Whether and How Do Mandarin Sandhied Tone 3 and Underlying Tone 2 Differ?

Yu-Jung Lin
Department of Linguistics,
Indiana University Bloomington,
U.S.A.
lin41@indiana.edu

Yu-Yin Hsu
Department of Chinese and Bilingual Studies
Hong Kong Polytechnic University,
Hong Kong
yyhsu@polyu.edu.hk

Abstract

One interesting linguistic question has been whether and how the characteristics of sandhied Tone 3 are similar to or different from an underlying Tone 2 in Mandarin. In this study, we took a micro view to investigate the vowel height, duration, and rime structure in syllables, and their influence on the degree of Tone 3 sandhi through native speakers' speech production and perception. It was found that vowel height and duration played important roles in shaping the F0 contour of the sandhied Tone 3 in production. As to perception, sandhied Tone 3 syllables with low vowels and longer duration were more likely to be perceived as Tone 3.

1 Introduction

Mandarin Tone 3 (T3) sandhi generally refers to a T3 syllable's tonal contour that has been changed from the original falling-rising tone to a contour perceived as Tone 2 (T2). This sandhi process takes place when a T3 syllable is followed by another T3 syllable (Chao, 1948). Many studies tested whether the sandhied T3 and the underlying T2 are truly the same. Some studies have discovered that acoustically, the pitch height of sandhied T3s (T3S) were lower than those of underlying T2s (e.g., Shen, 1990; Xu, 1997; Peng 1996; Kuo, Xu, & Yip, 2007; Yuan & Chen, 2014). Concerning tone perception, Wang and Li (1967) and Peng (1996) argued that the difference between T3S and underlying T2 cannot be perceived, but Speer and Xu (2008), with eye-tracking data, reported that listeners who are native Mandarin speakers can perceive the differences between T3S and underlying T2.

Assuming that the intrinsic acoustic properties of vowels can influence their syllables' tonal contours and in turn may influence listeners' categorical perception of tone types, in the current study, we focused on effects of vowel height, rime structures, and syllable duration in the production and the perception of T3S and underlying T2 items. Several studies have pointed out that higher vowels tend to have higher fundamental frequencies than low vowels (i.e., a phenomenon referred to as intrinsic F0; Whalen & Levitt, 1995; Whalen, Gick, & LeSourd., 1999), and Shi and Zhang (1986) reported that such F0 characteristics were found in all four citation tones in Mandarin Chinese, although they did not report whether and how such characteristics influence the tonal contour.

With respect to duration, the consensus is that T3 is usually the longest among the four Mandarin tones (e.g., Xu, 1997; Liu & Samuel, 2004; Wu & Kenstowicz, 2015), as demonstrating a falling-rising contour requires more time than demonstrating a level or a unidirectional tone change. In perception studies, some have reported that longer duration is needed for syllables with lower F0 to be perceived as the same length than for syllables with higher F0 (Yu, Lee, & Lee, 2014).

Interestingly, Peng (1996) reported conflicting results between the production and the perception of T3 sandhi, that is, the rising slope of both T3S and T2 were found to be shallower at the fast speed and that the difference between T2 and T3S was maintained even in fast speech; however, such differences could not be perceived by listeners even when speech rate was slow and when the syllable duration was long.

Chen, Zhu and Wayland (2017) examined the effect of duration and its interaction with vowel quality on the perception of Mandarin rising and falling tones; vowel quality was reported to significantly contribute to the tonal differences, although the interaction between tone direction and duration varied, i.e., they found that in general, the

32nd Pacific Asia Conference on Language, Information and Computation
Hong Kong, 1-3 December 2018

longer duration was required for perception of both raising and falling tones with low vowels than with high vowels.

Related to the issue of timing and syllable duration, Clements and Keyser (1983) argued that each segment occupied some timing slot, and therefore syllable structures like those with nasals (CVN) and offglides (CVG) should be longer than a structure like CV. However, Duanmu (2007) maintained the view that all Mandarin stressed syllables (i.e., syllables carrying lexical tones) occupy two rime slots, that is, although a simple rime, V in a stressed CV is counted as occupying two rime slots, and in this view, the length of V in a stressed CV syllable is considered the same as the length of VG and VN in CVG and CVN syllables. Xu (1998) examined the consistency of tone-syllable alignment across different rime structures and concluded that F0 contours for all four tones retain the same alignment to the syllables that carry them regardless the rime structure.

