Oriental Language Recognition (OLR) 2021 Challenge Summary

Qingyang Hong
Xiamen University

2022.01.14

Outline

- Challenge Organization
- Tasks, Data and Baseline Systems
- Popular Technologies
  - OLR-LID Tasks
  - OLR-ASR Tasks
- Challenge Results
- Summary

Challenge Organization

OLR 2021 Challenges

Organization Committee

Qingyang Hong, Xiamen University **Lin Li**, Xiamen University
Binling Wang, Xiamen University **Wenxuan Hu**, Xiamen University **Jing Li,** Xiamen University
Dong Wang, Tsinghua University
Ming Li, Duke-Kunshan University
Xiaolei Zhang, Northwestern Polytechnical University
Ke Li, SpeechOcean **Cheng Yang**, SpeechOcean

Special thanks to: **Qiulin Wang**, **Yiming Zhi, Feng Wang** at XMU Speech Lab

OLR 2021 Workshop

2016

- 7 languages
- Transcriptions

9/9

2017

- 10 languages
- Transcriptions
- Shorter segments

19/31

2018

- 10 languages
- Short-utterance
- Confusing-language
- Open-set

17/25

2019

- 13 languages
- +3 languages
- Short-utterance
- Cross-channel
- Zero-resource

20/45

2020

- 16 Languages
- +3 Dialects
- Short-utterance
- Cross-channel
- Open-set dialect
- Noisy-test

20/58

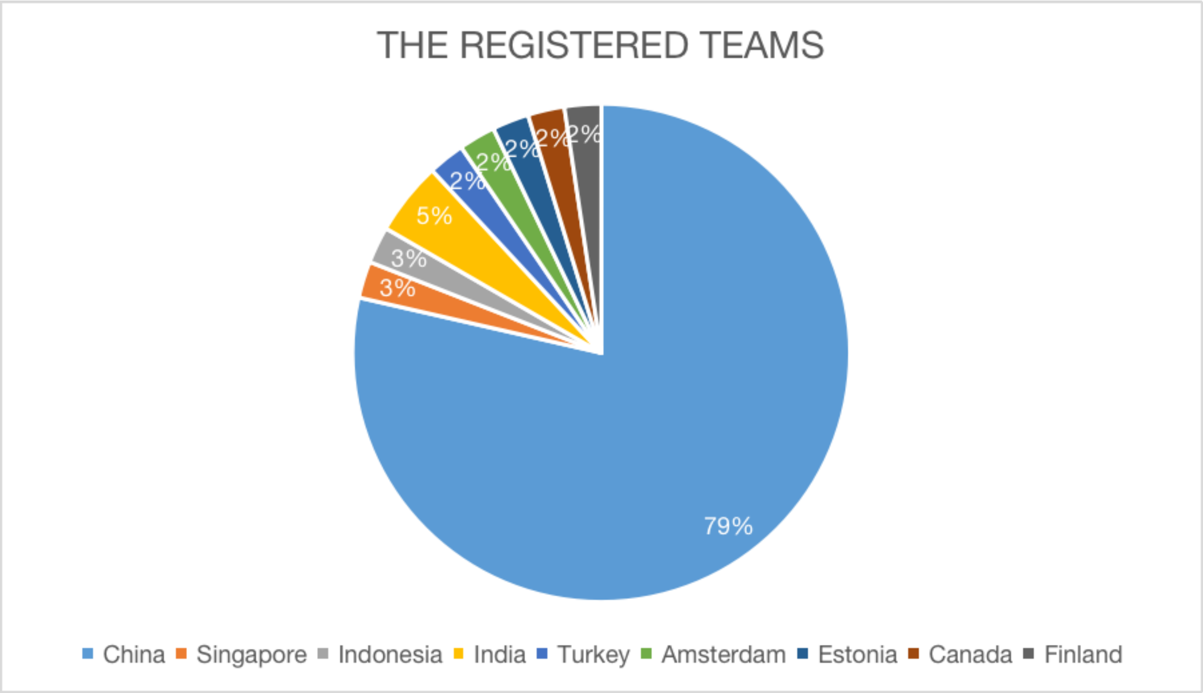
2021

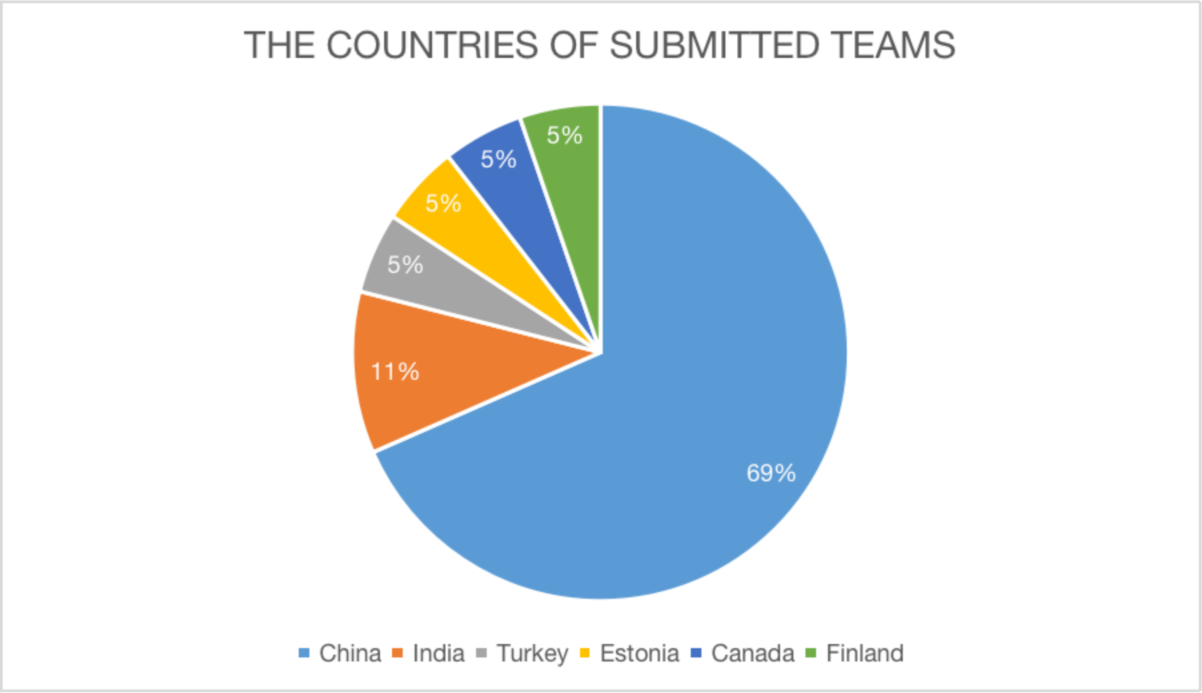
- 18 languages/dialects
- 280h+ Training data
- Cross domain
- Multilingual ASR tasks
- Unconstrained tasks
- Wild data for test
- Real-time leaderboards

