The Iambic/Trochaic Law in Mandarin Tone Sandhi

Sijia Zhang,
Gemma Huang,
and Yijie Yan

Abstract

According to the Iambic/Trochaic Law (i.e. ITL), there is a universal perceptual bias that a sequence of (non-)speech signals varied in duration leads to iambic rhythmic grouping, whereas signals with contrasting intensities result in trochaic grouping. While ITL has been tested cross-linguistically in previous studies, it remains mostly unexplored how ITL might be applied in tone languages such as Mandarin Chinese, where native speakers may use their phonological knowledge of tone sandhi in addition to duration and intensity cues for lexical grouping. This study examines how Tone 3 sandhi cues interact with duration cues, and to what extent both cues might influence Mandarin listeners' perception of grouping and prominence, as well as naturalness judgement. In the experiment, participants hear streams of repetitions of existing bisyllabic Mandarin words with manipulated duration and with either the underlying Tone 3 or the sandhi Tone 2 on each syllable. The results from all the three tasks suggest that the tone sandhi cues are stronger than the duration cues in its effect on grouping and prominence judgement. Duration does have some effects on prominence, especially when duration cues agree with sandhi cues. In terms of naturalness judgement, sequences with sandhi cues are judged to be significantly more natural than sequences without sandhi cues. However, our current results are insufficient for any conclusion regarding duration's effects on the naturalness rating.

1 Introduction

This study investigates how phonetic and phonological cues interact with each other and what roles they play in listeners' perceptions of grouping and prominence. According to the Iambic/Trochaic Law (i.e. ITL), there is a universal perceptual bias that a sequence of speech or non-speech signals varied in duration leads to iambic rhythmic grouping, whereas the signals with contrasting intensities result in trochaic grouping (Hay & Diel, 2007). Even though previous research on ITL has tested the perceptual grouping of native speakers of English and French (Hay & Diehl, 2007), French and German (Bhatara et al., 2013), Spanish and English (Crowhurst & Olivares, 2016), English and Japanese (Iversen et al., 2008), the application of ITL in tone languages, however, remains mostly unknown. In tone languages spoken in Asia such as Mandarin Chinese, tones are associated to each syllable in a word and they are used to make lexical contrasts, such as in monosyllabic words ma (T1) 'mom', ma (T2) 'hemp', ma (T3) 'horse', ma (T4) 'scold'. The type of a tone may alternate depending on the types of adjacent tones or the prosodic and/or morphosyntactic position in which the tone occurs (Chen, 2000). This phonological process of alternation of tones in different contexts is called tone sandhi. In tone sandhi processes, citation syllable refers to the syllable on which the tone does not change, and sandhi syllable is the one on which the tone type alters. There are two types of tone sandhi concerning the relative positions of the citation syllable and the sandhi syllable in a disyllabic unit. Right dominant tone sandhi refers to tonal changes observed on the left syllable whereas left dominant tone sandhi on the right. Yiu (2019) points out that the distinction between left and right tonal prominence displayed in bidirectional tone sandhi resembles the metrical distinction between iambic and trochaic rhythm. The application of ITL in Southern Min is investigated in Yiu (2019), and it is found that the metrical prominence predicted by the Iambic/Trochaic Law aligns with tonal prominence. Specifically, the production test shows that duration contrast occurs in the right dominant tone sandhi type, with longer duration on the right citation syllable, whereas intensity asymmetry occurs in the left dominant tone sandhi type, with higher intensity on the left citation syllable. Yiu concluded that the tone sandhi in Southern Min matches the pattern of the iambic trochaic law, such that the left dominant tone sandhi type matches with the trochaic words, and the right dominant tone sandhi type matches with the iambic words. Citation syllables of different rhythmic types are more prominent in their corresponding phonetic cues, while sandhi syllables are less prominent.

Although some production tests have been conducted by Yiu (2019) on Southern Min, it is still unexplored how ITL might be applied to other tone languages like Mandarin in a perception study. Based on the findings in Yiu (2019), it is expected that the tone sandhi in Mandarin Chinese, which only has the right dominant sandhi type, will pattern with the iambic rhythm and has duration asymmetry. When tone sandhi is applied in certain existing words, Mandarin speakers may perceive words with an iambic rhythm as more natural than those with a trochaic rhythm, if Yiu (2019)'s hypothesis is correct. Additionally, in the case where phonetic cues and tone sandhi cues are both present, it is uncertain whether the acoustic variations will influence the perceptual effect that tone sandhi has on word segmentation. More specifically, since Mandarin tone sandhi is claimed to be right dominant (i.e. pattern with iambs) and has a strong phonological cue to grouping, the acoustic cue of duration contrasts may also affect Mandarin listeners' perception.