In brief, different views and results of duration, rime structures, and vowel quality were reported in terms of their impact on tone contours and on the production and the perception of tones. In the current study, we focused on the effects of these qualities of syllables on the production and the perception of sandhied T3 and underlying T2. We measured syllable duration, mean F0, F0 difference (i.e., the initial fall of the F0 from onset of the tone to the point where the F0 starts to rise), and the turning point (i.e., the time point where the contour turns, see Figure 1).

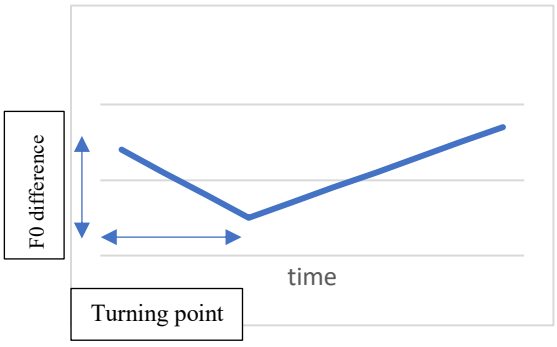

Fig 1. F0 difference and the turning point schematized for T2/T3 contour.

According to Chao (1948), T2 was described as 3-5 and T3 as 2-1-4 on a subjective 5-point scale with 5 as the highest level. A categorization like this indicates that typical T2 and T3 occupy pitch ranges with little overlap. If the categorial differences between T2-T3 can be represented through pitch range, mean F0 can be used to observe such differences (see also Peng, 1996). F0 difference (dif-F0) and the timing of the turning point were reported to be most reliable in perception and production when distinguishing T2 and T3 (Shen & Lin, 1991; Moore & Jongman, 1997). The larger the dif-F0 is, the more likely the tone is perceived as a T3; the later the turning point is, the more likely a tone is perceived to be T3-like.

With these three measures, we examined whether and how vowel height, rime structures and syllable duration might help to distinguish sandhied T3 from underlying T2 and their impact on the identification of these two tone categories.

Regarding the environment that triggers T3 sandhi, Shih (1997) agreed with Cheng (1968), who suggested that it is the following low pitch that motivates the target T3 to be changed to a T2-like sound. Cheng (1968) reported that in code-switching contexts, T3S happens even when a T3 syllable is followed by a non-Chinese word bearing an initial pitch lower than the pitch of its second syllable. For example, *hao* 'good', an underlying T3, sounds like a T2 in *hao professor* 'good professor'), but remains T3 when followed by a word where its first syllable has pitch relatively higher than that of its second syllable (e.g., *hao* in *hao student* 'good student'). We consider this an important factor triggering T3S and implemented it in our production task.

Furthermore, following Lin (2007), we categorized low vowels and non-low vowels based on the vowel height of the nuclear vowel of that syllable; in our stimuli, low vowel [a], and non-low vowels (mid vowels [e] and [o] and high vowels [i] and [y]) were included (cf. Chen et al. 2017). To examine how rime structures might influence the production and the perception of sandhied T3, we compared syllables of the structure CV with CVN (with nasal) and CVG (with offglide) in contexts of sandhied T3 and underlying T2.

We conducted one production and one perception experiment. We first examined how the vowel height, rime structure and duration influenced the mean F0 of the syllable (mean F0), the F0 difference from onset to the turning point (dif-F0), and the timing of the minF0. We then

examined whether and how these factors influenced native speakers' perception rating of tone types.

The experiments reported below show that in a sentential T3 sandhi context, monosyllabic T3 syllables containing a high vowel were more likely to be produced as a T2-like contour, and to be perceived as T2 more often than T3 syllables with low vowels were. We also observed that longer syllable duration, more complex rimes, and later timing of the turning point made the F0 contour less T2-like in the production data and that among factors under investigation, vowel height and duration correlated with tonal classification.

2 Experiment 1: Production study

The predictions for the production study were that a) T3S syllables with low vowels should be produced more T3-like, b) T3S syllables with longer duration should be produced more T3-like, and c) a null hypothesis was assumed for syllables with different rimes.

