Case Study Voice Recognition

Objective

The objective is to gain a qualitative and quantitative understanding of the recognition performance of a fully open source voice recognition engine without specific hardware acceleration devices on phone-quality audio.

Students should be able to answer the following questions:

- How complicated is it to deploy a working pre-trained voice recognition engine?
- Is the engine able to approach near real-time performance? (lag less than 10 seconds)
- What is the hardware chosen?
- What is the CPU and memory cost of this recognition on the hardware they chose to use?
- How good is the recognition?
    - Qualitatively, can you illustrate performance with well-chosen perfect recognitions and illustrate typical mistakes?
    - Quantitatively, what is the Word Error Rate? How does it compare with state-of-the-art engines on the same dataset?

References

- DeepSpeech open source implementation: https://github.com/mozilla/DeepSpeech
- DeepSpeech english pre-trained models: https://github.com/mozilla/DeepSpeech/releases/tag/v0.9.3
- Switchboard phone-quality recordings for testing: https://deepai.org/dataset/switchboard-hub500