Music Perception and Cognition Final Project

The Pitch of Speech between Mainland China and Taiwan China

Abstract

Introduction

- ○ Motivation:

It has been so long observed that there are pitch level differences of speech between Chinese mandarin speakers and Taiwanese mandarin speakers in general. Chinese is a tonal language and mandarin, the modern standard Chinese that has been spoken in both Mainland China and Taiwan, has 4 tones. Community within a linguistically enclosed geographical area should have converged into a consistent speech pitch distribution that has stabilized over the time. Despite the same language spoken across the Taiwan strait, we conjectured that because of the long-term geographical isolation the overall pitch levels of speech might differ across societies.

- ○ Literature

Diana Deutsch has conducted an experiment to study this phenomenon in two Chinese villages that were located in a mountainous region with a travel time of several hours. The study's results confirmed the conjecture as mentioned above that ***"the pitch ranges of the subjects' speech clustered within each village, but differed overall across villages."*** This provided evidence that the overall pitch range of a person's speaking voice reflects long-term exposure to the speech of others – at least in the case of tone language speakers.

- ○ Hypothesis

This project investigates the pitch level differences among Chinese and Taiwanese mandarin speakers in an attempt to generalize Diana's theory to a larger geographical scale and to both female and male speaking voices. **We would expect that Taiwanese pitch level are generally higher than of mainland participants regardless of genders.**

Methodology
- Experimental Design

This experiment has 4 female participants and 2 male participants with half of them from Taiwan and the other half from Mainland China. (That is, 2 female Taiwanese, 2 female Chinese, 1 male Taiwanese and 1 male Chinese.)

This is an extremely small sample pool that would result in inevitable biases due to significant individual differences as it shows on experiment results.

- Task and Manipulation

Subjects were given the same text to read out of roughly 4 minutes. They were asked to control the tempo at which they read out every line of text to roughly 84 bpm by following along the metronome clicks played in their headphones during the recording. The text we chose was a classical Chinese poem that followed a particular pitch and rhythm pattern as the writing format, and every line of poem are same long as made up of 7 characters. The coherent rhythm characteristic of the chosen text allowed us to maximize the tempo control across all participants at the best we can. The consistent pitch patterns carried along throughout the poem writing provided a more clustered/concentrated pitch samples rather than a sparse, unconsolidated random distribution if provided with a general essay texts.

Note: the purpose of controlling participants' reading tempo is as a way to achieve the control over sample size so that hopefully the amounts of pitch samples we took from one participant to another didn't vary too much.