2.1 Method

2.1.1 Stimuli

In this study, 20 items were used : each of 10 monosyllabic units was associated with an underlying T2 and an underlying T3 Mandarin monosyllabic word, e.g., *lan* [lan] 'basket' and *lan* [lan] 'lazy'. In each tone group, there were five syllables with low vowels (*lan* [lan] 'basket, lazy', *wan* [wan] 'play, bowl', *ma*[ma] 'hemp, horse', *wa* [wa] 'doll, tile', *ya* [ja] 'tooth, dumb') and five with non-low vowels (*li* [li] 'pear, plum', *mei* [mej] 'coal, beauty', *yi* [ji] 'aunt, chair', you [jow] 'oil, friend', *yu* [ɥy] 'fish, rain'). Among the five low vowels, two were in a rime structure composed of vowel and nasal (VN) and three syllables were in the vowel-only rimes (V). Among the five non-low vowels, two were in rimes composed of vowel and offglide (VG) and three were in vowel-only rimes (V). Onglides were analyzed as part of the onset (Duanmu, 2007). All 20 target items and 52 filler items (36 of tone 1 and tone 4, and 16 of tone 2 and tone 3 with irrelevant vowels) were embedded in a carrier sentence: 请将___点出来。 [tɕʰjəŋ tɛjɑŋ ___ tjɛn tʂʰu laj] "Please point at ____." in which the word 点 'point' after the target item is T3, providing a T3 sandhi context for target items.

2.1.2 Participants and Procedure

Eight female native speakers of Mandarin from the Northern provinces of China, who were students in Hong Kong Polytechnic University, participated in this study (mean age: 24.0). Participants received a $50 coffee coupon after they had completed all recordings.

The experiment was conducted in a sound-attenuated room. Chinese characters for these items were placed in the carrier sentence, showed in simplified characters, and were displayed in a randomized list on a computer screen. Participants were instructed to read the sentences one by one as natural and clearly as possible. Participants were instructed to make no pauses within a sentence while reading it. The whole list of items was repeated once. Participants were asked to repeat a sentence only when mispronounced words or paused while reading the sentence, they would be asked to repeat that sentence again. Recordings were made in .wav format at a sampling rate of 44.1kHz and a 16-bit quantization. The whole experimental session lasted about 20 minutes.

2.1.3 Analysis

Both the target and control items were segmented using a customized-written script, ProsodyPro (Xu, 2013) for Praat (Boersma & Weenink, 2018). Syllable boundaries were determined by spectrogram, waveform, and auditory input. For [m], the onset starts at the nasal murmur. For [l], the onset starts at the release of the tongue. For vowels or glides, the onset starts at the point where the clear and steady F1 and F2 formants for that sound appear. A total of 320 syllables (20 syllables X 2 repetitions X 8 participants) were extracted from the carrier sentences, of which three were discarded due to the obvious creakiness; thus, 317 syllables were analyzed. ProsodyPro generated the intrinsic duration of each syllables and turned all the syllable durations into a normalized time. It also generated the F0 values at the 10 time points with the same intervals.

Generalized Linear mixed-effects models were fit to the data using the lme4 package (Bates, Maechler, Bolker, & Walker, 2015) in R (R Development Core Team, 2016, version 3.3.2). P-values were obtained from likelihood-ratio tests of the models, with and without fixed effects and their interactions, which were included only when doing so yielded a better fit ($\alpha = 0.05$). Dependent

variables used in the analyses were (a) mean F0, (b) dif-F0, and (c) timing of the turning point. The three fixed effect variables were VOWELHEIGHT, RIMETYPE, and DURATION of the syllable, and PARTICIPANTS and ITEMS were random intercepts.

2.2 Results

As shown in Figure 2, the shapes of the T2 and T3S F0 were both deep curves. This could be related to the carryover and the anticipatory effects. According to Xu (1997), the carryover effect refers to a phenomenon by which the onset of a tone is influenced by its previous tone's offset, and the anticipatory effect refers to a phenomenon by which the F0 of a tone is raised because of the low onset of the following tone. In our study, both T3 target items and T2 control items were preceded by a T1 (high level tone) syllable, and were followed by a T3 (falling-rising low tone) syllable. We suppose that it is this environment that made both the T3S target and the T2 control started relatively higher in F0 and exhibited deep F0 curves.

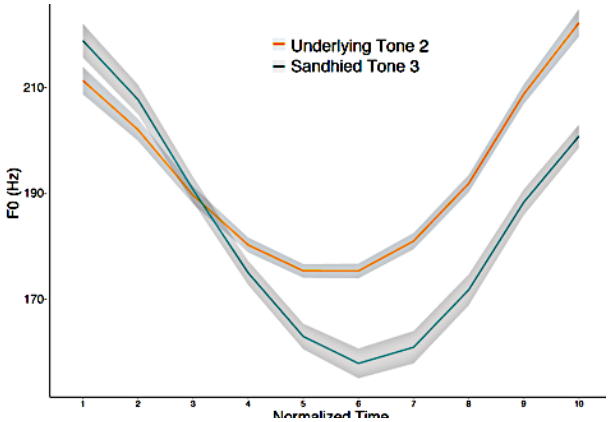

Fig 2. F0 contours of the underlying T2 and the sandhied T3. Lines indicate the mean F0 and the grey area represents the standard error of the mean.

