

# Final Project Outline

---

## 1. Group members:

Heidi Shi, Jung-ah Lee, Jeongim Jin

## 2. Description of the data:

We have two sets of linguistic data collected from three native speakers (NS) of Korean and three non-native speakers (NNS) of Korean (shared L1 English). During the original data collection, the six speakers were asked to produce eight Korean vowels orally, and each vowel is repeated three times for audio recording. The acoustic data is later quantified by two Hz-based acoustic measurements: F1 (reflects vowel height) and F2 (indicates vowel backness). For each vowel, F1 and F2 always occur in pairs so as to provide locational information of the resonance in a speaker's vocal cavity. The exact value of F1 and F2 are obtained through software called Praat, which is widely used in phonetic measurements. Specifically, our dataset's basic structure can be illustrated by the following table. For instance, in the NNS dataset, under the Korean vowel "i," there are three F1 values and three F2 values associated with each speaker. What we are trying to explore from these datasets is the differences (or similarities) between the two speaker groups in their Korean vowel production. Through comparing and visualizing F1 and F2 values, we aim to discover the "challenging" and "easy" vowels for L1 English speakers of Korean. The result of our analysis can be applied in Korean language pedagogy or phonetic studies in Korean linguistics.

		i		
	repetition	1	2	3
NNS 1	F1	347.0051	399.5299	307.0789
	F2	2656.345	2514.29	2393.816
NNS2	F1	361.5158	345.7214	334.6823
	F2	2606.105	2602.082	2699.489
NNS3	F1	406.4256	402.6381	411.3147
	F2	2437.481	2509.16	2467.863

Table 2.1

### 3. Preparatory work to be done:

First, to meet the requirement of the final project, our original data needs to be cleaned into the tidy version. We might have to clean the outliers and then make statistics-driven decisions concerning the mean, median, max and min values. Second, we also need to discover the correlations between the two speaker groups and present them in the form of summarized tables. For instance, for which vowel, there is a significant difference between the NS and NNS? For this specific vowel, is the inter-group or inter-speaker discrepancy reflected by F1 or F2? Third, through analyzing the data through R, we also wish to discover some unrevealed NNS patterns that hindered by our previously adopted conservative methodology. Therefore, to better understand research questions, we are eager to visualize the data from multiple aspects. We will include but not limited to the following visual tools: (1) a general vowel space comparison between NNS and NS based on mean values, (2) a speaker-based comparison facet by vowels, (3) a vowel-based comparison facet by speakers.