JASA ARTICLE

Pre-low raising in Cantonese and Thai: Effects of speech rate and vowel quantity[a)]

Albert Lee,[1,b)] Santitham Prom-on,[2] and Yi Xu[3]

[1]*The Education University of Hong Kong, Hong Kong*

[2]*King Mongkut's University of Technology Thonburi, Bangkok, Thailand*

[3]*University College London, London, United Kingdom*

ABSTRACT:

Although pre-low raising (PLR) has been extensively studied as a type of contextual tonal variation, its underlying mechanism is barely understood. This paper explored the effects of phonetic vs phonological duration on PLR in Cantonese and Thai and examined how speech rate and vowel quantity interact with its realization in these languages, respectively. The results for Cantonese revealed that PLR always occurred before a large falling excursion (i.e., high-low); in other tonal contexts, it was observed more often in faster speech. In the Thai corpus, PLR also occurred before large falling excursions, and there was more PLR in short vowels. These results are discussed in terms of possible accounts of the underlying mechanism of PLR. © *2021 Acoustical Society of America*.

https://doi.org/10.1121/10.0002976

(Received 11 March 2020; revised 3 December 2020; accepted 8 December 2020; published online 7 January 2021)

[Editor: Benjamin V. Tucker] Pages: 179–190

I. INTRODUCTION

Pre-low raising (PLR) refers to the raised realization of the high target in a high-low sequence compared to that in high-high. It is a type of contextual tonal variation that has been extensively studied across languages. However, despite its ubiquity, the cause and the underlying mechanism of this phenomenon have hardly been explored. As no language has been reported to defy PLR, a good understanding of how it occurs is of both theoretical and practical importance. Understanding how PLR occurs not only contributes to a better understanding of the division of labour between phonetics and phonology in speech prosody, it is also useful to areas such as speech synthesis and speech-understanding systems. In this paper, we explored the role of duration in PLR realization in both its phonetic (speech rate) and phonological (vowel quantity) senses through two languages that have a rich tonal inventory, namely, Cantonese and Thai.

A. What is PLR?

PLR is a well-known phenomenon in contextual tonal variation, which has been widely reported across languages. Often known otherwise as anticipatory dissimilation (Gandour *et al.*, 1994; Xu, 1997) or anticipatory raising (Connell and Ladd, 1990; Xu, 1999), it is a local anticipatory tonal variation where the f_o of a high tone (H_1) is higher in a H_1L sequence than in a H_1H_2 sequence. Since all the languages reported to show PLR have different lexical prosody, perhaps the only thing they have in common is that the first of two consecutive syllables (henceforth syllable 1) contains a high pitch point, whereas syllable 2 contains a low pitch point. See Lee and Mok (2021) and Xu and Lee (2021) for a review.

Despite extensive reports on the tonal contexts in which PLR occurs, little is known about its underlying mechanism. Franich (2015) found that increased cognitive load was associated with greater PLR but had no effect on carryover tonal variation. This seems to suggest that under normal cognitive load, speakers may have successfully suppressed some of the dissimilatory effect. However, little else is known that might shed light on the underlying mechanism of PLR. This lack of understanding poses a problem when there is a suspected case of PLR, where one tone category might potentially be the PLR-induced allotone of another [cf. Lee *et al.* (2017) on the case of Japanese]—without understanding its cause, it is difficult to provide a reliable diagnosis. This paper attempts to fill this gap by investigating variation of PLR in different speech rate conditions, which is a natural starting point for exploratory studies in speech production.

B. Possible causes of PLR

Although we know of no previous study that has directly investigated the underlying mechanism of PLR, numerous possibilities have been suggested or are conceivable. They can be broadly categorized into articulatory, perceptual, and anatomical accounts.

Based on the findings in his production experiment, Xu (1997) offered two suggestions on the possible causes of PLR. First, PLR might be seen as a strategy to aid reaching a low pitch target, which is articulatorily difficult. Normal

---

[a)] Some of the results of experiment 1 of this work, based on different statistical tests, were reported at the 5th International Symposium on Tonal Aspects of Languages (TAL 2016), Buffalo, NY, USA, May 24–27, 2016.