Moreover, the naturalness and the perceptual prominence of the existing words may be affected when the duration cue is manipulated. We hypothesize that tone sandhi cue has a stronger effect on grouping than the duration cue, yet when only the duration cue is present, the listeners may still perceive the rhythm as iambic. We also hypothesize that both sandhi and duration cues can influence perceived naturalness and prominence judgement. We conducted a perception study to examine to what extent sandhi cues and duration cues might influence Mandarin speakers' perception of grouping and prominence, as well as naturalness judgement.

2 Methodology

2.1 Participants

Sixteen native speakers of Mandarin took part in the experiment. They were all students at McGill University and lived in Montreal when the experiments took place. Most of them were fluent L2 learners of English, and some had knowledge of French. In addition, we recorded each participant's years of musical experience through a music questionnaire as musical experience evidently aids in perceiving lexical stress (Kolinsky et al., 2009; Bhata, 2013).

2.2 Stimuli

The third tone sandhi in Mandarin was tested in this experiment. The third tone sandhi refers to the change of Tone 3 (213) to Tone 2 (35) when preceding another Tone 3, for example, the word ke3.kou3 'delicious' becomes ke.2kou3 in speech. We selected existing bisyllabic Mandarin words with underlying Tone 3 on each syllable. Additionally, when the order of the two syllables is reversed, they still form an existing word albeit with a different meaning. We selected a total of four reversible bisyllabic words as follows:

1. ke3.kou3 - kou3.ke3 ("delicious" – "thirsty")
2. nü 3.zi3 - zi3.nü 3 ("woman" – "offspring")
3. dian3.jiu3 - jiu3.dian3 ("iodin" – "9 o'clock")
4. cai3.shui3 - shui3.cai3 ("to tread water" – "watercolour")

We recorded those item pairs in three different tonal combinations: T3T3 (underlying form), as well as T2T3 and T3T2 (two surface forms in reversed orders). The intensity of those recordings was scaled to the same level (i.e. 78 dB) and the duration of each syllable was adjusted to the same length (i.e. 425ms). The duration contrast in the two syllables was manipulated in two steps (step one: length of one syllable -0.075s; step two: length of the other syllable +0.075s) for each tonal combination, resulting in four duration conditions, namely, "short-short", "short-long", "long-long", "long-short". We generated sequences of those item pairs in repetition with a Praat script, for instance, the sequences generated from ke3kou3-kou3ke3 item pair are shown as follow:

"ke3kou3ke3kou3ke3kou3..." in short-short, short-long, long-short, long-long
"ke2kou3ke2kou3ke2kou3..." in short-short, short-long, long-short, long-long
"kou2ke3kou2ke3kou2ke3..." in short-short, short-long, long-short, long-long

Each recording lasted twelve seconds, and a shading noise of around 3.3s was added at the beginning for each stimulus. To counterbalance the order effect, we generated two sets of sequences of both syll1 syll2 and syll2 syll1 for each item, and they were coded as "Condition1" and "Condition2". To sum up, 4 item pairs, each with 12 different combinations of tones and duration manipulations plus 2 different orders, resulted in a total of 96 sequences of stimuli for the entire experiment.

2.3 Procedure

Participants were seated in a quiet room, and the stimuli were presented at a comfortable listening level using MatLab software on a desktop in Prosody Lab at McGill. Participants were asked to listen carefully to the recordings of Mandarin word sequences first, and then they were tasked to answer the following three questions for each trial in order: (1) select the word they heard in the recording, (2) how natural this word sounded in Mandarin on a scale from 1 to 8 (1=very unnatural and 8= very natural), and (3) which syllable was more prominent in the word they heard. The instruction and tasks were all presented in Mandarin Chinese. The recording cannot be replayed at the question page. In the instruction page, we listed out the three tasks that they will be asked to do during the experiment so that the participants will know what to pay attention to in the audio recordings.

