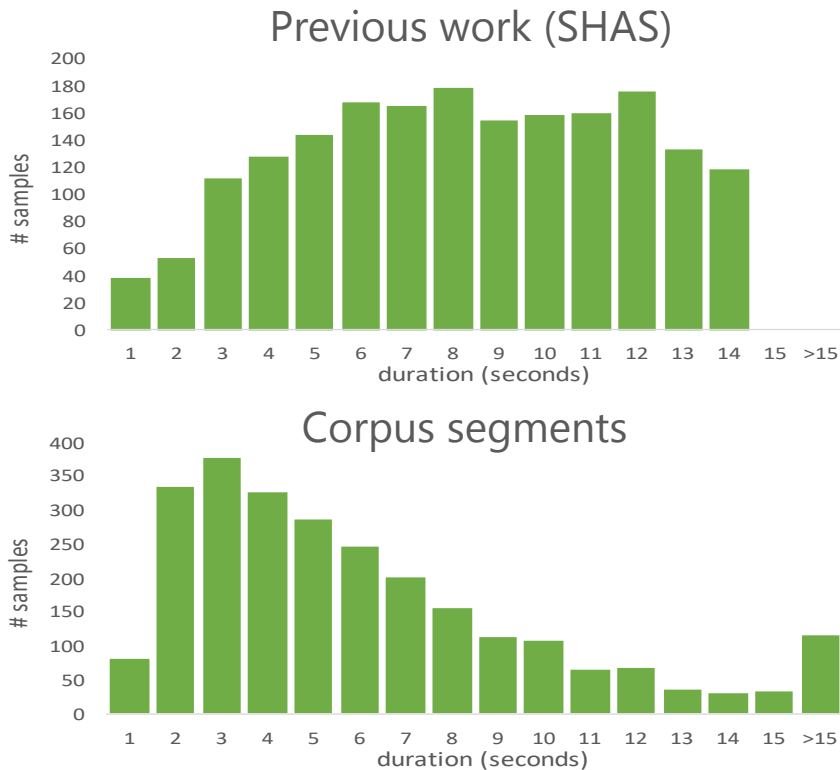
- Pause-Length Hybrid

- **Model-based**
  - Identify segment boundaries using a classifier model

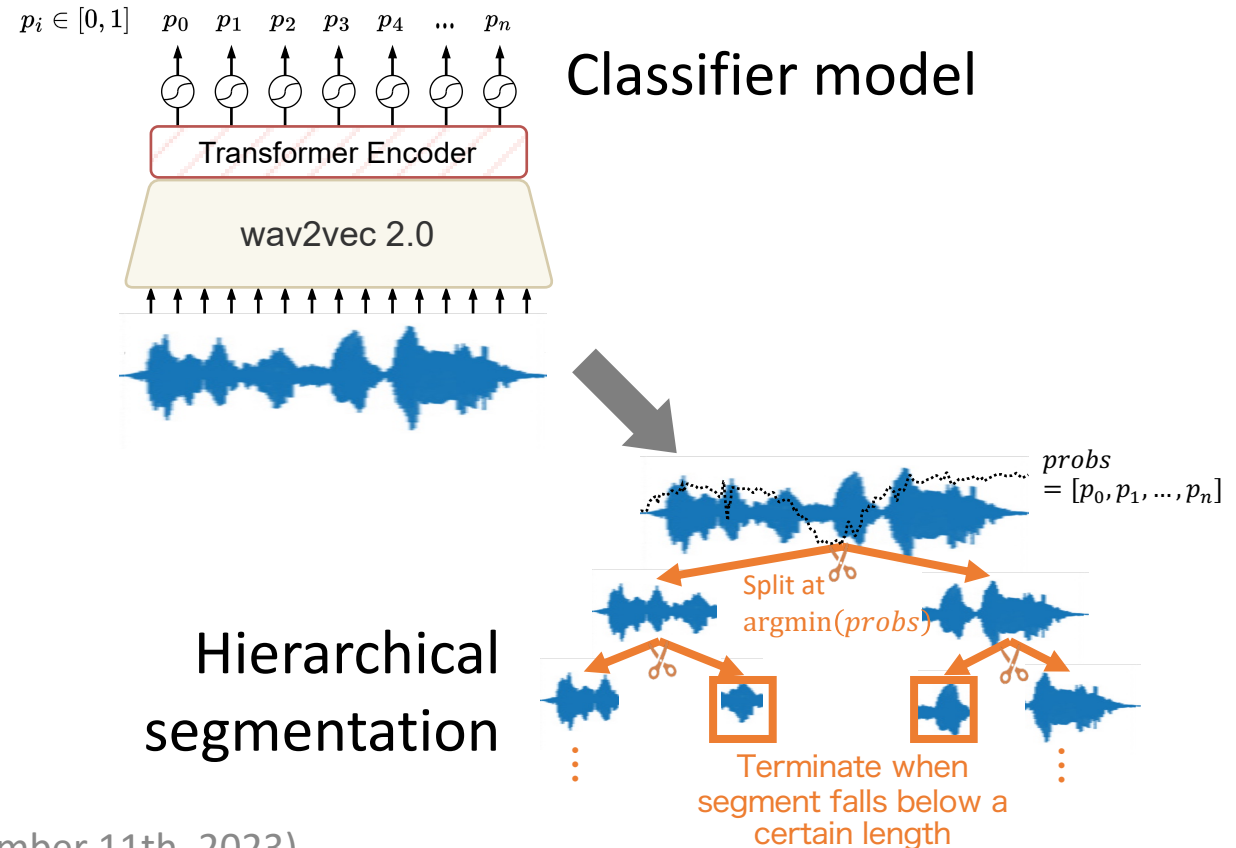


