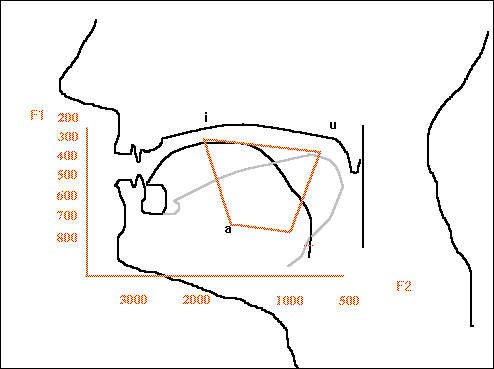
**Outline**

**What is this data set?**

* This dataset consists of 12 young adults’ (6 non-native speakers and 6 native speakers) phonetic performance in Mandarin. Each subject was required to read nine different sentences with nine target vowels ([i], [y], [u], [e], [o], [ə], [ɤ], [a], [ɑ]) twice. Praat, a speech analysis software, was used to record the subjects’ voice and further analyze their sound frequency. The outcome measures are F1 (the first formant) and F2 (the second formant). F1 value indicates the height of the tongue and F2 value shows the frontness or backness of the tongue.

 figure [1]

Non-native speakers are presumed to produce the nine vowels in Mandarin differently due to the influence of their native language (English). Therefore, the purpose of the study is to investigate if non-native Mandarin speakers apply their native patterns in English vowels to Mandarin vowels. If so, do the patterns show any similarity or difference than native Mandarin speakers’?

**What preparatory work needs to be done?**

The raw data is in an excel format. We will first import the raw data and transform it into a tibble format using {rio} and {here}. The data frame will then be organized in a way that each row contains one observation using {tidyverse}. Each unique vowel sound needs to be renamed so the program can recognize it. To summarize the mean value of F1 and F2 in each group, the {dplyr::filter, ::summarize} will be used. To explore the data, we will make a F1-F2 plot similar to figure [1] for each group using {ggplot2}. If a different pattern is noticed, a F1-F2 point plot with both groups showing different colors will be made.

**How will the requirements of the final project be met?**

Our team had an initial meeting on 10/19/2018 7pm. We make a decision regarding which dataset will be used. Each team member will complete their roles by reaching each check point. Steffi and Jun (both from EALL dep.) are the owner of the dataset. They collected the data from 6 native Chinese speakers and 6 non-native speakers who study Mandarin Chinese as a second language. They agreed to make this dataset public and as the material of the final project. They will fill in the phonetic knowledge and any related information about this dataset to the other two members. Teresa (HPHY dep.) drafted this outline and will be in charge of taking the meeting notes and monitoring the project’s progress. Ting-fen (CDS dep.) is the organizer of the group Github and she will keep an eye on any merge conflicts and notify the group members if issues come up. All four members will contribute to the R script.

**Timeline**

1. Day 1 (10/19)
   * 😊 agreement on the dataset
   * ☹ unable to figure out how to upload the dataset to group Github after trying for 2 hours
   * ☹ no one knows how to use GitKraken exactly

**Reference**

[1] <http://www.uni-bielefeld.de/lili/personen/vgramley/teaching/HTHS/acoustic_2010.html>. Date: 10/19/2018