We used LatinSquare design to ensure each participant will see one order condition from each item set, and an equal number of trials from each condition across items, so each participant needed to complete 48 trials in the experiment. This was done by creating two playlists with a selection of trials from the two conditions for the two orders. Each playlist was randomized for each participant, and the same stimulus was not repeated more than once.

3 Results

The data were analyzed and visualized in R using ggplot2 packages. In the plots below, for each condition on the horizontal axis, the first manipulation is always on Syllable 1, and the second manipulation is on Syllable 2. For example, tonal condition 23 indicates Tone2 on Syllable 1 and Tone3 on Syllable 2; duration condition short-long means Syllable 1 is short and Syllable 2 is long.

3.1 What word did you hear?

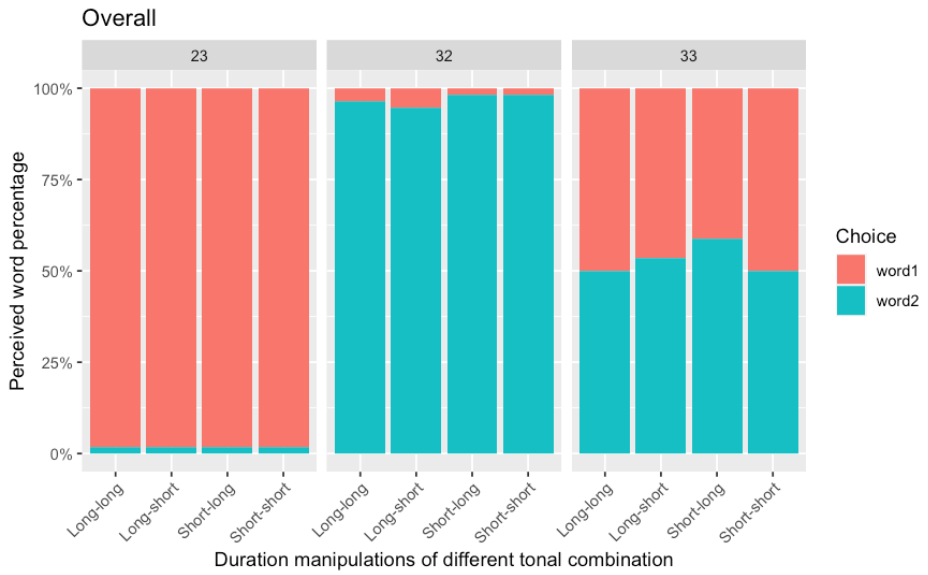
Predictions:

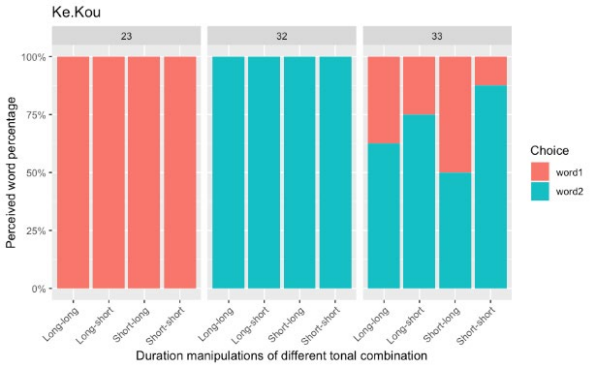
If the sequence of stimuli consists of the T2T3 or T3T2 tonal combination (i.e. combination that resembles tone sandhi process), then we predict that regardless of the duration manipulations, the participants will most likely hear the word that corresponds to the surface

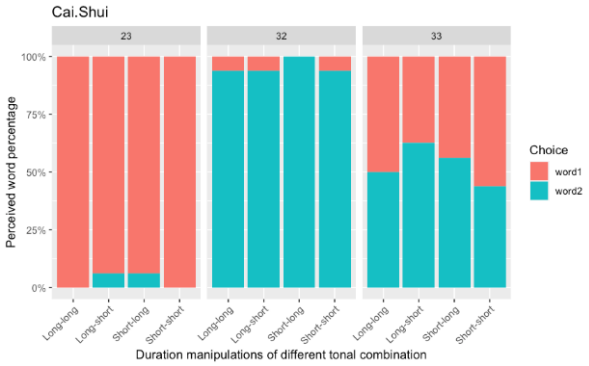
form after tone sandhi. If the sequence of stimuli consists of the T3T3 tone combination (i.e. the underlying form), then there is a 50% chance that the participants will hear either word when there is no duration contrast; they will group the sequence iambically when the duration contrast is present.

