Capstone Project

Pitch Point Distribution in Syllables of Spoken Questions

JongHee Lee, Sejun Song

Introduction

In our capstone, we're looking at a dataset that captures the different aspects of how people speak, especially the tone of voice used when they ask questions in audio recordings. Our dataset, provided by Frances Blanchette, PhD, the Department of Psychology offers a collection of audio recordings. This data includes various details about the speakers and how they talk. Our aim is to find patterns in how the voice's pitch changes across different parts of speech, particularly in syllables within questions. In linguistics, how we use tone and pitch is important. They help us express feelings and meanings in what we say.

Data Collection and Variables

The data set includes 95 samples, for each of which 32 variables were collected. Each sample is observations from an audio file.

1. FileName: Unique identifier for each audio file/question token (48 A-Type & 47 B-Type).
2. Speaker: Unique identifier for each speaker that produced audio files.
3. FinalSyll: The shape of the final syllable in each audio file (neutral, fall, rise).
4. Sex: The biological sex of each speaker (female or male).
5. Question: What the speaker is saying in each audio file.
6. PitchMin: The lowest pitch point across the entire audio file.
7. PitchMax: The highest pitch point across the entire audio file.
8. numPoints: The number of times the pitch fluctuates (goes markedly up or down) across the audio file.
9. PtXTime: The time point in the audio file at which a point of pitch fluctuation occurs (with up to 9 points, resulting in potentially 9 columns for this variable).
10. syllableStart.X: The time point in the audio file at which a syllable begins (each audio file has 7 syllables, resulting in 7 columns for this variable).
11. syllableEnd.X: The time point in the audio file at which a syllable ends (each audio file has 7 syllables, resulting in 7 columns for this variable).

Objective

The objective of our analysis is to determine the relationship between pitch points and syllables in spoken questions. We aim to identify if there are specific syllables where pitch fluctuations occur more frequently. By doing so, we can better understand the rhythmic patterns of speech and how pitch contributes to the auditory signature of a question. This knowledge can have implications in several fields, from speech synthesis and recognition to linguistic studies of question intonation patterns across different languages or within a specific language group.

Understanding the distribution of pitch points across syllables can help in refining speech recognition systems, improving the naturalness of text-to-speech engines. The insight gained from this analysis can enhance our understanding of how intonation variations are utilized in natural language to perform various communicative functions.

Data Transformation and Preprocessing

To streamline our dataset for analysis, we took a few key preprocessing steps. We simplified the identifiers for audio files, categorizing them into 'A Type' and 'B Type' for better clarity. For instance, a file labeled "CORRECTPAEnglish_013b_ExpAudio_B_05_ER_P13_1" was renamed to 'B'. The categorization into 'A Type' and 'B Type' simplifies file identification without losing track of individual file characteristics. We also removed the 'Pt9Time' column due to it containing only NA values, which cleans up the dataset for more accurate analysis. Lastly, we created new variables 'syllable.X.ptNum' to link the pitch points 'Pt.X.time' directly with the corresponding syllable numbers, providing a clear view of where and how many times pitch fluctuations occur within the syllables. These transformations are essential for a focused analysis on pitch distribution.

Exploratory Data Analysis

Figure 1. Pitch Points Frequency Across Each Syllable (1-7)