# Problem

- Inefficiency due to *passive* segmentations

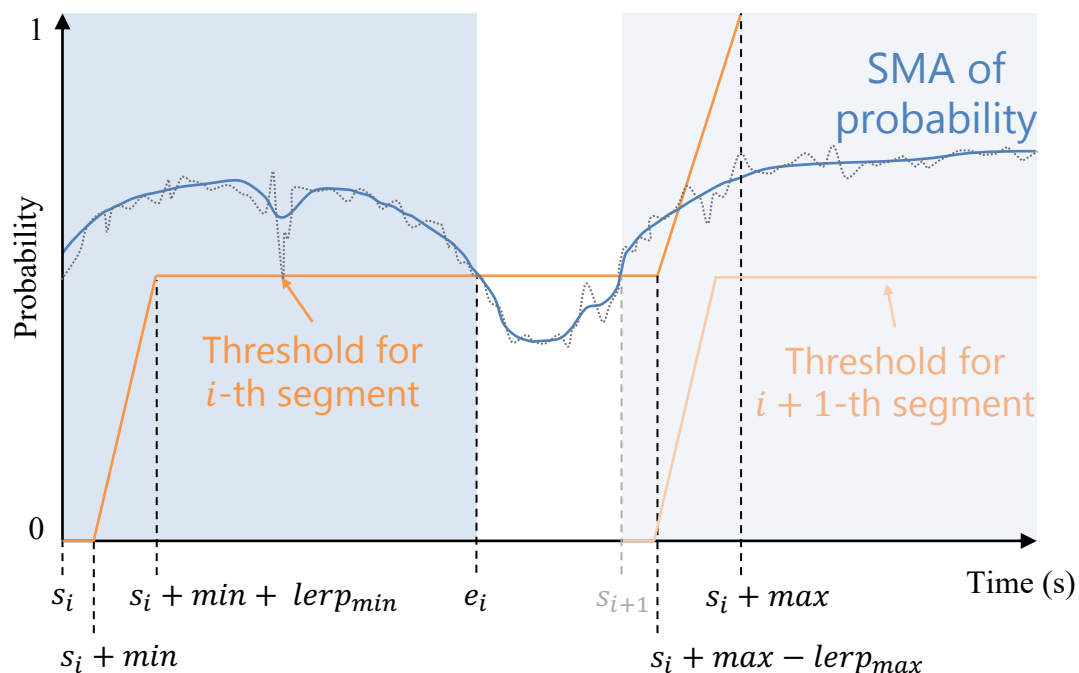


- Probabilistic divide-and-conquer [Tsaimas+ 2022 Interspeech]

