

Oriental Language Recognition (OLR) 2020 Challenge Summary

Qingyang Hong
Xiamen University

2020.12.18

Outline

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

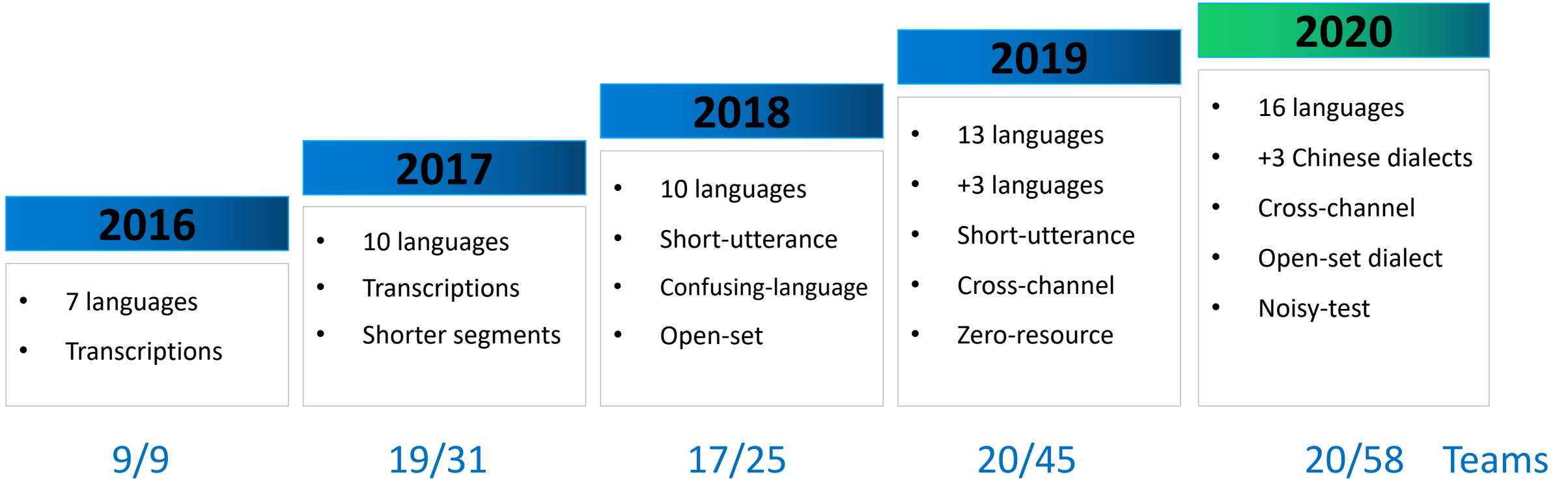
OLR 2020 Challenges

Organization Committee

Qingyang Hong, Xiamen University **Lin Li**, Xiamen University **Zheng Li**, Xiamen University
Dong Wang, Tsinghua University **Zhiyuan Tang**, Tsinghua University **Ming Li**, Duke-Kunshan University
Xiaolei Zhang, Northwestern Polytechnical Uni. **Liming Song**, SpeechOcean **Cheng Yang**, SpeechOcean



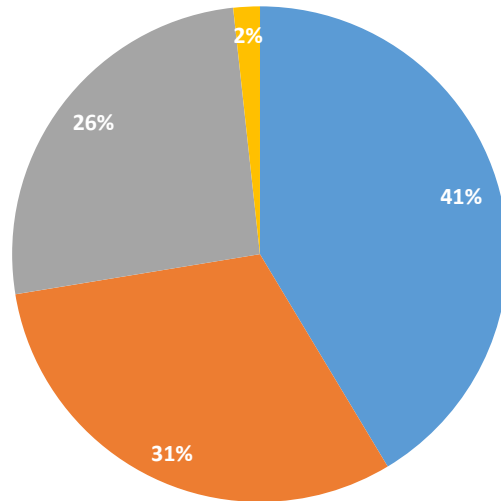
Special thanks to: **Jing Li, Yiming Zhi, Binling Wang** at XMU Speech Lab



