

Tone Sandhi Domain is not a Foot in Standard Mandarin: Evidence from the Perceptual Side

Naiyan Du
Michigan State University

1 Introduction

There is a long debate on whether or not tone sandhi domain in Standard Mandarin should be treated as metrical foot. In promoting the idea, Duanmu assumes that tone sandhi domain is metrical foot and proposes a foot-building theory where morpho-syntactic information is transferred to stress according to a general theory of non-head stress (Duanmu, 2007), and then foot is built according to trochaic rhythm. The theory of non-head stress is stated in (1):

(1) Non-Head Stress

In the syntactic structure [X XP] (or [XP X]), where X is the syntactic head and XP, the syntactic nonhead, XP should be stressed.

The rationale behind the non-head stress theory is information structure. Since there are more options choosing modifiers (non-head) than choosing head, modifiers carry more information, and therefore should be stressed according to information-stress principle by Duanmu (2007).

(2) Information-Stress Principle

A word or phrase that carries more information than its neighbor(s) should be stressed.

Despite its complexity of this model and many theoretical problems with it, this theory offers an elegant explanation for why utterances that have the same syntactic structure can have different tone sandhi domain patterns at the post-lexical level. An example is shown in (3):

(3) a.	horse very easy raise	b.	dog than horse good
	X X		X X
	[[ma][xən xau] [iaŋ]]		[kou[[bi ma] xau]]
	“Horse is very easy to raise.”		“Dog is better than horse.”
	UR 3 3 3 3		UR 3 3 3 3
	Foot Building (3)(2 3) 3		Foot Building (2 3)(2 3)
	Free Syllable Joining (3)(2 2 3)		
	Foot Merge (3 2 2 3)		
	SR 3 2 2 3		SR 2 3 2 3

Note: ‘X’ indicates the position of stress predicted by Duanmu’s theory

Although this theory is successful in accounting for tone sandhi patterns. No experimental results fully support this theory (Jia, 2011; Lai et al., 2010; Shen et al., 2013; Yi, 2016). Previous studies focus on disyllabic minimal pair of Modifier + Noun and Verb + Object. The two groups of utterances have identical segments and tones, and differ only in morpho-syntactic structures. The stress should fall on the first syllable of modifier for the former group and fall on the second syllable of object for the second group according to Non-Head Stress by Duanmu. However, no previous production or perceptual studies can fully support this prediction. For example, the perceptual study by Jia (2011) shows that native speakers perceive the first syllable to be more stressed in both groups.

* I would like to thank Karthik Durvasula, Yen-Hwei Lin and audience of AMP 2020 for comments and feedbacks.

However, Duanmu's theory can be saved by a constraint for disyllabic utterances in Standard Mandarin. Since in Standard Mandarin two Tone 3 syllables being adjacent to each other is not allowed (Chen, 2000), i.e. there is a constraint stating *T3T3, disyllabic utterance must be in one tone sandhi domain. And since tone sandhi domain is a foot according to Duanmu, disyllabic utterance must form a foot, since Mandarin is a trochaic language, stress must fall on the first syllable. So Jia's result is actually not a surprising one. The way to test if stress can be inferred from tone sandhi domain is to go up to tri-syllabic utterance with different tone sandhi domain patterns to see if stress actually falls on different positions.

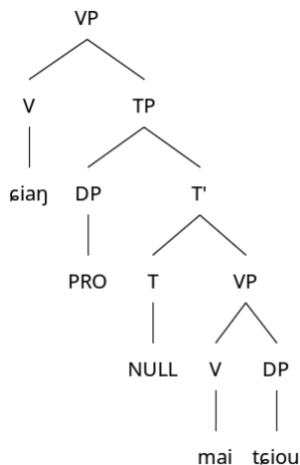
2 The Current research

2.1 Summary The current research seeks to probe the question using a new experiment paradigm involving tri-syllabic utterances. According to Duanmu, by assuming the strong correlation between tone sandhi domain and metrical foot, tone sandhi domain pattern can function as a cue for positions of stress. However, this study shows that, by just changing the tone sandhi domain patterns, stress pattern remains unchanged perceptually for native speakers Standard Mandarin. The result conforms to previous claim that F0 movement cannot be used to indicate stress in tonal languages (Gandour, 1983; inter alia). Therefore, tone sandhi domain pattern is at least not a strong cue for stress in Standard Mandarin. As a result, the analysis that tone sandhi domain is a foot cannot be supported phonetically.