2.2.1 Vowel height

Figure 2 above shows that the overall F0 contours of T2 and T3S were similar – both exhibited clear falling and rising F0 contours. This seems to suggest that T3S items and T2 control should be similar in terms of tonal characteristics. However, Figure 3 demonstrates that it is not the full story. Although T3 syllables with non-low vowels were patterned very similarly to T2 syllables with low vowels, T3S syllables of low vowels in T3 showed much higher initial F0 and much deeper F0 curves than syllables in the other three conditions.

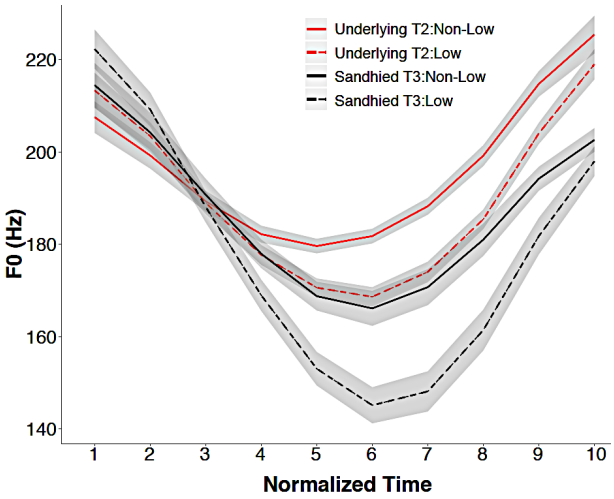

Fig 3. F0 contours of underlying T2 and sandhied T3 with syllables of low and non-low vowels. Lines indicate the mean F0 of different conditions, and the grey area represents the standard error of the mean.

The observations illustrated in Figure 3 were confirmed by statistical analysis (Table 1): among the T3 items, non-low vowels were significantly different from low vowels, having a higher mean F0 ($p < .001$), smaller dif-F0 ($p < .01$) and earlier timing of the turning point ($p < .001$).

Table 1. Statistical tests with linear mixed-effect model in Experiment 1

Dependent variable	Variables	Estimate	t value	Pr(>|z|)
Mean F0	Non-low	10.50298	2.556	<0.05*
	Duration	-0.01924	-0.761	0.4483
	VG	1.55290	0.356	0.7221
	VN	5.27760	1.261	0.2093
Dif-F0	Non-low	-20.2347	-2.466	<0.05*
	Duration	0.3013	6.548	<0.001*
	VG	1.2168	0.138	0.8925
	VN	-18.8652	-2.220	0.0529
Turning point	Non-low	-1.081559	-4.078	<0.001*
	Duration	0.005832	4.710	<0.001*
	VG	0.949226	3.304	<0.001*
	VN	-0.865258	-3.067	<0.001*

*p<0.05.

2.2.2 Duration

For T3S, the minimum and the maximum durations were 153.64 ms and 590.49 ms, the mean was 293.07 ms and the standard deviation was 92.40 ms. According to Peng (1996), a faster speech rate leads

to a shallower falling contour, which might make the mean F0 of shorter syllables higher than that of longer ones, and the dif-F0 of shorter syllables smaller than that of the longer ones.

The statistical analysis represented in Table 1 confirmed that when the duration was shorter, the mean F0 tended to be higher although this difference was not significant. The dif-F0 was confirmed to be significantly smaller when the duration was shorter. The turning point occurred significantly earlier when the syllable duration was shorter. All these observations pointed to the conclusion that a shorter T3S was more likely to sound like a T2.

Because vowel height had a main effect on all three dependent variables, we divided T3S syllables according to vowel height and rime type. Table 2 shows the averages of duration and standard deviation of the three rime types and the two levels of vowel height. Among them, syllables with non-low vowels were shorter than syllables with low vowels, and rime type V with non-low vowels were the shortest.

Row	Average of duration (ms)	Standard deviation of duration (ms)
Low	**318.73**	**92.43**
V	319.83	87.74
VN	317.15	100.20
Non-low	**277.34**	**88.23**
V	256.00	81.88
VG	309.36	88.93

Table 2: The means and standard deviations of the durations of syllables with low vowels, non-low vowels, and different rime types V, VN and VG.