Observations:

For the sequence of stimuli consisting of T2T3 tone combination, the overall plot in Figure 3.1 shows a high percentage of choosing the surface forms after tone sandhi regardless of duration manipulations as expected. When the sequence of stimuli consists of the T3T3 tone combination, however, the overall plot shows that duration contrast does not have a significant effect on grouping when there is no sandhi process. In addition, in some cases, the duration contrast has a slightly trochaic effect. From the plots of caishui and kekou in Figure 3.1, Word 1 with the short-long manipulation has less percentage of selection than that of Word 2.







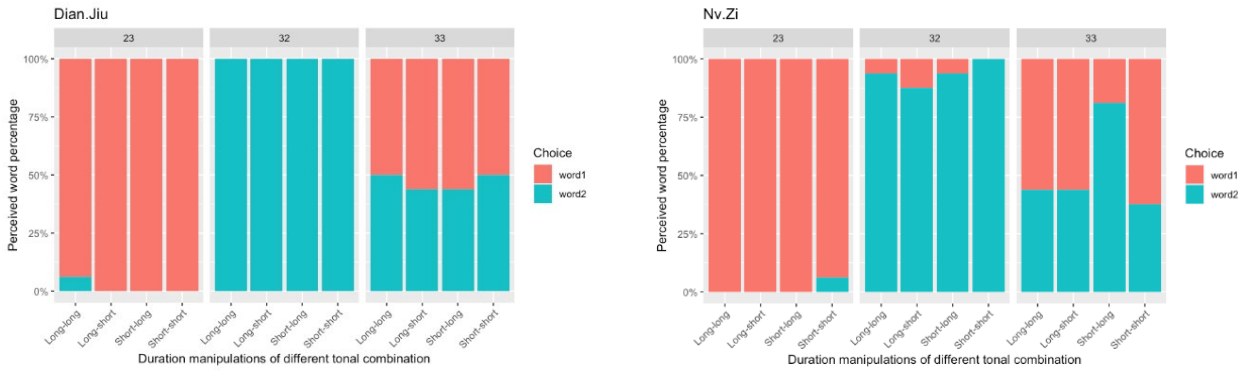


Figure 3.1 Plots for grouping question results. The overall results as well as results for each item pair are presented. The x-axis shows duration manipulations of different tonal combinations, and the y-axis exhibits the percentage of perceived words by Mandarin speakers.

3.2 Naturalness

Predictions:

It is expected that the stimulus with sandhi rules applied and surfacing as T2T3 or T3T2 will be judges as more natural than those with underlying T3T3 tones by Mandarin speakers. Following Yiu (2019)'s finding, Mandarin third tone sandhi should pattern with iambic rhythm. Among conditions of 23 or 32 tonal combinations, participants should perceive stimulus that have longer duration on the syllable with T3 as more natural, and stimulus that have longer duration on the syllable with T2 as less natural, according to the rhythmic nature of iambs.

Observations:

Our results exhibit that words consisting of the T2T3 or T3T2 tone combination are more natural compared to words with T3T3 underlying forms as shown in Figure 3.2. When tone sandhi cues are present, there is a slight trend that short-long condition of T2T3 combination has higher-rated naturalness than long-short condition, and long-short condition of T3T2 combination has higher-rated naturalness than short-long condition. When tone sandhi cues are absent (i.e. T3T3 group), sequences with short-long contrast are perceived as more natural than those with long-short contrast by a small margin. Nevertheless, the naturalness judgement is mostly based on the third-tone sandhi cues instead of the duration cues.