2.2 Stimuli It has been claimed by Duanmu (2007: 293) that for complex sentences containing subordinates clauses. The verb that takes subordinate clause as complement can either be stressed or unstressed, which results in different tone sandhi domain patterns in the surface. For the current study, I only used utterances with this structure. An example is shown in (4). The utterance [ɕiaŋ mai tɕiəu] 'want to buy wine', which is underlyingly "3 3 3", can either be pronounced as "2 2 3" or "3 2 3". For "2 2 3", the initial syllable and the last syllable are stressed, and for "3 2 3", only the syllable in the middle is stressed. The corresponding syntactic structure is shown in (5).

(4) a.	X	X	b.	X
	ɕiaŋ	mai tɕiəu		ɕiaŋ mai tɕiəu
	UR	3 3 3		UR
	SR	2 2 3		SR
	Tone Sandhi Domain	(2 2 3)		Tone Sandhi Domain
				3 (2 3)

(5)



There are a total of 6 such utterances used in this studies as shown in Table 1:

IPA	Underlying Tone	Surface Tone
ɛiaŋ mai tɛiou	333	223 or 323
ɛiaŋ kan ma	333	223 or 323
tɛiaŋ ta ʂu	333	223 or 323
tɛiaŋ tɕ ^h iaŋ ʂuei	333	223 or 323
ni ʂau kou	333	223 or 323
ni iaŋ niau	333	223 or 323

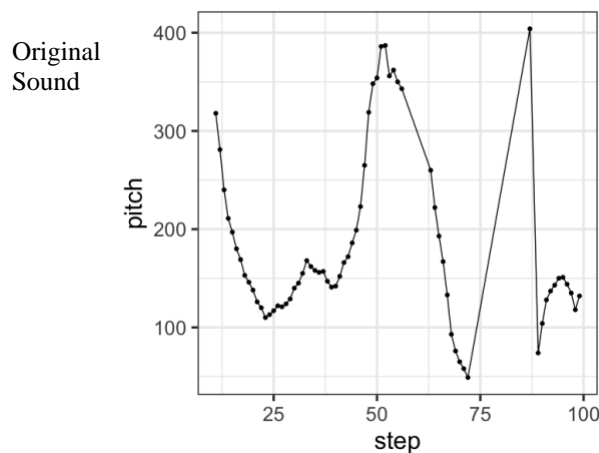
Table 1: Summary of Stimuli

The stimuli with the underlying tone “3 2 3” and “2 2 3” are naturally produced by a female native speaker (age 50) of Standard Mandarin for this study. To ensure that tone sandhi domain pattern, i.e. the difference of tones on the initial syllables, is the only cue that is manipulated, I did a tone manipulation on the first syllables of all the stimuli. The initial Tone 3 in manipulated “3 2 3” utterances that are presented to participants are manipulated from either an original Tone 3 in an original “3 2 3” utterance or an original Tone 2 in an original “2 2 3” utterance. Similarly, The initial Tone 2 in manipulated “2 2 3” utterances that are presented to participants are manipulated from either an original Tone 2 in an original “2 2 3” utterance or an original Tone 3 in an original “3 2 3” utterance. By doing so, a balance can be achieved by making all the stimuli presented to the participants manipulated anyway. Also, if the manipulation that yields the same lexical tone is called “small” manipulation and the manipulation that yields a different lexical tone is called “big” manipulation, then for all manipulated “3 2 3” utterances that are presented to the participants, an equal number of “small” manipulation and “big” manipulation is involved, and for all manipulated “2 2 3” utterances that are presented to the participants, also an equal number of “small” manipulation and “big” manipulation is involved. The tone manipulation conditions are summarized in Table 2:

Original tone patterns	Manipulated tone patterns
223	323 or 223
323	323 or 223

Table 2: Tone Manipulation Condition

To do this phonetically, for the naturally produced “3 2 3”, I made the initial Tone 3, which is a low tone, to either become either a lower Tone 3 and keep the same contour, or a rising Tone 2, which has comparative contour and a higher overall pitch compared to the following original Tone 2 of the same utterance. In Mandarin, if there are two consecutive Tone 2s, the first one has a higher pitch due to the influence of intonation. An example of [ɛiaŋ mai tɛiou] ‘want to buy wine’ is shown in Figure 1 with the order of original sound, manipulated “3 2 3”, manipulated “2 2 3”.



