TAILOR WP7 second workshop (3-June-2021)

Valencian Research Institute for Artificial Intelligence
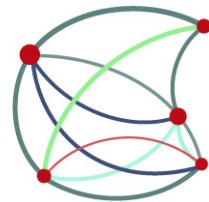
VRAIN

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Streaming Cascade-based Speech Translation leveraged by a Direct Segmentation Model

Javier Iranzo-Sánchez, Javier Jorge, Pau Baquero-Arnal, Joan Albert Silvestre-Cerdà, Adrià Giménez, Jorge Civera, Albert Sanchis and Alfons Juan
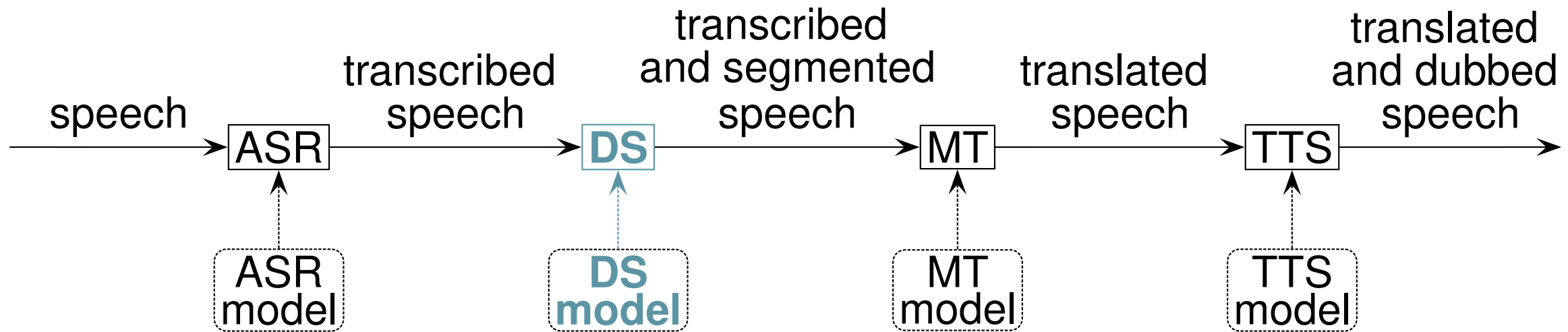
MLLP | Machine Learning and Language Processing

Streaming Cascade-based Speech Translation

A continuous stream of transcribed speech is split by a *Direct Segmentation (DS)* component into sentence-like chunks. In this way, simultaneous MT components trained at sentence level can be used to deliver high-quality translations with low latency.

Streaming cascade-based speech translation with segmenter

- ASR output is generated as the speech becomes available
- DS decides at word x_7 whether to split ASR output considering next two words ($d = 2$)
- MT input are those words on which a split decision has been taken
- The wait-k simultaneous MT system translates $k = 4$ words behind its input
- In other words, y_3 is output when x_7 becomes available