Submitted results/Registered

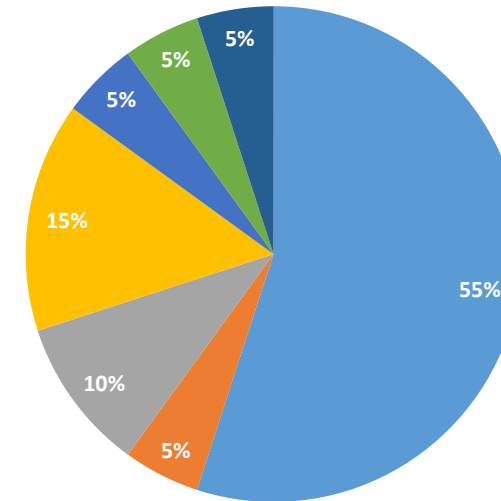
THE REGISTERED TEAMS

■ College ■ Industry ■ Overseas ■ Individual



THE COUNTRIES OF SUBMITTED TEAMS

■ China ■ Japan ■ India ■ Singapore ■ Czech Republic ■ France ■ Turkey



Tasks, Data and Baseline Systems

Task Definition:

Task 1: Cross-channel LID, which is a close-set identification task of six languages (**Cantonese, Indonesian, Japanese, Russian, Korean and Vietnamese**), but utterances were recorded with different channels.

Task 2: Open-set dialect identification, which is an open-set identification task, in which three nontarget languages (**Mandarin, Malay and Thai**) are added to the test set with the three target dialects (**Hokkien, Sichuanese, Shanghainese**).

Task 3: Noisy LID, where noisy test data of the 5 target languages (**Cantonese, Japanese, Russian, Korean and Mandarin**) are provided.

Training Data:

In OLR 2020, the permitted resources were several specified data sets, including AP16-OL7, AP17-OL3, AP17-OLR-test, AP18-OLR-test, **AP19-OLR-test**, **AP20-OLR-dialect**, and THCHS 30.

Those data sets included sixteen oriental languages, which were Mandarin, Cantonese, Indonesian, Japanese, Russian, Korean, Vietnamese, Kazakh, Tibetan, Uyghur, **Catalan**, **Greek**, **Telugu**, **Hokkien**, **Sichuanese**, **Shanghainese**.

Test Data: (AP20-OLR-test)

Task	Dataset	Language code	Language	Speakers	Male	Female	Duration (hours)	Utterances	Channel	Environment
1	King-ASR-454	CT-CN	Cantonese in China Mainland	6	3	3	3.11	2,394	Cross-channel	Quiet
	King-ASR-061	ID-ID	Indonesian in Indonesia	6	3	3	3.2	1,800		
	King-ASR-415	JA-JP	Japanese in Japan	6	3	3	2.21	2,254		
	King-ASR-183	RU-RU	Russian in Russia	6	3	3	3.87	1,800		
	King-ASR-174	KO-KR	Korean in Korea	6	3	3	2.3	1,800		
	King-ASR-074	VI-VN	Vietnamese in Vietnam	6	3	3	3.25	1,800		
2	King-ASR-406	Sichuan	Sichuanese in China	6	3	3	2.31	1,800	Mobile-channel	Quiet
	King-ASR-407	Shanghai	Shanghainese in China	6	3	3	2.26	1,800		
	King-ASR-427	Minnan	Hokkien in China	6	3	2	3.23	1,998		
	King-ASR-098	MS-MY	Malay in Malaysia	6	3	2	3.72	2,000		
	King-ASR-042	ZH-CN	Mandarin in China	6	3	3	2.57	1,800		
	King-ASR-062	TH-TH	Thai in Thailand	6	2	3	1.83	2,000	Cross-channel	
3	King-ASR-241	CT-CN	Cantonese in China Mainland	6	2	3	2.71	1,965	Mobile-channel	Noisy
	King-ASR-269	JA-JP	Japanese in Japan	6	3	3	1.99	1,862		
	King-ASR-124	RU-RU	Russian in Russia	6	7	6	3.61	1,944		
	King-ASR-137	KO-KR	Korean in Korea	6	7	6	2.26	1,925		
	King-ASR-042	ZH-CN	Mandarin in China	6	3	3	2.53	1,800		

Baseline Systems*:

Kaldi

1. i-vector
2. x-vector (E-TDNN)

Pytorch -- asv-subtools

1. x-vector (E-TDNN)