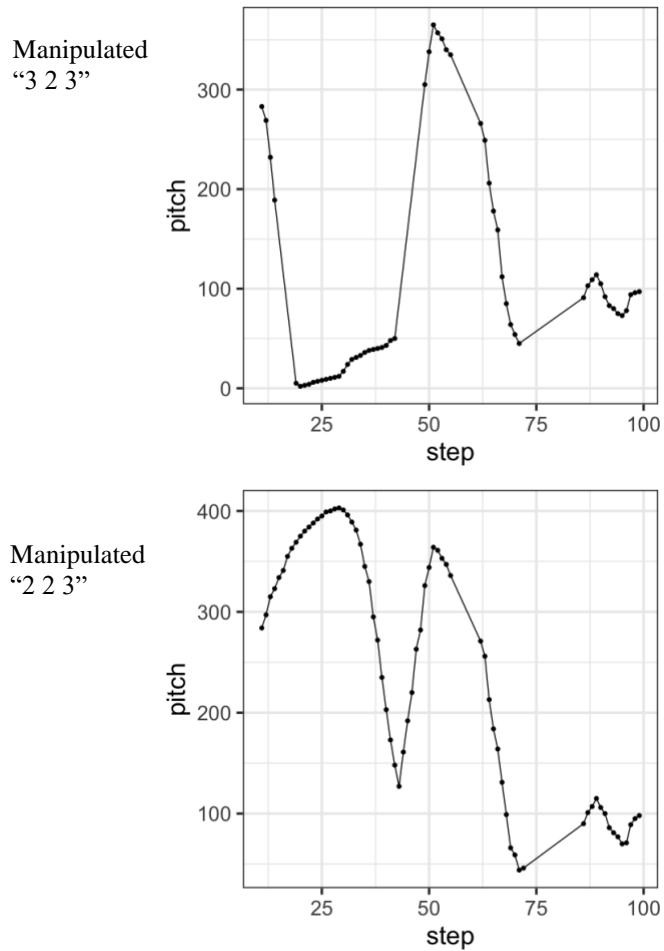


Figure 1: Examples of tone manipulation with the original sound of "3 2 3"

Similarly, for the naturally produced "2 2 3", I make the initial Tone 2, which is a rising tone, to either become either a higher Tone 2 of the same contour, or a low Tone 3, which has comparative contour and a higher pitch compared to the final original Tone 3 of the same utterance. The higher pitch of manipulated Tone 3 is also according to intonation. An example of [ɕiaŋ mai tɕiou] 'want to buy wine' is shown in Figure 2 with the order of original sound, manipulated "2 2 3", manipulated "3 2 3".

