

Unsupervised and User Feedback Based Lexicon Adaptation for Foreign Names and Acronyms

André Mansikkaniemi and Mikko Kurimo

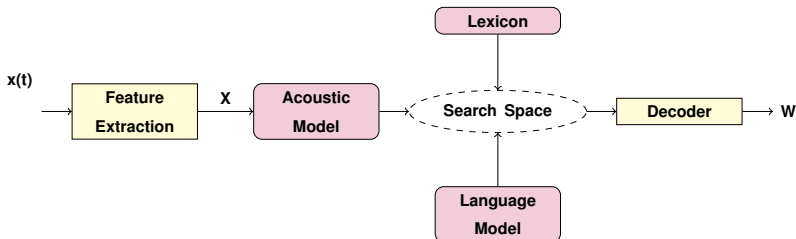
Aalto University School of Electrical Engineering
Department of Signal Processing and Acoustics

SLSP 2015 Budapest, Hungary

ASR lexicon

- **Lexicon in automatic speech recognition (ASR):**

- Ties together the acoustic and language model
- Stores phoneme-level pronunciation rules for words
- Pronunciation rules either handcrafted (i.e English) or automatically generated (i.e Finnish)

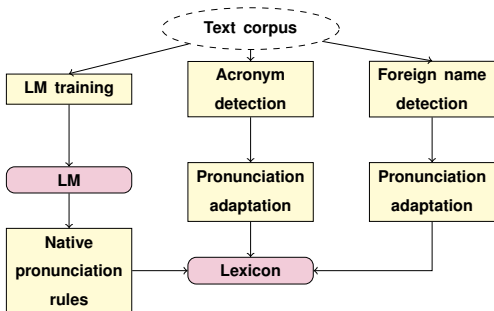


ASR lexicon

- **Foreign proper names (FPNs) and acronyms (ACRs) in ASR:**
 - Occurrence rate in speech at 1-5%
 - High information value
 - Correct recognition improves overall intelligibility of ASR output
- **Recognition challenges for FPNs and ACRs:**
 - Non-standard pronunciation rules
 - Many pronunciation variations between speakers
 - Often poorly estimated by language models
- **Solutions presented in this work:**
 - Detect FPN and ACR words in LM training text
 - Add adapted pronunciation rules to lexicon
 - Discover pronunciation variants in speech

Unsupervised lexicon adaptation

- Adaptation framework:



Detecting special vocabulary units

- **Foreign proper names**

- Words starting in uppercase letter selected as FPN candidates
- Letter n-gram perplexity

$$ppl(word) = \sqrt[n]{\prod_{i=1}^n \frac{1}{P(l_i|h)}} \quad (1)$$

- Case-frequency

$$case(word) = \frac{\sum uppercase(word)}{\sum uppercase(word) + \sum lowercase(word)} \quad (2)$$

- FPN selection score

$$score(word) = ppl(word) * case(word) \quad (3)$$

- **Acronyms**

- All words composed of only uppercase letters selected as ACR candidates
- Most frequently used ACR candidates selected for adaptation

Pronunciation adaptation

- **Foreign proper names**

- Adapted pronunciation rules generated for selected FPNs
- Data-driven G2P converter used (Sequitur)
- Trained on list of foreign words found in Finnish text corpus

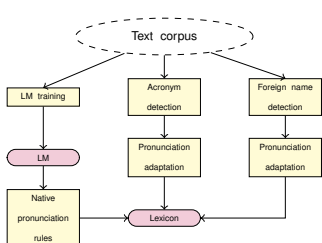
- **Acronyms**

- Two pronunciation variants generated for each ACR
 - Phonetic pronunciation
 - Alphabetic pronunciation

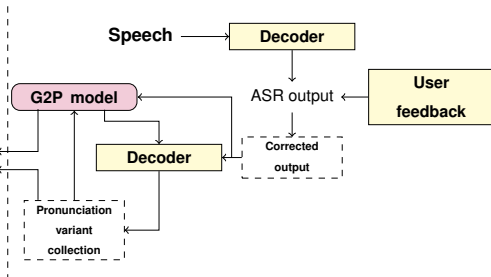
User feedback based adaptation

- **Find FPN pronunciation variants from speech data**
 - Generate multiple pronunciation variants for corrected word
 - Use forced alignment and decoder to find the most likely pronunciation variant

Unsupervised lexicon adaptation



User feedback based adaptation



Experiments

- **System and models**

- Aalto ASR system
- Morph-based LM (varigram LM, $n=10$) trained on 140 million word Kielipankki corpus (newswire texts) and smaller 7 million word Web corpus (online news articles)
- 46k morph lexicon trained on same corpus
- Letter ngram ($n=2$) trained on Kielipankki word list
- G2P pronunciation model trained on list of 2000 foreign names with Sequitur

- **Speech data**

- Finnish radio and TV news segments from 2011-2012
- Development set
 - 5 hours (35,056 words)
 - FPN rate 4.6%, ACR rate 1.3%
- Evaluation set
 - 5 hours (36,812 words)
 - FPN rate 3.8%, ACR rate 1.1%

Development set

- **Unsupervised lexicon adaptation**

- WER = Word error rate, FER = Foreign word error rate, ACER = Acronym error rate
- F = Threshold for FPN selection, A = Threshold for ACR selection

| F[%] | <i>ppl</i> | | <i>case</i> | | <i>ppl*case</i> | |
|------|-------------|-------------|-------------|-------------|-----------------|-------------|
| | WER[%] | FER[%] | WER[%] | FER[%] | WER[%] | FER[%] |
| 0 | 27.6 | 60.6 | 27.6 | 60.6 | 27.6 | 60.6 |
| 4 | 27.5 | 58.3 | 27.6 | 60.4 | 27.5 | 58.1 |
| 8 | 27.5 | 57.2 | 27.7 | 60.1 | 27.5 | 56.0 |
| 12 | 27.5 | 56.0 | 27.7 | 58.7 | 27.4 | 55.3 |
| 16 | 27.4 | 55.2 | 27.5 | 56.4 | 27.4 | 54.9 |
| 20 | 27.5 | 55.1 | 27.5 | 55.0 | 27.4 | 54.0 |
| 24 | 27.5 | 54.2 | 27.5 | 54.2 | 27.5 | 54.3 |

| A[%] | <i>freq</i> | |
|------|-------------|-------------|
| | WER[%] | ACER[%] |
| 0 | 27.6 | 85.6 |
| 1 | 26.7 | 40.9 |
| 2 | 26.8 | 40.0 |
| 3 | 26.8 | 40.0 |
| 4 | 26.8 | 40.0 |

Development set

- **Unsupervised lexicon adaptation**
 - Combined FPN and ACR adaptation

| F[%] | A[%] | <i>ppl*case,freq</i> | | |
|------|------|----------------------|-------------|-------------|
| | | WER[%] | FER[%] | ACER[%] |
| 0 | 0 | 27.6 | 60.6 | 85.6 |
| 20 | 2 | 26.5 | 53.7 | 40.4 |

- **User feedback based adaptation**

| Adaptation | WER[%] | FER[%] | ACER[%] |
|-------------------------------------|--------------|---------------|--------------|
| - | 27.6 | 60.6 | 85.6 |
| Lexicon | 26.5* | 53.7* | 40.4* |
| Lexicon + Feedback | 26.5* | 51.9** | 40.9* |
| Lexicon + Feedback + G2P-retrain | 26.4* | 51.9** | 41.1* |

- * = Significant improvement compared to baseline, ** = Significant improvement compared to lexicon adaptation

Evaluation set

| Adaptation | WER[%] | FER[%] | ACER[%] |
|-------------------------------------|--------------|--------------|--------------|
| - | 29.2 | 59.9 | 79.1 |
| Lexicon | 28.2* | 53.9* | 31.9* |
| Lexicon + Feedback | 28.2* | 53.4* | 31.9* |
| Lexicon + Feedback + G2P-retrain | 28.1* | 53.8* | 31.6* |

Conclusions

- **Unsupervised lexicon adaptation**
 - Successfully lowers error rates for both foreign names and acronyms
 - Average WER is also lowered
 - ACR adaptation most successful (around 50% rel. improvement)
- **User feedback based adaptation**
 - Only slight improvement in FPN recognition
- **Future work**
 - Further research how to improve FPN recognition
 - Error analysis
 - Combine with more advanced language modeling