



Zero-resource Language Recognition

Jiawei Yu, Jinsong Zhang

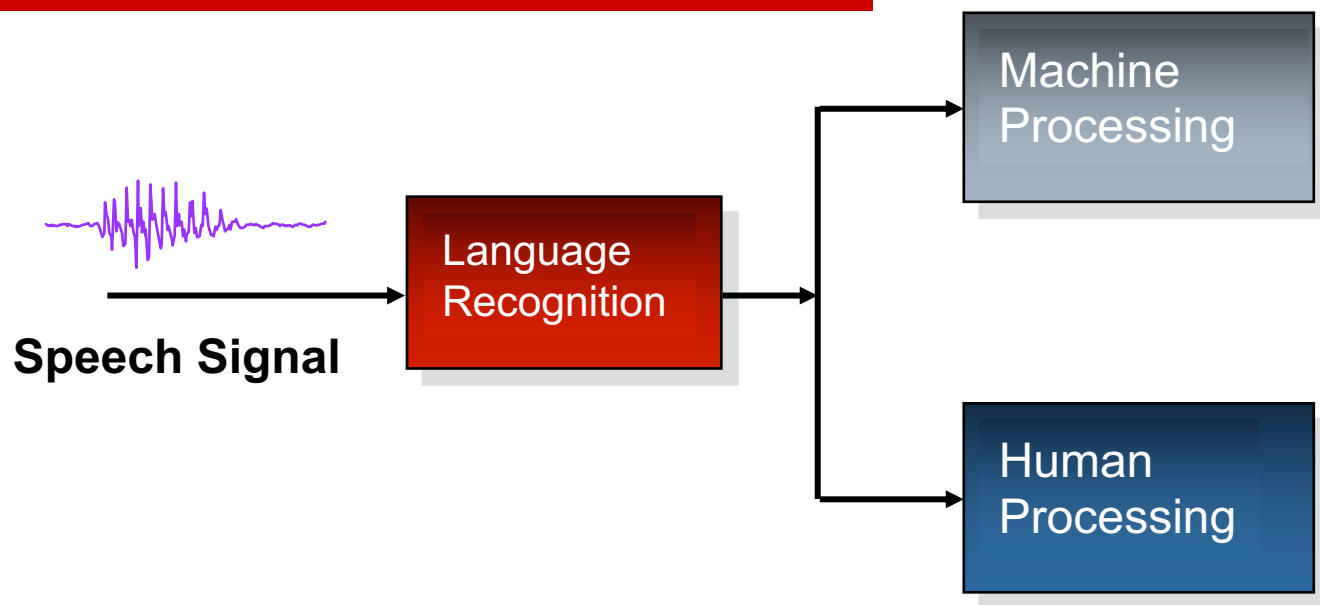
Speech Acquisition and Intelligent Technology Lab (SAIT)
Beijing Language and Culture University

Nov 21st 2019

Table of Contents

- Introduction
- Zero-resource LRE System
- Experimental Setup
- Results
- Conclusion

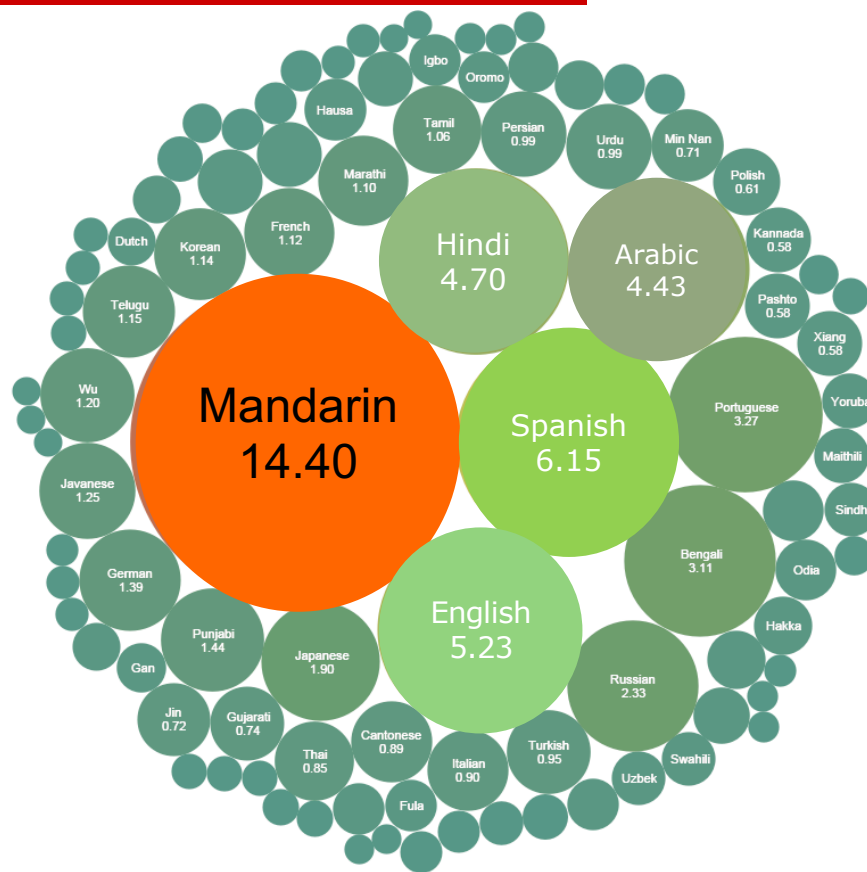
Introduction: Applications of LRE



- **multilingual speech processing applications**
 - spoken language translation
 - multilingual speech recognition

