The CHiME-7 DASR Challenge: Distant Meeting Transcription with Multiple Devices in Diverse Scenarios

Samuele Cornell[1] Taejin Park[2] Steve Huang[2] Christoph Boeddeker[3] Xuankai Chang[1] Matthew Maciejewski[4] Matthew Wiesner[4] Paola Garcia[4] Shinji Watanabe[4]

[1]Carnegie Mellon University, USA [2]NVIDIA, USA [3]Paderborn University, Germany [4]Johns Hopkins University, USA

Motivation & Novelties

- Comparing domain-optimized (NOTSOFAR-1 Task 2) vs "Generalist" approach (DASR, Task 1).
- Increasing the scenario diversity, especially regarding the number of speakers and recording devices
  - NOTSOFAR-1 is included as an additional scenario
- Addressing some limitations of C7DASR by re-annotating Mixer 6 dev set and providing official training and dev partitions for DiPCo and Mixer 6.
- Spur more innovative and practically viable approaches via a jury-award mechanism

"Core" Datasets

Participants systems are evaluated on 4 "core" scenarios.

External datasets and models as listed in the website are also allowed for training and validation

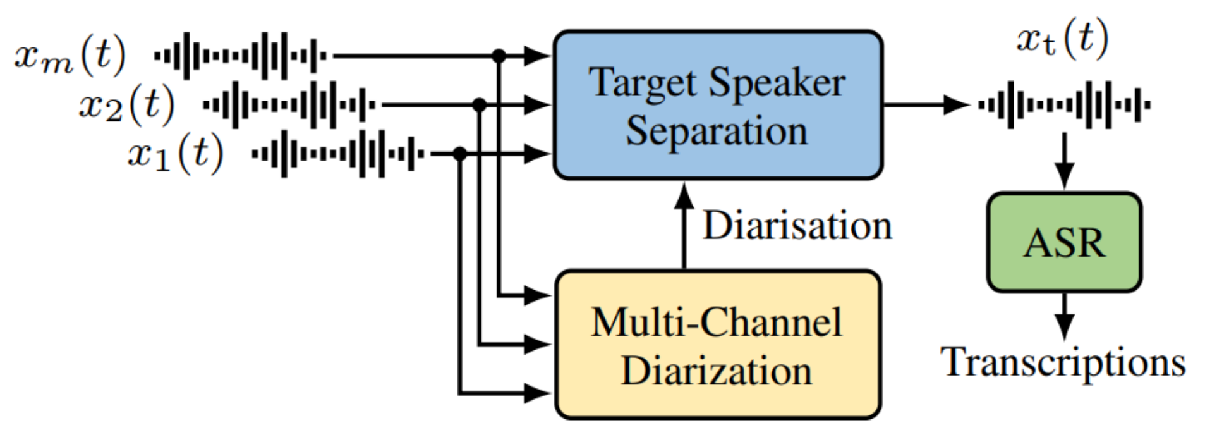
Table 1: *C8DASR scenarios diversity overview.*

Scenario	Setting	Number of Speakers	Recording Setup	Avg. Duration
CHiME-6	dinner party	4	6 linear arrays	∼2h-2h 30 mins
DiPCo	dinner party	4	5 circular arrays	∼20-30 mins
Mixer 6	1-to-1 interview	2	10 heterogeneous devices	∼15 mins
NOTSOFAR-1	office meeting	4-8	1 circular array	∼6 mins

Table 2: *CHiME-8 DASR core datasets statistics overview. We report the number of utterances, speakers, and sessions, as well as silence (sil), single-speaker speech (1-spk) and overlapped speech (ovl) ratios over the total duration.*

Scenario	Split	Size (h)	Utts	Spk.	Sess.	sil (%)	1-spk (%)	ovl (%)
CHiME-6	train	40:05	79967	32	16	22.6	52.7	24.7
	dev	4:27	7437	8	2	13.1	43.4	43.5
	eval	5:12	11028	8	2	21.3	52.0	26.7
DiPCo	train	1:12	1379	8	3	8.3	72.0	19.6
	dev	1:31	2294	8	2	7.4	61.9	30.6
	eval	2:36	3405	16	5	9.4	65.7	24.9
Mixer 6	train calls	36:09	27280	81	243	–	–	–
	train intv	26:57	29893	77	189	–	–	–
	train	6:13	3785	19	24	8.6	73.3	18.0
	dev	8:56	5903	22	35	8.4	72.1	19.5
	eval	5:45	5115	18	23	2.4	83.6	13.9
NOTSOFAR-1	train	14:43	101301	14	379	6.0	62.3	31.7
	train_sc	53:43	139913	14	526	5.9	62.4	31.7
	dev	13:25	24238	11	130	15.6	67.7	16.7
	eval	16:29	38662	12	160	5.6	64.7	29.6

Baseline Systems



Two baseline systems are provided. Both share the same structure but differ greatly in the diarization component.