19/42

Submitted results/Registered teams

OLR 2021 Workshop





Tasks, Data and Baseline systems

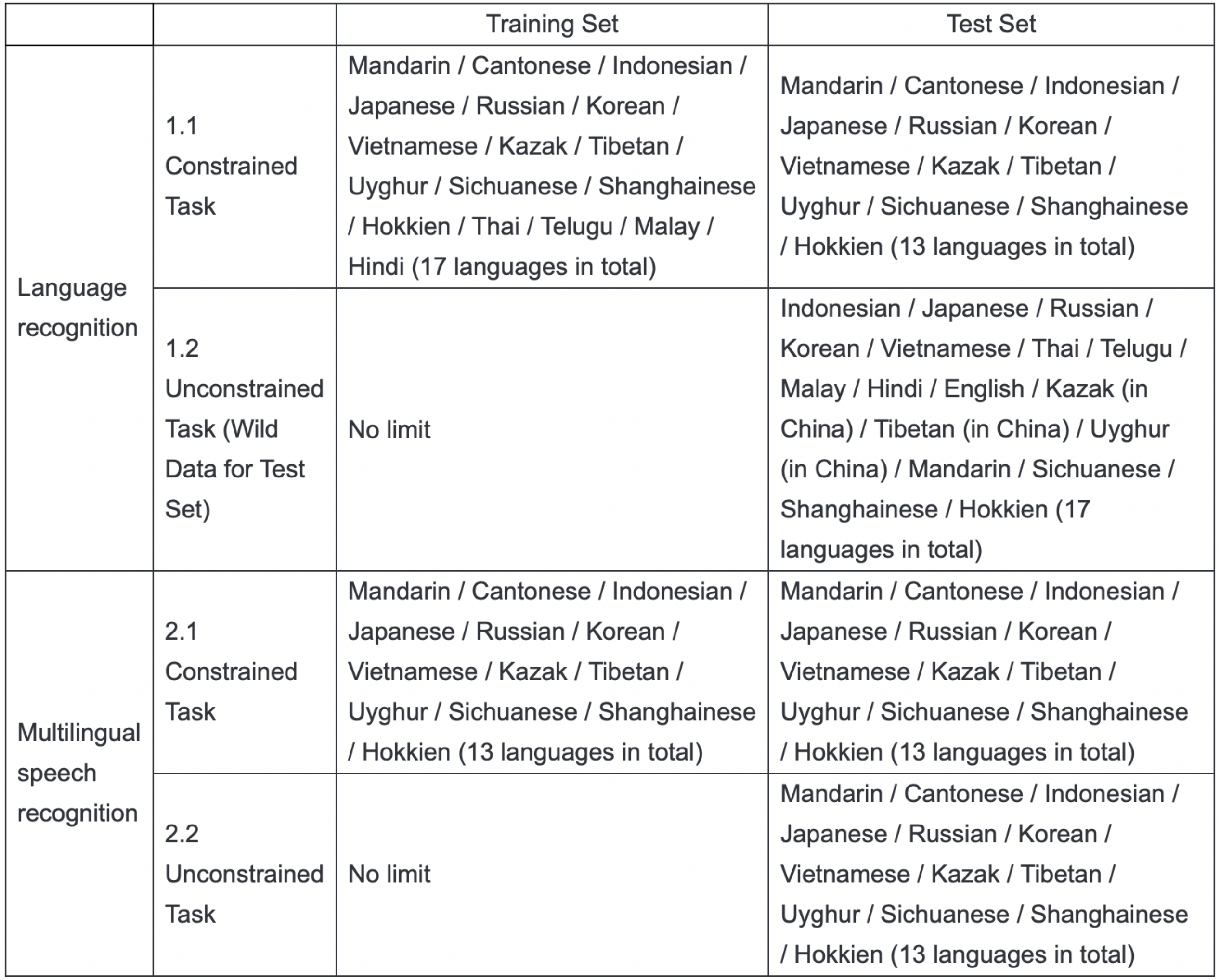
Tasks

OLR-LID Tasks

- Task 1: Constrained LID Task (Cross Domain)
- Task 2: Unconstrained LID Task (Wild Data)

OLR-ASR Tasks

- Task 1 : Constrained ASR Task
- Task 2 : Unconstrained ASR Task

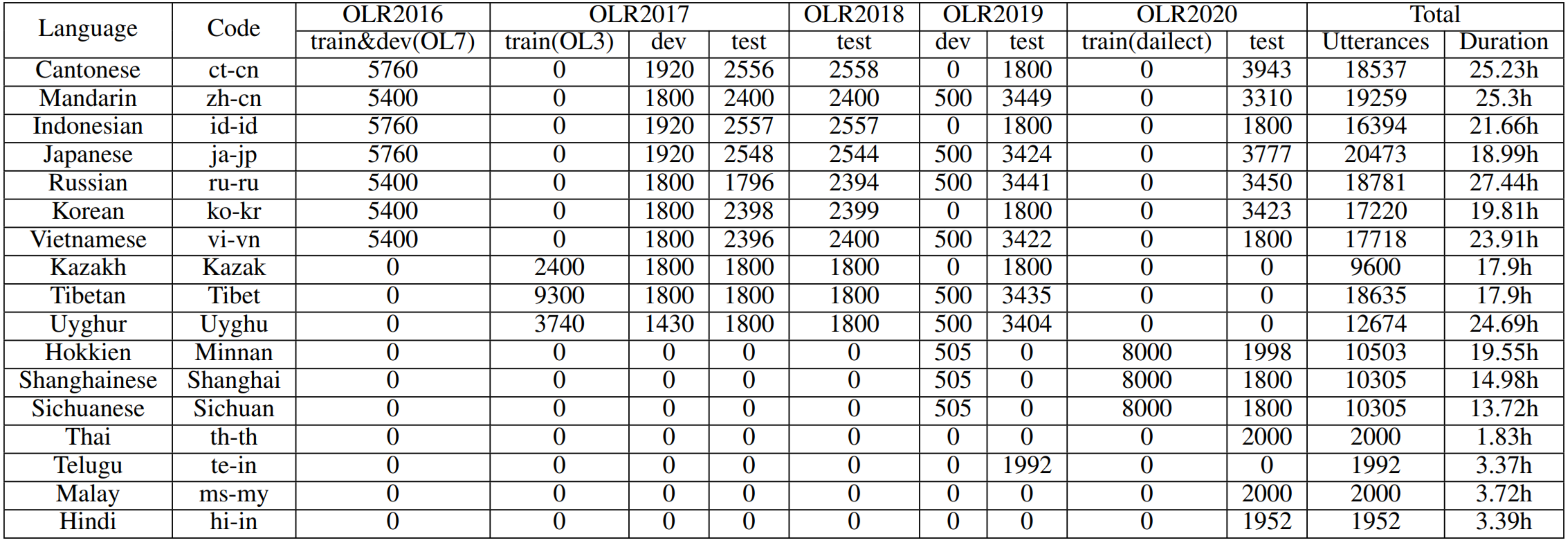


		Training Set	Test Set
Language recognition	1.1 Constrained Task	Mandarin / Cantonese / Indonesian / Japanese / Russian / Korean / Vietnamese / Kazak / Tibetan / Uyghur / Sichuanese / Shanghainese / Hokkien / Thai / Telugu / Malay / Hindi (17 languages in total)	Mandarin / Cantonese / Indonesian / Japanese / Russian / Korean / Vietnamese / Kazak / Tibetan / Uyghur / Sichuanese / Shanghainese / Hokkien (13 languages in total)
	1.2 Unconstrained Task (Wild Data for Test Set)	No limit	Indonesian / Japanese / Russian / Korean / Vietnamese / Thai / Telugu / Malay / Hindi / English / Kazak (in China) / Tibetan (in China) / Uyghur (in China) / Mandarin / Sichuanese / Shanghainese / Hokkien (17 languages in total)
Multilingual speech recognition	2.1 Constrained Task	Mandarin / Cantonese / Indonesian / Japanese / Russian / Korean / Vietnamese / Kazak / Tibetan / Uyghur / Sichuanese / Shanghainese / Hokkien (13 languages in total)	Mandarin / Cantonese / Indonesian / Japanese / Russian / Korean / Vietnamese / Kazak / Tibetan / Uyghur / Sichuanese / Shanghainese / Hokkien (13 languages in total)
	2.2 Unconstrained Task	No limit	Mandarin / Cantonese / Indonesian / Japanese / Russian / Korean / Vietnamese / Kazak / Tibetan / Uyghur / Sichuanese / Shanghainese / Hokkien (13 languages in total)

Data: Official Data

Training Data Released up to 280h.