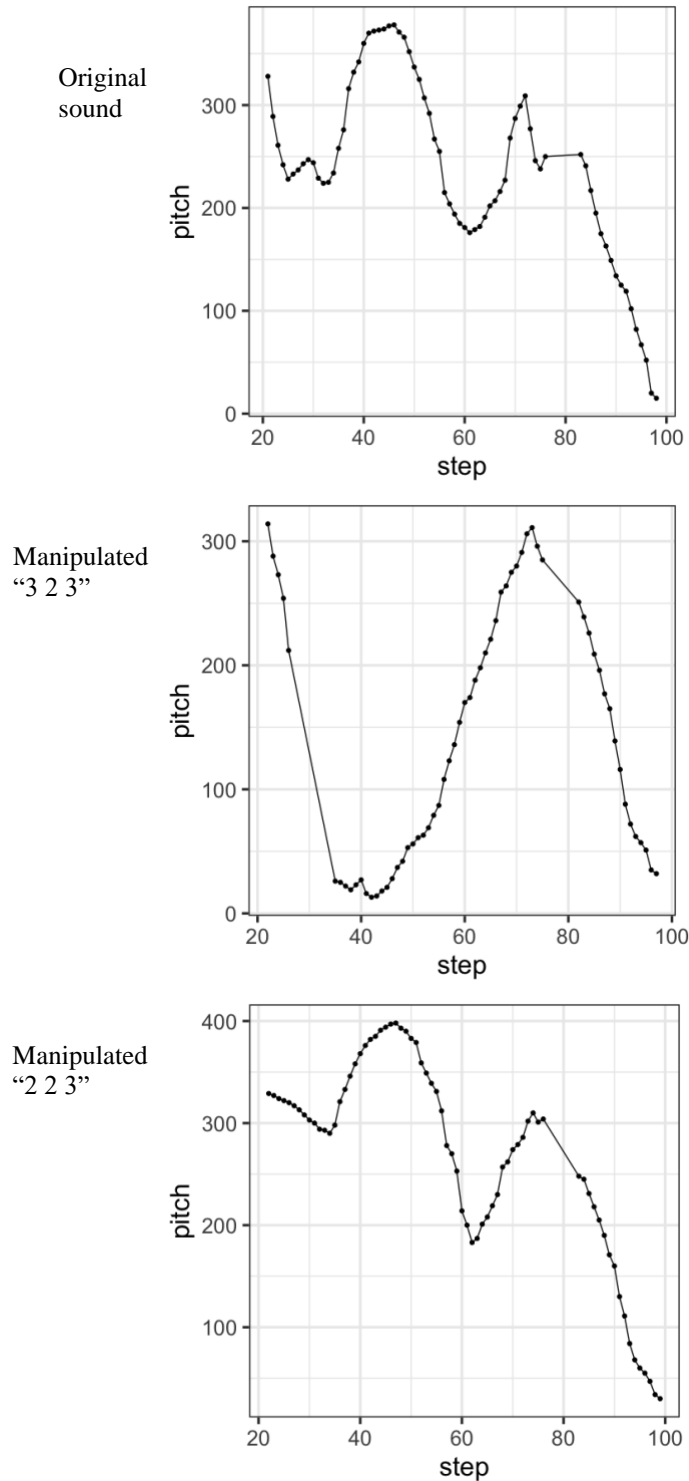


Figure 2: Examples of tone manipulation with the original sound of "2 2 3"

2.3 Linking Hypothesis To avoid metalinguistic knowledge completely, participants cannot be asked directly to judge the positions of stress. The current study used an AX task, which means a strong cue of stress needs to be manipulated to function as a diagnosis of the positions of stress. I chose duration, which

is generally agreed be a strong cue for stress in Standard Mandarin (Chrabaszc Anna et al., 2014; Shen, 1993), and linking hypothesis is stated as follows:

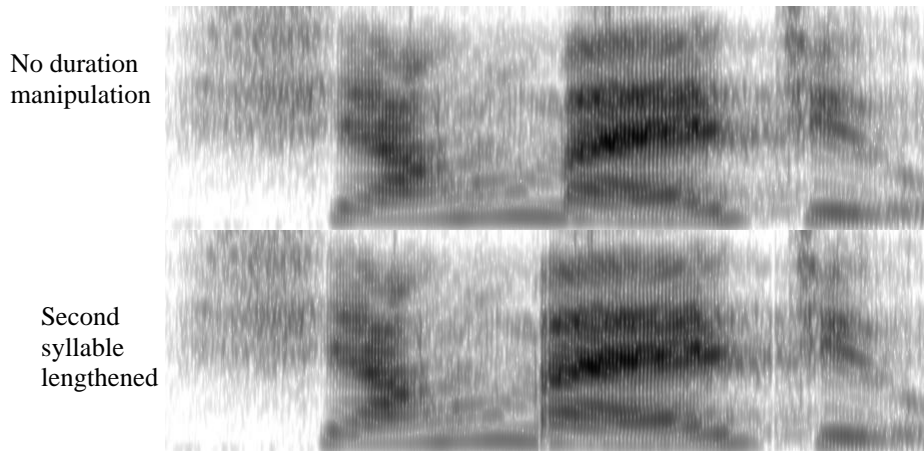
- (6) a. If a syllable is stressed, native speakers are more sensitive when it is shortened than when it is lengthened.
 b. If a syllable is unstressed, native speakers are more sensitive when it is lengthened than when it is shortened.
 c. But if the final syllable is unstressed, due to the effect of final lengthening, native speakers are not sensitive to lengthening or shortening of it.