- ESPnet: derived from last year baseline.
- NeMo: based on last year NeMo team submission.

Baseline Systems Results

Table 4: *Top panel: CHiME-8 DASR ESPnet and NeMo baselines overall results in terms of cpWER (%) and tcpWER (%). We highlight best figures between the two baselines for each scenario. Bottom panel: figures obtained by last year ESPNet C7DASR baseline.*

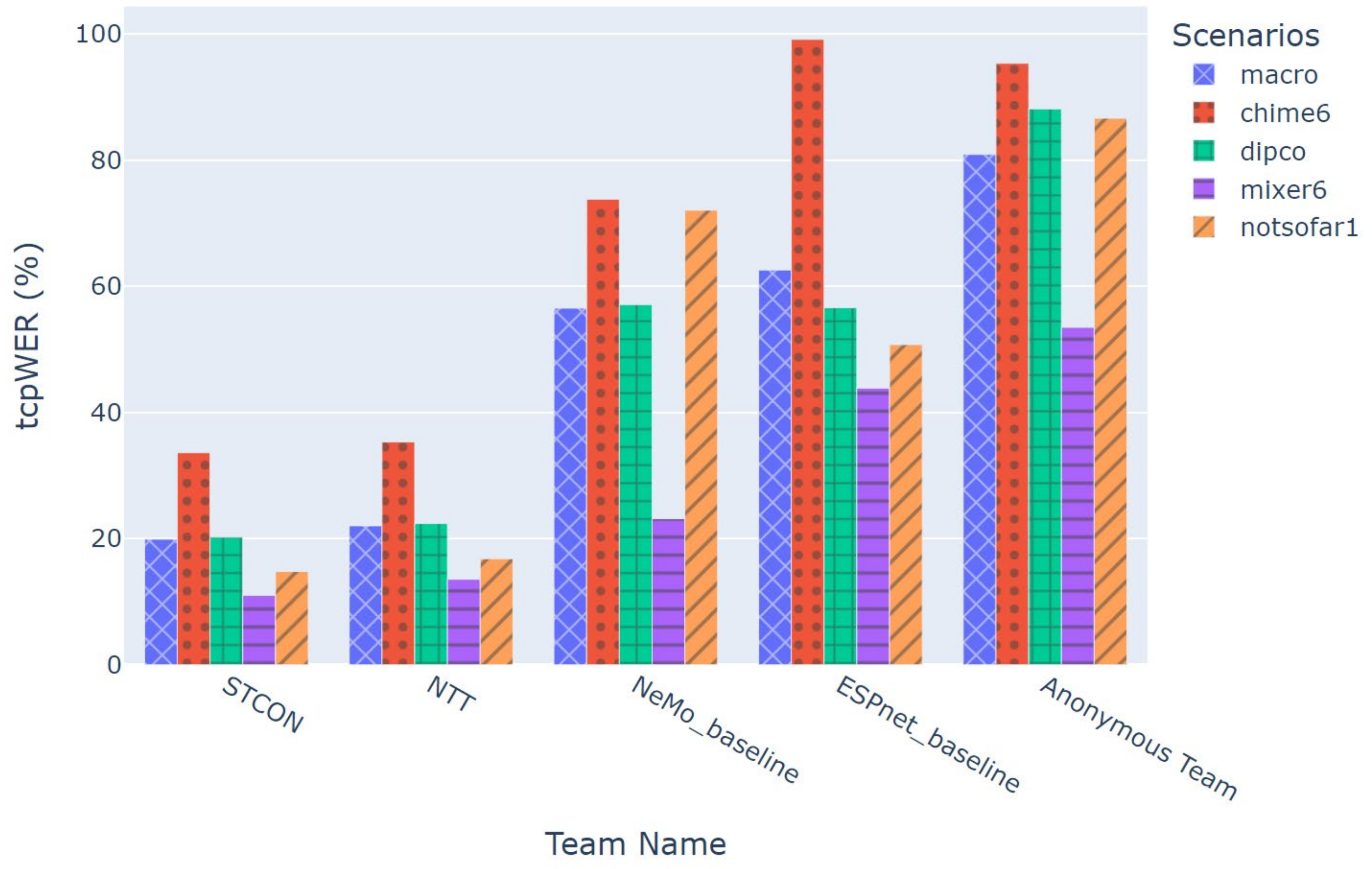
Baseline		**Dev** WER (%)		**Eval** WER (%)	
System	**Scenario**	cp	tcp	cp	tcp
ESPnet	CHiME-6	79.2	88.6	91.8	99.1
	DiPCo	90.9	98.3	52.8	**56.6**
	Mixer 6	**23.4**	**23.9**	42.0	43.8
	NOTSOFAR-1	**42.4**	**46.2**	**48.5**	**50.7**
	Macro	59.0	64.2	58.8	62.6
NeMo	CHiME-6	**52.2**	**56.5**	**67.7**	**73.8**
	DiPCo	**72.3**	**75.8**	**54.6**	57.1
	Mixer 6	17.9	19.4	**22.3**	**23.1**
	NOTSOFAR-1	55.5	61.0	67.2	72.0
	Macro	**49.6**	**53.2**	**52.9**	**56.5**
C7DASR	CHiME-6	60.8	65.7	73.7	85.2
	DiPCo	38.0	38.9	52.4	58.4
	Mixer 6	20.7	21.5	31.7	32.2

There is a significant degradation w.r.t. last year results, especially on CHiME-6 and Mixer 6 scenarios.

