

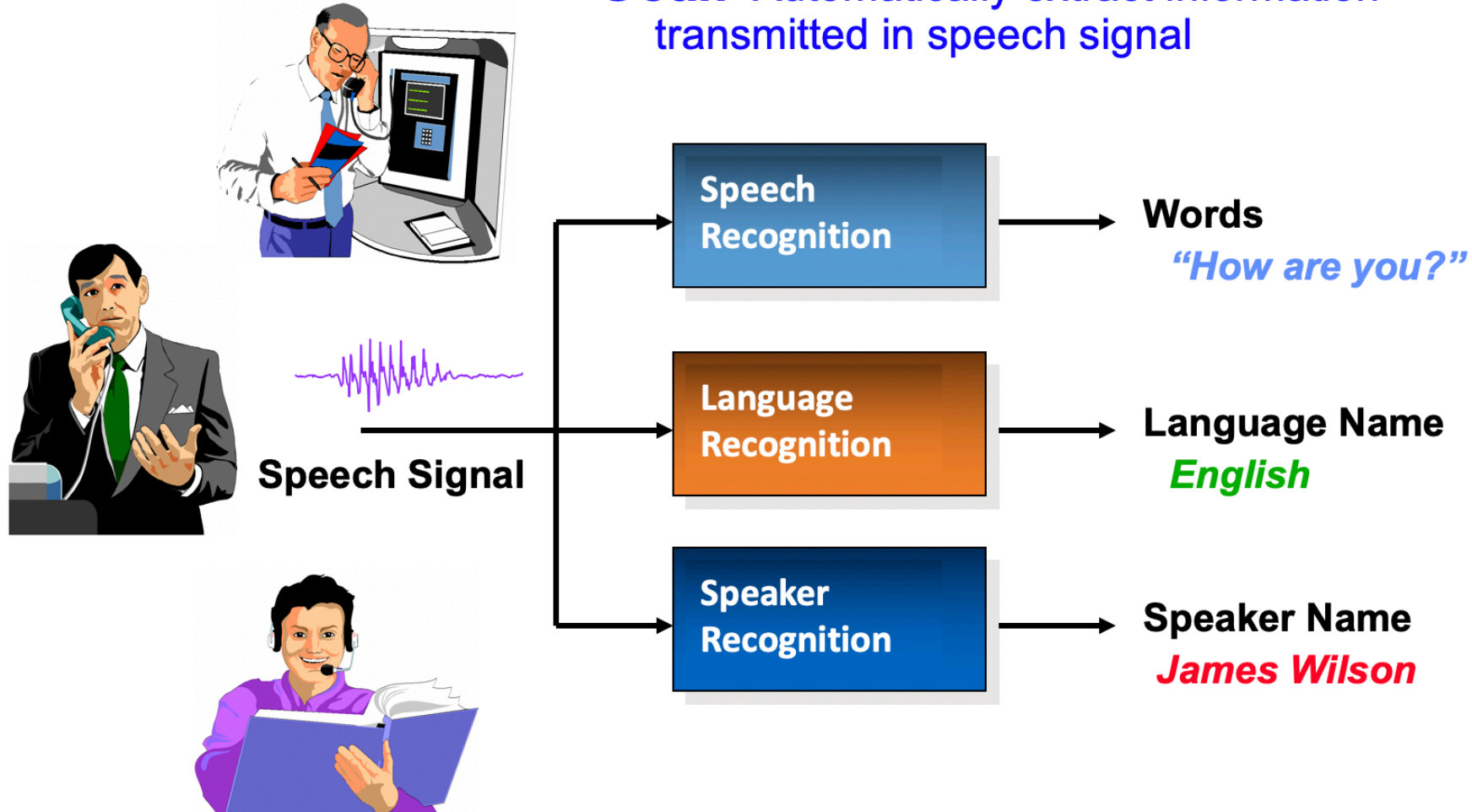
# Zero-resource Language verification and identification