[b)] Electronic mail: albertlee@eduhk.hk, ORCID: 0000-0002-3224-5788.

speech typically operates just above the floor of one's over two-octave total pitch range (Honorof and Whalen, 2005), which means that the articulation of the low tones would often push one's low pitch limit. The effect of approaching the low limit can be seen in the absence of carryover or anticipatory effects in the low offset of a tone in Xu's production data—one's lower pitch range is much less flexible than its upper counterpart. Physiologically, to raise pitch, one mainly needs to contract the cricothyroid (CT) muscles, which are the only muscles that lengthen the vocal folds (Zemlin, 1988). To lower pitch, however, one needs to both (i) relax CT to unstretch the vocal folds and (ii) lower the larynx so as to increase the effective mass of the vocal folds (Ohala, 1978). The lowering involves contracting multiple extrinsic laryngeal muscles to drag the cricoid cartilage across a spinal curvature in the neck to further shorten the vocal folds (Honda *et al.*, 1999). Therefore, unlike pitch-raising that typically goes well below one's pitch ceiling in normal speech, reaching a low pitch target is articulatorily more difficult. One way to push toward the pitch floor is to generate a high downward velocity, and this can be helped by increasing the distance of the pitch-lowering movement. This is similar to a tennis player first pulling back his/her arm in order to hit the ball hard during a serve or strike (Lee and Mok, 2021; Xu and Lee, 2021). In preparation for an upcoming low target that is articulatorily more difficult to produce, PLR may therefore serve to allow extra distance (by raising f_o peak) for acceleration so as to achieve a higher maximum f_o velocity. This account seems to make good sense, as it is compatible with our current understanding in physics, although how far a principle for free body movement can be extended to f_o control still requires careful examination.

Xu's (1997) second suggestion was that PLR might serve to counteract declination, which can potentially blur contrasts of tone categories. From the perceptual perspective, PLR may be useful for enhancing contrasts between otherwise similar-sounding tones. This echoes the cross-linguistic tendency that languages with more types of stop consonants tend to disperse voice onset time (VOT) values along the VOT continuum (Cho and Ladefoged, 1999). Enhancing tonal contrasts with PLR would be particularly useful for languages like Cantonese, in which most tones are clustered in the lower half of one's tone space, and which is undergoing tone-merger (Mok *et al.*, 2013). Moreover, the perception of level tones is known to strongly depend on context [e.g., Zhang *et al.* (2012)]. In Wong and Diehl (2003), for example, it was reported that a higher preceding context led to more low-tone identification responses. It thus follows that PLR can serve as a useful secondary cue to lexical tones. However, PLR is also present in languages where tone categories are not ambiguous, like in two-tone languages, such as Yoruba, or non-tonal languages, such as English. Therefore, enhancement of perceptual contrasts cannot be taken as the (main) underlying mechanism of PLR.

A related question is whether PLR might be a clear speech strategy [see review in Smiljanić and Bradlow (2009)], as it can expand f_o range. Adult native speakers of English have been found to use a number of strategies when trying to speak clearly (Hazan and Baker, 2011), including higher pitch (median) and larger pitch range, which are reminiscent of PLR. In their data, the exact strategies a speaker used depended on task type (read vs conversational) and listening condition (no barrier vs challenging). The difficulty with this account is that there is no mirror phenomenon of anticipatory lowering before a high pitch target (Xu, 1997, 1999). While it may be true that PLR is part of a communicative strategy to enhance the clarity of speech when needed [cf. Lindblom (1990)], there must be something special about the low pitch articulation that is absent in the articulation of the high pitch.

Finally, a more speculative account concerns speech anatomy. It is known that in mammals, CT is supplied by the external superior laryngeal nerve, whereas all other intrinsic laryngeal nerves are supplied by the recurrent laryngeal nerve. The left branch of the recurrent laryngeal nerve passes under and around the aorta on its way to the larynx, whereas the right recurrent laryngeal nerve passes under and around the subclavian artery. Compared with both branches of the recurrent laryngeal nerve, the external superior laryngeal nerve takes a more direct route to the larynx. If it is the case that neural impulses take less time to reach CT than to other laryngeal muscles, then functions associated with CT contraction (e.g., PLR) may stand out in very fast speech when other muscles (that are supplied by the recurrent laryngeal nerve) cannot keep up to maintain balance. In such a scenario, pitch-raising CT stands out before antagonistic muscles can keep up, leading to PLR. In turn, one would predict that there is more PLR in faster speech than otherwise. There is some evidence pointing in this direction. For example, Udaka *et al.* (1988) reported shorter mean response times for CT (around 23 ms) than lateral cricoarytenoid (LCA) muscles (37.5–42 ms) upon auditory stimulation. However, this difference appears to be too small to motivate this anatomical account. Moreover, although the length of nerves can determine muscle latency (Sims *et al.*, 1996), there are also physical and histological confounding factors that prevent direct testing of this account (Prades *et al.*, 2012).