TABLE I
DATA ALLOWED FOR CONSTRUCTING SYSTEMS

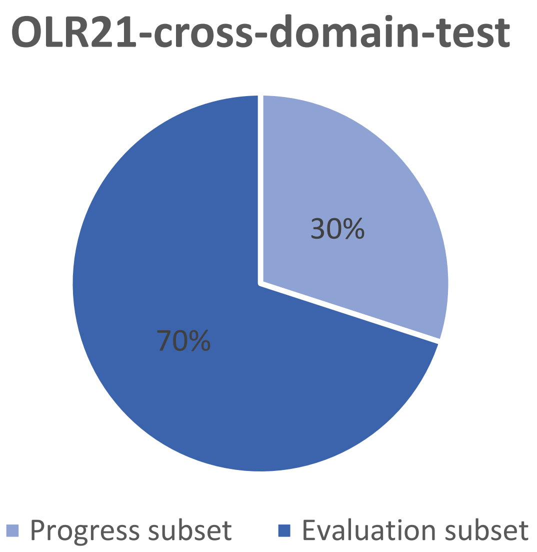


Language	Code	OLR2016	OLR2017			OLR2018	OLR2019		OLR2020		Total	
		train&dev(OL7)	train(OL3)	dev	test	test	dev	test	train(dailect)	test	Utterances	Duration
Cantonese	ct-cn	5760	0	1920	2556	2558	0	1800	0	3943	18537	25.23h
Mandarin	zh-cn	5400	0	1800	2400	2400	500	3449	0	3310	19259	25.3h
Indonesian	id-id	5760	0	1920	2557	2557	0	1800	0	1800	16394	21.66h
Japanese	ja-jp	5760	0	1920	2548	2544	500	3424	0	3777	20473	18.99h
Russian	ru-ru	5400	0	1800	1796	2394	500	3441	0	3450	18781	27.44h
Korean	ko-kr	5400	0	1800	2398	2399	0	1800	0	3423	17220	19.81h
Vietnamese	vi-vn	5400	0	1800	2396	2400	500	3422	0	1800	17718	23.91h
Kazakh	Kazak	0	2400	1800	1800	1800	0	1800	0	0	9600	17.9h
Tibetan	Tibet	0	9300	1800	1800	1800	500	3435	0	0	18635	17.9h
Uyghur	Uyghu	0	3740	1430	1800	1800	500	3404	0	0	12674	24.69h
Hokkien	Minnan	0	0	0	0	0	505	0	8000	1998	10503	19.55h
Shanghainese	Shanghai	0	0	0	0	0	505	0	8000	1800	10305	14.98h
Sichuanese	Sichuan	0	0	0	0	0	505	0	8000	1800	10305	13.72h
Thai	th-th	0	0	0	0	0	0	0	0	2000	2000	1.83h
Telugu	te-in	0	0	0	0	0	0	1992	0	0	1992	3.37h
Malay	ms-my	0	0	0	0	0	0	0	0	2000	2000	3.72h
Hindi	hi-in	0	0	0	0	0	0	0	0	1952	1952	3.39h

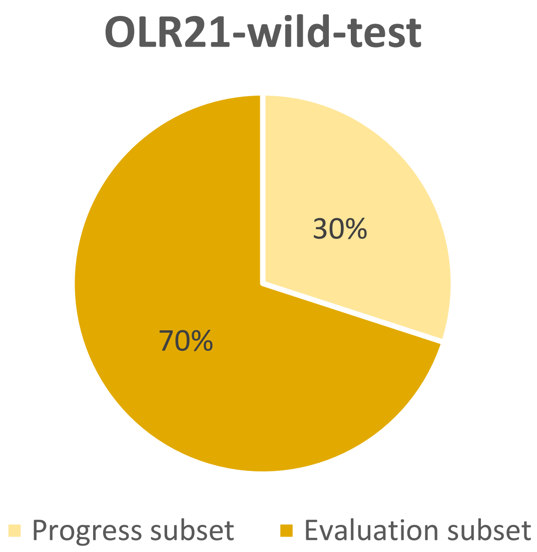
Male and Female speakers are balanced.
All data in the table except the last column is the number of utterances.
The number of total utterances might be slightly smaller than expected, due to the quality check.

Data: Official Data

Two standard test sets for the OLR 2021 challenge.



OLR21-cross-domain-test

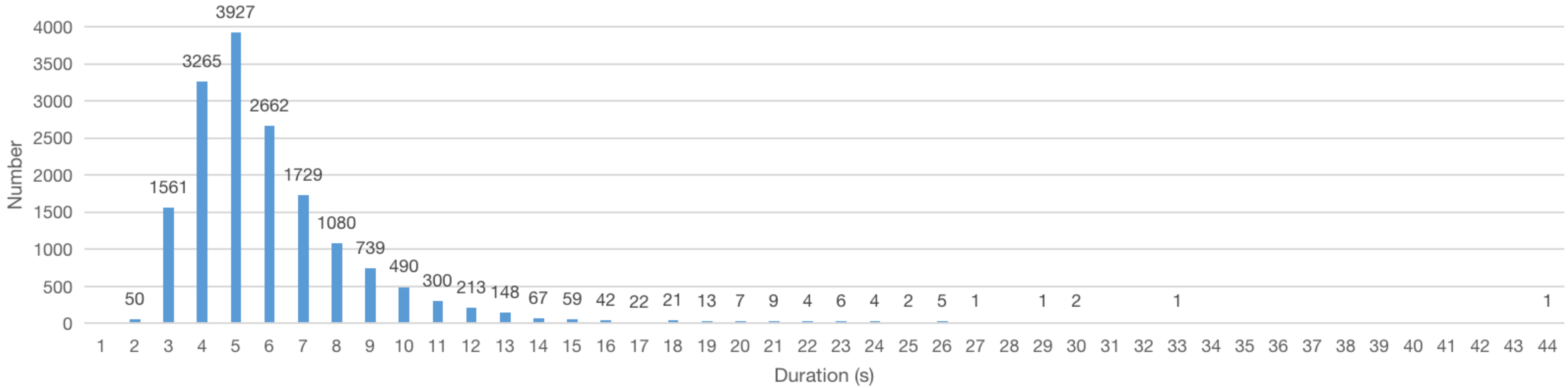


OLR21-wild-test

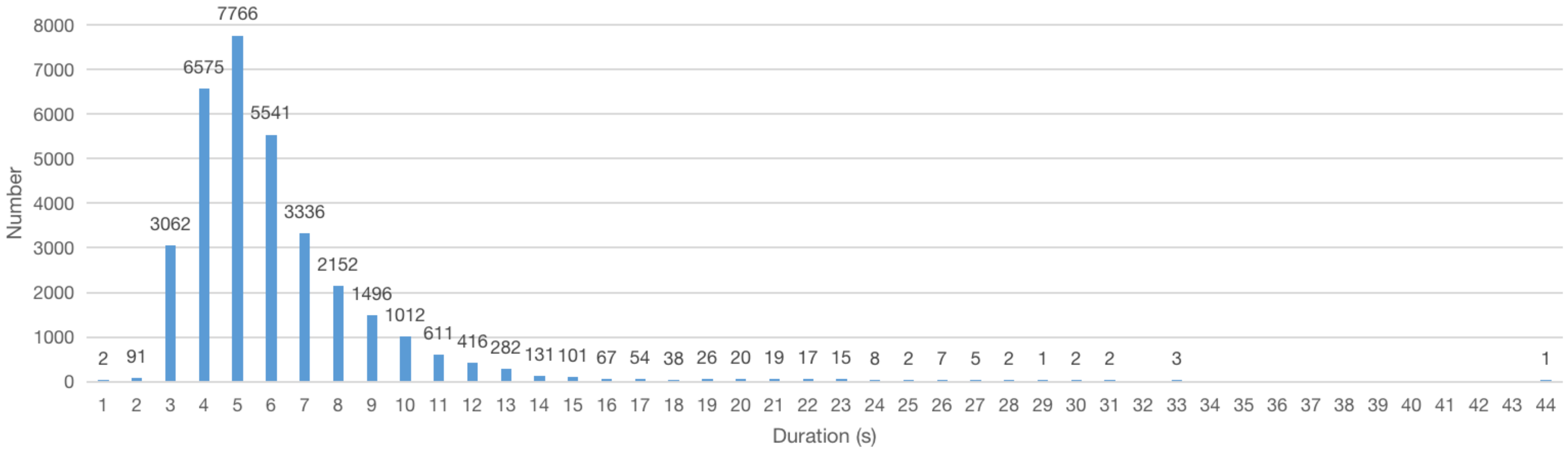
- Test set for Task1, Task3, Task4;
- Contains 13 languages;

