BSTC: A Large-Scale Chinese-English Speech Translation Dataset

Ruiqing Zhang, XiyangWang, Chuanqiang Zhang, Zhongjun He, Hua Wu, Zhi Li, Haifeng Wang, Ying Chen, Qinfei Li

Baidu Inc.

Automatic Speech Translation

Wireless Speech Translator

Simultaneous Interpreation

Speech Translation Datasets

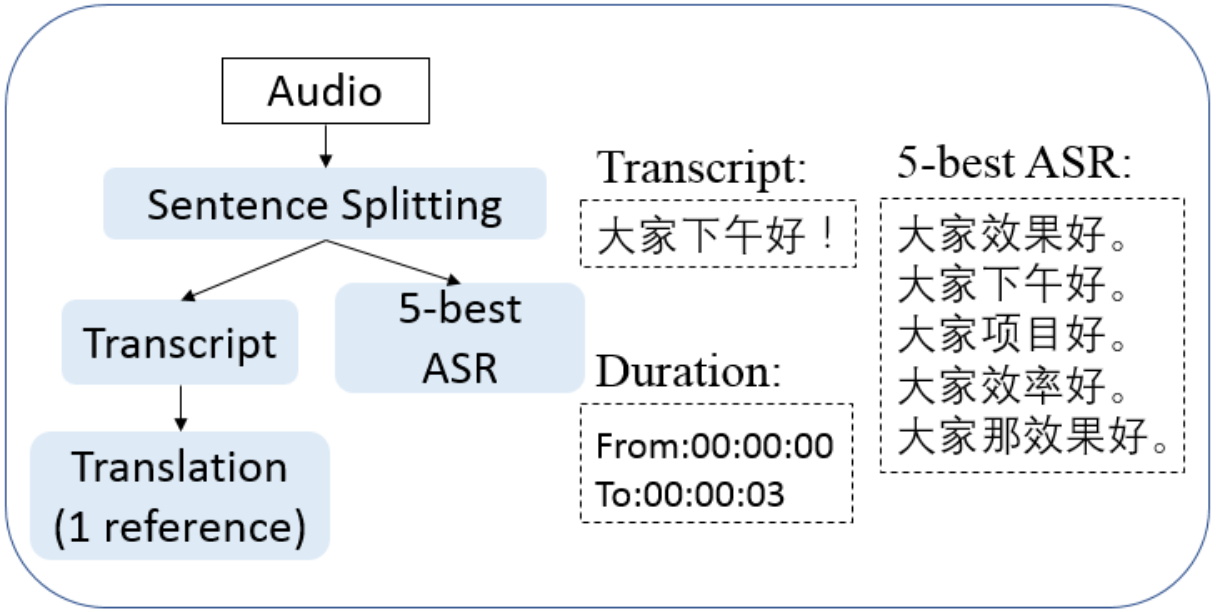
Speech Translation	Languages	Hours
F-C (2013)	Es→En	38
KIT-Disfluency (2014)	De→En	13
BTEC (2016)	En→Fr	17
MSLT V1.0 (2016)	En↔Fr/De	23
	En→Zh/Jp	6
MSLT V1.1 (2017)	Zh→En	5
	Jp →En	9
Travel (2017)	Am→En	8
Aug-LibriSpeech (2018)	En→Fr	236
MuST-C (2019)	En→8 Euro langs	3617
Europarl-ST (2020)	9 Euro langs	1642
Covost (2020a; 2020b)	En↔21 langs	2880
Simultaneous Translation	**Languages**	**Hours**
CIAIR (2004)	En↔Jp	182
EPPS (2009)	En↔Es	217
Simul-Trans (2014)	En↔Jp	22
BSTC (ours)	Zh→En	68