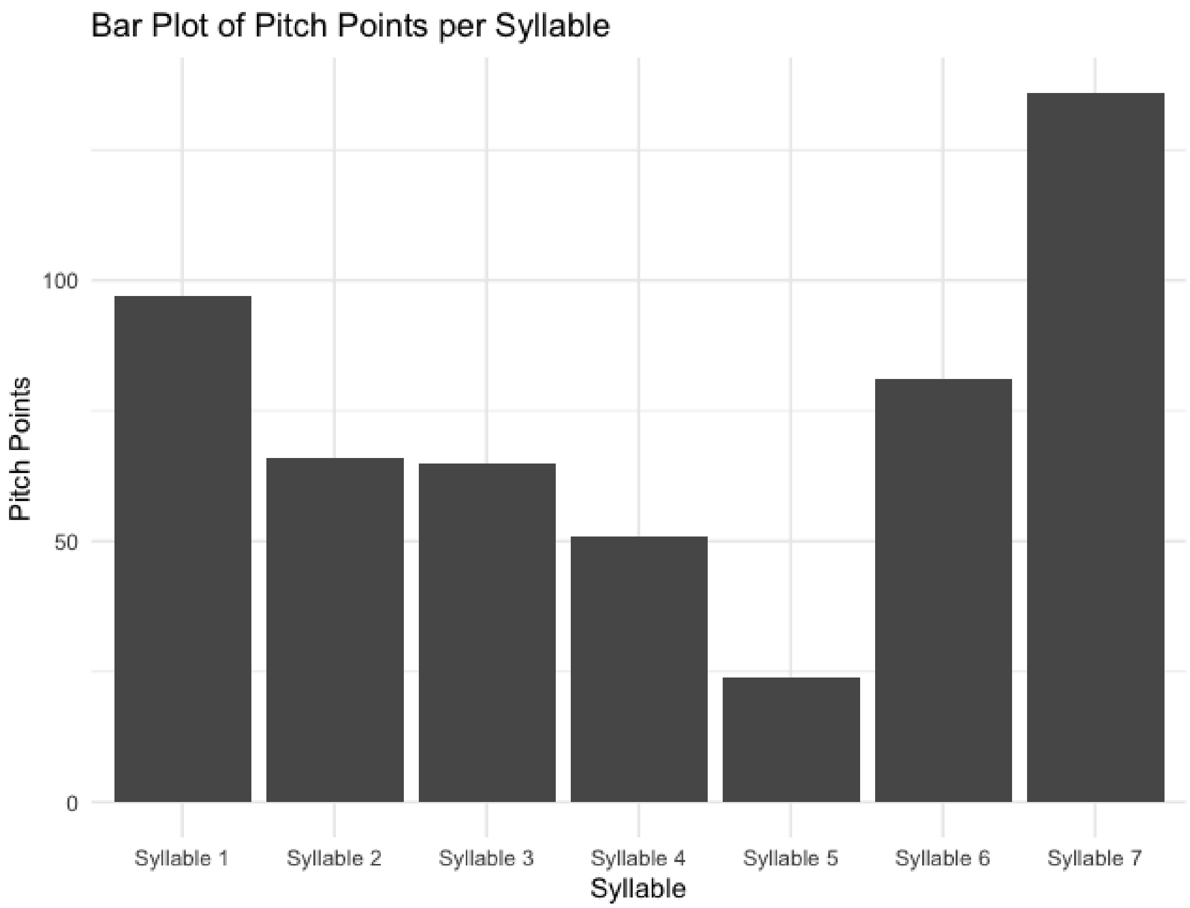

Figure 1 presents a bar plot illustrating the frequency of pitch points across each of the seven syllables. The height of each bar reflects the number of pitch points for a given syllable, providing a comparison of their occurrence. In Syllable 1, the higher bar suggests a frequent presence of lower pitch points at the beginning of phrases. Conversely, the highest bar for Syllable 7 indicates a higher pitch point towards the end of phrases.