- Test set for Task2;
- Contains 17 languages;

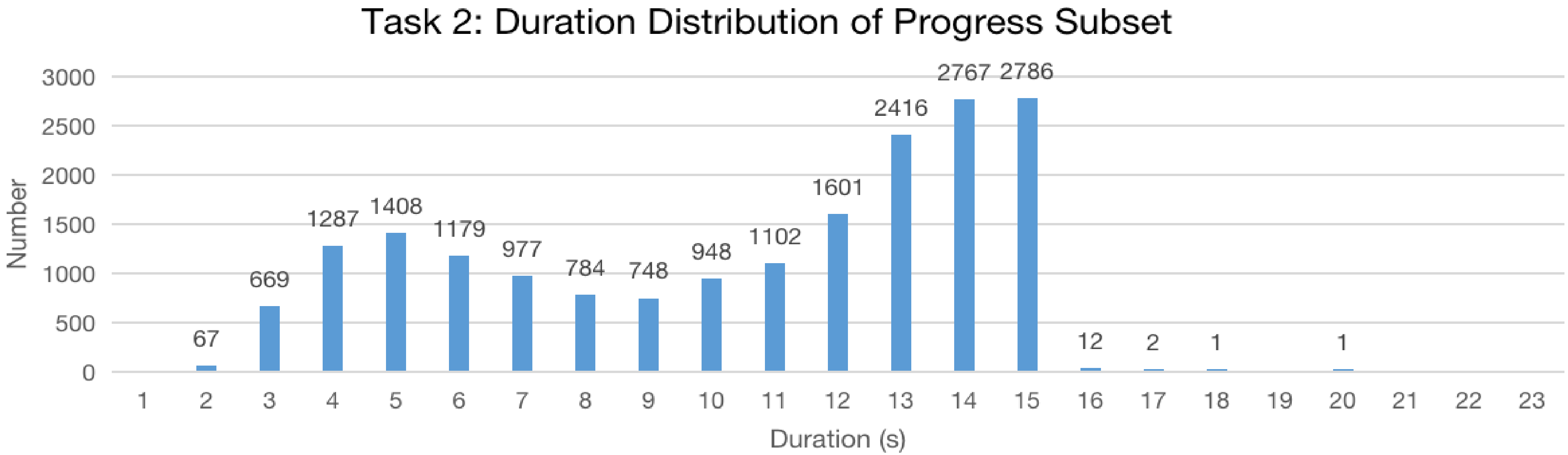
For the OLR 2021 Challenge, the trials of the four tasks are divided into two subsets respectively: a progress subset (30%), and an evaluation subset (70%).

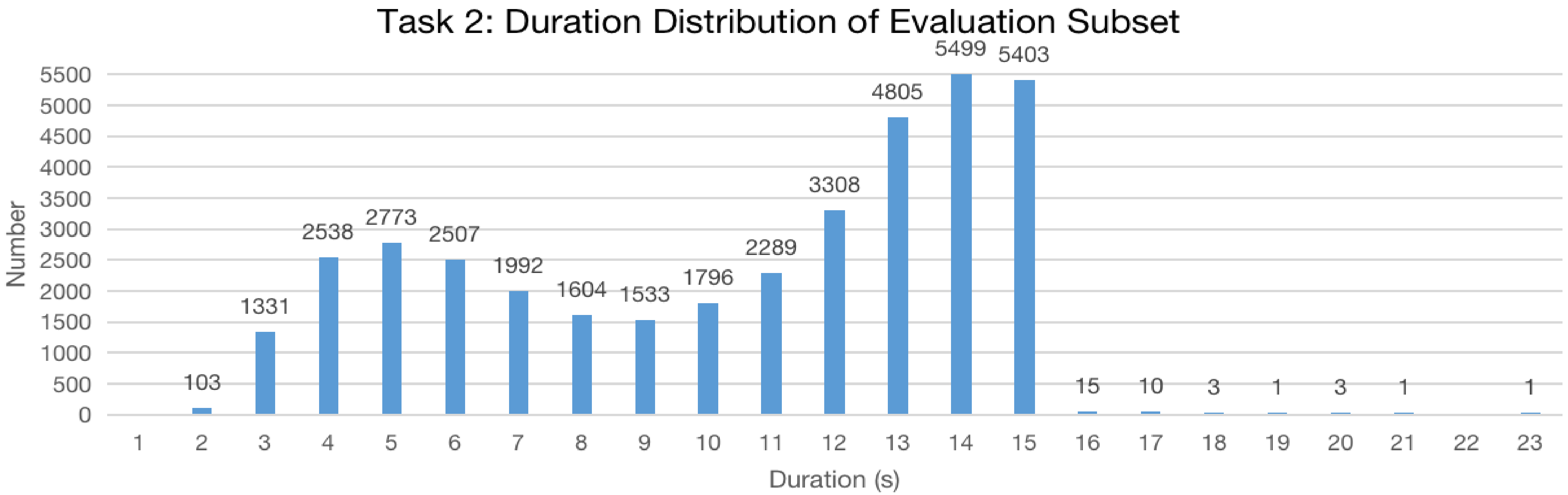


Task 1/3/4: Duration Distribution of Progress Subset



Task 1/3/4: Duration Distribution of Evaluation Subset


Task 2: Duration Distribution of Progress Subset
Number
3000
2500
2000
1500
1000
500
0
67
669
1287
1408
1179
977
784
748
948
1102
1601
2416
2767
2786
12
2
1
1
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
Duration (s)


Task 2: Duration Distribution of Evaluation Subset
Number
5500
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
103
1331
2538
2773
2507
1992
1604
1533
1796
2289
3308
4805
5499
5403
15
10
3
1
3
1
1
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
Duration (s)

Data: Additional Data (publicly available)

Data Name	Download Link
VoxLingua107	http://bark.phon.ioc.ee/voxlingua107/
OpenSLR 22/28/40/63/66/97/102/103	http://www.openslr.org/
Common Voice	https://commonvoice.mozilla.org/zh-CN/datasets
Librispeech	http://www.openslr.org/12/
WenetSpeech	https://wenet-e2e.github.io/WenetSpeech/
GigaSpeech	https://github.com/SpeechColab/GigaSpeech

Baseline System*

LID system

- An extended TDNN x-vector model, constructed with ASV-Subtools.

- The back-ends were conducted with Kaldi.