30 hours in total

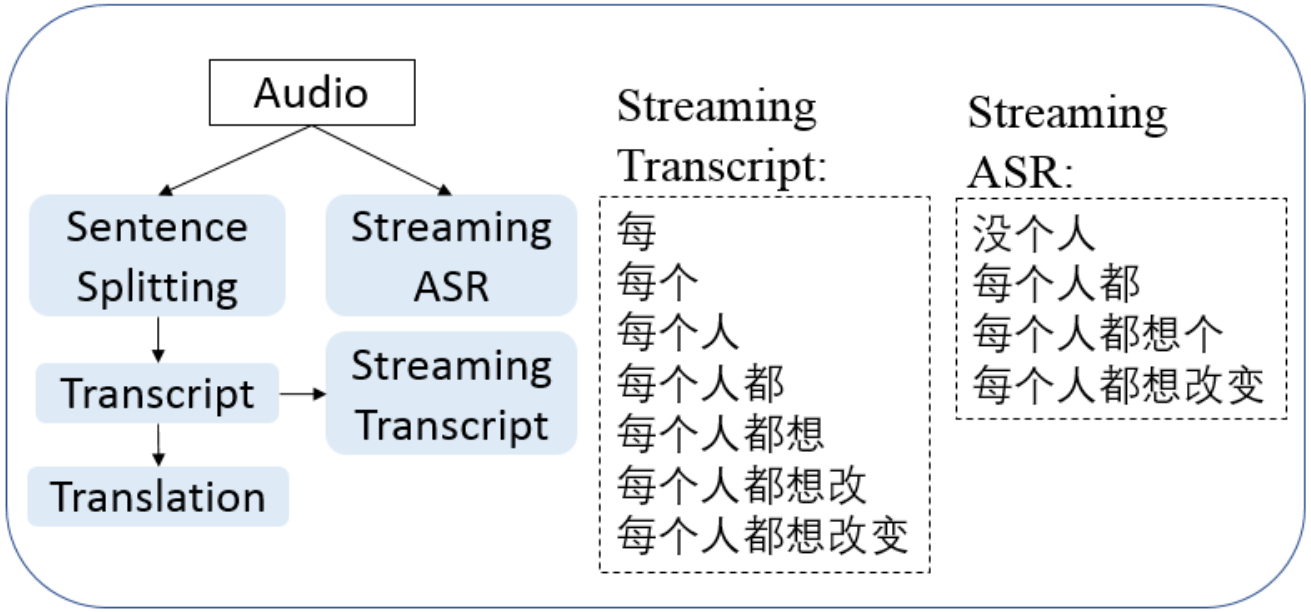
BSTC Datasets

- The first large-scale (68 hours) Chinese-English Speech Translation corpus.
- We organized two simultaneous interpretation tasks in *the 1st and 2nd workshop on Automatic Simultaneous Translation*.
- BSTC can also be taken as:
  - A *Chinese Spelling Correction* (CSC) corpus <ASR, Transcription>
  - A Zh->En *Document Translation* corpus <Document, Translation>

BSTC Training set



(a). Training set



(b). Development set and Testset

Dataset	Talks	Utterances	Transcription (characters)	Translation (tokens)	Audio (hours)	WER(1-best)
Train	215	37,901	1,028,538	606,584	64.57	27.90%
Dev	16	956	26,059	75,074	1.58	15.21%
Test	6	975	25,832	70,503	1.46	10.32%

BSTC Dataset

	BLEU	AP	Omissions
A	24.20	83.0%	53%
B	17.14	62.8%	47%
C	25.18	76.5%	53%

Table 4: Comparison of the simultaneous interpretation results of three interpreters (A, B, and C) on the BSTC test set. "AP" is the Acceptability and the "Omissions" indicates the proportion of missing translation in all translation errors.

{
"offset": "105.975",
"duration": "3.287",
"wav": "2.wav",
"transcript": "但是你们的每个人都有多个设备，啊有手持设备，有手机。",
"Streaming ASR":
Type: partial 但是
Type: partial 但是你们
Type: partial 但是你们的没
Type: partial 但是你们的没个人都
Type: partial 但是你们的没个人都有多个
Type: final 但是你们的没个人都有多个设备
Type: partial 啊有
Type: partial 啊有首
Type: partial 啊有手持摄
Type: final 啊有手持设备
Type: partial 首
Type: partial 手机

"translation": "In fact, every one of you has multiple digital devices, handheld devices and mobile phones.",
"interpreter A": "But actually you own several devices, mobile devices, mobile phones.",
"interpreter B": "But every of you have multiple equipments with you hand held equipment like phone, smartphone.",
"interpreter C": "But every one of you have multi devices, we have mobile phones."
}

Benchmark for Speech Translation

ASR system: Baidu ASR results

MT: Big Transformer

Systems	Test on Transcript		Test on ASR	
	Dev	Test	Dev	Test
pre-train on WMT	20.78	35.13	18.22	33.32
Finetune on <transcript, translation>	**23.47**(2.69↑)	**41.14**(6.01↑)	19.68(1.46↑)	35.71(2.39↑)
Finetune on <ASR, translation>	22.53(1.75↑)	39.23(4.1↑)	**19.82**(1.6↑)	**36.89**(3.57↑)