Introduction: The LRE Problems

- Exist r
langua
- 85% c
only 1
2010)

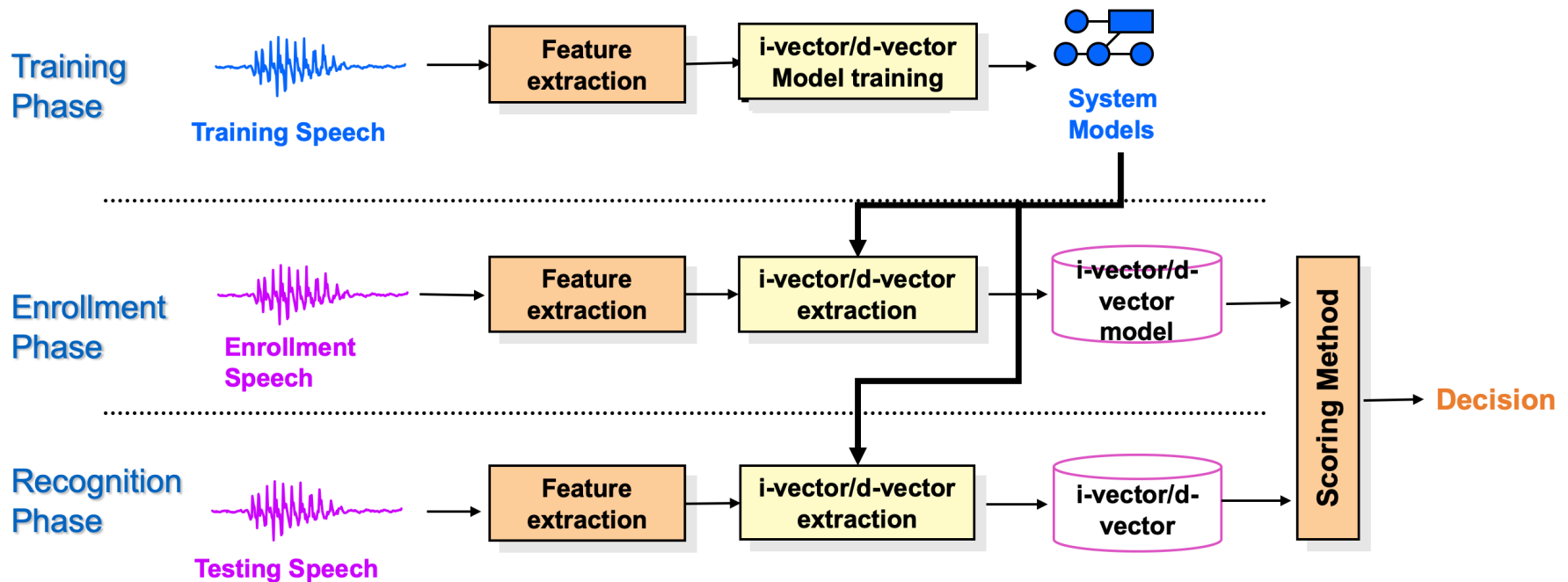


en
2011]
peak
es(until

Introduction: The LRE Problems

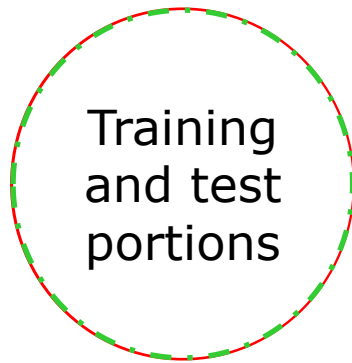
- ❑ The majority of the world's languages do not have enough speech resource.
- ❑ Need massive amounts of speech data.

Zero-resource LRE System



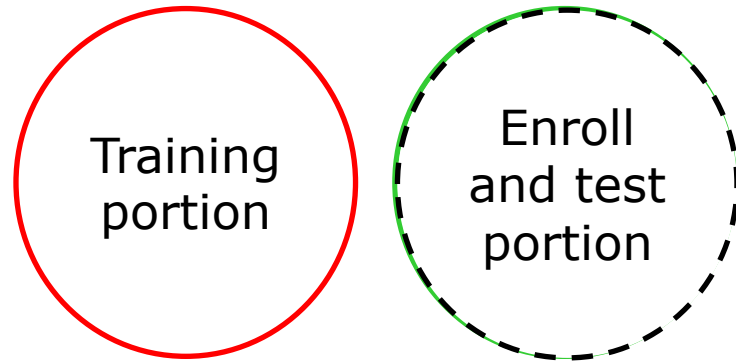
Difference between traditional LRE and zero-resource LRE