A plausible account of PLR should be able to explain its occurrence as well as non-occurrence. Considering the articulatory account and the unique properties of CT as reviewed above, as a starting point here, we investigated the effect of speech rate on PLR.

II. EXPERIMENT 1: CANTONESE

A. Introduction

1. Tones in Cantonese

Hong Kong Cantonese was chosen in this study because of its rich tonal inventory (see Fig. 1). Table I describes the contour of the six contrasting tones with their respective tone letters (Chao, 1930). The highest tones are T1 and T2,

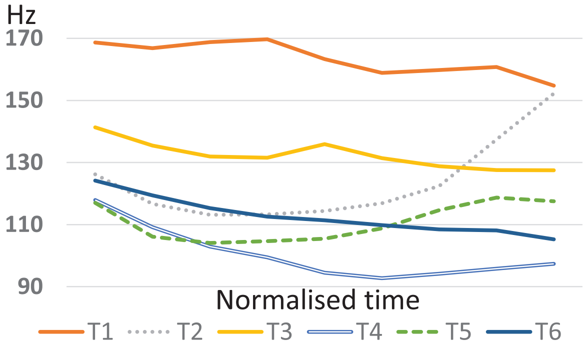

FIG. 1. (Color online) Time-normalized f_o contours of the six lexical tones of Cantonese (carrier syllable /ma/) produced by a male native speaker.

while T4 is the lowest. Presumably, PLR would likely take place in the higher tones T1 and T2, whereas the lowest T4 would likely give rise to it in the preceding syllable, though Gu and Lee (2009) reported otherwise, as will be reviewed below.

2. PLR in Cantonese

Gu and Lee (2009) presented a comprehensive production study on contextual tonal variation in Cantonese. They recorded three native speakers of Hong Kong Cantonese, of which two were professional announcers. Their stimuli were the disyllable *jau wai* in all $6 \times 6 = 36$ tone combinations, spoken under broad focus or with narrow focus on either of the target syllables. Based on visual inspection of mean f_o curves and t tests on mean f_o, they concluded that PLR occurred on T1, T2, and T5, with T2 showing the largest effect. These findings led them to suggest that PLR more likely takes place in rising tones than in level tones. On a side note, Gu and Lee (2009) also reported downstep and post-low bouncing after a low tone that follows syllable 1 and discussed the link between these articulatory phenomena from the point of view of laryngeal muscle coordination.

Although Gu and Lee (2009) offered a clear picture of where PLR could occur in Cantonese, many questions remained unclear. First, while the effect of narrow focus on contextual tonal variation has been investigated, speech rate is another effect that can shed light on this phenomenon. Second, although they mentioned that PLR might be due to the antagonistic forces of pitch-raising CT and pitch-lowering extrinsic laryngeal muscles, exactly how these forces are related to PLR was not discussed. Third, with two out of three of their participants being professional announcers who might produce highly articulate speech, it would be interesting to verify their findings with speakers less trained in enunciating.

TABLE I. Cantonese words contrasting six lexical tones on open syllable /ji:/ [based on Bauer and Benedict (1997) with the high falling tone removed].

Tone	Lexical item	Tone contour	Value
T1	衣 "clothes"	High level	55
T2	椅 "chair"	High rising	25
T3	意 "idea"	Mid level	33
T4	疑 "suspicious"	Mid-low falling	21
T5	耳 "ear"	Mid-low rising	23
T6	二 "two"	Mid-low level	22

Against this backdrop, this study has three goals: (i) verify the claim of Gu and Lee (2009) that only rising tones can serve as PLR hosts; (ii) examine whether speech rate has an effect on PLR (e.g., whether a lower general f_o register associated with slow speech would provide a better trigger for PLR); and (iii) offer an account of the cause of PLR. Here we test two hypotheses. First, (H1) PLR can occur in T1 too—as PLR has been extensively reported in languages without a rising tone [e.g., Laniran and Clements (2003) for Yoruba], it is unlikely that PLR does not occur in the high level tone in Cantonese. Second, (H2a) more PLR can be observed in slower speech. This follows from the fact that one's pitch register is lower in slower speech; thus, a lower syllable 2 would lead to more PLR [cf. Lee *et al.* (2017) for Japanese]. An alternative to this would be that (H2b) there is more PLR in faster speech. This stems from the articulatory account above: to reach a high velocity within a short time, more distance is needed (cf. pulling one's arm further back in order to hit the tennis ball harder). With a better understanding of how PLR interacts with tone shape and speech rate, we would be in a better position to postulate its cause(s).