2019/5/29

于嘉威

# 语音中包含很多信息

**Goal:** Automatically extract information transmitted in speech signal



# 什么是自动语种识别？

- 通过计算机识别给定语音是什么语言

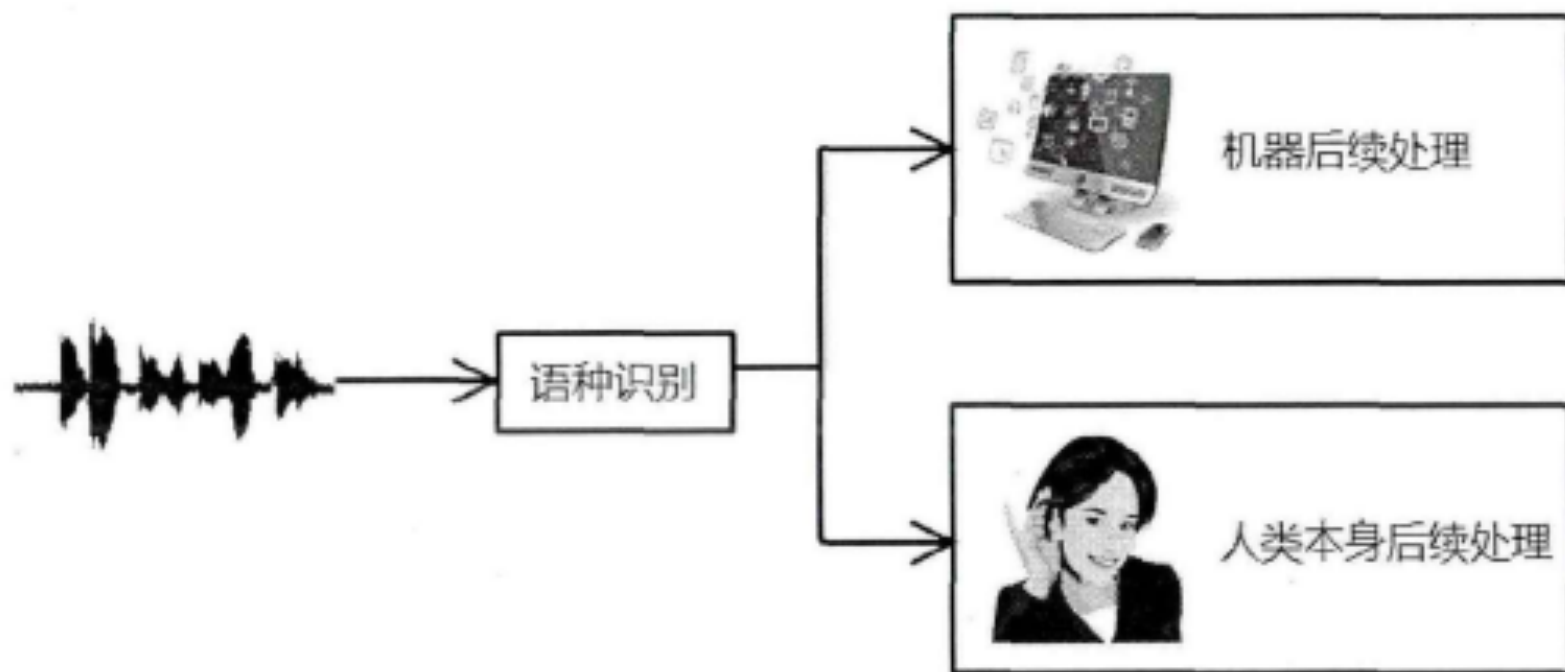
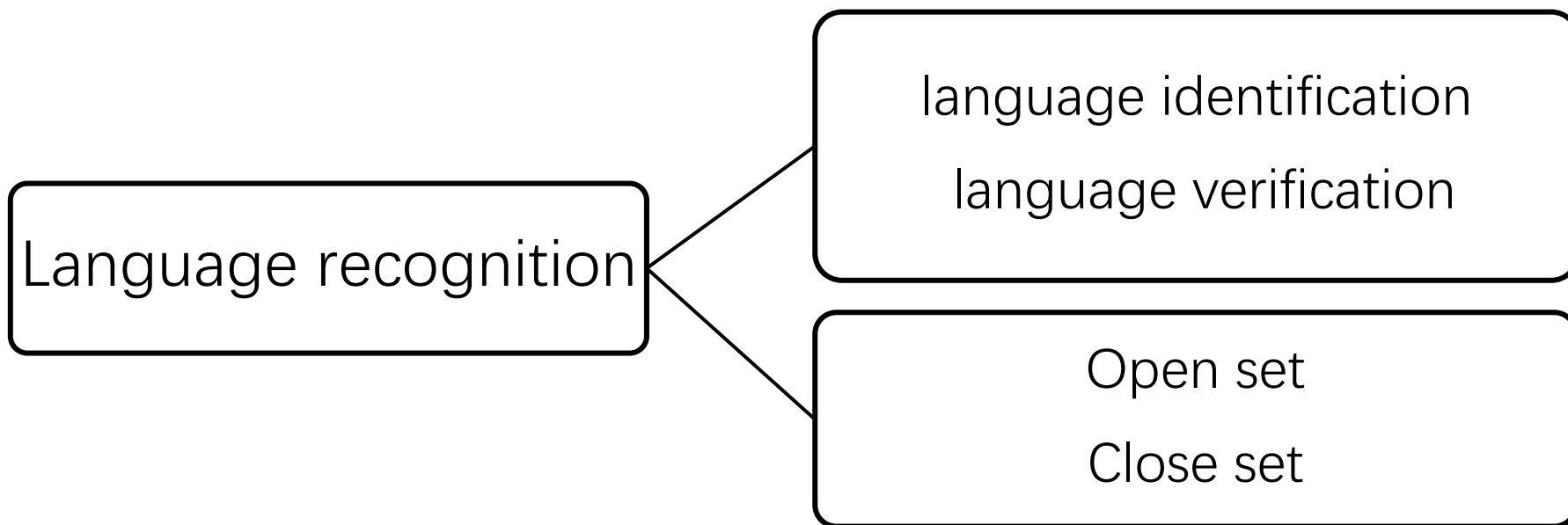