Crucially, for stimulus in this experiment, unstressed syllables and stressed syllables always appear alternatively as shown in (7), a stressed syllable should be relatively longer than neighboring syllables with regard to duration but cannot be shorter. Therefore lengthening a stressed syllable will not make speakers sensitive but shortening will. Therefore, if a syllable is stressed, native speakers are more sensitive when it is shortened than when it is lengthened. Similarly, an unstressed syllable should be relatively shorter than neighboring stressed syllables with regard to duration but cannot be longer. Therefore shortening a stressed syllable will not make speakers sensitive but lengthening will. Therefore, if a syllable is unstressed, native speakers are more sensitive when it is lengthened than when it is shortened.

(7) a.	X	X	b.	X				
	eiɑŋ	mai	teiou	eiɑŋ	mai	teiou		
	UR	3	3	3	UR	3	3	3
	SR	2	2	3	SR	3	2	3
	Tone Sandhi Domain	(2	2	3)	Tone Sandhi Domain	3	(2	3)

However, if the final syllable is unstressed, to compensate the effect of final lengthening, native speakers are not sensitive to lengthening of it. Therefore if the final syllable is unstressed, native speakers will not be sensitive to duration manipulation. A piece of evidence support this linking hypothesis comes from the experimental results in Jia (2011): for disyllabic utterance, the final syllable can be longer than the initial syllable, but native speaker still perceived the initial syllable to be stressed.

For the duration manipulation, I made the second syllable and the third syllable (final syllable) to be 30% longer or shorter. 30% is above the just noticeable difference and will not induce ceiling effect according to the judgment of the author. An example of [eiɑŋ mai teiou] ‘want to buy wine’ of “2 2 3” that is manipulated from original “2 2 3” is shown in Figure 3 with the following order: sound without duration manipulation, manipulated sound with the second syllable being lengthened, manipulated sound with the second syllable being shortened, manipulated sound with the third syllable being lengthened, manipulated sound with the third syllable being shortened.



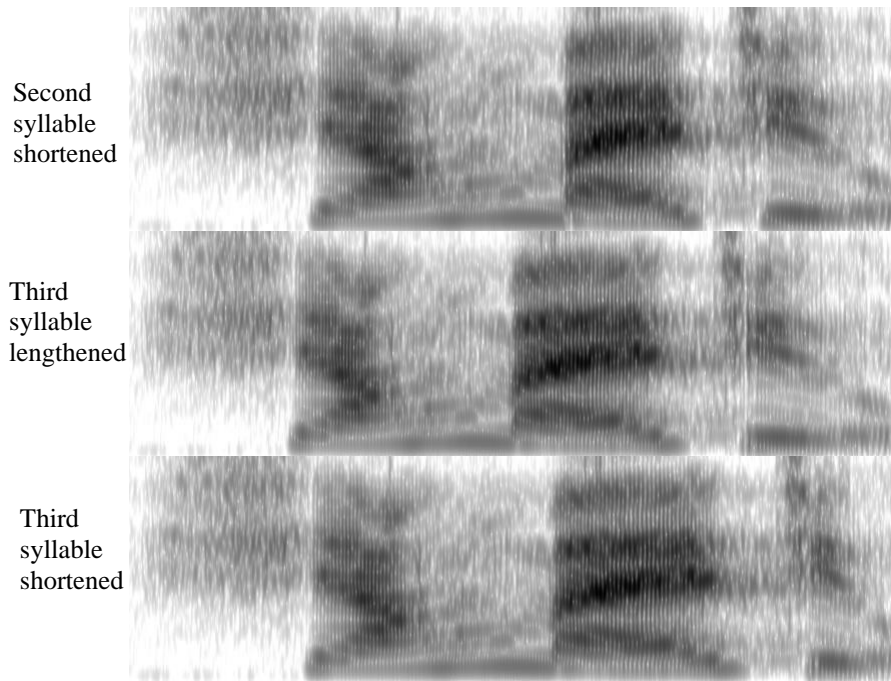


Figure 3: An example of duration manipulation

2.4 Procedure and participant For the AX task, stimuli are composed of three sets. For Set 1, pairs of identical stimuli without duration manipulation are used to set up the base line, for example, if the A is [ɕiaŋ mai tɕiəu] with the surface tone pattern of “2 2 3” without duration manipulation, the X should also be the [ɕiaŋ mai tɕiəu] with “2 2 3” without duration manipulation.

