

Voice Recognition

Seroussi Yishay [☆]

Bismuth Samuel ^{*}

Shaag Yehonatan [♦]

Deep Learning Assignment 1

Abstract

Given a voice recording, by using deep learning, we are able to determine (With high percentages of certainty) from what country the accent of the speaker coming, and then, make assumption about the origin of the speaker.

To achieve such a project, we need a strong data-set built from the website provided by our lecturer. Our data-set includes each one of the next languages : French, Hebrew, USSR and English (UK and USA), five different recording.

Each recording is the next text : " Please call Stella. Ask her to bring these things with her from the store: Six spoons of fresh snow peas, five thick slabs of blue cheese, and maybe a snack for her brother Bob. We also need a small plastic snake and a big toy frog for the kids. She can scoop these things into three red bags, and we will go meet her Wednesday at the train station. " read by a subject

[☆] Student of Computer Science (third year), Ariel University, Ariel 40700, Israel.
Id : 305027948. Email: seroussi1@gmail.com

^{*} Student of Computer Science (third year), Ariel University, Ariel 40700, Israel.
Id : 342533064. Email: samuelbismuth101@gmail.com

[♦] Student of Computer Science (third year), Ariel University, Ariel 40700, Israel.
Id : 308357953. Email: yoshago@gmail.com

1 Data-set

The data set of the website is as the next table:

Adding to each recording, for each subject we got information, such as the next table 1.

birth place	st. laurent d'onay, france (map)
native language	french (fra)
other language(s)	spanish
age, sex	20, female
age of english onset	12
english learning method	academic
english residence	usa
length of english residence	0.4 years

Table 1: Biographical Data

Adding of course the audio.

We decide to take only the native language of the fifth regions (France, USSR, UK, USA and Israel) and the audio that we convert into a WAV file. The rest of the data will be used in order to test our data-set.

An important note is about the audio: Of course the computer don't understand audio such as human does, so, we have to translate the audio into numbers understood by a computer. To do this, and by Lecturer's advice, we use Mel-frequency (mfcc) to handle the audio. To do this, we downloaded all the audio in a WAV files, and we use the module "python speech features" to transform the audio file into a matrix of numbers understanding by the computer.

Then, our actual data set is composed of a CSV table (ready to be read by a python script and easily alterable) with two columns, the first is the region and the second is the link to the WAV file.