Figure 2. Pitch Points Across Each Syllable (1-7) By File Types

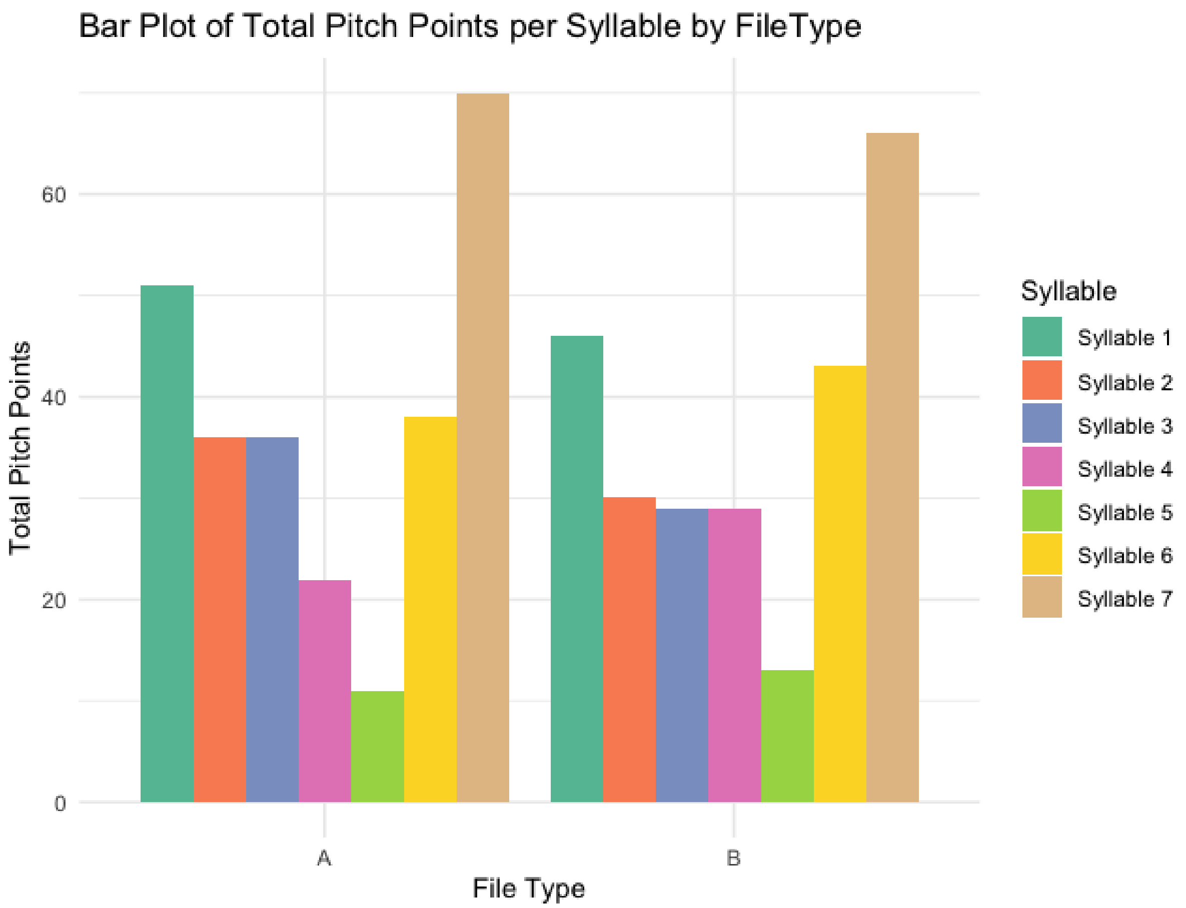

Figure 2 presents a comparative bar plot showcasing the total pitch points counted for each syllable across two types of audio files, A and B. In the dataset, there are 48 A-Type and 47 B-Type files. The bars reveal a pattern in the distribution of pitch points: both A and B file types exhibit a higher count in the first syllable, with the seventh syllable showing the most number of pitch points. This trend potentially reflects a linguistic pattern of rising intonation towards the end of phrases.