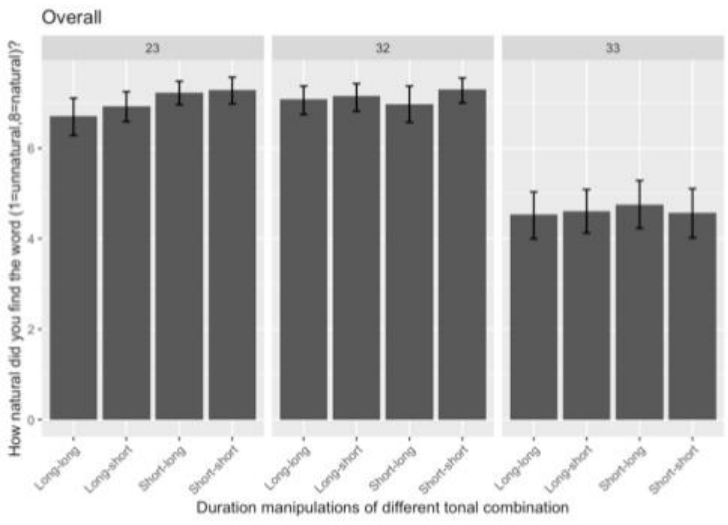


Figure 3.2 The plot for overall naturalness question responses. The plots for responses of individual item pairs are shown in the appendix. The x-axis shows duration manipulations of different tonal combinations, and the y-axis exhibits the mean values of rated naturalness.

3.3 Which syllable is more prominent?

Predictions:

Duration variations may serve as a cue to prominence. Syllables with longer duration might be judged as more prominent by the listeners. In addition, syllables with T3 (i.e. the citation syllable) might be judged as more prominent than syllables with T2 (i.e. the sandhi syllable).

Observations:

1) Duration effects on prominence

In terms of duration's effects on prominence, we observe that longer syllables are judged to be slightly more prominent than shorter syllables in Figure 3.3 for overall results. In the T3T3 group, when tone sandhi cues are not available, Syllable 1 which has longer duration in long-short condition is more prominent than Syllable 2 (Syllable 1 70%: Syllable 2 30%). However, when the duration of Syllable 2 is longer than Syllable 1 (short-long condition) in the T3T3 group, there is almost no difference in prominence between two syllables (Syllable 1 50%: Syllable 2 50%). Therefore, the duration effect on prominence judgement is a little more random than what we expect.

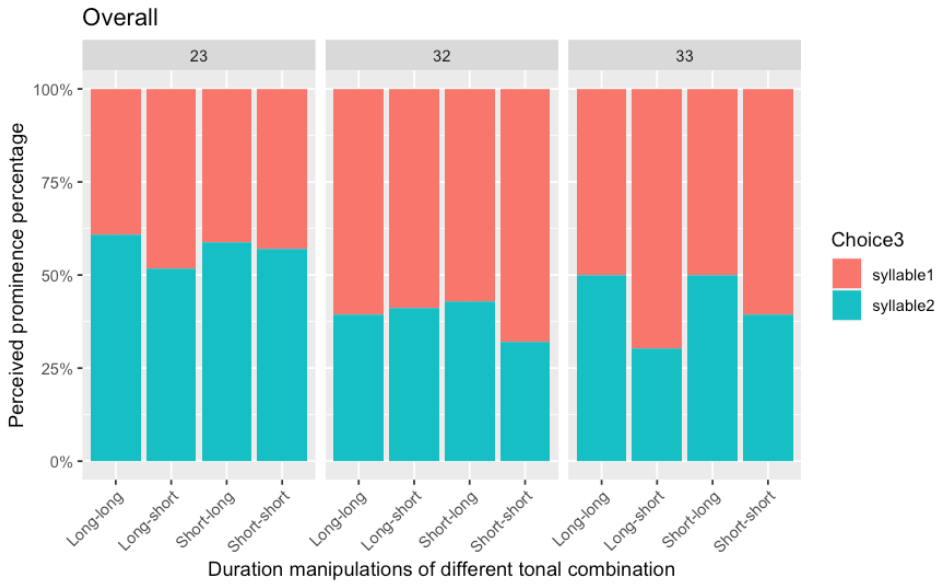


Figure 3.3 The plot for overall prominence question responses. The plots for responses of individual item pairs are shown in the appendix. The x-axis shows duration manipulations of different tonal combinations, and the y-axis exhibits the overall percentage of perceived prominence.

2) Tone sandhi effects on prominence

In terms of the tone sandhi effects on prominence, we observe that syllables with T3, the underlying tone, are more prominent than syllables with T2, the sandhi tone, as shown in Figure 3.3. When Syllable 2 has underlying T3 and Syllable 1 has sandhi T2 (i.e. T2T3 group), the percentages of perceived prominence of Syllable 2 are in general more than 50%. In other words, Syllable 2 with T3 has an overall higher percentage of prominence than Syllable 1 with T2. Similarly, When Syllable 1 has underlying T3 and Syllable 2 has sandhi T2 (i.e. T3T2 group), the prominence percentages of Syllable 1 are higher than those of Syllable 2.