Overlap of languages



Traditional LRE system

Overlap of languages



Zero-resource LRE system

Experimental Setup

- Baseline: traditional LRE
- Zero-resource LRE
 - Different test utterance duration
 - Different enroll utterance duration
- Model
 - I-vector (400 dim)
 - D-vector(650 dim)

Data set

□ AP18-OLR(tang, 2018)

■ Baseline

	languages	utterance	hours
Training	10	93285/92285	193h/191h
test	10	1000/2000	2h/4h

■ Zero-resource

	languages	utterance	hours
Training	10	94285	196h
enroll	8	80/160	10min/20min
test	8	1184/1104	2.4h/2.3h

Results: Baseline

EER(%) RESULTS OF THE CLOSE-SET LRE SYSTEMS.

			EER%		
Total Numbers ^a	Systems	Scoring	test_1s ^b	test_3s ^b	test_all ^b
1000	I-vector	Cosine	13.90	4.50	2.10
		LDA	13.20	4.00	2.00
		PLDA	12.30	3.70	1.70
	D-vector	Cosine	7.70	6.20	6.10
		LDA	0.50	0.20	0.10
		PLDA	1.90	0.90	0.60
2000	I-vector	Cosine	13.71	4.00	2.05
		LDA	12.86	3.95	2.20
		PLDA	12.01	3.60	2.00
	D-vector	Cosine	8.20	7.10	7.05
		LDA	0.80	0.10	0.10
		PLDA	1.50	0.80	0.60

- D-vector based method better than i-vector based method

- LDA plays an important role for d-vector system

^a The Total Number represents the total number of utterances in the test set.

^b Test_1s, test_3s and test_all represent that the test sentence is 1 second, 3 second and full length (about 7 seconds).

Results: Different test duration

EER(%) RESULTS OF THE ZR-LRE SYSTEMS.

Enrollment ^a	Systems	Scoring	EER%		
			test_1s	test_3s	test_all
10	I-vector	Cosine	21.71	14.02	10.64
		LDA	26.60	19.76	16.47
		PLDA	34.46	29.73	26.60
	D-vector	Cosine	19.93	17.40	15.88
		LDA	20.27	15.37	13.85
		PLDA	28.89	25.59	22.97
20	I-vector	Cosine	18.03	10.69	8.70
		LDA	23.91	14.86	12.77
		PLDA	33.79	30.34	27.45
	D-vector	Cosine	19.66	17.57	16.76
		LDA	18.21	14.22	12.41
		PLDA	30.80	26.90	24.91

^a The Enrollment represents the number of utterances enrolled in each language.

- The more number of enrollment utterance, the more better system performance.

- The i-vector based method are better than the d-vector based method

Results: Different enroll duration

EER(%) RESULTS OF THE ZR-LRE SYSTEMS.

Enrollment ^a	Systems	Scoring	EER%		
			test_1s	test_3s	test_all
10	I-vector	Cosine	21.71	14.02	10.64
		LDA	26.60	19.76	16.47
		PLDA	34.46	29.73	26.60
	D-vector	Cosine	19.93	17.40	15.88
		LDA	20.27	15.37	13.85
		PLDA	28.89	25.59	22.97
20	I-vector	Cosine	18.03	10.69	8.70
		LDA	23.91	14.86	12.77
		PLDA	33.79	30.34	27.45
	D-vector	Cosine	19.66	17.57	16.76
		LDA	18.21	14.22	12.41
		PLDA	30.80	26.90	24.91

^a The Enrollment represents the number of utterances enrolled in each language.

EER(%) RESULTS ON THE DIFFERENT DURATION OF ENROLLMENT UTTERANCES ZR-LRE SYSTEMS.

Enrollment ^a	Systems	Scoring	EER%		
			enroll_1s ^b	enroll_3s ^b	enroll_all ^b
10	I-vector	Cosine	13.21	11.82	10.64
		LDA	23.60	19.76	16.47
		PLDA	30.46	26.73	26.60
	D-vector	Cosine	16.39	16.22	15.88
		LDA	16.81	15.21	13.85
		PLDA	23.90	23.48	22.97
20	I-vector	Cosine	12.21	9.59	8.70
		LDA	18.93	14.58	12.77
		PLDA	30.17	27.53	27.45
	D-vector	Cosine	16.85	16.76	16.76
		LDA	13.50	12.50	12.41
		PLDA	22.83	25.27	24.91

^a The Enrollment represents the number of utterances enrolled in each language.

^b Enroll_1s, Enroll_3s and Enroll_all represent that the enrollment utterances is 1 second, 3 second and full length (about 7 seconds).

- The i-vector based method are still better than the d-vector based method
- Enroll utterance duration has little effect compare to the condition of different test duration

Conclusion

- ☐ **ZR-LRE system can effectively recognize the “unseen” language of training set.**
- ☐ **I-vector method is better than d-vector method in ZR-LRE system when test utterance duration is long enough (more than 3s).**
- ☐ **Enroll utterance duration has little effect in ZR-LRE system when the number of enroll utterance is enough (more than 10 utterance)**

Thank you very much!

Questions?