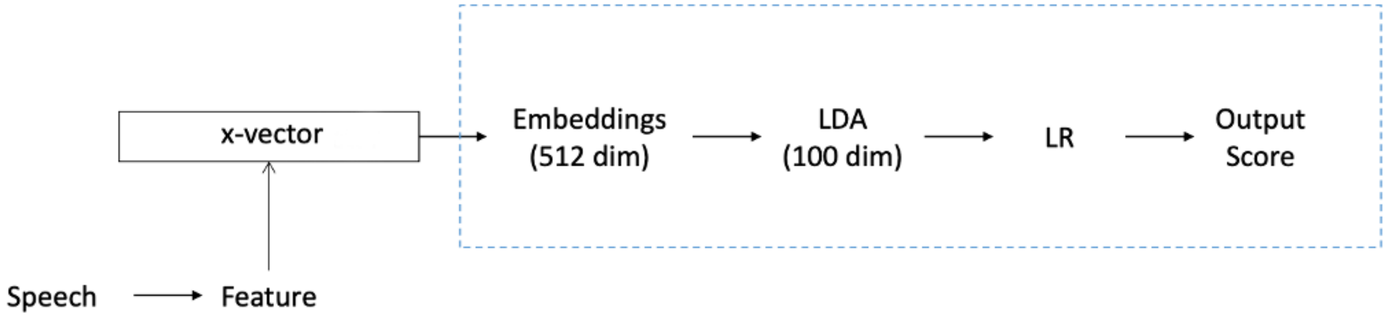
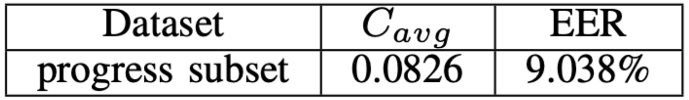


TABLE IV
C_{avg} AND EER RESULTS ON THE PROGRESS SUBSET



Dataset	C_{avg}	EER
progress subset	0.0826	9.038%

*https://github.com/Snowdar/asv-subtools#3-olr-challenge-2021-baseline-recipe-language-identification
B. Wang, W. Hu, J. Li, Y. Zhi, Z. Li, Q. Hong, L. Li, D. Wang, L. Song, and C. Yang, “Olr 2021 challenge: Datasets, rules and baselines,” arXiv preprint arXiv:2107.11113, 2021.

Baseline System

ASR system

- Built with ESPnet
- Transformer-based end-to-end model
- Language-independent architecture
- Characters based

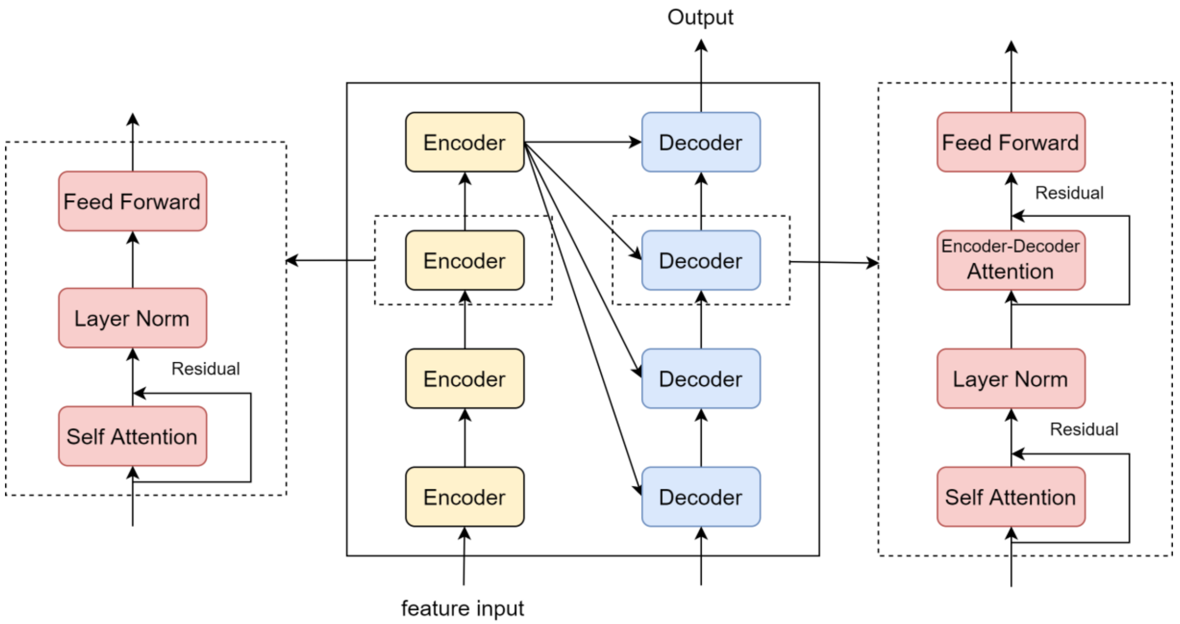
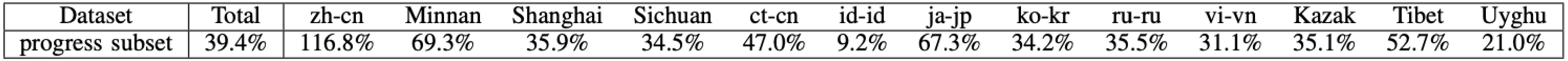


TABLE V
CER RESULTS ON THE PROGRESS SUBSET



Dataset	Total	zh-cn	Minnan	Shanghai	Sichuan	ct-cn	id-id	ja-jp	ko-kr	ru-ru	vi-vn	Kazak	Tibet	Uyghu
progress subset	39.4%	116.8%	69.3%	35.9%	34.5%	47.0%	9.2%	67.3%	34.2%	35.5%	31.1%	35.1%	52.7%	21.0%

Popular Technologies

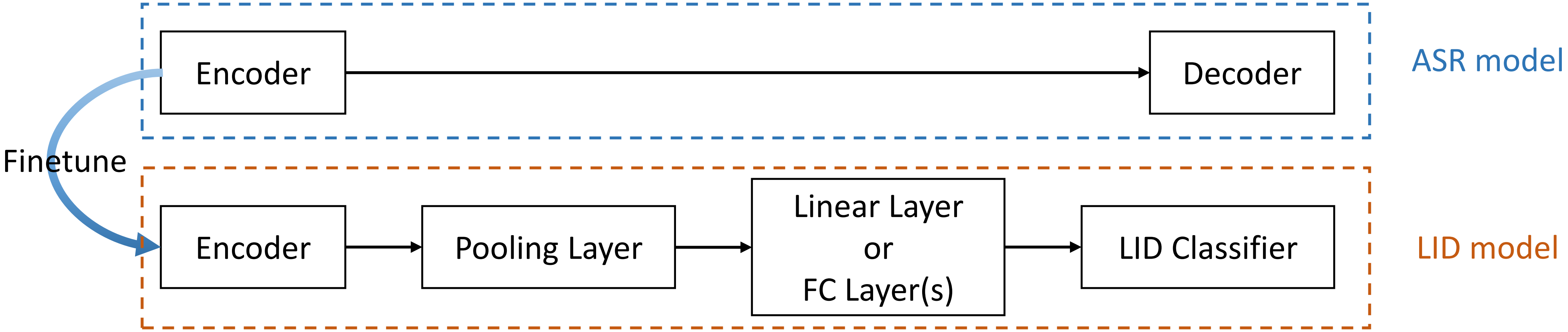
OLR-LID Tasks

Popular technologies

- **Feature**
  MFCC, FBank, PLP, Spectrum, **BNFs** ...
- **Augmentation**
  SpecAugment, speed/volume perturbations, noise from training data, white noise, gaussian noise, nonspeech, random artificial band-pass filters, **mp3/mp4a codec** ...
- **Structure/Optimization**
  VAD, no-VAD, TDNN, E-TDNN, F-TDNN, ECAPA-TDNN, ResNet, CNN, SE,
  ASR(conformer-transformer), Wav2vec2.0, E2E, multi-task,
  attention, attention-based fusion of features (PLP/MFCC),
  attentive pooling, multi-head attention pooling, global multi-head attention pooling ...
- **Loss**
  CE, AM, **AAM**, **KL** ...
- **Auxiliary task/multi-task**
  Phonetic aware, CTC ...
- **Scoring backend**
  Cosine, LDA, Logistic Regression (LR), PLDA ...
- **Model fusion**
  average fusion, greedy fusion …
- **Platform**
  Kaldi, PyTorch, ASV-Subtools(PyTorch), ESPnet, ESPNet2, Wenet, SpeechBrain ...