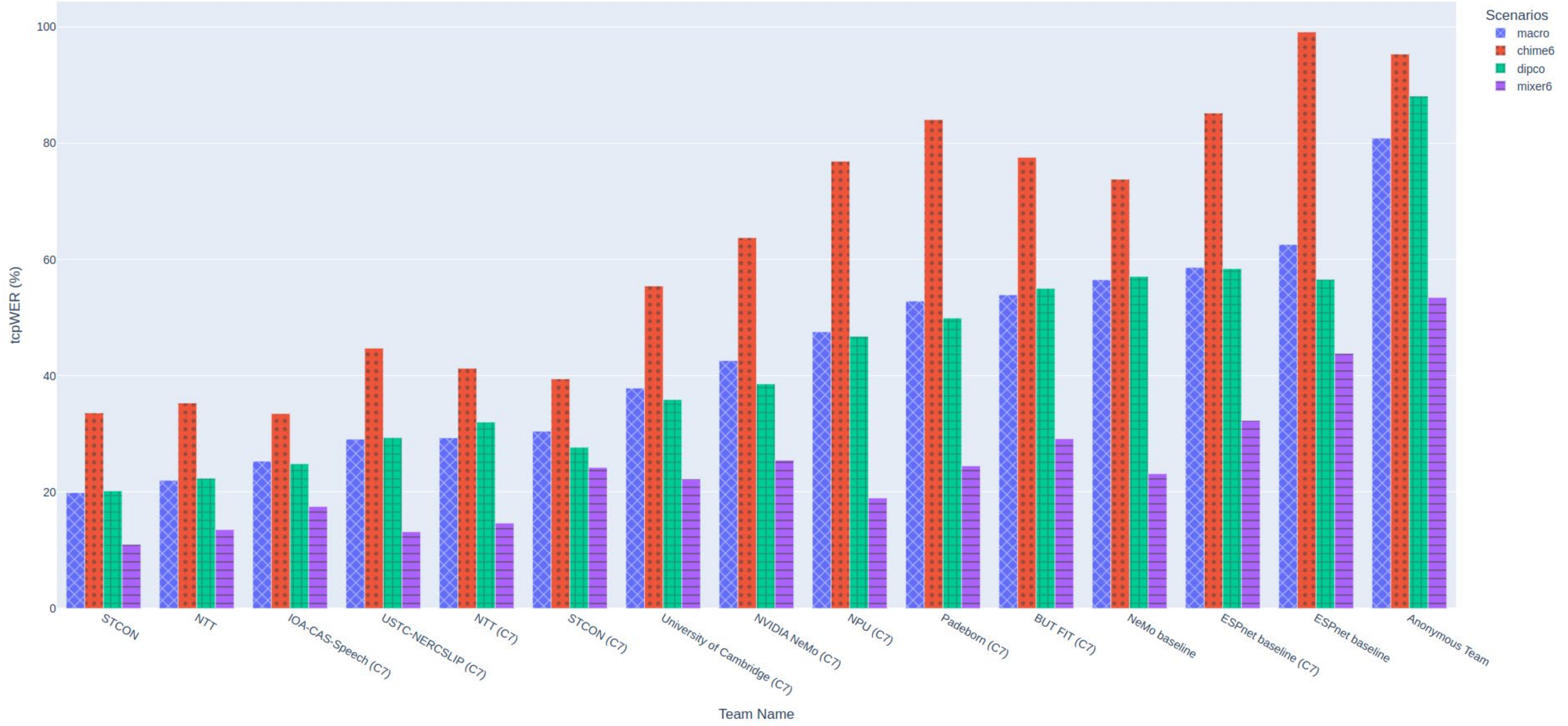
- The addition of NOTSOFAR-1 complicates the speaker counting.
- NeMo system, fares overall better (NOTSOFAR-1 results are however significantly worse)

Challenge Results

- Participants systems are ranked according to tcpWER macro-averaged across scenarios.
- CHiME-6 and DiPCO scenarios are still the hardest ones, despite the lower number of speakers w.r.t NOTSOFAR-1 and the higher amount of devices.
- Remarkably, STCON and NTT virtually also place 2nd and 3rd on NOTSOFAR-1 Task 2.



CHiME-8 + 7 Overall Results



Both STCON and NTT improve w.r.t. their last year submissions

- This is notable as their systems have also to handle in C8 the NOTSOFAR-1 scenario.

-.