图 1.1: 语种识别应用分类示意图

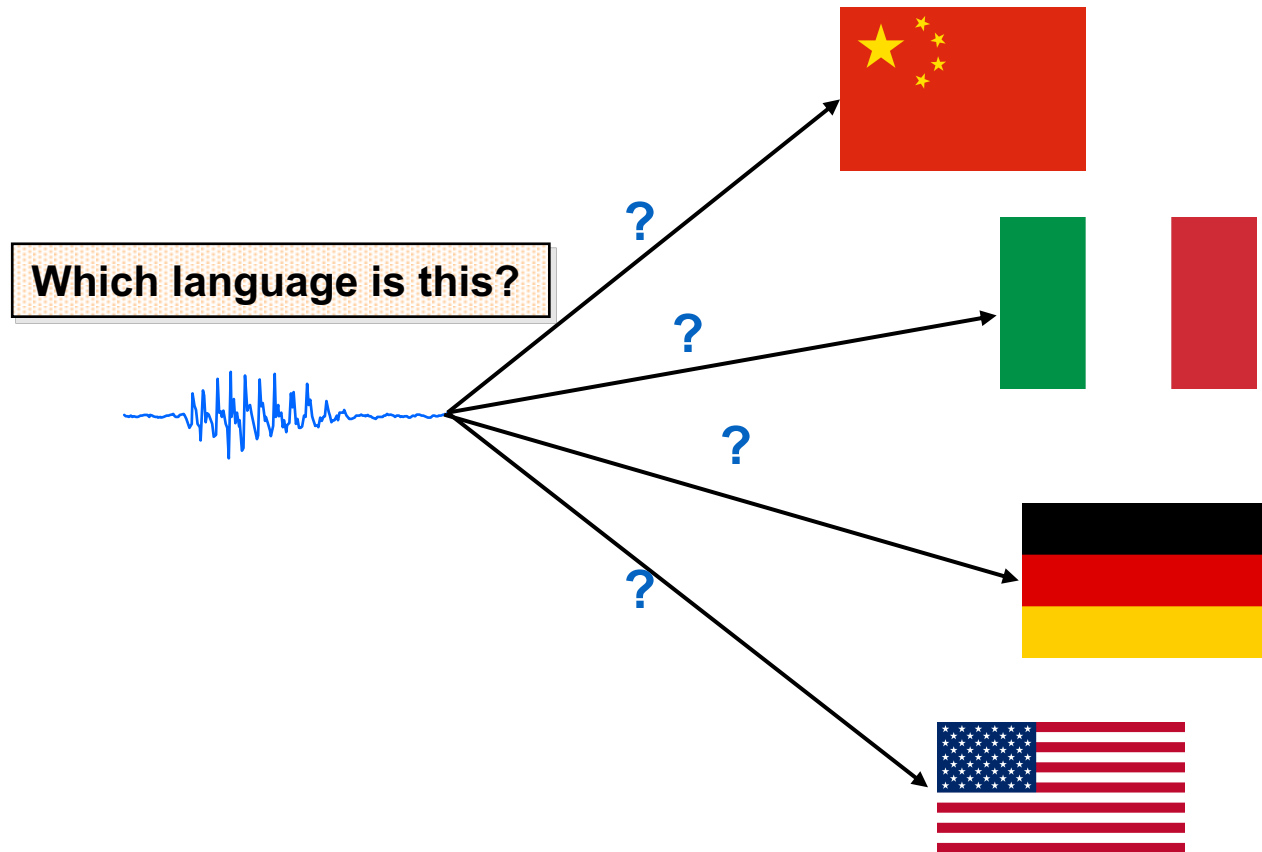
- 多语言环境下的自动语音识别
- 自动转接
- 国家安全

# 分类

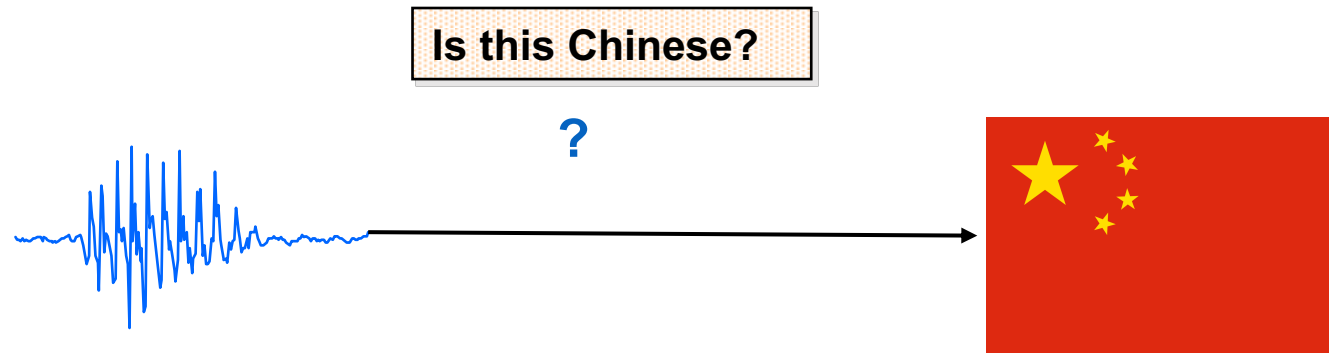
- 基于文本的语种识别
- 基于语音的语种识别



# Language identification



# Language verification

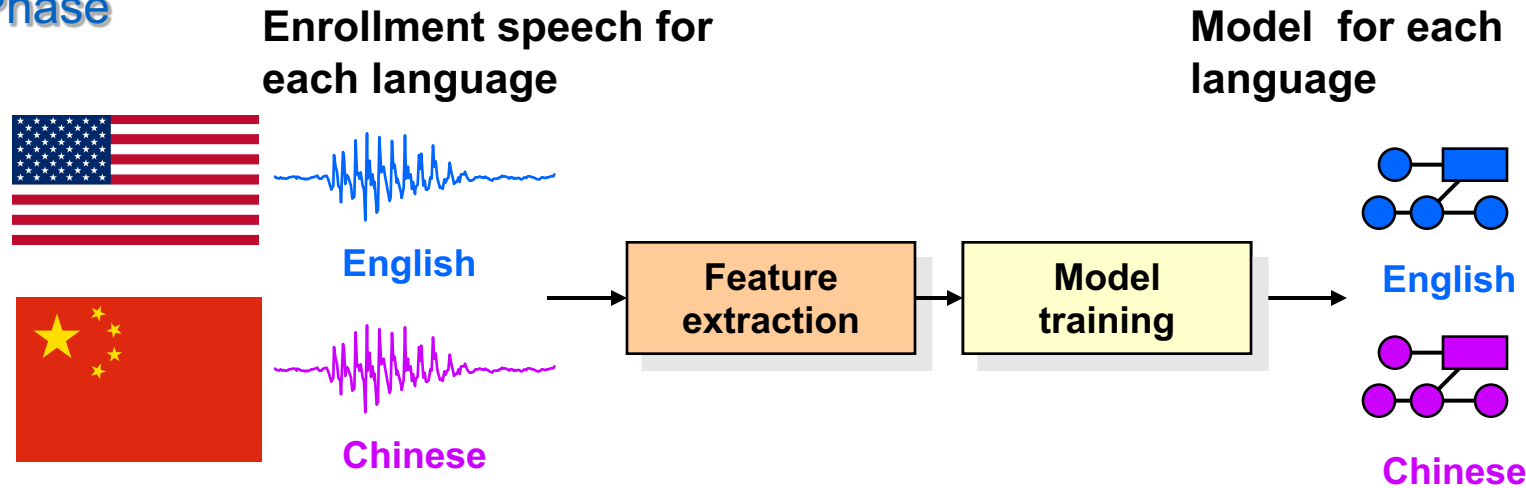


# Open set v.s. zero-resource

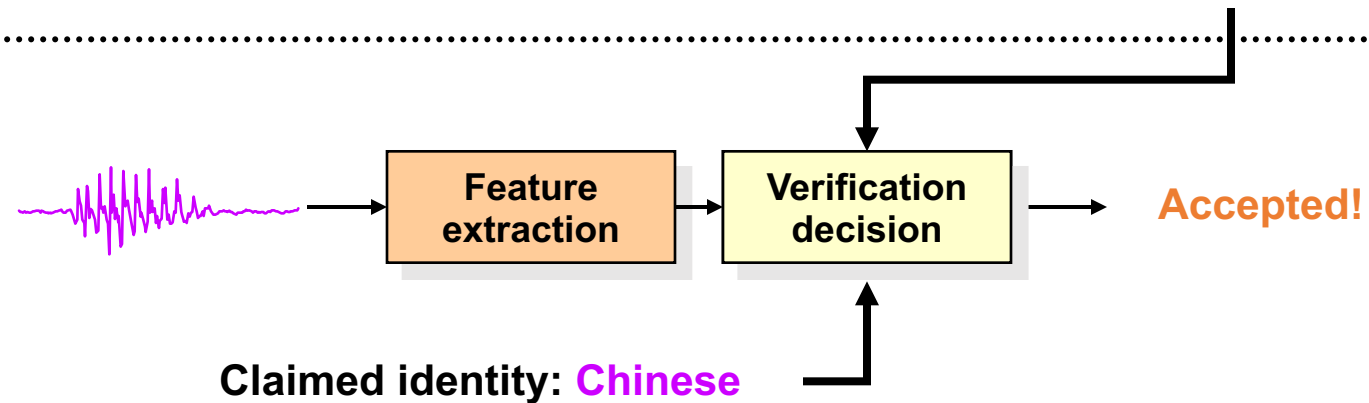
- Open set将训练集之外的数据都标识为unknown，识别目标还是以训练集中的语种为主。
- Zero-resource 则是类似于说话人识别的过程，训练集之外的数据都有自己所属的国别标签，训练目标是识别这些集外的语种。

# zero-resource language verification

## Enrollment Phase



## Verification Phase





# Why we do this ?

- 对于一些低资源的语种，如果有这样一个语种识别系统，我们便可以在有丰富资源的语言上训练好一个语种识别模型，然后让用户注册几句话，以后我们便可以识别这一低资源的语言。

# Experimental set-up

- Data set: AP18-OLR