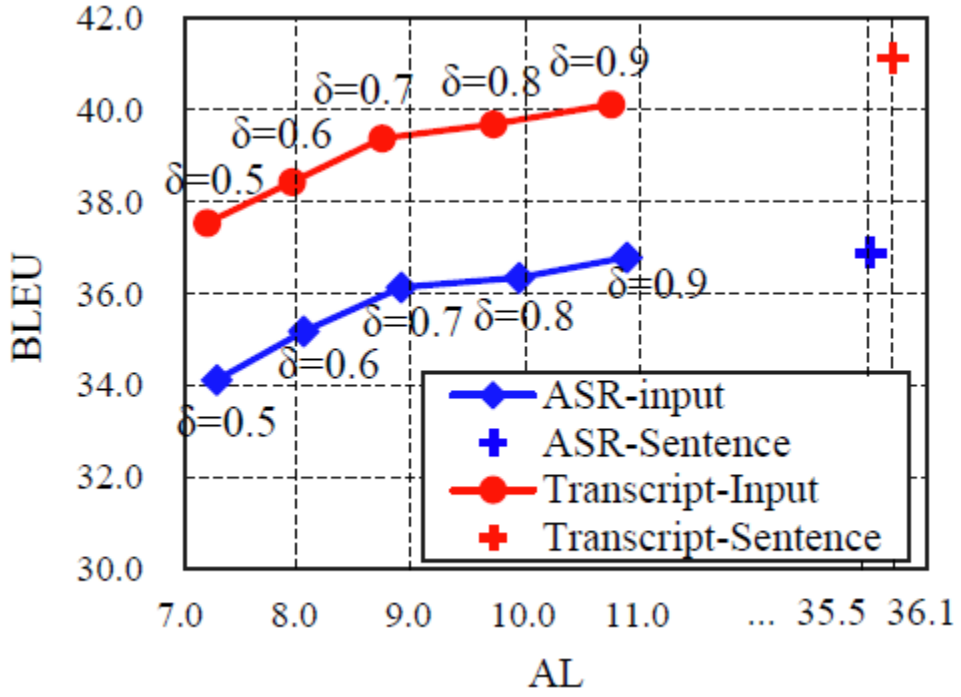
Benchmark for Speech Translation

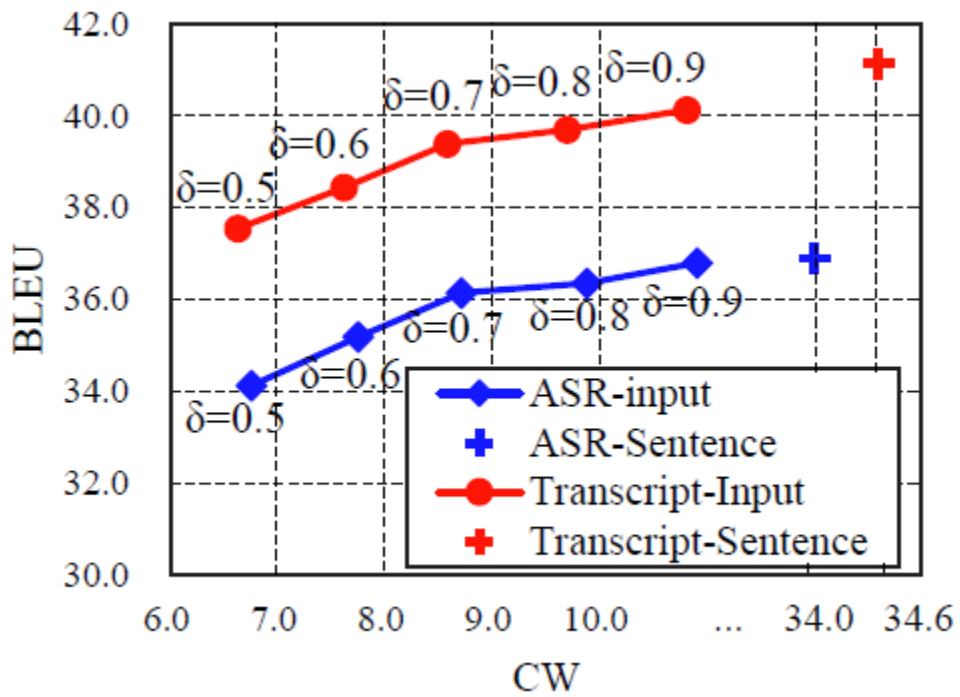
ASR system: Baidu ASR results

MT: Big Transformer

Policy: Meaningful Unit

,	。	?	MU	none





Performance on the BSTC testset.

Thank you.