In brief, compared with T3S syllables with high vowels, T3S syllables with low vowels tended to have longer durations, which might allow their F0 contours to be more like T3 than T2.

2.2.3 Rime type

Rime type did not show main effects on mean F0 and dif-F0 but had main effects on the timing of the turning point. In terms of the normalized time, the turning point of Vowel-Nasal (VN) was 5.59, the turning point of Vowel (V, two vowel height levels combined) was 5.73, and the turning point of Vowel-Offglide (VG) was 6.28.

As summarized in Table 1, VG had a significantly later turning point than V, and VN had a significantly earlier turning point than V. The ranking regarding the timing of the turning point was VN<V<VG, which held true within and across vowel height, indicating that the characteristics of the rime types influence the turning point. Figure 4 summarizes the breakdown patterns of rime types with different vowel heights.

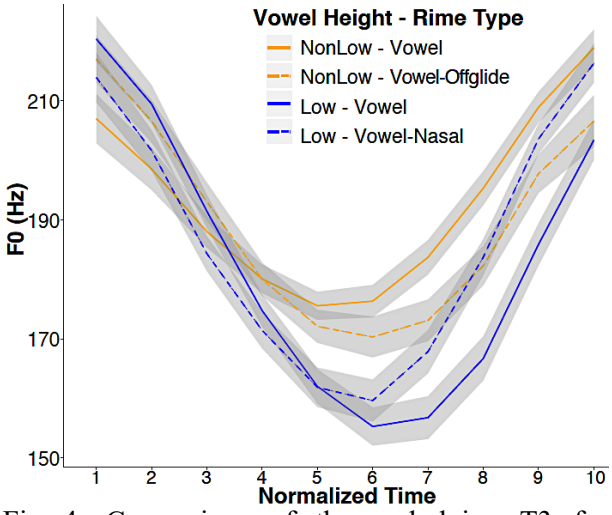

Fig 4. Comparison of the underlying T3 for syllables with different syllable heights and rime structures in a T3 sandhi context. The grey area represents the standard error of the mean.

With regard to vowel height, when VN with low vowel and V with low vowel were compared, VN was significantly more T2-like according to all criteria examined due to the characteristic of the rime type – the earlier the turning point, the smaller the dif-F0, and the higher the meanF0. Further research is needed to untangle the relationship between the rime structure and the turning point of the F0.

2.3 Discussion

Our results regarding the F0 contours and the difference of mean F0 between T2 and T3S were consistent with those of previous production studies on T3 sandhi (Zee, 1980; Peng, 1996). Peng (1996) observed that the rising slope "became shallower as speech rate increased" (p.24), and we found that when the syllable duration was shorter, both the rising and falling slope tended to be shallower.

The shape of T2 and T3S also supported Xu (1997). The contextual variation had influence on

our three dependent variables: it made the T2 and T3S overlap greatly in terms of F0 range, and made the initial fall of both T2 and T3S deeper, thus enlarging their dif-F0s and timing of turning point. However, even in this context when the T2-T3S difference was reduced, we still observed significant effects of vowel height, duration, and rime type.

These results regarding the influence of vowel height confirmed that low vowels have lower F0 than those with high vowels (Shi & Zhang, 1986), and suggested that this phenomenon holds true even in T3S context. In addition, the current results further indicated that not only the mean F0, but also the F0 difference and turning point were changed.

Our results regarding duration were generally consistent with the predictions – the longer the syllables are, the more T3-like they become. This finding is consistent with the previous findings about vowel height in that syllables with non-low vowels were generally shorter than those with low vowels. It is compatible with our findings that T3S syllables with non-low vowels seemed to show more T2-like features and those with low vowels showed more T3-like features.

Regarding rime types, our results of mean F0 and F0 difference supported Xu's (1998) argument that the tonal alignment "is not affected by the internal structure of the syllable in any direct way, further confirming the hypothesis that the syllable is the only relevant unit for the alignment of lexical tones in Mandarin"(p.200). However, in terms of turning point, VG had a significantly later turning point than V, and VN had significantly earlier turning point than V. Further research can more clearly reveal the relationship between the turning point and the intrasyllabic structure.