Technical highlights

- **ASR Structure**
  - Encoder-conformer + Decoder-transformer
  - Wav2vec2.0
- **Language Structure**
  - **Encoder-conformer + Pooling Layer + Classifier**
    - **1st:** ASR encoder+ attentive statistical pooling + a linear layer (batch normalization and nonlinear activation) + softmax classifier + LDA (language categories - 1)
    - **2nd:** ASR encoder + pooling layer (3) + 2 FC + classifier + no LDA
  - **Wav2vec-encoder + Statistics Pooling Layer + FC + Classifier**
    - Finetuned on different data set



Popular Technologies

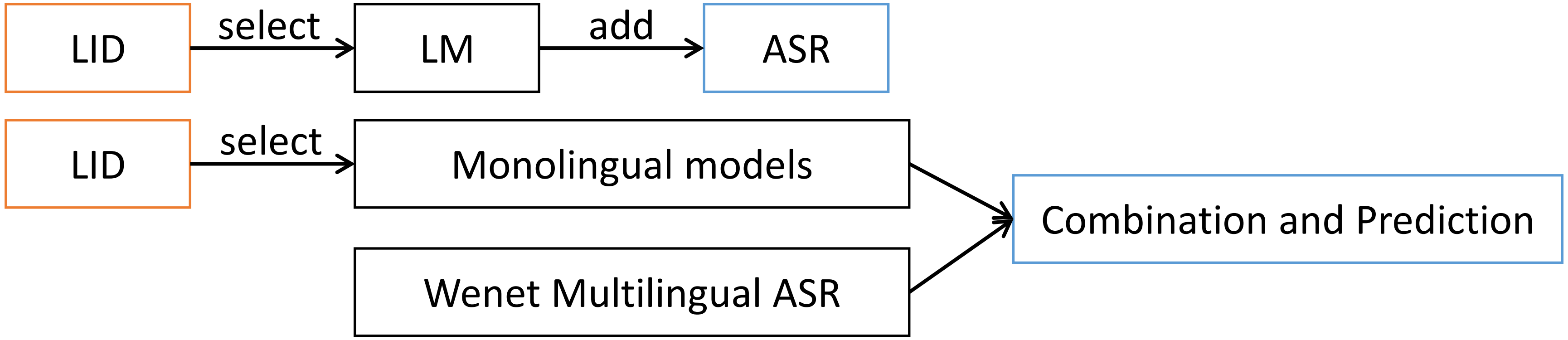
OLR-ASR Tasks

Popular technologies

- **Label**
  Language token was added to the beginning of the text transcripts, text transcripts
- **Feature**
  FBank + Pitch， MFCC + Pitch ...
- **Augmentation**
  SpecAugment, speed/volume perturbations, noise from other datasets or corpus, white noise, gaussian noise ...
- **Structure/Optimization**
  E2E Multilingual ASR, Hybrid Monolingual ASR
  E2E: Conformer(most) or Transformer, **Wav2vec2.0**,
  Multi-task ,
  Chain(LF-MMI) ...
- **Loss / Evaluate**
  CE, CTC, CE + CTC, Edit distance ...
- **LM(Language Model)**
  Add LM to E2E to improve performance...
- **Method**
  Model fusion: Use LID to identify language, then perform speech recognition,
  Pre-train and Finetune…
- **Platform**
  Wenet, ESPnet, ESPnet2, Kaldi, ...

Technical highlights

- **ASR Structure**
  - Hybrid Monolingual ASR
  - Wav2vec2.0
- **Method**
  - **Model fusion: LID + ASR**
    - Use LID model to identify language, then perform speech recognition system



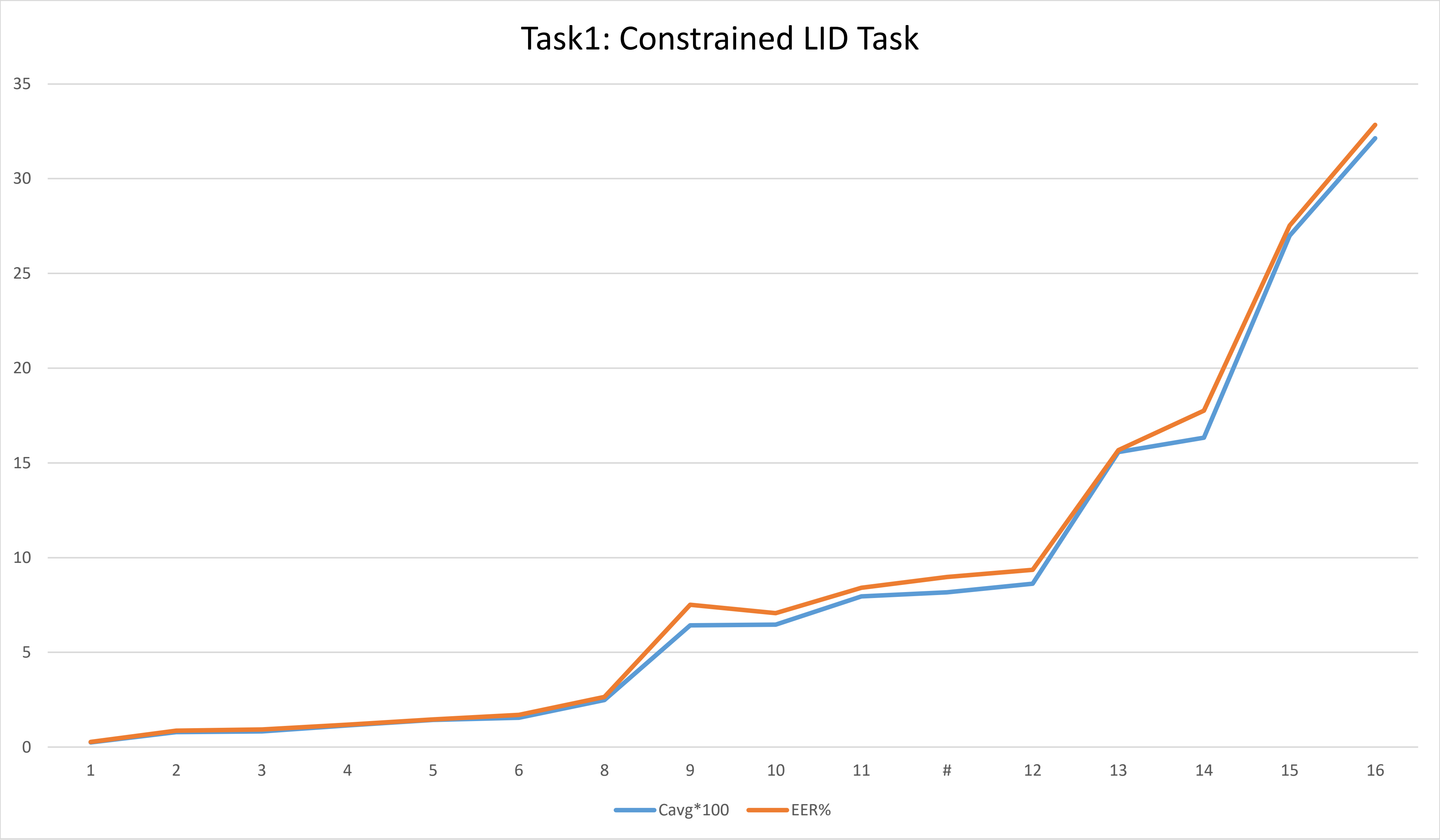
  - **Model fusion: Multi ASR**
    - Based on **ROVER**, implement a "voting" or re-scoring process
  - **Use pre-trained Wav2vec2.0 model**
    - XLSR-53
    - Finetuned on different data set