Figure 3. Pitch Points Frequency Across Each Syllable (1-7) By Sex

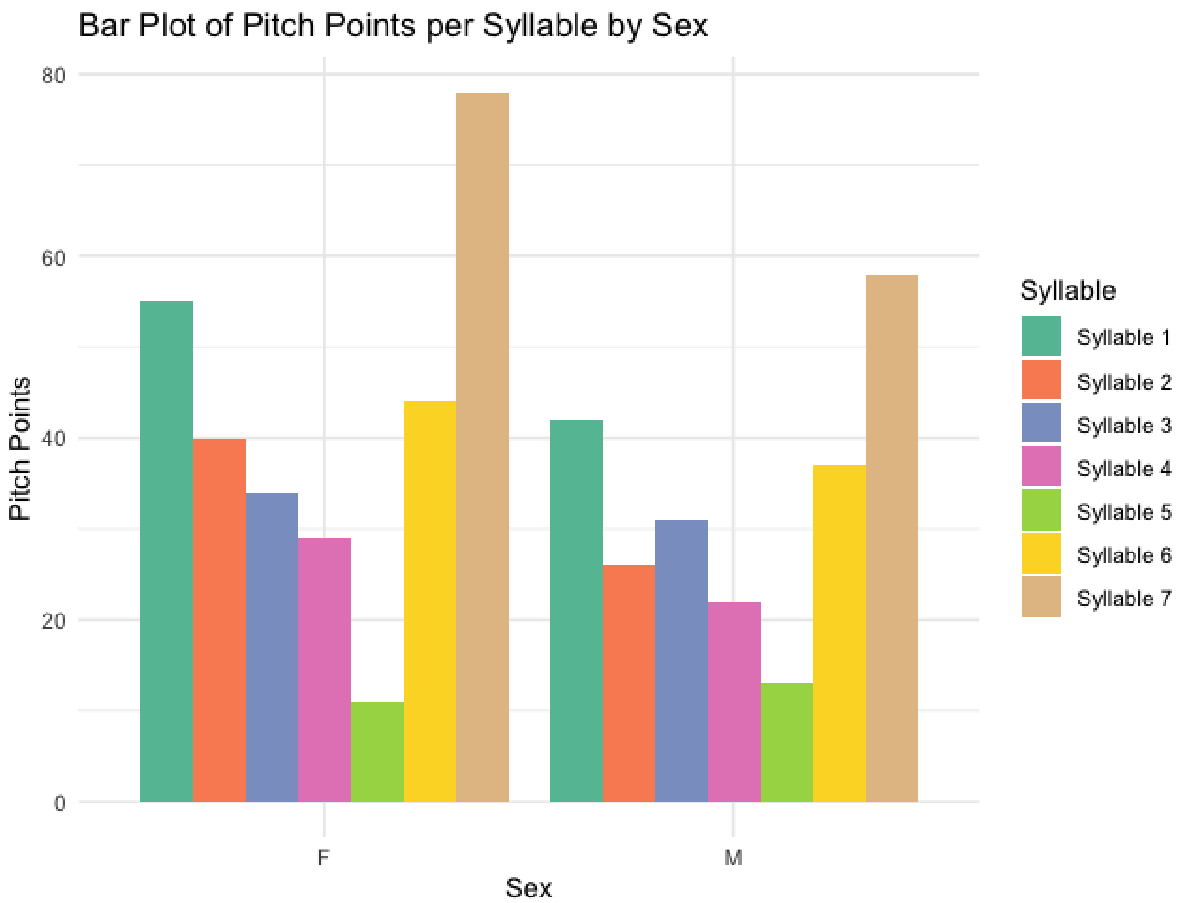

Figure 3 offers a bar plot that delineates the pitch point distribution across syllables, separated by gender, with 'F' representing female and 'M' representing male speakers. The dataset features a comparable number of female (52) and male (43) speakers. Upon examination, both genders exhibit a similar distribution pattern of pitch points throughout the syllables, with no marked gender-specific trends. Notably, the first and the last syllables consistently show higher pitch point counts for both genders.

Figure 4. Mean Pitch Points Frequency Across Each Syllable (1-7) By Final Syllable Type

