# Non-native Speech in SDS(es?)

Florian Braun

# Overview

- Basic overview of speech recognizer

# Overview

- Basic overview of speech recognizer
- Where does non-native speech cause problems

# Overview

- Basic overview of speech recognizer
- Where does non-native speech cause problems
- How to address issues

# Speech Recognizer

- Consists of 2 main parts

# Speech Recognizer

- Consists of 2 main parts
    - The acoustic model

# Speech Recognizer

- Consists of 2 main parts
  - The acoustic model
    - The model for the phones of the input

# Speech Recognizer

- Consists of 2 main parts

  - The acoustic model

    - The model for the phones of the input

  - The language model

# Speech Recognizer

- Consists of 2 main parts
    - The acoustic model
        - The model for the phones of the input
    - The language model
        - The model for grammatical structure

# Speech Recognizer

- Consists of 2 main parts
    - The acoustic model
        - The model for the phones of the input
    - The language model
        - The model for grammatical structure

- Both parts work together to allow for recognition

# Speech Recognizer

- Consists of 2 main parts

  - The acoustic model

    - The model for the phones of the input

  - The language model

    - The model for grammatical structure

- Both parts work together to allow for recognition

  - The LM backs the acoustic model to trim possible matches

# Problems for Non-native Speakers

- They don't have native acoustics

# Problems for Non-native Speakers

- They don't have native acoustics

  - Native trained acoustic model is inadequate to model their acoustics

# Problems for Non-native Speakers

- They don't have native acoustics

  - Native trained acoustic model is inadequate to model their acoustics

  - Different non-native populations have different acoustics

# Problems for Non-native Speakers

- They don't have native acoustics

  - Native trained acoustic model is inadequate to model their acoustics

  - Different non-native populations have different acoustics

    - A model for one group does not cover a group of speakers which share a different L1

# Problems for Non-native Speakers

- They don't have native acoustics

  - Native trained acoustic model is inadequate to model their acoustics

  - Different non-native populations have different acoustics

    - A model for one group does not cover a group of speakers which share a different L1

- They do not share the same idiomatic expressions

# Problems for Non-native Speakers

- They don't have native acoustics

  - Native trained acoustic model is inadequate to model their acoustics

  - Different non-native populations have different acoustics

    - A model for one group does not cover a group of speakers which share a different L1

- They do not share the same idiomatic expressions

  - A native LM often time misses vocab from L2 speakers

# Problems for Non-native Speakers

- They don't have native acoustics

  - Native trained acoustic model is inadequate to model their acoustics

  - Different non-native populations have different acoustics

    - A model for one group does not cover a group of speakers which share a different L1

- They do not share the same idiomatic expressions

  - A native LM often time misses vocab from L2 speakers

  - L2 speakers have much larger variation

# Problems for Non-native Speakers

- They don't have native acoustics

  - Native trained acoustic model is inadequate to model their acoustics

  - Different non-native populations have different acoustics

    - A model for one group does not cover a group of speakers which share a different L1

- They do not share the same idiomatic expressions

  - A native LM often time misses vocab from L2 speakers

  - L2 speakers have much larger variation

  - Coach them to get it right!

# How to Address These Problems

- More Data!

# How to Address These Problems

- More Data!
  - Just get L2 speakers into the Acoustic model

# How to Address These Problems

- More Data!

  - Just get L2 speakers into the Acoustic model

  - Make the LM more diverse.

    - Allow for more variation

    - Allow for more ungrammatical constructions

# How to Address These Problems

- More Data!
    - Just get L2 speakers into the Acoustic model
    - Make the LM more diverse.
        - Allow for more variation
        - Allow for more ungrammatical constructions
- Not so fast!

# How to Address These Problems

- More Data!
    - Just get L2 speakers into the Acoustic model
    - Make the LM more diverse.
        - Allow for more variation
        - Allow for more ungrammatical constructions
- Not so fast!
    - More data often just confuses the models.

# How to Address These Problems

- More Data!
    - Just get L2 speakers into the Acoustic model
    - Make the LM more diverse.
        - Allow for more variation
        - Allow for more ungrammatical constructions
- Not so fast!
    - More data often just confuses the models.
    - Makes the models too complex

# How to Address These Problems

- More Data!
    - Just get L2 speakers into the Acoustic model
    - Make the LM more diverse.
        - Allow for more variation
        - Allow for more ungrammatical constructions
- Not so fast!
    - More data often just confuses the models.
    - Makes the models too complex
    - Especially for the acoustic model

# Effects of Adding More Data

- Adding more data to the LM allows:

# Effects of Adding More Data

- Adding more data to the LM allows:

    - Better parsing, fewer OOV words, still improves L1 recognition.

# Effects of Adding More Data

- Adding more data to the LM allows:

  - Better parsing, fewer OOV words, still improves L1 recognition.

  - Works well on transcribed speech

# Effects of Adding More Data

- Adding more data to the LM allows:

    - Better parsing, fewer OOV words, still improves L1 recognition.

    - Works well on transcribed speech

    - Without some added boost to the acoustic model does not really impact overall performance

# Effects of Adding More Data

- Adding acoustic data

# Effects of Adding More Data

- Adding acoustic data
  - A little bit of data helps, even more, not so much

# Effects of Adding More Data

- Adding acoustic data

  - A little bit of data helps, even more, not so much

  - Adding everything to one model can be detrimental

# Effects of Adding More Data

- Adding acoustic data

    - A little bit of data helps, even more, not so much

    - Adding everything to one model can be detrimental

    - Cluster substitutions that occur together to form multiple acoustic models

# Effects of Adding More Data

- Adding acoustic data

  - A little bit of data helps, even more, not so much

  - Adding everything to one model can be detrimental

  - Cluster substitutions that occur together to form multiple acoustic models

  - Use results from most confident model

# Effects of Adding More Data

- Adding acoustic data
    - A little bit of data helps, even more, not so much
    - Adding everything to one model can be detrimental
    - Cluster substitutions that occur together to form multiple acoustic models
    - Use results from most confident model
        - Problem becomes picking most confident model

# Effects of Adding More Data

- Adding acoustic data
    - A little bit of data helps, even more, not so much
    - Adding everything to one model can be detrimental
    - Cluster substitutions that occur together to form multiple acoustic models
    - Use results from most confident model
        - Problem becomes picking most confident model
        - Just using raw recognition scores are not a good indicator

# Conclusions

- There need to be improvements in modeling to allow L2 speakers to use systems just like L1 speakers would

# Conclusions

- There need to be improvements in modeling to allow L2 speakers to use systems just like L1 speakers would

- The improvements need to affect both the acoustic and language models

# Conclusions

- There need to be improvements in modeling to allow L2 speakers to use systems just like L1 speakers would

- The improvements need to affect both the acoustic and language models

- There are still challenges in picking the right data to supplement these models with

# Conclusions

- There need to be improvements in modeling to allow L2 speakers to use systems just like L1 speakers would

- The improvements need to affect both the acoustic and language models

- There are still challenges in picking the right data to supplement these models with

- There are still challenges in making the extra data not degrade performance

# Conclusions

- There need to be improvements in modeling to allow L2 speakers to use systems just like L1 speakers would

- The improvements need to affect both the acoustic and language models

- There are still challenges in picking the right data to supplement these models with

- There are still challenges in making the extra data not degrade performance

- There are still challenges in choosing the correct result from recognition

# Questions

- Can coaching system users really help?

- Would non-standard synthesized voices make speakers of non-standard varieties more likely to use an SDS?

  - Can building those types of systems allow us to infer how to better build systems to recognize non-standard speakers?