

Acoustic data driven pronunciation lexicon for speech recognition



Edinburgh – Cambridge – Sheffield

Liang Lu

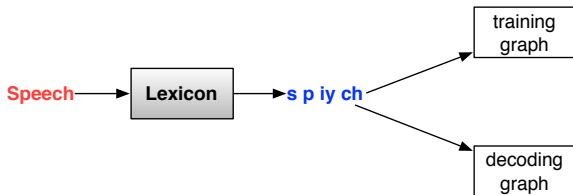


UNIVERSITY OF
EDINBURGH

23 May 2013

Motivation

- A lexicon is one of the key components for ASR
- Building a lexicon is expensive
- Maintain a lexicon is hard
 - New words, terms, name entities,
- Can we learn the lexicon automatically?

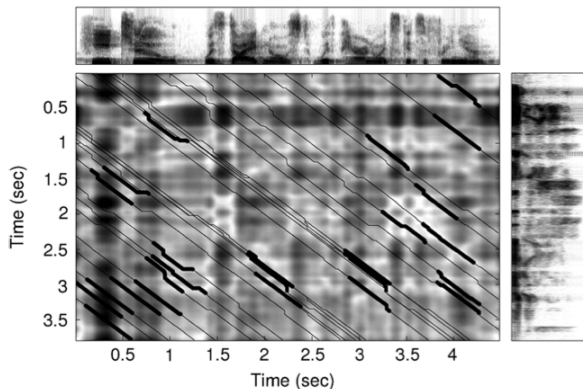


Grapheme to phoneme (G2P) conversion

- G2P conversion model is one of the standard methods
- But it requires large initial pronunciation lexicon
- Do not make use of the acoustic data

Grapheme	s	p	ee	ch
Phoneme	s	p	iy	ch

Purely acoustic data driven approach



Park and Glass, "Unsupervised pattern discovery in speech", IEEE TASLP, 2008

Combination of the two methods

- A weak G2P model trained with limited samples
- Pronunciation refinement using the acoustic data



McGraw et.al "Learning new word pronunciations from spoken examples", IEEE TASLP, 2013.

Pronunciation mixture model

- A standard ASR framework:

$$\hat{\mathbf{W}} = \arg \max_{\mathbf{W}} \underbrace{p(\mathbf{O}|\mathbf{W})}_{AM} \underbrace{P(\mathbf{W})}_{LM} \quad (1)$$

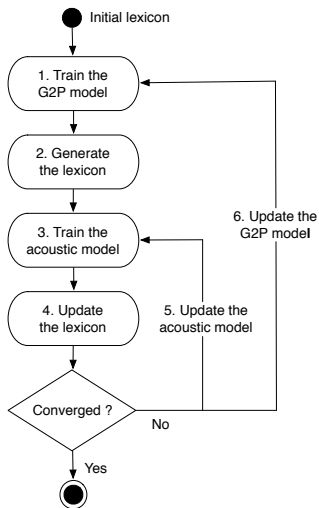
- With explicit pronunciation mixture model:

$$\hat{\mathbf{W}} = \arg \max_{\mathbf{W}} \underbrace{P(\mathbf{W})}_{LM} \sum_{\mathbf{B}} \underbrace{p(\mathbf{O}|\mathbf{B})}_{AM} \underbrace{P(\mathbf{B}|\mathbf{W})}_{PM} \quad (2)$$

- EM algorithm can be used to learn the pronunciation weights

Joint acoustic and lexicon model training

- Train the G2P model
- Generate the lexicon
- Train the acoustic model
- Update the lexicon



System setup & results

- Switchboard corpus
 - Training data: 110h
 - Initial lexicon: 5K
 - Expert lexicon: 30K
- MFCC+LDA_MLLT
- ML-GMM system
- No SAT

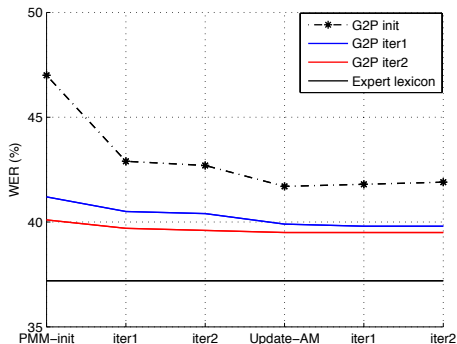


Table: Switchboard results with 286h data.

System	WER
110h-derived lexicon	35.7
+ lexicon update	35.1
Expert lexicon	34.2

Conclusions

- Low-resource pronunciation modelling
- G2P model + acoustic information
- Make the lexicon adaptable rather than fixed
- Unsupervised initialisation — get rid of the expert lexicon