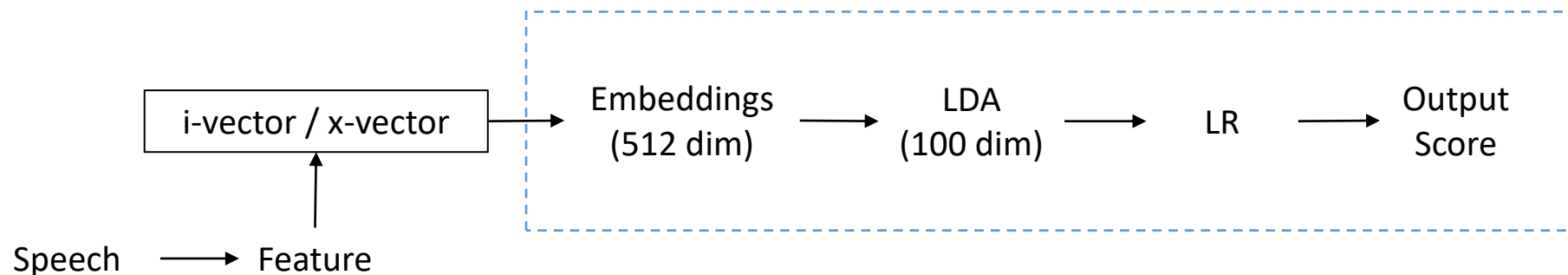


TABLE II
 C_{avg} AND EER RESULTS ON THE REFERENCED DEVELOPMENT SETS

Task	Cross-channel LID		Dialect Identification	
Enrollment Set	AP20-ref-dev-task1		AP20-OLR-dialect	
Test Set	AP19-OLR-channel		AP19-OLR-dev&eval-task3-test	
	C_{avg}	EER%	C_{avg}	EER%
[Kaldi] i-vector	0.2965	29.12	0.0703	9.33
[Kaldi] x-vector	0.3583	36.37	0.0807	14.67
[Pytorch] x-vector	0.2696	26.94	0.0849	12.40

TABLE III
 C_{avg} AND EER RESULTS ON THE AP20 EVALUATION SETS

Task	Cross-channel LID		Dialect Identification		Noisy LID	
Enrollment Set	AP20-ref-enroll-task1		AP20-OLR-dialect		AP20-ref-enroll-task3	
Test Set	AP20-OLR-channel-test		AP20-OLR-dialect-test		AP20-OLR-noisy-test	
	C_{avg}	EER%	C_{avg}	EER%	C_{avg}	EER%
[Kaldi] i-vector	0.1542	19.40	0.2214	23.94	0.0967	9.77
[Kaldi] x-vector	0.2098	22.49	0.2117	22.25	0.1079	11.12
[Pytorch] x-vector	0.1321	14.58	0.1752	19.74	0.0715	7.14

*<https://github.com/Snowdar/asv-subtools#2-ap-olr-challenge-2020-baseline-recipe-language-identification>

Zheng Li, Miao Zhao, Qingyang Hong, Lin Li, Zhiyuan Tang, Dong Wang, Liming Song and Cheng Yang, "AP20-OLR Challenge: Three Tasks and Their Baselines", APSIPA ASC 2020.

Popular Technologies

Popular technologies:

- **Feature**
Augmentation (speed, volume perturbations, noise from training data), SpecAugment, VAD, MFCC, FBank, PLP, Spectrum, Waveform with convolution, **no-VAD**, white noise, **random artificial band-pass filters, down/up sampling, pseudo-labeled** ...
- **Structure/Optimization**
TDNN, E-TDNN, F-TDNN, ResNet, CNN, **Squeeze-Excitation (SE)**, PTN, BLSTM, GRU, attention, attentive pooling, E2E, **GCNet, NetVLAD** ...
- **Loss**
CE, AM, AAM, **MMD, n-pair loss** ...
- **Auxiliary task/multi-task**
BNF, Phonetic aware, **CTC**, ...
- **Scoring backend**
Cosine, SVM, LDA, Logistic Regression (LR), PLDA, **GMM**...
- **Model fusion**
average fusion, greedy fusion, **FoCal toolkit** ...
- **Platform**
Kaldi, PyTorch, **ASV-Subtools(PyTorch), ESPnet**, TensorFlow, Matlab, **WebRTC** ...