Challenge Results

Task1: Constrained LID Task

Ranking	Team Name	Institute	Country	Cavg	EER%
1	X-Voice	Machine Intelligence Department, Security BG, Ant Group	China	0.0025	0.2708
2	TalTech	Tallinn University of Technology	Estonia	0.0079	0.8642
3	funspeech	Beijing Live Data speech	China	0.0083	0.9311
4	Anonymous	-	China	0.0114	1.184
5	Huawei_AMS	Huawei Amsterdam Research Center	China	0.0144	1.461
6	nisp_speech	Yidun AI lab, Netease(Hangzhou) Network Co., Ltd. China.	China	0.0155	1.698
7	RoyalFlush	Hithink RoyalFlush AI Research Institute, Zhejiang	China	0.0209	2.55
8	OLR_BIT	Beijing Institute of Technology	China	0.0248	2.653
9	Anonymous	-	Canada	0.0643	7.513
10	Anonymous	-	China	0.0646	7.066
11	Wind_Talker	SpeakIn Technologies Co.,Ltd.	China	0.0795	8.405
#	Baseline	-	-	0.0817	8.977
12	XMU-Automation	Automation Department, School of Aeronautics and Astronautics, Xiamen University	China	0.0863	9.357
13	SpeechGroup@MANAS_Lab	Indian Institute of Technology Mandi, Himachal Pradesh, India	Indian	0.1557	15.67
14	Anonymous	-	China	0.1633	17.76
15	IITDH-IIITDH-Armsofttech-Speechgroup	IIT Dharwad, IIIT Dharwad and Armsoftech.air, India	Indian	0.2699	27.52
16	Anonymous	-	Turkey	0.3214	32.84

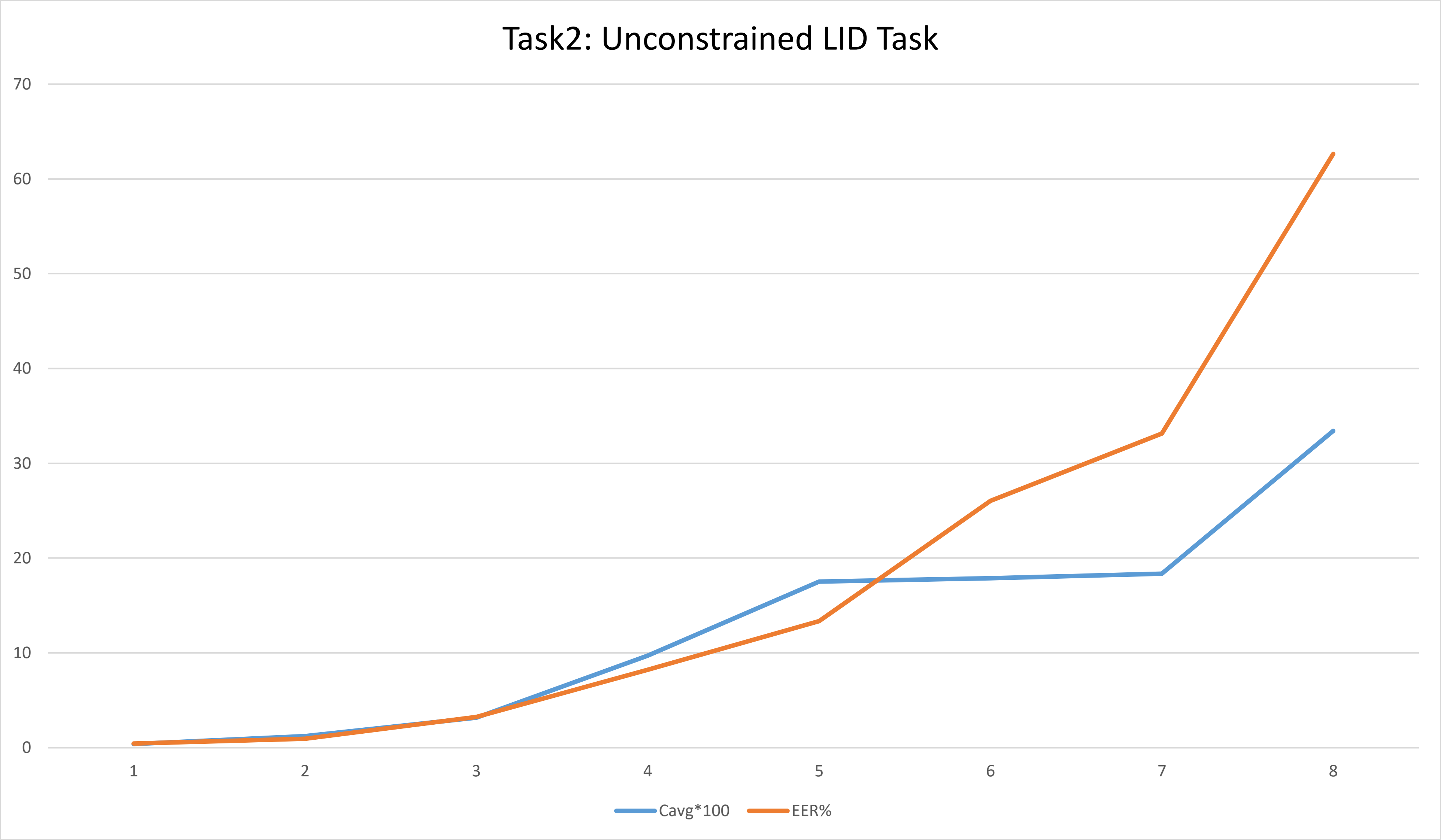
OLR 2021 Workshop



Task2: Unconstrained LID Task

Ranking	Team Name	Institute	Country	Cavg	EER%
1	X-Voice	Machine Intelligence Department, Security BG, Ant Group	China	0.0039	0.4212
2	TalTech	Tallinn University of Technology	Estonia	0.0122	0.9383
3	nisp_speech	Yidun AI lab, Netease(Hangzhou) Network Co., Ltd. China.	China	0.0316	3.228
4	funspeech	Beijing Live Data speech	China	0.097	8.229
5	Anonymous	-	China	0.1751	13.34
6	RoyalFlush	Hithink RoyalFlush AI Research Institute, Zhejiang	China	0.1788	26.05
7	XMU-Automation	Automation Department, School of Aeronautics and Astronautics, Xiamen University	China	0.1835	33.14
8	SUKI	The University of Helsinki	Finland	0.3342	62.64

