Motivation

There is a great research effort in looking for medical solutions for Alzheimer’s disease, while significantly less in creating solutions for caregiving post-diagnosis. The motivation of this project is to provide patients with a tool to recognize familiar people in common situations.

The solution implements a text independent speaker recognition system using deep learning. As input, the model receives spoken utterances to output a predicted identity profile.

Data

VCTK Corpus composed by speech data uttered by 109 English speakers. Every speaker reads around 400 sentences that were specifically selected to maximize contextual and phonetic coverage. Each speaker reads a different set of sentences, which is good for a text-independent model.

Models

DNN Model
2 FC Layers - Sigmoid activation functions

CNN Model
3 Conv. Layers - 3 Pooling Layers - 2 Dropout layers - 1 FC Layer - ReLU activation functions

Discussion

- DNN Model has low variance and high bias. A different architecture could help improve.
- CNN Model presented high variance and low bias on the training set. This suggests that there is still overfitting. The model did not converge on 80 epochs, adjusting learning rate and regularization techniques could lead to better performance.
- There is no sufficient data to choose a “best” model.

Future Work

1. Implement a hyperparameter tuning module.
2. Deploy in AWS instance with GPU to train CNN until it convergences.
3. Perform data augmentation to add background noise.
4. Explore the performance in RNN and Residual architectures.
5. Test with Alzheimer’s patients in common settings. Develop a proof of concept product.

References

5. Philippe Remy - Deep Speaker
6. Manish Pandit - Speaker Recognition