***Bold** represents new technologies this year

Conclusion of technologies:

- **Feature**
 1. **Augmentation** is widely used.
 2. **FBank** and **MFCC** are two most frequently used acoustic features.
- **Structure**

TDNN and **ResNet** are two mostly used model structures.
- **Auxiliary information**
 1. The introduction of **ASR** to help language recognition is investigated by TOP teams (2/5 TOP teams used E2E ASR technologies).
 2. We believe the **phonetic information** based on BNFs or multi-task learning can be further investigated to improve the performance of LID.
- **Loss**

CE and **AM** softmax are two mostly used loss functions.
- **Scoring backend**

Most teams use LDA + **LR**; One team uses GMM to enhance the performance of same channel.
- **Model fusion**

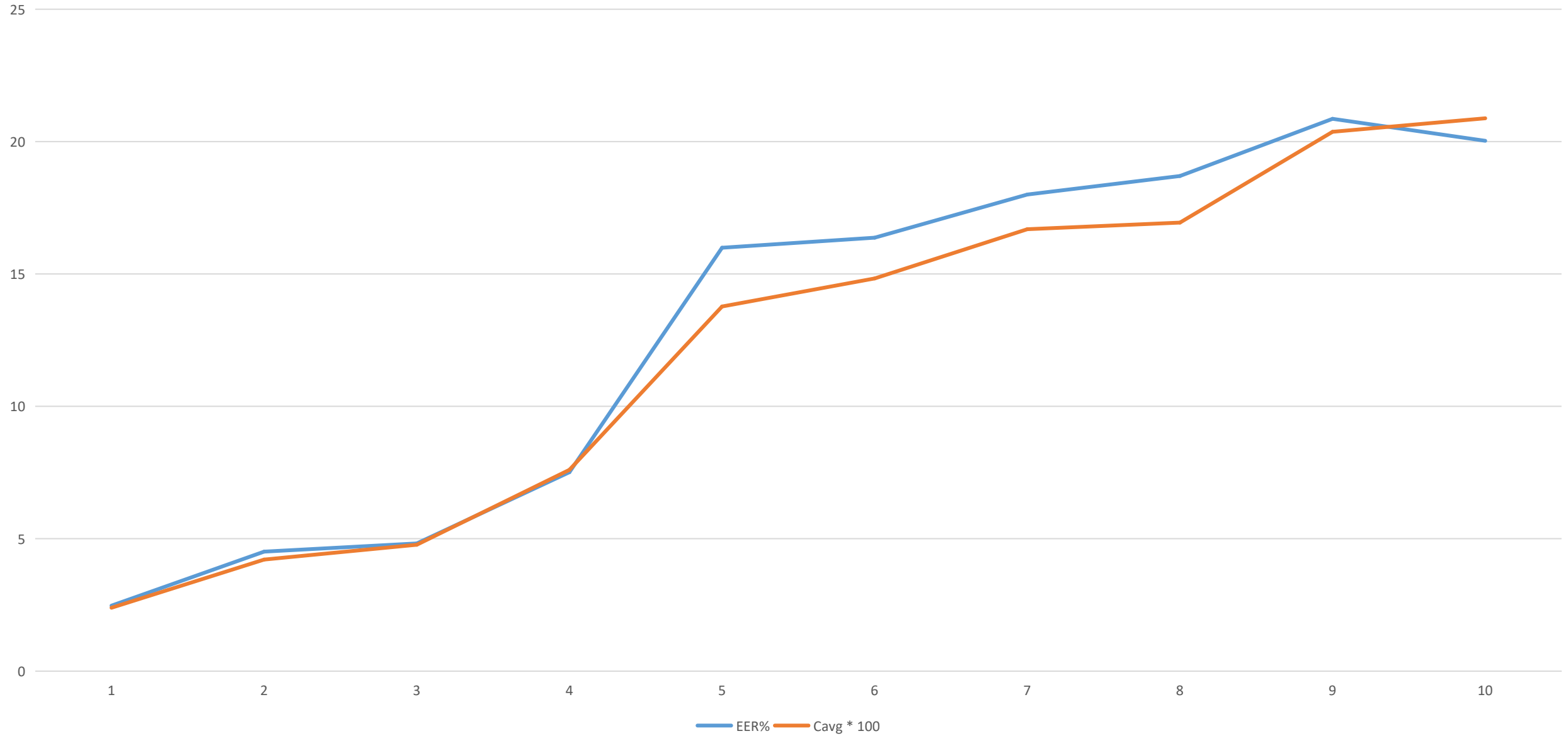
Most of submitted systems are the **score fusion** of many sub-systems.