B. Methods

1. Participants

Six native speakers (three male, including A. L.) of Hong Kong Cantonese were recruited in London for this experiment. They were age 22–30 yr [standard deviation (SD) 4.49 yr] at the time of recording. No one reported any (history of) speech or hearing impairment. All participants were briefed about the experiment and granted written consent before the recording commenced. Five of the speakers were remunerated a small sum for their time.

2. Target sentences

The disyllable *lau man* was chosen for this study. There is a six-way contrast for each of the two syllables, which yielded all 36 (6×6) possible tone combinations. Also, with sonorant initial consonants, these two syllables ensured that continuous f_o contours could be tracked. Target words were framed in the carrier 再講____嗰對字 [*zoi3 gong2* ____ *go2 deoi3 zi6*] "Say the disyllable ___ again." See Table II for details.

Not every Cantonese word can be written with a Chinese character that is known to the average native speaker. For example, for the syllable *man3*, we used the character ⿰亻敏, which is not commonly used. As such, during the experiment occasionally the experimenter had to remind the participants of the pronunciation of this character by

TABLE II. Target sentences of the Cantonese corpus. Transliteration follows the Jyutping convention, in which the number denotes tonal category. The tone values of tones 1–6 are, respectively, 55, 25, 33, 21, 23, and 22 (Bauer and Benedict, 1997).

Carrier	Syllable 1	Syllable 2	Carrier
zoi3 gong2 再講	lau1 褸	man1 蚊	go2 deoi3 zi6 嗰對字
	nau2 扭	man2 抿	
	lau3 嘍	man3 𠍊	
	lau4 留	man4 民	
	lau5 柳	man5 吻	
	lau6 漏	man6 問	

showing words associated with this character (i.e., 𠍊邊 and 𠍊水) on a card without saying them aloud.

Although the character 扭 "twist" is pronounced [nau2], as a result of the /n/-/l/ merger, it is equally natural to pronounce it [lau2] in Hong Kong Cantonese. This merger is an old one, with examples such as the place name 南丫島 [naam4 aa1 dou2] officially translated as *Lamma Island*.

3. Recording procedures

Recording took place in a quiet room at University College London, using a RØDE NT1-A microphone. The sampling rate was 44 100 Hz. Speakers were seated in front of a computer screen, which displayed the stimuli in a randomized order. Speakers were instructed to say each sentence twice, first at normal speed, followed by slow speed. Though speech rate was not stipulated in actual terms, subjects were instructed to speak more slowly in the second production. In this corpus, mean syllable duration was 180.2 ms (SD = 50.3 ms) for normal speech and 309.2 ms (SD = 59.3 ms) for slow speech. Altogether, 6 speakers × 2 speech rates × 36 tone combinations × 5 = 2160 utterances were elicited. Seven utterances (0.32%) were subsequently discarded due to mispronunciation.

4. Data extraction

Sound files were then annotated using ProsodyPro [Xu (2013), version 5.5.1]. Segmentation was done at the level of the syllable. Markings of vocal pulses were manually checked and rectified to ensure accurate tracking of f_o. Apart from the target word itself, the syllable before (*gong2*) as well as the one after (*go2*) were also labelled during annotation, so as not to neglect any carryover effect that extends from or into the target word. Other parts of the carrier sentence were not analyzed in the present study. ProsodyPro then generated acoustical measurements including time-normalized f_o values and f_o velocity for statistical analysis. ProsodyPro calculates f_o velocity according to Eq. (1),

$$f_0^{'} = ((f_{\mathrm{osti}} + 1) - (f_{\mathrm{osti}} - 1))/((\mathrm{t}_i + 1) - (\mathrm{t}_i - 1)). \tag{1}$$

Occasionally, some velocity values generated by ProsodyPro were physiologically implausible [cf. Xu and Sun (2002)]. We discarded any value greater than ±1000 ST/s, accounting for 0.62% ($N = 533$) of the velocity data. For each speaker, all raw f_o values (Hz) were converted into semitones (ST) with the overall mean f_o of that speaker as the reference.