TAILOR

VRAIN

UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA
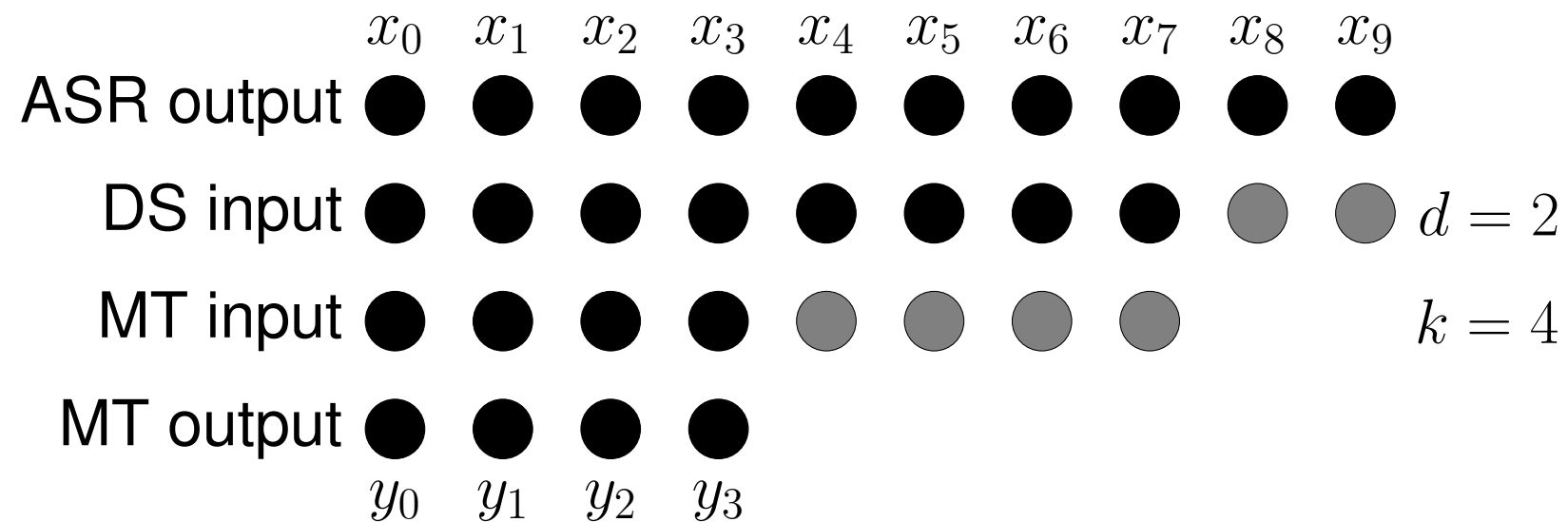
Trade-off between translation quality (BLEU) and latency
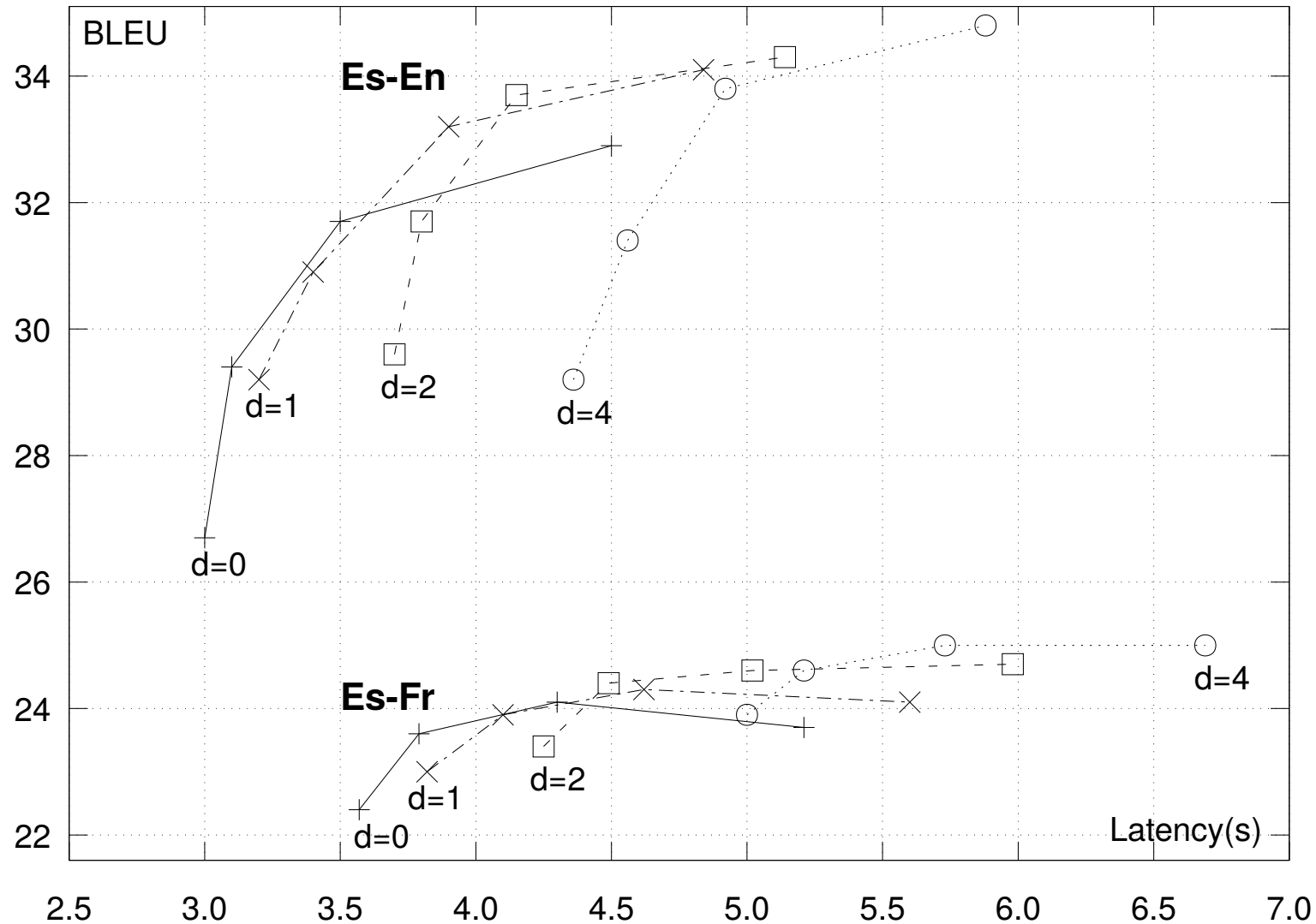
BLEU vs latency for Spanish-English (top) and Spanish-French (bottom) with future window length $d = \{0, 1, 2, 4\}$ and points on each curve from left to right representing increasing values of $k = \{1, 2, 4, 8\}$ in the wait-k MT system.

TAILOR

VRAIN UNIVERSITAT POLITÈCNICA DE VALÈNCIA