3 Experiment 2: Perception Study

The research question for the perception study was whether the results found in our production study also aligned with the perception results from listeners who were native speakers of Mandarin. Our predictions were based on the production results: (a) longer syllables tend to be less T2-like than shorter ones, (b) syllables with low vowels are less T2-like than syllables with non-low vowels, and (c) rime type does not play an important role during tone identification among T2 and T3S tokens.

3.1 Method

3.1.1 Stimuli

All the syllables analyzed in Experiment 1, in total 317 syllables, were played to the listeners, who were instructed to pay attention only to whether each audio input sounded like tone 2 or tone 3. For cases in which they could not make a decision, participants were instructed to give a mark of 2.5, resulting in three types of answers: 2 (for T2), 2.5 (between T2 and T3), and 3 (for T3). We allowed three options in the current task because according to our observation in the production experiment, the results of T3 sandhi rule exhibits ambiguity, and we wanted to allow the participants to express their uncertainty.

3.1.2 Participants and Procedure

Four native listeners of Mandarin participated in this study. Items were played to the participants in a quiet room. Participants sat in front of a computer listening to the audio input through a headset. Once an item had been played, and a mark had been given to it, the participant then clicked a key to play the next sound file and provided their categorical judgment. The experiment was conducted at participants' own pace, and the whole session lasted about 25 to 30 minutes.

3.1.3 Analysis

Cumulative Link Mixed Model with random intercepts for SUBJECT and ITEMS was fitted with the predictors VOWELHEIGHT (Low, Non-Low), RIMETYPE (V, VG, VN) and DURATION using the clmm() function in the ordinal package (Christensen, 2018) in R (R Development Core Team, 2016, version 3.3.2), based on likelihood ratio tests of the model comparison. The dependent variable was the selection of perceived tone types (2, 2.5, 3). Post-hoc comparisons were conducted by the *lsmeans* package in R (Lenth, 2016).

3.2 Results

1268 tokens (317 tokens X 4 listeners) were analyzed. VOWELHEIGHT and RIMETYPE were examined first. Table 3 below shows the percentage of the 4 types of VOWELHEIGHT - RIMETYPE combinations under each rating score for items in sandhied T3 condition and in the underlying T2

condition respectively. In the "underlying T2" condition, 69.8% ~ 85.9% of the sound files in each combination were identified as T2. In the "Sandhied T3" condition, however, only 41.8% ~ 63% of the tokens from each combination were identified as T2. Furthermore, the difference between low vowels and non-low vowels was obvious. In the sandhied T3 condition, more tokens were identified as either "between T2 and T3" or "T3" when T3S tokens were with low vowels (i.e., 58.2% of the Low-V syllables and 50% of the Low-VN syllables), while T3S syllables with non-low vowels were mostly identified as T2, and only 37% of the Non-low-V and 43% of the Non-Low-VG tokens were identified as either "between T2 and T3" or "T3."

In the underlying T2 category, while the majority of the tokens were identified as T2, 30.2% of the T2 Low-V syllables were identified as either "between T2 and T3" or "T3", much higher than T2 tokens in other conditions.

Vowel Height	Rime Type	Sandhied T3			Underlying T2		
		2	2.5	3	2	2.5	3
Non-Low	VG	57	18	25	85.9	10.9	3.1
	V	63	12	25	76.1	12.8	11.2
Low	V	41.8	22.3	35.9	69.8	21.9	8.3
	VN	50	19.5	30.5	82.8	8.6	8.6

Table 3: T2/T3 identification results in two conditions. The numbers in the table are the percentage of each combination of "rime type and vowel height" under each rating score.

Statistical analyses confirmed these observations. With respect to sandhied T3 items, the analysis indicated significant effects of VOWELHEIGHT (β = .56, *SE* = .20, p = .006), where syllables with non-low vowels were more often identified as T2 than as T3. As verified by a post-hoc test, the effect of the difference between vowel height on tone identification was significant (p = .006). Regarding the effects of VOWELHEIGHT in different RIMETYPE, the analysis revealed that in the T3 condition, low vowels in rime structure V were significantly different from non-low vowels in rime structure V (β = .72, *SE* = .259, p = .005), whereas other rime structures with different vowel heights combination did not show significant effects (i.e., VG-nonLow: β = 0.053, *SE* = .279, p = .847; VN-low: β = 0.389, *SE* = .277, p = .159). The post-hoc tests confirmed that rime structure V with low vowels differed significantly from rime structure V with non-low vowels (p = .027), as well as from rime structure VG with non-low vowels (p = .045)

For the sandhied T3 items, DURATION, was reported to show significant effects (p = .001) on the identification results. As Figure 5 shows, kernel probability density indicates that the longer the syllable duration, the more frequent the identification of T3, whereas shorter syllables tended to be identified as T2.