  - training set: 94285 utterances, 10 languages
  - enroll and test set: 8 languages different from the training set, total 1264 utterances (enroll 80 and 160)
  - Baseline test set: 10 languages, 1000 utterances and 2000 utterances
- Model
  - I-vector
  - D-vector

# Results

- Baseline EER

when test set have 1000 utterances,

dvector	test_1s	test_3s	test_full	ivector	test_1s	test_3s	test_full
cosine	6.90	6.40	6.10		13.90	4.50	2.10
lda	0.50	0.20	0.10		13.20	4.00	2.00
plda	2.50	0.90	0.60		12.30	3.70	1.70

when test set have 2000 utterances,

dvector	test_1s	test_3s	test_full	ivector	test_1s	test_3s	test_full
cosine	8.20	7.10	7.05		13.71	4.00	2.05
lda	0.80	0.10	0.10		12.86	3.95	2.20
plda	1.50	0.80	0.60		12.01	3.60	2.00

# Results

- Zero-resource

Dvector and ivector EERs when enroll have 10 utterances per language (total number of utterances in enroll set is  $8 \times 10 = 80$  )

dvector	test_1s	test_3s	test_full		ivector	test_1s	test_3s	test_full
cosine	19.93	17.40	15.88		21.71	14.02	10.64	
lda	20.27	15.37	13.85		26.60	19.76	16.47	
plda	28.89	25.59	22.97		34.46	29.73	26.60	

Dvector and ivector EERs when enroll have 20 utterances per language (total number of utterances in enroll set is  $8 \times 20 = 160$  )

dvector	test_1s	test_3s	test_full		ivector	test_1s	test_3s	test_full
cosine	19.66	17.57	16.76		18.03	10.69	8.70	
lda	18.21	14.22	12.41		23.91	14.86	12.77	
plda	30.80	26.90	24.91		33.79	30.34	27.45	