

LaTeX submissions are mandatory. Submitting your assignment in another format will be graded no higher than R.

1 Lab Book A5

In this Lab Book assignment you will investigate the interaction of fundamental frequency and lexical tone in Mandarin Chinese ¹

In Chinese f_0 has at least two linguistic functions: to mark lexical tone, where different pitch patterns distinguish different words, and intonation, when phrases that semantically differ might be distinguished based on pitch. But what happens when different pitch requirements of tone and intonation happen to be on a single word? To investigate this you have 24 sentences in Mandarin Chinese: 4 minimal question / statement sentence pairs produced by 3 speakers. The last word of each sentence has either a high rising tone (tone 2) or a high falling tone (tone 4).

1.1 Submission

For this submission you will need to commit the LaTeX file as usual, as well as plots of f_0 . Therefore, the completed commit for this assignment includes:

1. LaTeX file that includes the plots, your hypotheses (e.g., about what you can see on the plots), statistical results regarding the question, and a very brief discussion of your conclusions.
2. Your code (you commit your code file alongside your LaTeX file).
3. Plots.

1.2 Assignment 1

[10pts] *Intonation and tone:*

Preparation:

In the dataset ([that can be found here](#)) the file names are formed as a combination of sentence type (st for statement, q for question), tone, word, speaker. E.g., “st_4_lu_S3” means it is a statement with the fourth-tone word *lu* (“deer”) on the last syllable, produced by Speaker 3. Detailed descriptions are in the tables below.

For the analysis part you should focus your attention on the last word of the sentence. For simplicity you can assume that the last 15% of each recording correspond to the last word.

Task:

- To address the problem you need to track f_0 (you can use the Praat script that is committed alongside this Lab Book - I checked it in Windows, works fine with standard Windows paths - or you can try using Robust Algorithm for Pitch Tracking, RAPT, from [pysptk^a](#)). You will have to plot the f_0 contours of the sentences and perform simple statistical analysis to answer the following question: **How intonation influences tone realisation in Mandarin Chinese?**
- To answer this question, first separately plot f_0 for questions and statements with different final tones: tones 2 and 4. ^b
- Examine the plots and think of diverse ways of assessing f_0 behaviour, such as range, mean, minimum, maximum. Think of hypothesis/hypotheses about f_0 .

¹The assignment is inspired by and based on the assignments and materials provided by Dr. Rebecca Scarborough in her Stanford course on phonetics from 2005. The data come from the dissertation of Jiahong Yuan, Yuan, J. (2004). *Intonation in Mandarin Chinese: Acoustics, perception, and computational modeling*. New York: Cornell University, p.416

- Comparing f_0 measurements of questions and statements using a simple ANOVAs or alternatives (e.g., in this task a Kruskal-Wallis test for non-parametric data seems to be a good choice) would be enough for this assignment.^c
- Briefly describe the results. Mention any limitations or difficulties you faced during the assignment.

You may do this task in whatever programming language you prefer: R, Python, or anything else. If you are struggling with plotting, you may use Praat to draw contours.

If you are using Python, you may consider using `numpy` or `pandas` for convenient operations with data sets, `scipy.stats` for statistical algorithms and `matplotlib` for making plots. At the end of the document you can find some Python code snippets to help you if you have chosen Python as a tool for analysis.

^aIf you are going to use RAPT, set hopsize for 80, min f_0 for 75, max f_0 for 600

^bIt is a good idea to normalize f_0 by dividing each f_0 measurement in each recording by mean f_0 of the same recording. This would allow us to ignore the f_0 difference between speakers and get nicely aligned pictures. To have the graphs match even better, try dividing the time by the length of the record, so every graph has a “normalised time” from 0 to 1.

^cThere are multiple sources you can look into for these tests, a famous and fun book about statistics is by Andy Field, Jeremy Miles and Zoe Field, and you can check it out [here](#).

Table 1: **Statements**

File name	Sentence	Literal translation
st_2_niu	Luo2yan4 li3bai4wu3 mai4 ye3niu2	Luoyan Friday sells wild cows. ("Luoyan sells wild cows on Friday.")
st_2_yang	Li3bai4wu3 Luo2yan4 yao4 mai3 yang2	Friday Luoyan will buy sheep. ("On Friday, Luoyan will buy sheep.")
st_4_lu	Luo2yan4 li3bai4wu3 mai4 ye3lu4	Luoyan Friday sells wild deer. ("Luoyan sells wild deer on Friday.")
st_4_la	Li3bai4wu3 Luo2yan4 yao4 mai3 la4	Friday Luoyan will buy candles. ("On Friday, Luoyan will buy candles.")

Table 2: **Echo questions**

File name	Sentence	Literal translation
q_2_niu	Luo2yan4 li3bai4wu3 mai4 ye3niu2	Luoyan Friday sells wild cows. ("Luoyan sells wild cows on Friday?")
q_2_yang	Li3bai4wu3 Luo2yan4 yao4 mai3 yang2	Friday Luoyan will buy sheep. ("On Friday, Luoyan will buy sheep?")
q_4_lu	Luo2yan4 li3bai4wu3 mai4 ye3lu4	Luoyan Friday sells wild deer. ("Luoyan sells wild deer on Friday?")
q_4_la	Li3bai4wu3 Luo2yan4 yao4 mai3 la4	Friday Luoyan will buy candles. ("On Friday, Luoyan will buy candles?")

Answer

1.3 Visualization

REPLACE THIS TEXT WITH YOUR ANSWER.

1.4 Hypothesis (e.g., based on the plots)

REPLACE THIS TEXT WITH YOUR ANSWER.

1.5 Statistical analysis

REPLACE THIS TEXT WITH YOUR ANSWER.

1.6 Brief discussion and conclusion

REPLACE THIS TEXT WITH YOUR ANSWER.

Python tips (for a simplistic solution)

1. Function to read results of pitch tracking from Praat, where “np” is imported numpy (`import numpy as np`): [can be found here](#).
2. Function to measure end tones, [can be found here](#).
3. Plotting tones together and measuring end tones [can be found here](#), where:
 - “pd” is imported pandas (`import pandas as pd`),
 - “plt” (`from matplotlib import pyplot as plt`),
 - `data_dir = Path('data')`,
 - “end_of_sentence_fraction” is a parameter defining which “end portion” of the sentence you analyse (it is for you to define).