

Retraining-Free Speech Recognition For Code-Switching

Zhen Huang, Xiaodan Zhuang, Daben Liu, Xiaoqiang Xiao, Yuchen Zhang, Sabato Marco Siniscalchi
Apple Inc.

Code-Switching

Code Switching (CS) refers to the phenomenon of switching languages within a sentence or between sentences. It represents a challenge for ASR systems deployed to support a single target language.

Typical CS scenarios:

- 下午3:00帮我schedule一个meeting.
Help me schedule a meeting at 3:00PM.
- 美国的Costco超市好不好?
Is the U.S. Costco supermarket any good?

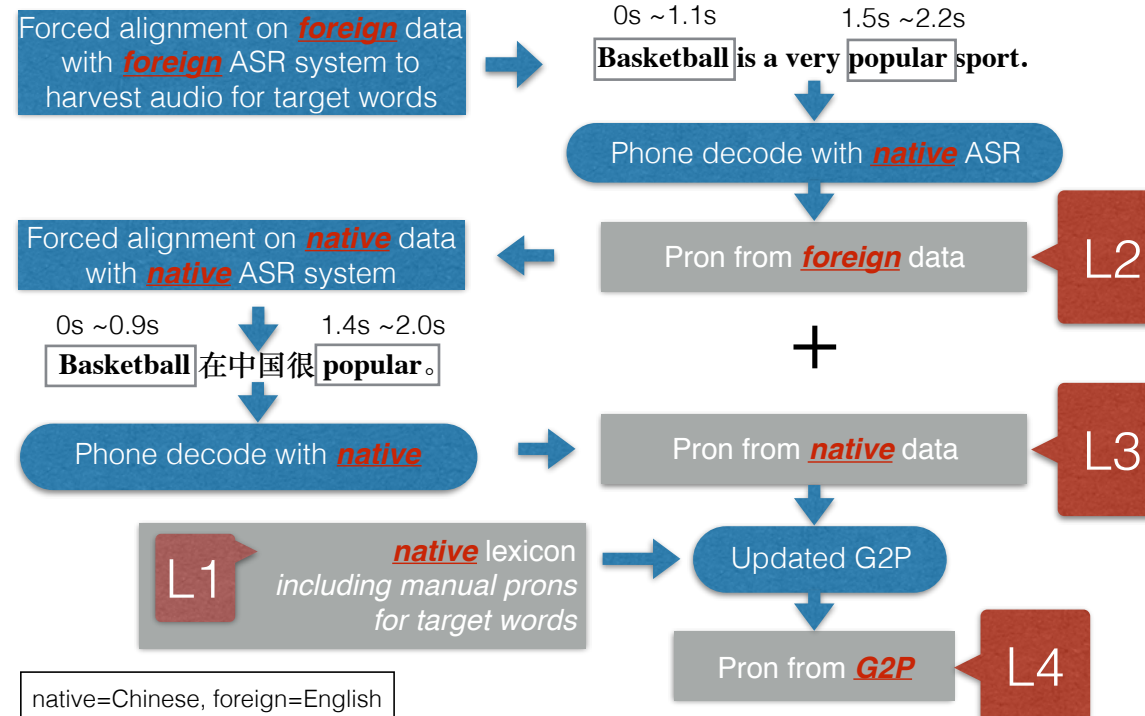
CS ASR Challenges

- Not easy to find enough CS training data.
- Excessive CS training data may degrade native monolingual systems.
- Hard to obtain high quality pronunciations for foreign words.

Retraining-Free CS ASR

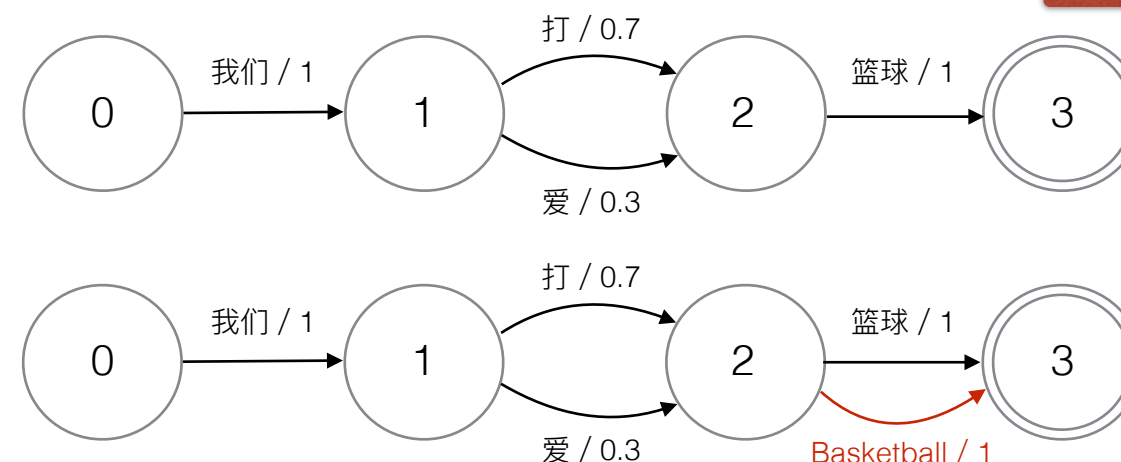
- Data driven pronunciation learning:**
 - Learning from foreign data
 - Learning from native data
 - Grapheme to Phoneme (G2P) model training
- LM enriching:**
 - Borrow statistics from native language model by using translated word pairs

Pronunciation Learning for Foreign Words



LM Enriching for Foreign Words

Borrow statistics from native language model by using translated native-foreign phrase pairs, e.g., 篮球 ↔ basketball



Experimental Results

WER on CS testing set
L0: no foreign word pron
L1: manually created prons
G0: no LM enriching

System	WER
L0 + G0 (Baseline)	34.4%
L1 + G1	24.3%
L2 + G1	24.3%
L3 + G1	20.7%
L4 + G1	18.8%
(L1+L2+L3+L4)+G1	15.3%

WER on general testing set

System	WER
L0 + G0	14.5%
(L1+L2+L3+L4)+G1	14.6%

Conclusion & Future Work

- Pronunciation learning for CS should be data driven and should learn from native speech.
- Borrowing statistics from native language model by using translated pairs saves LM retraining.
- Future work: to tackle word-reordering issue between foreign and native languages, and expand this method to other language pairs.