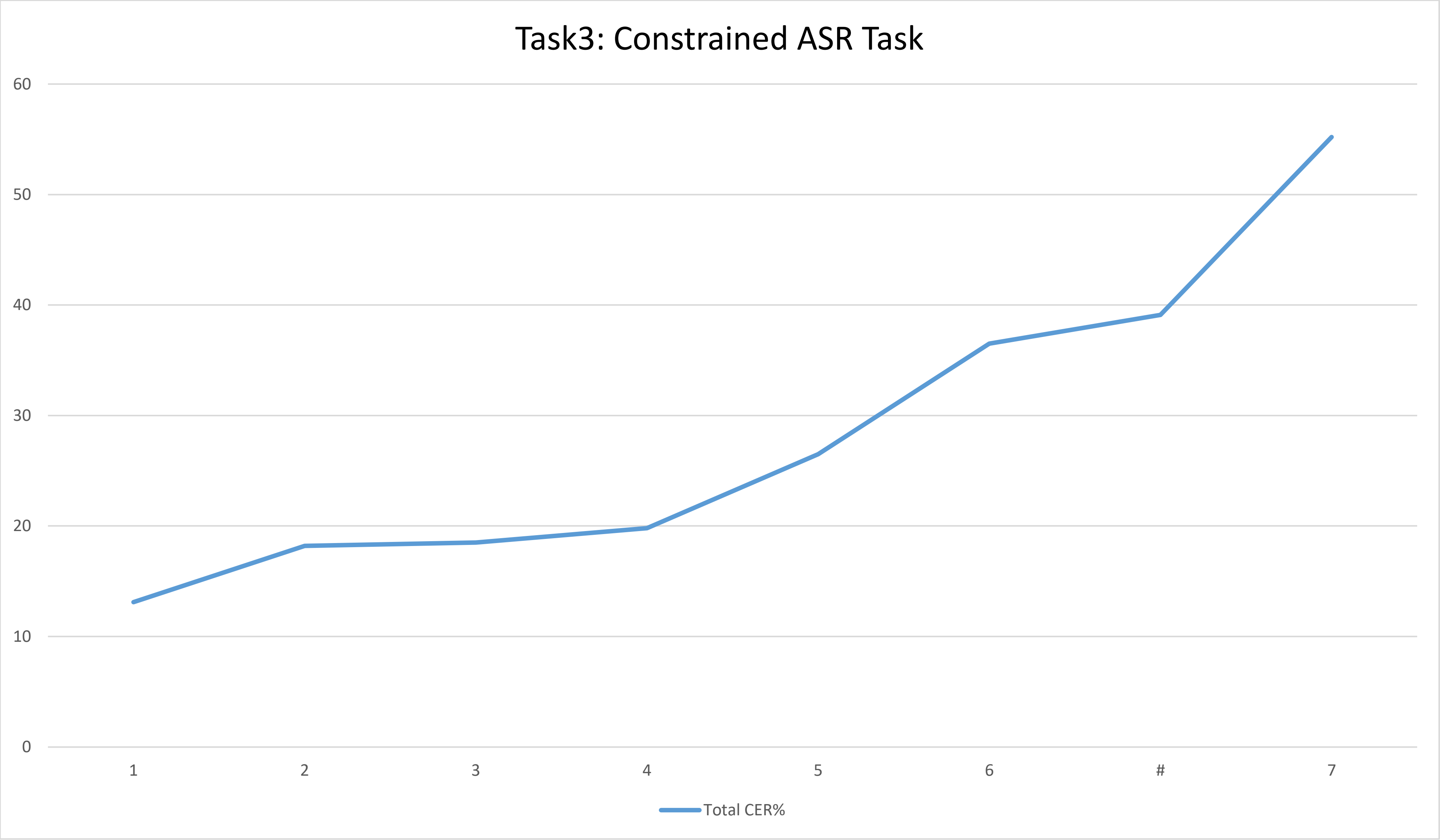
OLR 2021 Workshop



Task3: Constrained ASR Task

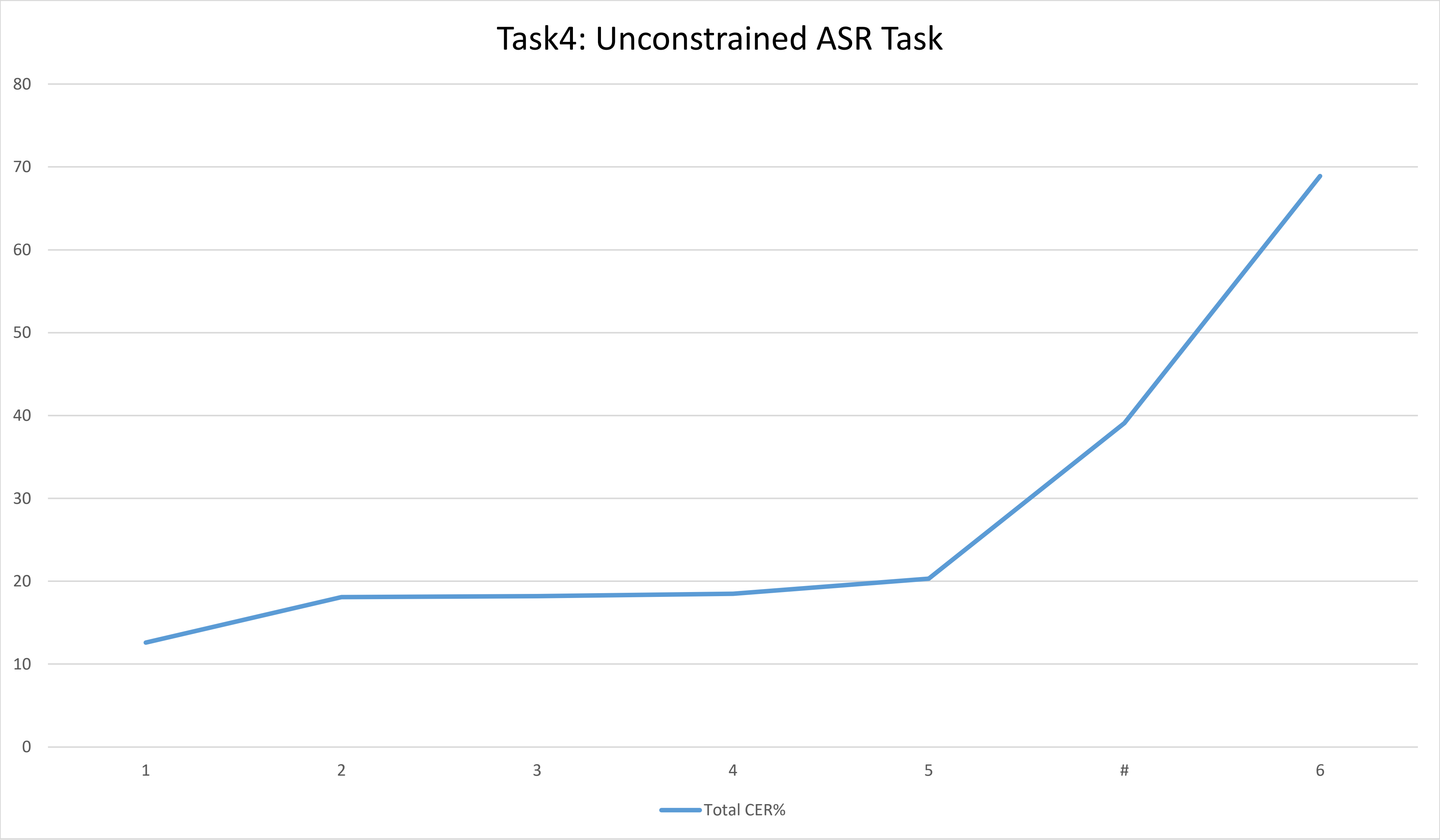
Ranking	Team Name	Institute	Country	Total CER%
1	CCDL	NetEase Games AI Lab	China	13.1
2	OLR_BIT	Beijing Institute of Technology	China	18.2
3	RoyalFlush	Hithink RoyalFlush AI Research Institute	China	18.5
4	Huawei_AMS	Huawei Amsterdam Research Center	China	19.8
5	nisp_speech	Yidun AI lab, Netease(Hangzhou) Network Co., Ltd.	China	26.5
6	funspeech	Beijing Live Data speech	China	36.5
#	baseline	-		39.1
7	BIIC	National Tsing Hua University	China	55.2

OLR 2021 Workshop



Task4: Unconstrained ASR Task

Ranking	Team Name	Institute	Country	Total CER%
1	CCDL	NetEase Games AI Lab	China	12.6
2	Huawei_AMS	Huawei Amsterdam Research Center	China	18.1
3	OLR_BIT	Beijing Institute of Technology	China	18.2
4	RoyalFlush	Hithink RoyalFlush AI Research Institute	China	18.5
5	nisp_speech	Yidun AI lab, Netease(Hangzhou) Network Co., Ltd.	China	20.3
#	baseline			39.1
6	IITDH-IIITDH-Armsofttech-Speechgroup	IIT Dharwad, IIIT Dharwad and Armsoftech.air	India	68.9


Task4: Unconstrained ASR Task
80
70
60
50
40
30
20
10
0
1
2
3
4
5
#
6
Total CER%

Summary

Summary

- This year’s challenges add new tasks of **wild LID** and **multilingual ASR**.
- The best systems of LID have achieved great improvements compared with the baseline systems, e.g. EER of constrained LID was reduced from 8.977% to 0.2708%. And wild LID also achieved very low EER of 0.4212%. **More evaluation of top systems in real-world applications is desired.**
- The best systems of ASR have achieved much improvements compared with the baseline systems, the CER was reduced from 39.1% to 13.1% in constrained ASR, and to 12.6% in unconstrained ASR.
- Most teams used LID model to identify language, and then perform speech recognition. Meanwhile, many teams used partial information on ASR systems to identify the categories of languages, especially adopted **ASR encoder** to boost the performance of LID.
- We can conclude that **LID and multilingual ASR complement each other**, and hope to see more studies in the future.

OLR 2022 Challenges

Looking forward to seeing you!