For Set 2, A should be the same, X is the same utterance with the second syllable of the utterance should be lengthened or shortened.

For Set 3, A should also be the same, X is the same utterance the third syllable of the utterance should be lengthened or shortened.

All the stimuli types for AX task are summarized in Table 3 with one example for each type.

Set	Condition	Category	Utterance	Example	
				Segment	Surface tone pattern
Set 1	Identical	1	A	ɕiaŋ mai tɕiəu	2 2 3
			X	ɕiaŋ mai tɕiəu	2 2 3
Set 2	Different with the second syllable long/short	1	A	ɕiaŋ mai tɕiəu	2 2 3
			X	ɕiaŋ mai(L/S) tɕiəu	2 2 3
Set 3	Different with the third syllable long/short	1	A	ɕiaŋ mai tɕiəu	2 2 3
			X	ɕiaŋ mai tɕiəu(L/S)	2 2 3
Set 1	Identical	2	A	ɕiaŋ mai tɕiəu	3 2 3
			X	ɕiaŋ mai tɕiəu	3 2 3
Set 2	Different with the second syllable long/short	2	A	ɕiaŋ mai tɕiəu	3 2 3
			X	ɕiaŋ mai(L/S) tɕiəu	3 2 3
Set 3	Different with	2	A	ɕiaŋ mai tɕiəu	3 2 3

	the third syllable long/short		X	ciaŋ mai teiou(L/S)	3 2 3
--	-------------------------------------	--	---	---------------------	-------

Table 3: Stimuli Types for AX Task

Note: the syllable that is lengthened is marked with (L),
the syllable that is shortened is marked with (S).

Both AX order and XA order are included in this experiment to counterbalance any artefact of order. So for each kind of utterance with the same segments and underlying tones, four categories of surface tones are created by tone manipulation; and for each sound that has undergone tone manipulation, $3 \times 2 = 6$ pairs of comparison are generated by duration manipulation and AX task paradigm. The total number of pairs of comparison are $6 \times 4 \times 6 = 150$. There are 30 pairs of fillers, so the total pairs of comparison each participant needs to complete is 180. The experiment is conducted through PsychoPy 3 (Peirce et al, 2011).

11 native speakers of Standard Mandarin (3 male, 8 female, age 20-30) participated in this experiment and they judged on a scale of 1-5, with 1 being identical and 5 being totally different. 3 excluded since they did not explore the whole range of judgement scale.

3 Results

The mean values for all tonal sandhi patterns conditions and all duration categories are calculated. The result is shown in Figure 4. Here the black bars represented “2 2 3” utterances (after tone manipulation), and the grey bars represented the “3 2 3” utterances (after tone manipulation). The X-axis represents different duration categories, “S2L” means the second syllables are lengthened to be compared to the sounds that do not undergo duration manipulation, “S2S” means the second syllables are shortened to be compared to the sounds that do not undergo duration manipulation, “S3L” means the third syllables are lengthened to be compared to the sounds that do not undergo duration manipulation, “S3S” means the third syllables are shortened to be compared to the sounds that do not undergo duration manipulation.

It is clear from the figure that “2 2 3” utterances show similar patterns with “3 2 3” utterances with regard to participants’ sensitivity to manipulation of duration. For the second syllable, participants are unanimously more sensitive to shortening than lengthening. For the third syllable, there is no difference of participants’ with regard to shortening or lengthening. And the values are low compared to the case of second syllables, which indicate native speakers are not sensitive to the lengthening or the shortening of the third syllable.