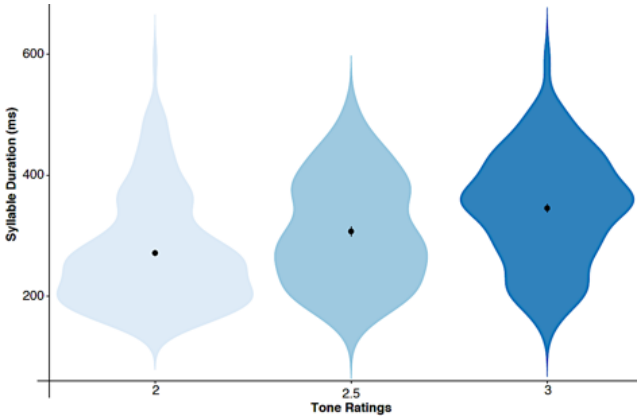

Fig. 5. Effects of DURATION on the identification of tone types. For any given violin plot, the width of the colored area represents the proportion of the data located there. The dot inside each set displays the standard error of the mean.

As for items in the underlying T2 condition, effects of VOWELHEIGHT and DURATION were not significant in terms of impact on identification (ps > .37). Interestingly, however, with regard to the RIMETYPE, syllables in the VG rime structure with non-low vowels showed a strong tendency to be identified as T2 (β = -.89, *SE* = .447, p = .046) than other rime-vowel height combinations. Post-hoc tests confirmed the effect where the VG rime structure with non-low vowels was significantly different from the V rime structure with low vowels (p = .043), whereas other pair-wise comparisons did not show significant differences.

3.3 Discussion

Our predictions for the perception task were that (a) longer syllables would tend to be less T2-like than shorter ones, (b) syllables with low vowels would be less T2-like than those with non-low vowels, and

(c) rime type does not play an important role in tone identification among T2 and T3S tokens.

All these predictions were confirmed. Furthermore, with regard to both rime type and vowel height, CV syllables with low vowels and CV syllables with non-low vowels tended to be classified as different tones. This finding is consistent with production results. Among the four VOWELHEIGHT - RIMETYPE combinations, CV syllables with non-low vowels and low vowels was the only pair that significantly different in all three parameters: mean F0, dif-F0, and timing of the turning point (see Table 1 and Fig 4).

Our results were not consistent with those of Wang and Li (1967) and Peng (1996), which found that native listeners could not distinguish T3S from T2. In terms of syllable duration, our results also contradicted with those of Peng (1996), which found that speech rate did not change perception responses. In our study, duration had a main effect. The rating score increased (towards 2.5 or 3) as the duration lengthened. However, Peng (1996) reported that listeners' response were at chance and did not vary across the speed. She did not provide the stimuli so we do not know if CV syllables with low vowels were included in her stimuli. Also, she did not provide the duration of the slow, mid and fast speeds so it is hard to judge whether the syllable durations in her study is comparable to the durations in our study.

Furthermore, disyllabic words were used in her study while we used monosyllabic words, which might have enabled the listeners to focus more on fine-grained details in the target syllables. Moreover, in our study, listeners listened to target syllables one by one and could not easily access the meaning of those monosyllabic words. Thus, the lexical meaning of the sounds was less likely to distract them. Finally, beside the option of the rating "2" (T2), we provided two other options "2.5 (between T2 and T3) and "3 (T3)." These two options encouraged the listeners to express their uncertainty about the T3S syllables when rating.

4 Concluding Remarks

Through two experiments (one speech production experiment and one perception judgment task), we examined native speakers' speech production to study how the vowel height, the rime structure and the duration influenced mean F0, F0 difference and turning point, and whether the differences reported in the production experiment influenced the perception of syllables' tone under a sandhi condition.

We showed in the production experiment that both duration and vowel height played important roles influencing how T2-like a T3S could be. Rime structure also significantly influenced the timing of the turning point. Furthermore, in the perception study, our sampled listeners used vowel height and syllable duration to identify T3; that is, syllables with low vowels and longer syllables were more likely to be perceived as T3 than when syllables carried other characteristics.

Acknowledgments

This research was made possible through the support of the research project (4-ZZHN) funded by the Department of Chinese and Bilingual Studies, Hong Kong Polytechnic University.