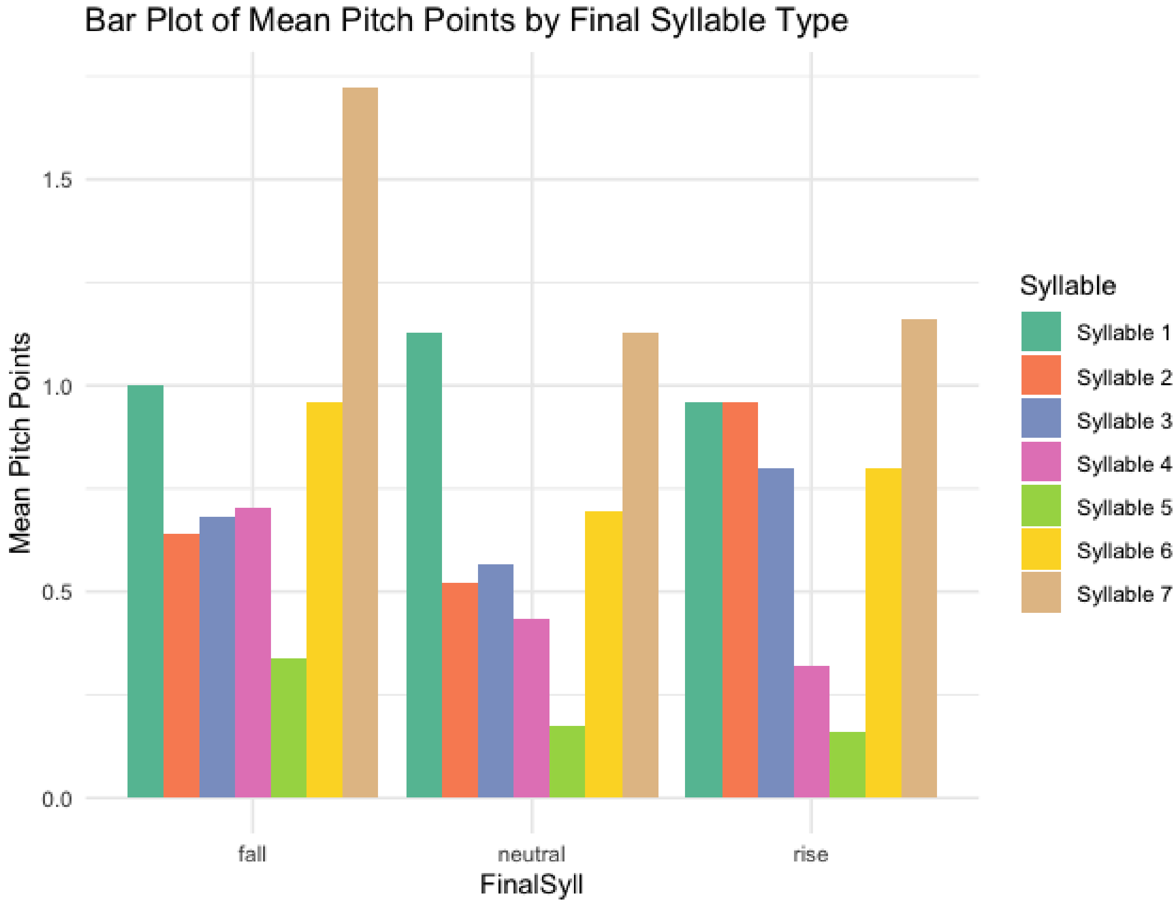

Figure 4 presents a bar plot that categorizes the distribution of mean pitch points by the type of final syllable intonation — "fall," "neutral," and "rise." The dataset records a mean of 47 pitch points for the "fall" intonation, 23 for the "neutral" intonation, and 25 for the "rise" intonation at the end of phrases. The bar plot reveals that the seventh syllable has the highest concentration of pitch points within the "fall" intonation category, towering over its counterparts in the "neutral" and "rise" categories. This pronounced peak in the "fall" pattern indicates that speakers tend to exhibit a significant increase in pitch variation at the end of phrases that finish with a falling intonation.

Methodology

In our study, the primary goal is to identify the specific syllables within spoken questions where pitch changes occur most frequently. Our methodology is designed to pinpoint these critical points in speech, using two main stages:

Point Estimation Analysis

We begin by analyzing the dataset to understand baseline characteristics of speech, focusing on the pitch. This step helps us identify general trends and variations in voice pitch, providing a foundation for more detailed analysis.

Hypothesis Testing

After establishing the baseline, we conduct hypothesis testing to compare pitch changes across different syllables in speech.

- ANOVA (Analysis of Variance): This is used for normally distributed data to compare pitch variations among different syllables within speech groups or conditions.
- Kruskal-Wallis Test: For data that does not follow a normal distribution, this test helps compare median pitch values across different syllables.

