## Sec 5: **Advanced topics**

**Utterance segmentation** 

**Multilingual ST** 

**Under-resourced languages** 

# Sec 5.1 Utterance Segmentation

### **Utterance segmentation - Problem**

### Mismatch between training and evaluation data

Training corpora: "sentence-level" split of continuous speech  $\bigcirc$ 



### **Utterance segmentation - Problem**

- **Mismatch between training and evaluation data** 
  - Training corpora: "sentence-level" split of continuous speech  $\bigcirc$



thisisanaudiosignalinthetrainingdataitwassplitusingstrongpunctuationthreesentencesintotal



Matusov et al.: "Automatic Sentence Segmentation and Punctuation Prediction for Spoken Language Translation", IWSLT`06











Advantage: silences as a proxy of sentence boundaries Drawback: variable segments' length (including very short and very long ones)















Advantage: uniform segment length Drawback #1: split points are likely to break the input in critical positions Drawback #2: non-speech frames are kept in the input

## Solution 3: Split on silences & segments' length

Potapczyk and Przybysz: "SRPOL's system for the IWSLT 2020 end-to-end speech translation task", IWSLT 2020







### three sentences in total

### Solution 3: Split on silences & segments' length

Potapczyk and Przybysz: "SRPOL's system for the IWSLT 2020 end-to-end speech translation task", IWSLT 2020



Advantages: closer to sentence-like splits, uniform segment length Drawback #1: manually-detected silences (non scalable/reproducible) Drawback #2: full audio required for splitting (not applicable to audio streams)



### **Utterance segmentation - An open problem**



Large room for improvement compared to manual segmentation



### **Utterance segmentation - An open problem**



### **Utterance segmentation - An open problem**