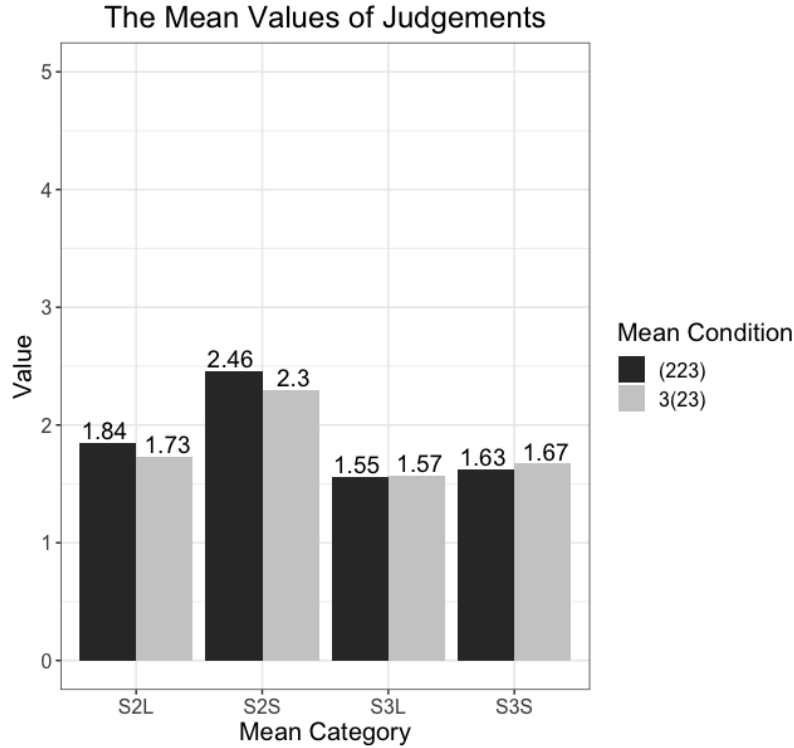


Figure 4: The mean values of judgement

This observation is confirmed by ANOVA as shown in Table 4:

Tonal patterns	Comparison	Result of ANOVA
(223)	S2L~S2S	$F(1,14) = 5.45, p=0.03$
3(23)	S2L~S2S	$F(1,14) = 3.48, P=0.08$
(223)	S3L~S3S	$F(1,14) = 0.12, P=0.73$
3(23)	S3L~S3S	$F(1,14) = 0.37, P=0.55$

Table 4: ANOVA Results

Recall the linking hypothesis as shown in (8):

- (8) a. If a syllable is stressed, native speakers are more sensitive when it is shortened than when it is lengthened.
 b. If a syllable is unstressed, native speakers are more sensitive when it is lengthened than when it is shortened.
 c. But if the final syllable is unstressed, due to the effect of final lengthening, native speakers are not sensitive to lengthening or shortening of it.

The second syllables are always stressed. For the third syllable, since participants are not sensitive to the shortening of it, the third syllable should be interpreted as unstressed. And since it is the final syllable, native speakers should take into consider of final lengthening. Therefore, it is not surprising participants are not sensitive to the lengthening of it.

4 Conclusion

This study shows that tone sandhi domain pattern is at least not a strong cue for stress in Standard Mandarin. Assuming the linking hypothesis of the diagnosis of duration, native speakers judge the positions of stress to be consistent with different tone sandhi domain patterns. This result posts a strong

challenge to the analysis that tone sandhi domain is a foot (Duanmu, 2007). Therefore, the nature of tone sandhi domain should be reconsidered.

References

- Chen, Matthew. 2000. *Tone Sandhi: Patterns across Chinese Dialects*. Vol. 92. Cambridge University Press.
- Chrabaszcz Anna, Winn Matthew, Lin Candise, and Idsardi William. 2014. "Acoustic Cues to Perception of Word Stress by English, Mandarin, and Russian Speakers." *Journal of Speech, Language, and Hearing Research* 57(4):1468–79.
- Duanmu, San. 2007. *The Phonology of Standard Chinese*. OUP Oxford.
- Gandour, Jack. 1983. "Tone Perception in Far Eastern Languages." *Journal of Phonetics* 11(2):149–75.
- Jia, Yuan. 2011. "Stress Patterns of Disyllabic Words Terms with Identical Pronunciation and Different Morph-Syntactic Structures in Standard Chinese." *Journal of Tsinghua University Science and Technology* 51(9):1307–12.
- Lai, Catherine, Yanyan Sui, and Jiahong Yuan. 2010. "A Corpus Study of the Prosody of Polysyllabic Words in Mandarin Chinese." in *Speech Prosody 2010-Fifth International Conference*.
- Peirce, Jonathan, and others. 2011. "PsychoPy—A Psychology Software in Python."
- Shen, Xiaonan Susan. 1993. "Relative Duration as a Perceptual Cue to Stress in Mandarin." *Language and Speech* 36(4):415–33.
- Yi, Hao. 2014. "Phrasal Stress in Beijing Mandarin Disyllabic Phrases."