Post hoc analysis is crucial in this phase to explore the specifics of where these pitch changes are most pronounced. Through this approach, we aim to pinpoint the syllables in speech, particularly in questions, that are most prone to pitch changes.

Table 1. Point Estimates with 95% Confidence Intervals

Point Estimates with 95% Confidence Intervals

variable	Mean	SD	n	CI
Counted Pitch Points in syllable 1	1.021	0.437	95	(0.932, 1.110)
Counted Pitch Points in syllable 2	0.695	0.547	95	(0.583, 0.806)
Counted Pitch Points in syllable 3	0.684	0.570	95	(0.568, 0.800)
Counted Pitch Points in syllable 4	0.537	0.598	95	(0.415, 0.659)
Counted Pitch Points in syllable 5	0.253	0.437	95	(0.164, 0.342)
Counted Pitch Points in syllable 6	0.853	0.583	95	(0.734, 0.971)
Counted Pitch Points in syllable 7	1.432	0.595	95	(1.310, 1.553)

In the point estimation phase of our study, we isolated and analyzed pitch points within each syllable. Table 1 shows that Syllable 1 has a mean pitch point count of 1.021 with a 95% confidence interval (CI) between 0.932 and 1.110. Syllable 7 stands out with the highest mean count of 1.432 and a 95% CI from 1.310 to 1.553. These two syllables, both exceeding a mean count of 1, indicate a higher frequency of pitch changes that may be particularly significant in the modulation of speech during questions.

Setting Hypothesis

As part of our analysis on pitch point variations across different syllables in spoken questions, we set forth the following statistical hypotheses:

For ANOVA (Analysis of Variance):

- H0: $\mu 1=\mu 2=\mu 3=\mu 4=\mu 5=\mu 6=\mu 7$
- HA: $\mu i \neq \mu j$ for at least one pair (i, j)

The null hypothesis posits that the average number of pitch points for all seven syllables is statistically equivalent. In other words, no syllable differs from the others in terms of its mean number of pitch points. The alternative hypothesis suggests that there is a difference in the average number of pitch points for at least one syllable compared to the others.

For the Kruskal-Wallis Test:

- H0: The median number of pitch points for syllables 1 through 7 is statistically equivalent. This implies that the central tendency of pitch points across the syllables does not vary.
- HA: The median number of pitch points for at least one syllable is significantly different from the others. This suggests that there is a variation in the pitch point distribution among the syllables.

The selection between ANOVA and the Kruskal-Wallis Test will be determined by the distribution characteristics of our data. The ANOVA is appropriate for data that can be assumed to follow a normal distribution, whereas the Kruskal-Wallis Test is suitable for data that do not follow a normal distribution. Following the hypothesis testing, if the null hypothesis is rejected, we will proceed with post hoc analyses to identify which syllables differ.

ANOVA Test with Checking Assumptions

Figure 5. Assumption of Normality

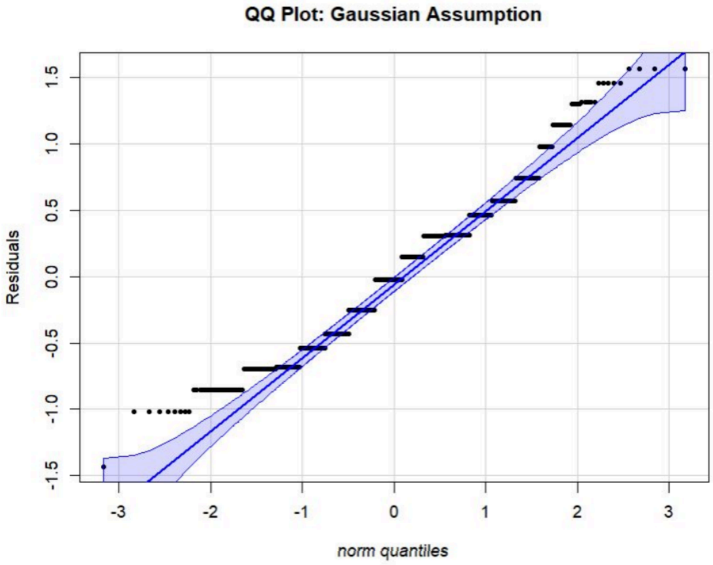

Figure 5 provides a Q-Q plot that illustrates that the data deviates from the expected normal distribution. Notably, the heavy tails and deviation from the reference line show an increased occurrence of extreme values.