- Measurement

Then we took pitch estimates of their speech at 5 ms intervals. Then for each subject we derived the fundamental frequency histogram for a range from 50hz to 400hz. **Figure 1** shows each historgram for all six speech samples. **Figure 2** aggregates each two female samples in both groups and plots one histogram for both female group respectively. **FIgure 3** contains the same information as Figure2 does except it plots two groups (Taiwan vs Mainland) on the same graph side by side as the comparison for both genders. **Figure 4** computes variances and means across groups with various partitions.
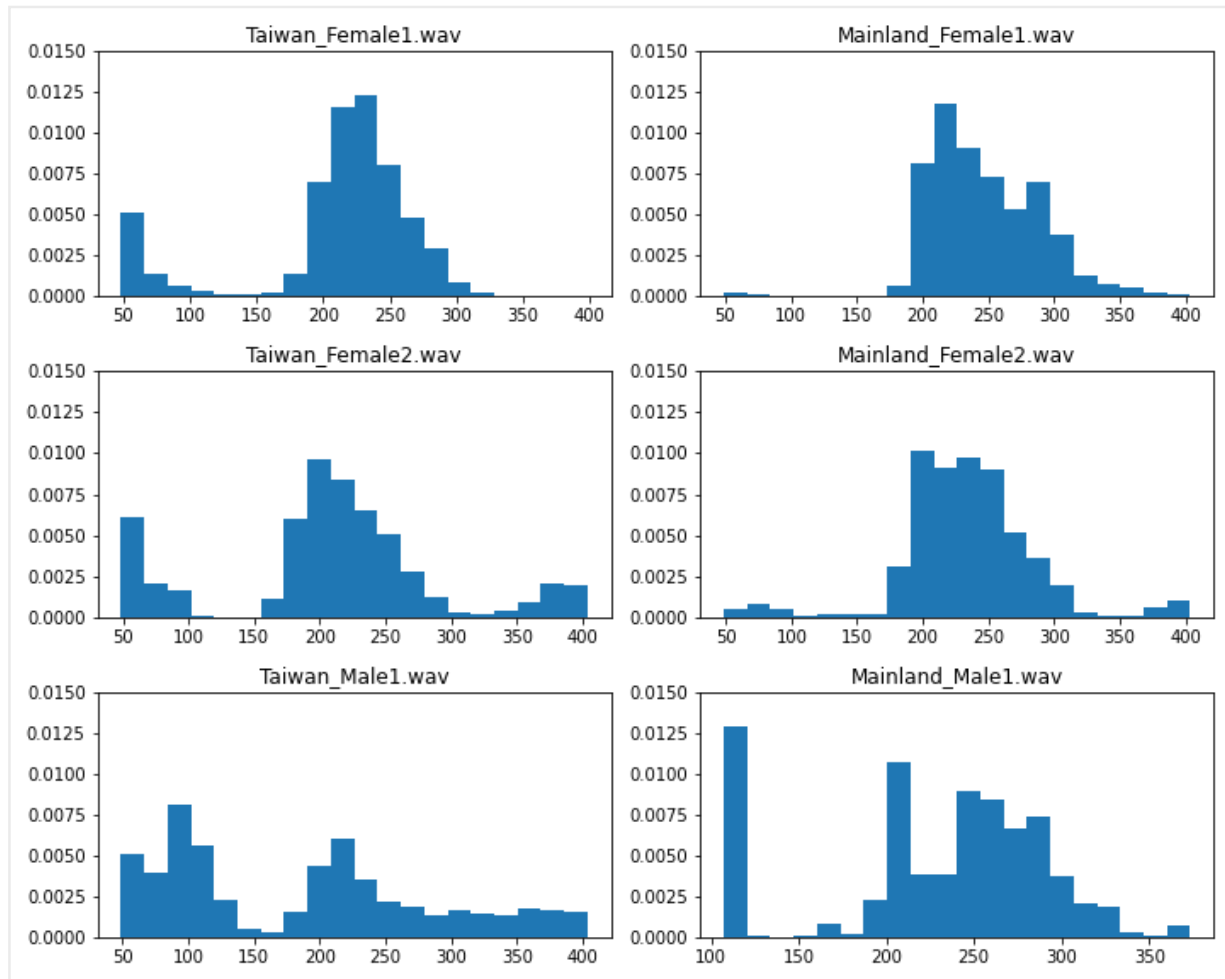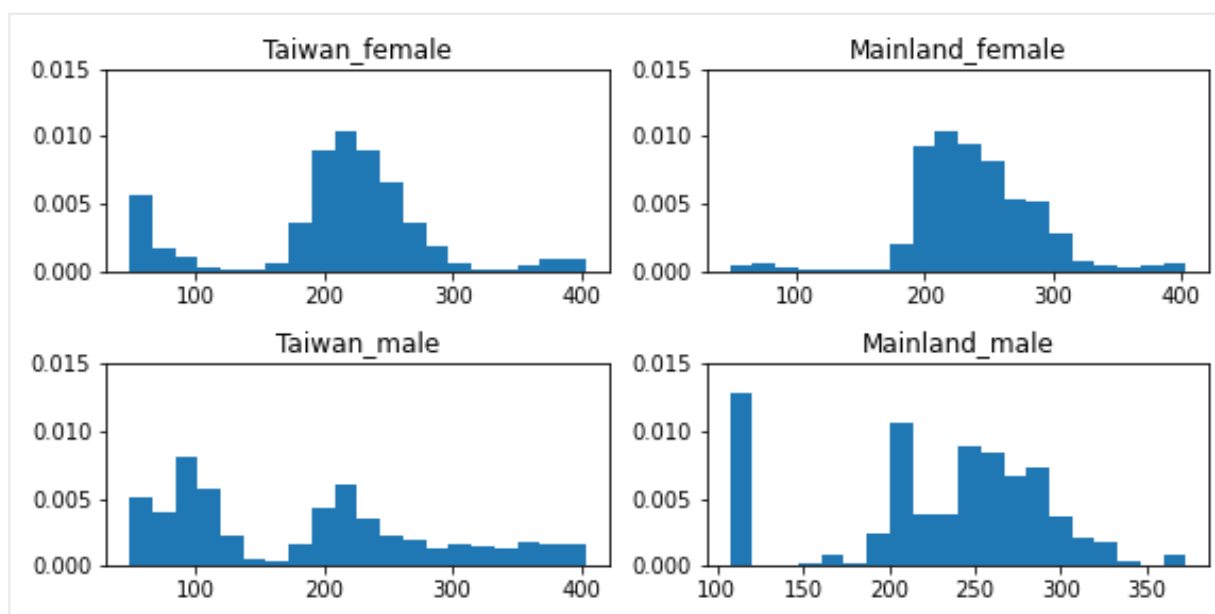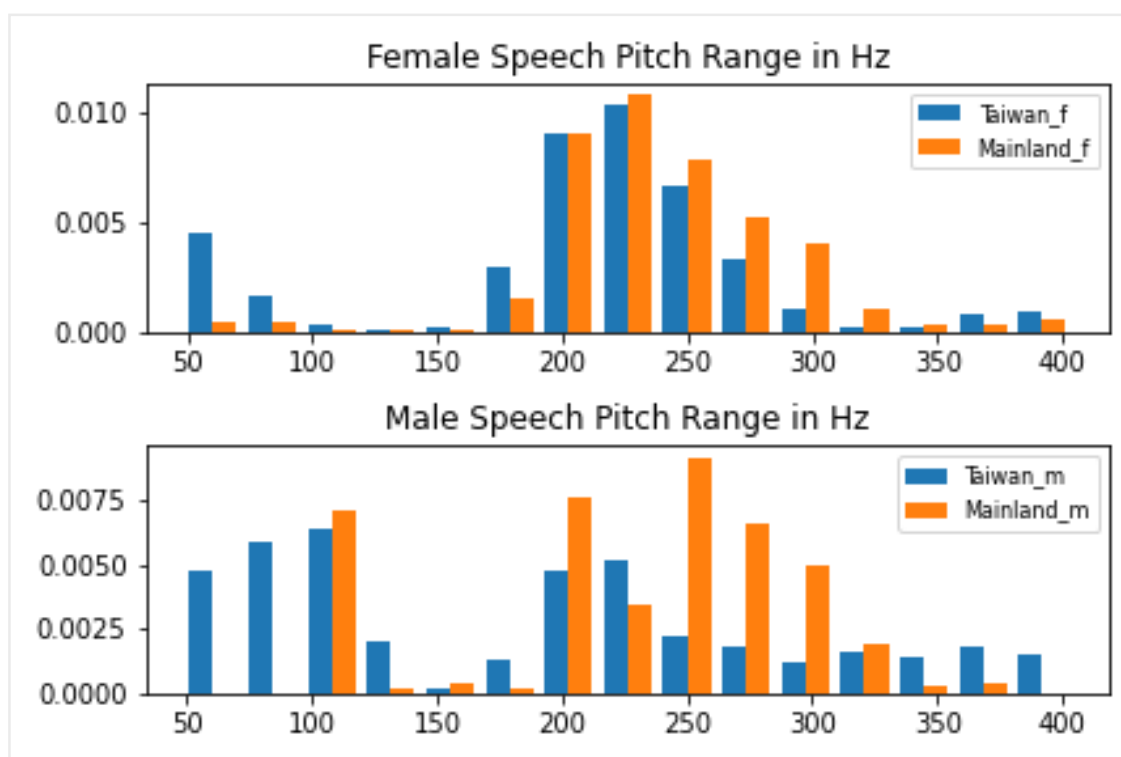
Figure 1:

Figure 2:

Figure 3:

Figure 4:

	TaiwanFemale	MainlandFemale	TaiwanMale	MainlandMale	Female	Male
variance	5422.0	2294.0	9925.0	3987.0	4167.0	8804.0
mean	209.0	240.0	186.0	230.0	223.0	212.0
number of samples	2.0	2.0	1.0	1.0	4.0	2.0

- o Variables
  - ◆ Independent variable(s):

Taiwannese vs Mainland mandarin speakers
  - ◆ Dependent variable(s):

Pitch levels of speech
  - ◆ Control variable(s):
    1. Genders,
    2. The reading text,
    3. Reading tempo
  - ◆ Confounding variable(s):
    1. The individual difference of speech in general (identified but couldn't controlled),
    2. The rate at which the pitch estimation algorithm was computed (controlled),
    3. The size of sample was also controlled indirectly by controlling the reading tempo and pitch estimation rate as sample size, or sample length = (text size / reading tempo) * pitch estimation rate
  - ◆ Nuisance variable(s):
    1. Pitch level differences between participants from different regions of Mainland China. Due to the small size of the samples we could collected we had to assume a general average speech pitch levels for all participants from mainland China. Thus, the internal differences might be neglected as an extraneous nuisance variable that is out of the interest of this study context as it aims to compare mainland against Taiwan rather than to find variance among of different regions of mainland. Also, every participant was speaking mandarin which should differ in pitches much less than if they were to speak in their regional dialects. **In simple words we assume that the variance within individual mandarin speakers is much smaller than difference between mainland and Taiwan**
    2. The personality of participants,

3. The time of a day when they recorded themselves,
4. The recording setting and equipment they used,
5. The mood they were in during the recording

- ○ Measuring the dependent variable
1. See the Methodology - measurement section above.
2. The python function used to compute the pitches is librosa.core.piptrack, from python library librosa package.

```
fre, mag = librosa.core.piptrack(d, hop_length=110,
fmin=50, fmax=400)
mask = (mag == mag.max(axis = 0)) & (mag != 0)
pitches = fre[mask]
```

- ○ Experimental interface
Every participant was aware of the goal of the experiment as to compare pitch level variances between two groups. And everyone was told to read in their normal voice as they would normally communicate.

Analysis
- ○ Summary description
  - ◆ See 4 graphs in the Methodology - measurement section
  - ◆ **Figure 1** shows each historgram for all six speech samples. **Figure 2** aggregates each two female samples in both groups and plots one histogram for both group respectively. **FIgure 3** contains the same information as Figure 2 does except it plots two groups (Taiwan vs Mainland) for each gender on the same graph side by side as the comparison.
  - ◆ The **number of data points** (pitch samples) across samples is controlled with the percent variance down to 3.3% on average. It is to make sure that every participant is equally represented in the study and that each set of pitch samples weigh the same amount in the aggregation of the corresponding subject group.

```
datasize = [49354, 50594, 50861, 51704, 56634, 52443]
avg_datasize = np.mean(datasize)
diff = np.abs(datasize - avg_datasize)
diff_percent = diff/avg_datasize
diff_percent.round(4).mean()
```

  - ◆ **Figure 4** shows the variances and means across different groups

according to various way of partition. Comparing 'TaiwanFemale' and 'MainlandFemale' we can see a smaller variance and a higher mean in Mainland female pitch levels. Based on this data though very limited in samples we may infer Taiwanese females may have a more variable tone of speech at a lower average pitch while Mainland females may have a less variable tone of speech at a higher average pitch. Meanwhile, Taiwanese male has a higher variance and lower mean than Mainland male. But this data is constrained to one sample for each subject so may not reflect the real pitch levels between two male groups.

Conclusions

- Results

Unfortunately, due to the very limited number of samples with female and male of 2 participants and 1 participant for each subject group, we didn't get what we expected about the pitch level differences. Rather, the experiment results appear to partly contradict with our hypothesis as it suggests both Mainland female and male speaking voices are higher than that of Taiwan female and male, which is in contrast to our hypothesis. However, on the other aspect, the bar plot graphs do show a clustered distribution of speech pitch levels in both group level and individual level. So in general, we might conclude that the theory saying ***the pitch ranges of the subjects' speech clustered within each community, but differed overall across communities*** *is true based on this experiment results*.