Challenge Results

Task1: Cross-channel LID Top 10

Ranking	Team Name	Institute	Country	Participants	Cavg	EER%
1	LORIA-Inria-Multispeech	LORIA	France	Raphaël Duroselle, Irina Illina, Denis Jouvét, Md Sahidullah	0.0239	2.47
2	NTU-XJU	TL@NTU (Temasek Laboratories @ Nanyang Technological University)	Singapore	Zhang Haobo, Peng Yizhou, Xu Haihua, Pham Van Tung, Lim Zhi Hao, Chng Eng Siong	0.0421	4.51
3	Malaxiaolongxia	MI-SPEECH	China	杨朔、王欣、蔡国都	0.0477	4.82
4	IBG_AI	Tencent Technology(Shenzhen) Co., Ltd	China	柯川(Jayke), 周飞飞(Feifei Zhou)	0.0760	7.51
5	youdao	Youdao Limited	China	Haiwei Wu, Wenwei Dong, Yanqing Sun	0.1377	15.99
6	RoyalFlush	浙江核新同花顺网络信息股份有限公司	China	胡新辉, 王鼎, 叶帅帅, 喻陈毅, 刘小康	0.1483	16.37
7	gz	中科院模式识别实验室	China	耿旺, 张大威	0.1669	18.00
8	Phonexia	Phonexia s.r.o.	Czech Republic	Jan Profant, Josef Slavicek, Michal Klco, Ondrej Novotny	0.1694	18.70
9	BJFU	Institute of Artificial Intelligence, Beijing Forestry University	China	Ju Minghang, Li Zimu	0.2037	20.86
10	Anonymous	Anonymous	Anonymous	Anonymous	0.2088	20.03