In an earlier section, we predicted that our data might not fit the classic bell curve, which would mean that standard tests like ANOVA might not be suitable. Figure 5 confirms this prediction, showing that our pitch data doesn't follow the expected pattern for such tests. Because of this, we're choosing the Kruskal-Wallis Test instead. This test works well for data that doesn't meet the normality assumption. By using this test, we can trust our analysis to give us a true picture of the pitch variations, even with the unusual data pattern we have.

Kruskal Wallis Test

Table 2.Summary of Kruskal-Wallis Test

Kruskal-Wallis Test Result			
Test	Chi_Squared	Degrees_of_Freedom	p_value
Kruskal-Wallis	191.947	6.000	0

Table 2 shows the Kruskal-Wallis test that was employed due to the non-normal distribution of our data and the potential for unequal variances among groups. The p-value from this test, confirms that the median pitch points for the syllables we examined are statistically significantly different from each other. This finding is crucial for our study as it validates the hypothesis that there is a variation in pitch among the syllables, a variation that is not attributed to random chance but to a meaningful difference in the audio data we have collected and analyzed.

Conclusion

Our capstone project explored pitch variations in syllables of spoken questions, discovering significant differences, especially in syllables 1 and 7, which are key to speech modulation. This capstone project improves understanding of speech intonation and has broad implications, including improving speech recognition technology. The findings also suggest future research directions, such as exploring the interaction of pitch variations with demographic factors. Overall, this study not only improves our comprehension of the complexities of human speech but also opens up new possibilities for practical applications and further research in speech technology, linguistics, and communication studies.

Code Appendix

```r
# 0. Set working directory and load data
setwd("~/Desktop/JH/PSU/23 Fall/STAT 470w/Case Study 3")
knitr::opts_chunk$set(echo = TRUE)

library(vroom)
library(tidyverse)
library(dplyr)
library(reshape2)
library(tidyr)
library(ggplot2)
library(stringr)
library(car)
library(gt)
library(dunn.test)
library(PMCMRplus)

Intonation <-  vroom("Intonation.csv", col_types = c(numPoints = "i"))


# 1. Data Preprocessing

# 1.1 Extract those characters from filenames (A or B)
df <- Intonation %>%
  mutate(FileName = gsub(".*_(A|B)_.*", "\\1", FileName))

# 1.2 Remove Pt9time column (NA)
df <- df %>% select(-Pt9Time)

# 1.3 Add Columns - syllable.x.ptNum
calculate_ptNum <- function(start, end, ptTimes) {
  sum(ptTimes >= start & ptTimes <= end, na.rm = TRUE) # start와 end 사이에 있는 ptTimes의 개수를 카운트합니다.
}

df <- df %>%
  rowwise() %>%
  mutate(
    syllable.1.ptNum = calculate_ptNum(syllableStart.1, syllableEnd.1, c(Pt1Time, Pt2Time, Pt3Time, Pt4Time, Pt5Time, Pt6Time, Pt7Time, Pt8Time)),
    syllable.2.ptNum = calculate_ptNum(syllableStart.2, syllableEnd.2, c(Pt1Time, Pt2Time, Pt3Time, Pt4Time, Pt5Time, Pt6Time, Pt7Time, Pt8Time)),
    syllable.3.ptNum = calculate_ptNum(syllableStart.3, syllableEnd.3, c(Pt1Time, Pt2Time, Pt3Time, Pt4Time, Pt5Time, Pt6Time, Pt7Time, Pt8Time)),
    syllable.4.ptNum = calculate_ptNum(syllableStart.4, syllableEnd.4, c(Pt1Time, Pt2Time, Pt3Time, Pt4Time, Pt5Time, Pt6Time, Pt7Time, Pt8Time)),
    syllable.5.ptNum = calculate_ptNum(syllableStart.5, syllableEnd.5, c(Pt1Time, Pt2Time, Pt3Time, Pt4Time, Pt5Time, Pt6Time, Pt7Time, Pt8Time)),
    syllable.6.ptNum = calculate_ptNum(syllableStart.6, syllableEnd.6, c(Pt1Time, Pt2Time, Pt3Time, Pt4Time, Pt5Time, Pt6Time, Pt7Time, Pt8Time)),
    syllable.7.ptNum = calculate_ptNum(syllableStart.7, syllableEnd.7, c(Pt1Time, Pt2Time, Pt3Time, Pt4Time, Pt5Time, Pt6Time, Pt7Time, Pt8Time))
  ) %>%
  ungroup()
```

```r
# 2. EDA section

# 2.1. Change the dataset into long format
df_long <- df %>%
  pivot_longer(
    cols = starts_with("syllable."),
    names_to = "Syllable",
    names_prefix = "syllable\\.",
    values_to = "ptNum"
  ) %>%
  mutate(Syllable = str_replace(Syllable, "\\D+", ""),
         Syllable = paste("Syllable", Syllable))

# 2.2. Figure 1 - Bar Plot
ggplot(df_long, aes(x = Syllable, y = ptNum)) +
  geom_bar(stat = "identity") +
  labs(title = "Bar Plot of Pitch Points per Syllable", x = "Syllable", y = "Pitch Points") +
  theme_minimal()


# 2.3. Figure 2
df_summary <- df_long %>%
  group_by(FileName, Syllable) %>%
  summarise(TotalPtNum = sum(ptNum, na.rm = TRUE)) %>%
  ungroup()

ggplot(df_summary, aes(x = FileName, y = TotalPtNum, fill = Syllable)) +
  geom_bar(stat = "identity", position = position_dodge()) +
  labs(title = "Bar Plot of Total Pitch Points per Syllable by FileType", x = "File Type", y = "Total Pitch Points") +
  theme_minimal() +
  scale_fill_brewer(palette = "Set2")