We would like to thank the three anonymous reviewers for their insightful comments and suggestions. We also want to thank Wang Xia, Kong Fanyu, Wang Zexing, and Ding Ning for their technical support. Any errors and inadequacies that remain are exclusively our own.

References

Bates, D., Maechler, M., Bolker, B, Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48.<doi:10.18637/jss.v067.i01>.

Boersma, Paul & Weenink, David (2018). Praat: doing phonetics by computer [Computer program]. Version 6.0.40, retrieved 11 May 2018 from http://www.praat.org/

Chao, Y. R. (1948). *Mandarin primer: An intensive course in spoken Chinese*. Harvard University Press.

Chen, S., Zhu, Y., & Wayland, R. (2017). Effects of stimulus duration and vowel quality in cross-linguistic categorical perception of pitch directions. *PloS one*, *12*(7), e0180656.

Cheng, C. C. (1968). English stresses and Chinese tones in Chinese sentences. *Phonetica, 18*(2) (77-88)

Christensen, R. H. B. (2018). ordinal - Regression Models for Ordinal Data. R package version 2018.4-19. http://www.cran.r-project.org/package=ordinal/.

Clements, G. N., & Keyser, S. J. (1983). CV phonology. a generative theory of the syllable. *Linguistic Inquiry Monographs Cambridge, Mass*., (9), 1-191.

Duanmu, S. (2007). *The phonology of standard Chinese*. Oxford University Press.

Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge University Press.

Kuo, Y. C., Xu, Y., & Yip, M. (2007). The phonetics and phonology of apparent cases of iterative tonal change in Standard Chinese. *Tones and tunes*, *2*, 211-237.

Russell V. Lenth (2016). Least-Squares Means: The R Package lsmeans. Journal of Statistical Software, 69(1), 1-33.<doi:10.18637/jss.v069.i01>

Lin, Y. H. (2007). *The Sounds of Chinese with Audio CD* (Vol. 1). Cambridge University Press.

Liu, S., & Samuel, A. G. (2004). Perception of Mandarin lexical tones when F0 information is neutralized. *Language and spech*, *47*(2), 109-138.

Moore, C. B., & Jongman, A. (1997). Speaker normalization in the perception of Mandarin Chinese tones. *The Journal of the Acoustical Society of America*, *102*(3), 1864-1877.

Peng, S. H. (1996). *Phonetic implementation and perception of place coarticulation and tone sandhi* (Doctoral dissertation, The Ohio State University).

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Shen, X. S. (1990). Tonal coarticulation in Mandarin, *Journal of Phonetics,* 18: 281-295.

Shen, X. S., & Lin, M. (1991). A perceptual study of Mandarin tones 2 and 3. *Language and speech*, *34*(2), 145-156.

Shi, B., & Zhang, J. (1986). Vowel intrinsic pitch in Standard Chinese. *Working Papers in Linguistics*, *29*(1).

Shih, C. (1997). Mandarin third tone sandhi and prosodic structure. *Linguistic Models*, *20*, 81-124.

Speer, S., & Xu, L. (2008). Processing lexical tone in third-tone sandhi. *Talk presented at Laboratory Phonology*, *11*.

Wang, W. S-Y. and Li, K. P. (1967). Tone 3 in Pekinese. *Journal of Speech and Hearing Research* 10: 629-236.

Whalen, D. H., & Levitt, A. G. (1995). The universality of intrinsic F0 of vowels. *Journal of phonetics*, 23(3), 349-366

Whalen, D. H., Gick, B., & LeSourd, P. S. (1999). Intrinsic FO in Passamaquoddy Vowels. *Algonquian Papers-Archive*, *30*.

Wu, F., & Kenstowicz, M. (2015). Duration reflexes of syllable structure in Mandarin. *Lingua*, *164*, 87-99.

Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of phonetics*, *25*(1), 61-83.

Xu, Y. (1998). Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica*, *55*(4), 179-203.

Xu, Y. (2013). ProsodyPro—A tool for large-scale systematic prosody analysis. Laboratoire Parole et Langage, France.

Yu, A. C., Lee, H., & Lee, J. (2014). Variability in perceived duration: pitch dynamics and vowel quality. In *Fourth International Symposium on Tonal Aspects of Languages*.

Yuan, J. H., & Chen, Y. Y. (2014). 3rd tone sandhi in standard Chinese: A corpus approach. *Journal of Chinese Linguistics*, *42*(1), 218-237.

Zee, E. (1980). A spectrographic investigation of Mandarin tone sandhi. *UCLA working papers in phonetics*, *49*(9).