Task1: Cross-channel LID
Top 10

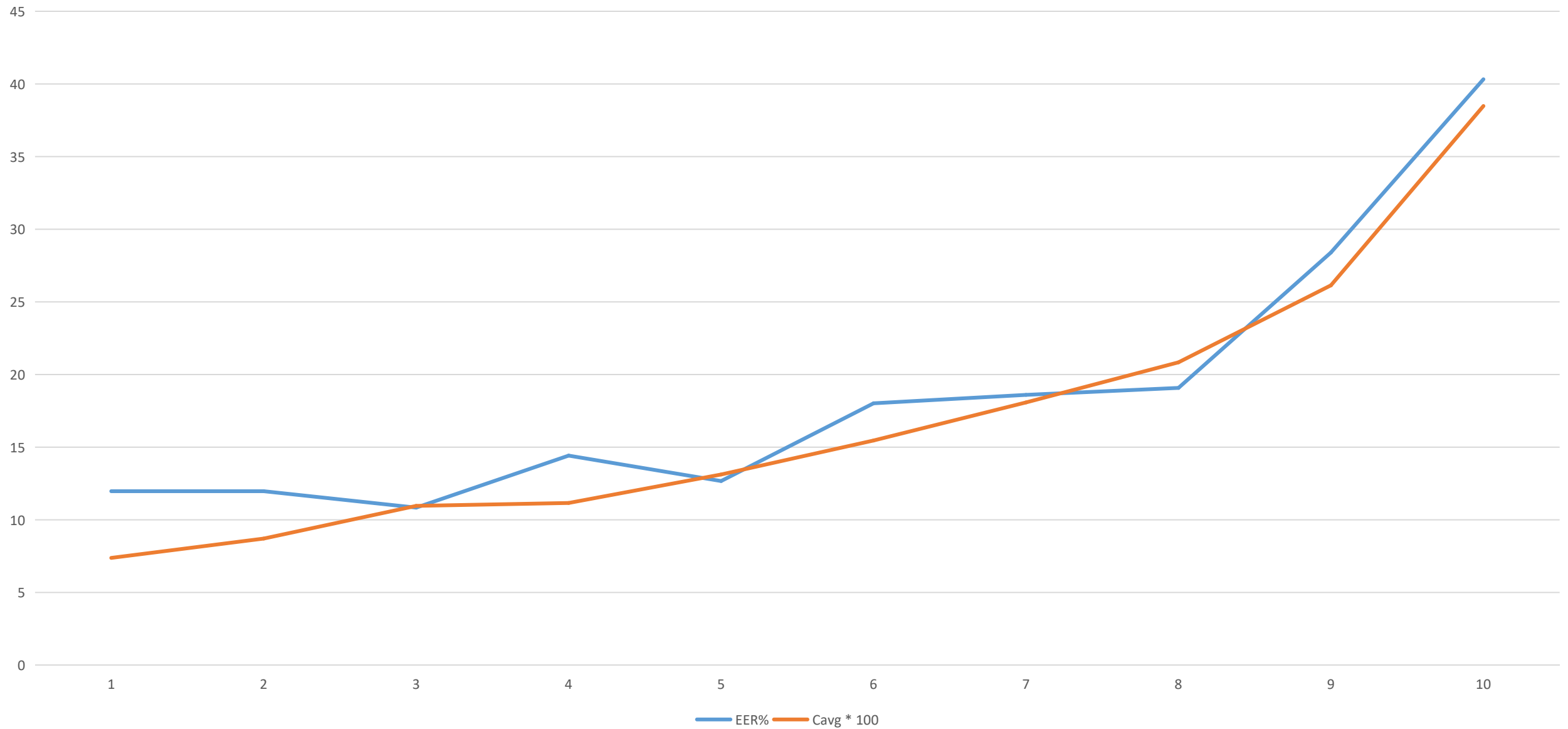


Task2: Open-set dialect identification
Top 10



Ranking	Team Name	Institute	Country	Participants	Cavg	EER%
1	Phonexia	Phonexia s.r.o.	Czech Republic	Jan Profant, Josef Slavicek, Michal Klco, Ondrej Novotny	0.0738	11.97
2	Royal-Flush	浙江核新同花顺网络信息股份有限公司	China	胡新辉, 王鼎, 叶帅帅, 喻陈毅, 刘小康	0.0871	11.97
3	IBG_AI	Tencent Technology(Shenzhen) Co., Ltd	China	柯川(Jayke), 周飞飞(Feifei Zhou)	0.1096	10.84
4	youdao	Youdao Limited	China	Haiwei Wu, Wenwei Dong, Yanqing Sun	0.1116	14.42
5	Anonymous	Anonymous	Anonymous	Anonymous	0.1312	12.67
6	NTU-XJU	TL@NTU (Temasek Laboratories @ Nanyang Technological University)	Singapore	Zhang Haobo, Peng Yizhou, Xu Haihua, Pham Van Tung, Lim Zhi Hao, Chng Eng Siong	0.1546	18.02
7	BJFU	Institute of Artificial Intelligence, Beijing Forestry University	China	Ju Minghang & Li Zimu	0.1808	18.60
8	gz	中科院模式识别实验室	China	耿旺, 张大威	0.2084	19.08
9	Anonymous	Anonymous	Anonymous	Anonymous	0.2614	28.40
10	ABSPlab_IIT_KGP	Indian Institute of Technology, Kharagpur	Indian	Dr. Goutam Saha, Spandan Dey	0.3848	40.32