3) Cooperating conditions

We also notice that when duration cues cooperate with tone sandhi cues, the prominence difference is more significant. Cooperation refers to the conditions wherein both tone sandhi and duration work towards the same direction on the same syllable, either together increasing or decreasing the prominence of the syllable, as discussed in Crowhurst & Olivares (2016). For example, according to the results above, both a longer duration and underlying T3 on the same syllable may be perceived as more prominent, and both a shorter duration and sandhi T2 on the same syllable may lead to a relatively lower percentage of prominence.

For the T2T3 group in the overall results, when Syllable 2 with the underlying T3 has a longer duration than Syllable 1 (i.e. short-long condition), the perceived prominence percentage of Syllable 2 is higher than Syllable 1 (i.e. 60%: 40%). This prominence difference is more significant than when Syllable 2 has T3 yet with a shorter duration, which shows a competing condition as illustrated in the next paragraph. Similarly, in the T3T2 group, when Syllable 1 has underlying

T3 and a longer duration than Syllable 2 (i.e. long-short condition), the perceived prominence percentage of Syllable 1 is higher than Syllable 2 (62%: 38%).

4) Competing conditions

Competition refers to the conditions wherein tone sandhi and duration work towards opposite directions on the same syllable (Crowhurst & Olivares 2016). For example, one cue serves to increase the prominence of the syllable while the other decreases the prominence. When duration cues compete with tone sandhi cues, tone sandhi wins to a small extent.

For the T2T3 group in the overall result, when Syllable 1 with sandhi T2 has a longer duration than Syllable 2 (i.e. long-short condition), the perceived prominence percentage of Syllable 1 is almost the same as Syllable 2 (50%: 50%). For the T3T2 group, when Syllable 1 with underlying T3 has a shorter duration than Syllable 2 (i.e. short-long condition), the perceived prominence Syllable 1 is higher than Syllable 2 (60%: 40%). Generally speaking, tone sandhi wins duration by a small margin when they compete with each other as prominence cues.

4 Discussion and Conclusion

The results from all the three tasks suggest that the tone sandhi cues are stronger than the duration cues to grouping and prominence. Sequences with the tone sandhi cues are judged to be significantly more natural than sequences without sandhi cues. We notice that when only duration cues are present, there is an overall trend that the naturalness ratings of sequences with short-long contrast are higher than sequences with long-short contrast. However, our current results are insufficient for any conclusion regarding duration's effects on the naturalness rating. Duration does have some effects on prominence, especially when duration cues cooperate with sandhi cues.

Potential methodological problems are discussed as follows, for which some improvement could be done in a further study. First of all, based on our observation, tone sandhi wins as the more prominent cue by a small margin when it competes with duration. The duration cue could potentially be perceived as more robust if the contrast is bigger. In a follow-up study, we could experiment with duration contrasts of different magnitudes and see how the effects differ.

In addition, the design of the question might have given the participants a "forced choice". In the first question, they were asked to select between two lexicons based on the stimuli that they heard. The question itself is making an assumption that they must group the sequence as bisyllabic words. However, for T3T3 sequences, it is possible that their grammar actually prefers having no grouping at all; instead, they might perceive individual T3 monosyllabic words ((T3)(T3)(T3)....). It is probably better to have three options, two being the lexicons and one being "no grouping".

Moreover, the perception test may not be able to capture the listeners' actual "competence". Compared to sandhi cues, the duration effect can be too subtle to be perceived by Mandarin speakers. It is likely that in a production experiment, speakers' spontaneous speech can better show their use of duration cues as stated in Yiu (2019). Additionally, differences in the frequency of usage of items may have also affected our results. An item that is not commonly used is likely to be less frequently grouped as a word when no sandhi cue is given. For example, cai3shui3 (to tread water) has a lower usage frequency in daily life scenarios than shui3cai3 (watercolour). The

results might be more accurate if we use word pairs with similar usage frequency. Besides, we currently use data from only 16 Mandarin speakers. More generalizable results could be obtained if we increase the number of participants and participants' age-range.

