

Assignment 3: Calculating phonotactic probability

Due: Sept. 30

The Jusczyk, Luce, and Charles-Luce study employed sets of monosyllables which were claimed to have high and low phonotactic probabilities in English. Your task is to check their claim, by computing the phonotactic probability of their test items. There is a file called `CelexWordsInTranscription.txt`, which contains a list of English words. Your task is to write a program that reads in this file, computes the probabilities of their items, by the criteria used in that study. (That is, by the “positional” probabilities). You will need to perform several sub-tasks:

- You will need to figure out how to break the syllables up into onsets, nuclei, and rhymes (a key to the symbols that are used is provided on the web site along with the file)
- You will need to calculate the probability of each phoneme in each position
- You will then need to find a way to translate these individual probabilities to a single score for the entire word
- I have created test files with the Jusczyk et al stimuli that you can run your program on to see what their scores are. These files are:
 - `Exp1-High.txt`
 - `Exp1-Low.txt`
 - `Exp3-High.txt`
 - `Exp3-Low.txt`

There is also a file `AlbrightHayes.txt`, which contains ratings of a batch of monosyllables created by myself and Bruce Hayes. These provide more fine-grained detail than the Jusczyk et al items, which are merely lumped into groups of “high” and “low” probability. As discussed in class, the numerical ratings in the `AlbrightHayes.txt` file can be used to calculate a correlation to test the predictions of your model.