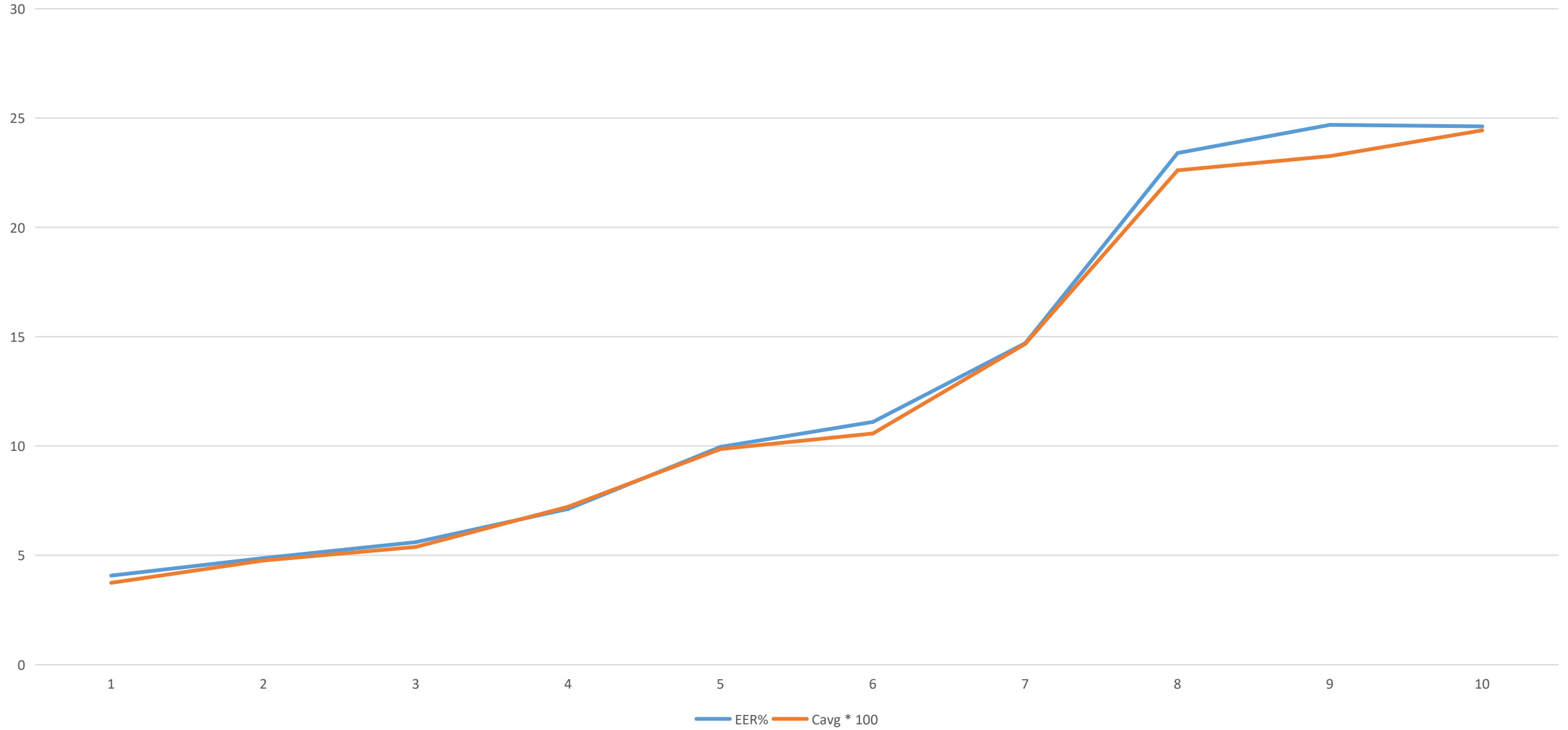
Task2: Open-set dialect identification
Top 10



Task3: Noisy LID Top 10

Ranking	Team Name	Institute	Country	Participants	Cavg	EER%
1	LORIA-Inria-Multispeech	LORIA	France	Raphaël Duroselle, Irina Illina, Denis Jouvét, Md Sahidullah	0.0374	4.07
2	NTU-XJU	TL@NTU (Temasek Laboratories @ Nanyang Technological University)	Singapore	Zhang Haobo, Peng Yizhou, Xu Haihua, Pham Van Tung, Lim Zhi Hao, Chng Eng Siong	0.0476	4.87
3	Malaxiaolongxia	MI-SPEECH	China	杨朔、王欣、蔡国都	0.0538	5.60
4	RoyalFlush	浙江核新同花顺网络信息股份有限公司	China	胡新辉, 王鼎, 叶帅帅, 喻陈毅, 刘小康	0.0722	7.12
5	Phonexia	Phonexia s.r.o.	Czech Republic	Jan Profant, Josef Slavicek, Michal Klco, Ondrej Novotny	0.0986	9.96
6	youdao	Youdao Limited	China	Haiwei Wu, Wenwei Dong, Yanqing Sun	0.1057	11.10
7	gz	中科院模式识别实验室	China	耿旺, 张大威	0.1467	14.70
8	BJFU	Institute of Artificial Intelligence, Beijing Forestry University	China	Ju Minghang & Li Zimu	0.2261	23.40
9	Anonymous	Anonymous	Anonymous	Anonymous	0.2326	24.69
10	Anonymous	Anonymous	Anonymous	Anonymous	0.2444	24.62

Task3: Noisy LID Top 10



Summary

- This year's challenges focus on three problems: cross-channel LID, open-set dialect identification, noisy LID.
- The best systems have achieved great improvements compared with the baseline systems, e.g. EER of cross-channel LID was reduced from 14.58% to 2.47%, EER of noisy LID was reduced from 7.14% to 4.07%.
- However, this kinds of systems are still not suitable for practical deployment.
- More studies should be conducted for the problems of LID. And more training data might be required to improve the robustness. We hope more research groups from around the world to attend the next challenge.

OLR 2021 Challenges

Looking forward to seeing you!