```

```r
# 2.4. Figure 3 - Updated to Bar Plot
df_summary2 <- df_long %>%
  group_by(Sex, Syllable) %>%
  summarise(TotalPtNum = sum(ptNum, na.rm = TRUE)) %>%
  ungroup()

ggplot(df_summary2, aes(x = Sex, y = TotalPtNum, fill = Syllable)) +
  geom_bar(stat = "identity", position = position_dodge()) +
  labs(title = "Bar Plot of Pitch Points per Syllable by Sex", x = "Sex", y = "Pitch Points") +
  theme_minimal() +
  scale_fill_brewer(palette = "Set2")


# 2.5. Figure 4 - Updated to Bar Plot
df_summary3 <- df_long %>%
  group_by(FinalSyll, Syllable) %>%
  summarise(TotalPtNum = mean(ptNum, na.rm = TRUE)) %>%
  ungroup()

ggplot(df_summary3, aes(x = FinalSyll, y = TotalPtNum, fill = Syllable)) +
  geom_bar(stat = "identity", position = position_dodge()) +
  labs(title = "Bar Plot of Mean Pitch Points by Final Syllable Type", x = "FinalSyll", y = "Mean Pitch Points") +
  theme_minimal() +
  scale_fill_brewer(palette = "Set2")
```

```r
# 5. Kruskal-Wallis test
kruskal_result <- kruskal.test(ptNum ~ Syllable, data=df_long)

kruskal_df <- data.frame(
  Test = "Kruskal-Wallis",
  Chi_Squared = kruskal_result$statistic,
  Degrees_of_Freedom = kruskal_result$parameter,
  p_value = kruskal_result$p.value
)

# Format the p-value
kruskal_df$p_value <- format(kruskal_df$p_value, digits = 4)

kruskal_table <- gt(kruskal_df) %>%
  tab_header(
    title = "Kruskal-Wallis Test Result",
    subtitle = ""
  ) %>%
  fmt_number(
    columns = everything(),
    decimals = 3
  )

# Print the table
print(kruskal_table)
```