Outline

Research Questions and Hypothesis

Research Questions

RQ1

- Can structures corresponding to vowel harmony classes be found through the vector representations learned by a *weakly-supervised* "deep" RNN

RQ2

- If yes (to RQ1), then what is the hierarchical position of the layer(s) that best learns to represent these structures w.r.t. the rest of the network
    - Do the results of Alishahi et al. (2017) replicate?

Hypothesis

- Alishahi et al. (2017) have shown that the vector representations learned by a deep RNN model do learn to represent phonemes with lower r-layers outperforming higher ones at phoneme discrimination
- if higher layers are able to learn more abstract structures (such as VH classes), then they should outperform the lower ones at a VH class discrimination task

Data and Methods

Dataset

- Aalto University DSP Course Conversation Corpus
- Finnish spontaneous conversations with force-aligned transcriptions (utterance, word, segment)
- 5200 utterances, 9.7hrs of audio from 218 male and 24 female speakers
- `http://urn.fi/urn:nbn:fi:lb-2017092133`

Model

Input

- MFCC vectors, utterance level

Layers

- 1 convolutional, 5 recurrent

Output

- Embedding layer, trains to project utterance encoding and image encoding to a joint vector space

Experiment

ABX discrimination task

- tuples of the form (A,B,X) where
  - A, B, and X are CV syllables
  - B and X vowels share harmonic class
- for each tuple, calculate $sign(dist(A, X) - dist(B, X))$, where $dist(i, j)$ is euclidean distance between model's vector representations of syllables *i* and *j*
- +ve sign indicates that the model has learned to discriminate the VH classes in A and B

References

Alishahi, A., Barking, M., and Chrupała, G. (2017). Encoding of phonology in a recurrent neural model of grounded speech. In *Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017)*, page nil.