Finally, according to ITL, both intensity and duration variations affect the grouping and prominence judgement of a sequence of syllables. However, in this experiment, only duration is manipulated. In a follow-up study, intensity can be manipulated to see its interaction with tone sandhi.

References

Bhatara, Anjali, et al. (2013). Native language affects rhythmic grouping of speech. *The Journal of the Acoustical Society of America*, 134(5), 3828-3843.

Chen, Matthew Y. (2000). Tone sandhi: patterns across Chinese dialects. Cambridge: Cambridge University Press.

Crowhurst, M. J., & Olivares, A. T. (2014). Beyond the Iambic-Trochaic Law: the joint influence of duration and intensity on the perception of rhythmic speech. *Phonology*, 51-94.

Hay, Jessica, and Randy Diehl. (2007). Perception of rhythmic grouping: testing the iambic/trochaic law. *Perception and Psychophysics,* 69(1), 113–122.

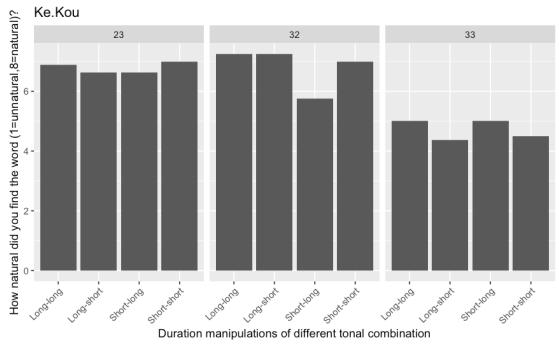
Iversen, J. R., Patel, A. D., & Ohgushi, K. (2008). Perception of rhythmic grouping depends on auditory experience. J. *Acoust. Soc. Am.*, 124, 2263-2271.

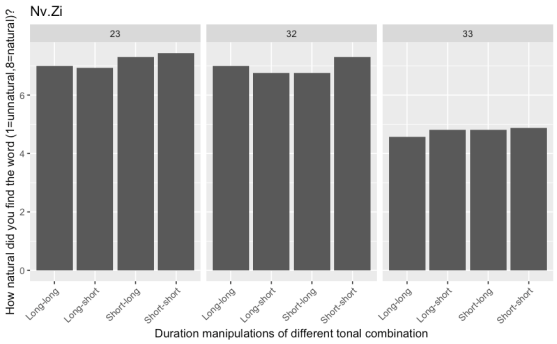
Kolinsky, R., Cuvelier, H., Goetry, V., Peretz, I., & Morais, J. (2009). Music training facilitates lexical stress processing. *Music Percept.*, 26, 235–246.

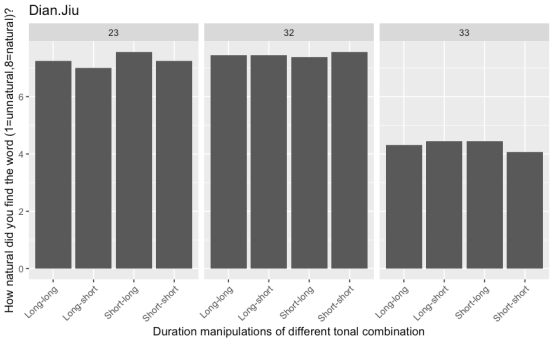
Yiu, Suki. (2019). Iambic and Trochaic Rhythm in Jieyang (Southern Min). *Proceedings of the Annual Meetings on Phonology*, 7.

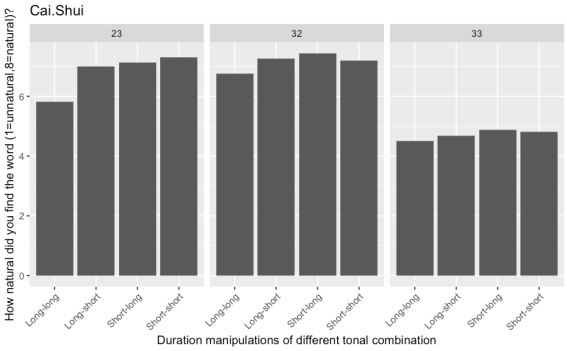
Appendix

Individual Plots for Question 2 Naturalness:









Individual Plots for Question 3 Prominences: