Non-native Pronunciation Modeling in a Command & Control Recognition Task: A Comparison between Acoustic and Lexical Modeling

Judith M. Kessens

TNO Human Factors, Department of Human Interfaces, Soesterberg, The Netherlands

judith.kessens@tno.nl

INTRODUCTION

source language = mother tongue of the non-native speaker

target language = language the non-native is trying to speak

- Safesound project: Non-native pilots speaking English commands
- Non-native accents deteriorate Automatic Speech Recognition (ASR)
- Pronunciation variation might improve ASR performance
- Non-native accents are modeled at three levels:
  1. *Acoustic model,* e.g. adaptation or sharing models
  2. *Lexicon,* e.g. adding non-native variants
  3. *Language model,* e.g. using variant-specific priors

➔ This study: all three levels + in various combinations

- Pronunciation variants can be obtained:
  1. *Knowledge-based,* e.g. pronunciation dictionaries, linguistic studies
  2. *Data-driven,* e.g. automatic/manual transcription of data

➔ This study: data-driven, manual transcriptions

DATA

- 100 English commands, on average 6 words per command
  e.g. "Frequency one one eight decimal nine"
- Three source languages:
  - 8 Italian, 12 French and 14 Dutch speakers
- Database divided in two independent sets
  - a development (dev) and a test set (test)
  - equally-sized, no overlap in speakers

RECOGNIZER

Loquendo ASR version 6.7:

- Hybrid Hidden Markow Model (HMM) and Artificial Neural Network (ANN) recognition system
- Stationary context-independent phones and diphone-transition coarticulation models
- Baseline US English models trained on Macrophone database (200,00 utterances of 5,000 speakers from all regions of the US)
- Baseline transcriptions automatically obtained

GOAL

- Improve recognition performance
- Compare the effect of modeling non-native accents at all three levels (acoustic models, lexicon, language model)

METHOD

Lexical modeling

- Manual phonetic transcriptions of dev set
- Variants are selected based on absolute frequency:

$$F_{abs} = 100\% \times \frac{\text{variant count}}{\text{total number of words}}$$

Acoustic model adaptation

- Linear Input Network for Neural Networks

Modeling at the level of the language model

- Use variant-specific prior probabilities
- Estimate priors on frequency of occurrence in dev set
- Reliable estimation: only priors for pronunciation variants of words with frequency >10

EXAMPLES OF NON-NATIVE ACCENTS

- Rules are derived by comparing the pronunciations from the manual transcriptions to the baseline transcriptions

Rule	Dutch	Italian	French
/ɪ/ → /i/	6.3%	7.6%	19.0%
/e/ → /ə/	5.1%	3.4%	6.7%
/d͡ʒ/ → /t͡ʃ/	7.6%	7.1%	4.6%
/ɑː/ → /ɔː/	4.1%	3.0%	4.2%
/ə/-deletion	5.4%	3.8%	-
/æ/ → /e/	11.3%	-	-
/v/ → /f/	5.1%	-	-
/d/ → /t/	4.2%	-	-
/ɑː/ → /ʌ/	3.6%	-	-
/t/-deletion	-	7.0%	-
/æ/ → /ɑː/	-	3.6%	-
/ə/ → /ɑː/	-	3.4%	-
/ʌ/ → /ɑː/	-	3.3%	-
/æ/ → /ʌ/	-	3.0%	-
/æ/ → /ə/	-	-	4.1%

Table 1: Most frequent non-native pronunciation rules with rule frequencies ("-" means that the rule frequency < 3.0%)

LEXICAL AND ACOUSTIC MODELING

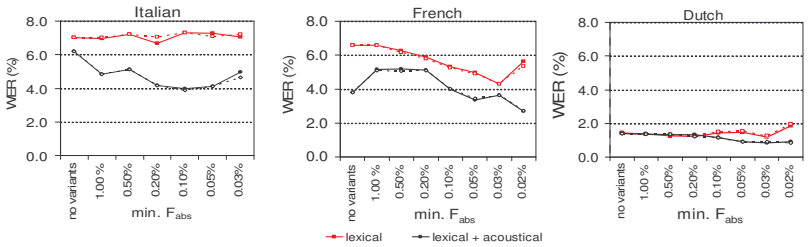

Figure 1: Effect of lexical modeling separately and in combination with acoustic modeling (dashed lines = WERs using variant-specific prior probabilities).

SUMMARY OF RESULTS

acoustic	no	no	yes	yes
lexical	no	yes	no	yes
Italian	7.1%	6.8%	6.2%	4.0%
French	6.5%	4.3%	3.8%	2.7%
Dutch	1.4%	1.2%	1.4%	0.8%

Table 2: Summary of WERs for all combinations of acoustic and lexical modeling

Summary of relative WER reductions (compared to baseline)	
• Lexical modeling:	4 - 34%
• Acoustic modeling:	0 - 42%
• Combination acoustic and lexical modeling:	43 - 58%
• Variant-specific priors:	no effect

CONCLUSIONS

- Results are source language dependent
- Best results for combination of acoustic and lexical modeling
- No improvement for using variant-specific prior probabilities

FUTURE WORK

- Automatic generation of non-native transcriptions
- Investigate dependency on amount of non-native speech material

(More details about work presented on this poster, see paper submitted to ICSLP'06)