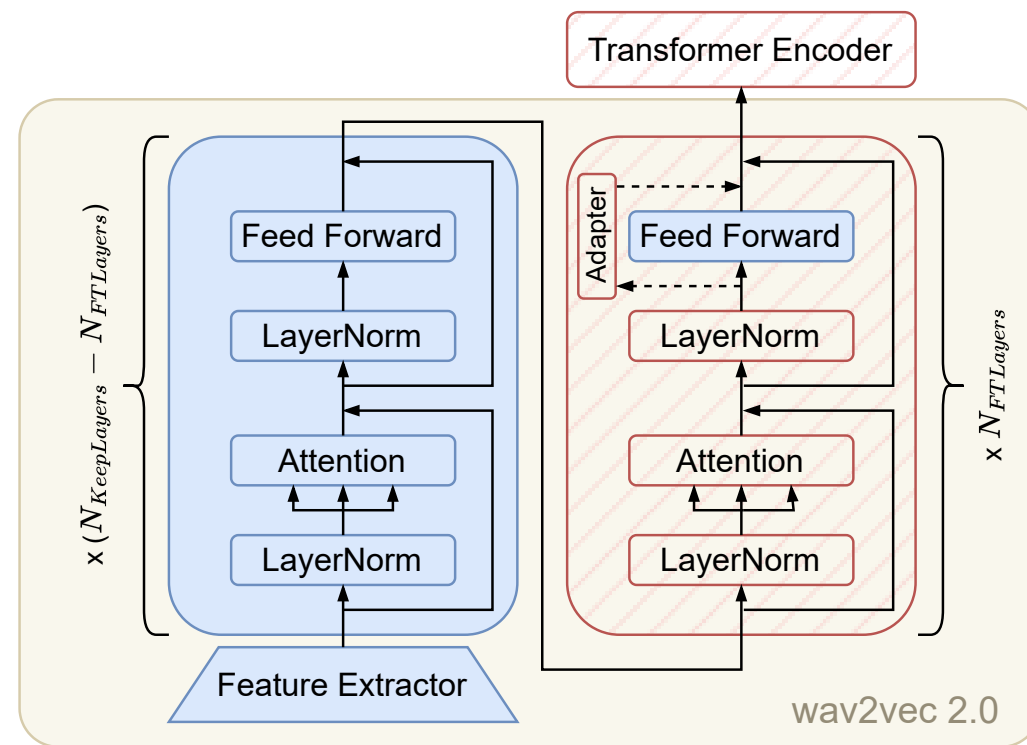


# Proposed Method

- Thresholding segmentation probability
  - Smoothed by moving average



- Fine-tuning wav2vec 2.0 Transformer layers



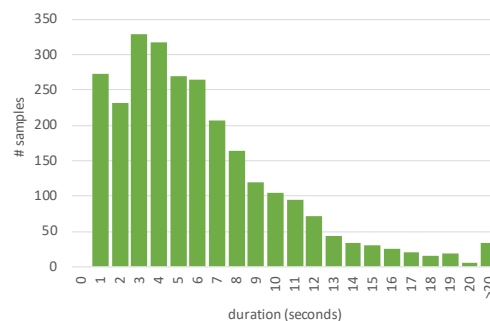
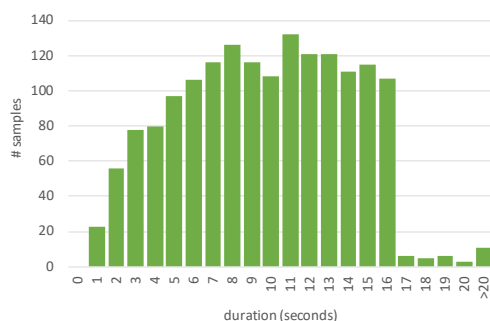


# Experimental Results

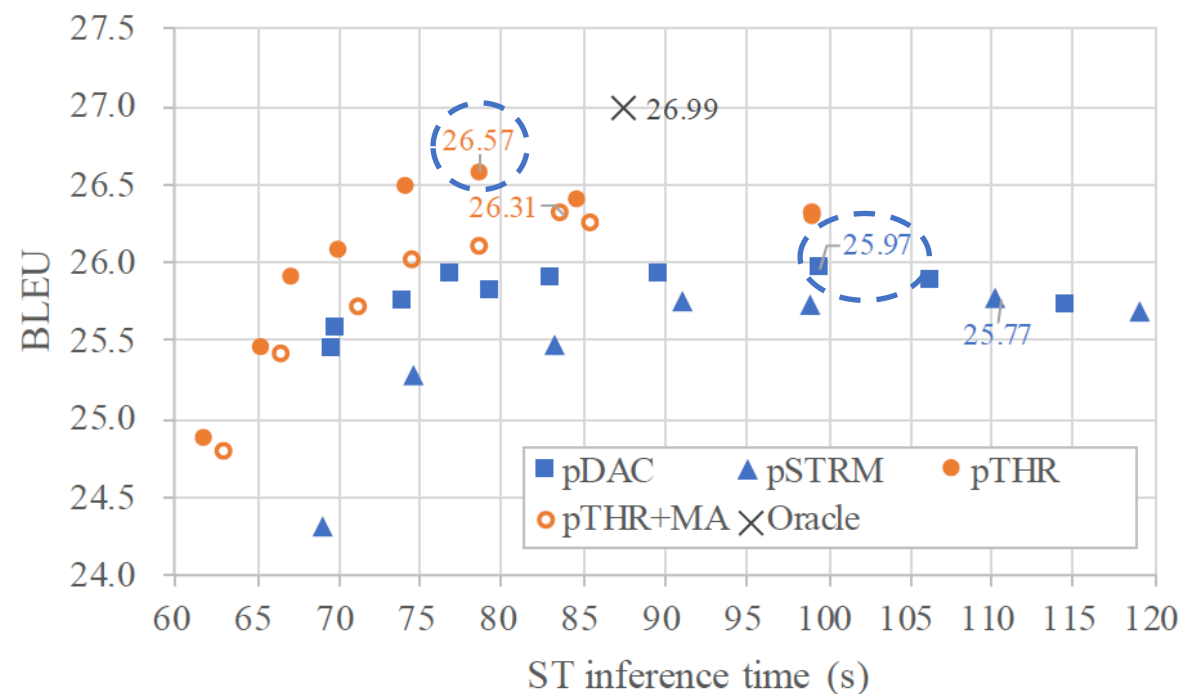
- Higher BLEU on MuST-C ende

MuST-C segmentation	26.99
SHAS [Tsiamas+ 2022]	25.67
Proposed	<b>26.30</b>

- Derives shorter segments



- More efficient than SHAS



# Summary

- NAIST's recent activity on simultaneous speech translation
  - Prototype SimulST system
  - Prefix Alignment
  - Average Token Delay
  - MQM-based human evaluation
- Recent results in IWSLT evaluation campaign
- Simultaneous Interpretation Corpus: NAIST-SIC