5. Data analysis

The resultant acoustic data were analyzed using growth curve models (Mirman, 2014) and smoothing spline analysis of variance (SS ANOVA) (Davidson, 2006; Gu, 2014). The former have the advantage of incorporating both time coefficients and subject-specific variation, whereas the latter allows us to assess (i) whether different lexical tones in syllable 2 cause significant differences in f_o contours in preceding syllable 1 and, if so, (ii) at which specific time points those differences can be found. These methods complement earlier studies [e.g., Gu and Lee (2009)] of which statistical analyses were based on static point measurements (e.g., maximum and mean f_o). The ST data were analyzed using both growth curve models and SS ANOVA, whereas only the latter was used to analyze f_o velocity, as we were mainly interested in differences at specific points in time.

We fitted a separate model for each lexical tone on syllable 1 using the *lme4* package [Bates *et al.* (2015), version 1.1–19]. We included both the linear and the quadratic time terms (orthogonal polynomials), the main effects of speech rate (contrast-coded) and lexical tone on syllable 2 (T1 as baseline), and their interactions. By-subject random intercepts and by-subject random slopes for speech rate were also included. The dependent variable was f_o (ST) at ten time points across syllable 1. For any model, if f_o is higher before a given lexical tone than before T1 on syllable 2, we take this as evidence of PLR. Although likelihood ratio tests (*anova()*) revealed that lexical tone on syllable 2 had a significant effect on f_o in all models ($p < 0.001$), it was only when syllable 1 was T1 or T2 where T4 on syllable 2 led to a significantly higher f_o compared to T1, i.e., PLR. This means that when syllable 1 bore T3, T4, T5, or T6, our speakers did not show evidence of PLR (i.e., f_o before the baseline T1 was significantly higher in preceding syllable 1 instead). Consequently, these subsets of data will be excluded from our analysis in Sec. II C.

SS ANOVA plots in Secs. II C and III C contain both averaged f_o curves (thin solid lines) and 95% Bayesian confidence intervals (width of the color ribbons) around the averaged curves. The x axis represents normalized time, and the y axis represents f_o or f_o velocity. At any point in time, if the confidence intervals of two conditions do not overlap, they are considered significantly different. See Davidson (2006) for a more detailed description.

C. Results

1. f_o contours

Table III presents the results of the growth curve analysis of the realization of T1 and T2 on syllable 1 (see results for other tones in the supplemental data[1]). A fixed effect is

TABLE III. Growth curve analysis on f_{o} realization of Cantonese T1 and T2 on syllable 1. Significant effects are in bold ($t < 2.0$).

	T1 on syllable 1				T2 on syllable 1			
	Fixed			Random by-speaker SD	Fixed			Random by-speaker SD
	β	s.e.	t		β	s.e.	t	
(Intercept)	2.868	0.257	**11.178**	0.618	0.228	0.245	0.933	0.588
Time (linear)	466.370	33.934	**13.743**		−1191.000	34.880	**−34.157**	
Time (quadratic)	−419.037		**−12.348**		1015.000		**29.100**	
Rate	1.884	0.554	**3.397**	1.339	3.331	0.462	**7.205**	1.108
T2 − T1	0.740	0.065	**11.389**		0.071	0.067	1.058	
T3 − T1	0.314		**4.827**		0.267		**3.993**	
T4 − T1	1.379		**21.229**		0.749		**11.197**	
T5 − T1	0.425		**6.546**		0.215		**3.208**	
T6 − T1	0.646	0.066	**9.850**		0.491		**7.314**	
Rate × (T2 − T1)	1.215	0.130	**9.349**		−0.215	0.134	−1.604	
Rate × (T3 − T1)	0.447		**3.436**		−0.395		**−2.955**	
Rate × (T4 − T1)	0.741		**5.701**		0.682		**5.096**	
Rate × (T5 − T1)	0.884		**6.802**		−0.290		**−2.170**	
Rate × (T6 − T1)	1.876	0.131	**14.311**		−0.021		−0.159	

considered significant if the absolute value of the t-statistic is greater than or equal to 2.0 (Gelman and Hill, 2007). To conserve space, here we focus on the main trends and discuss the interactions in detail in the SS ANOVA analysis to follow. The positive estimates for speech rate [T1: $\beta = 1.884$, standard error (s.e.) $= 0.554$, $t = 3.397$; T2: $\beta = 3.331$, s.e. $= 0.462$, $t = 7.205$] indicate that syllable 1 f_{o} was higher at the normal speech rate than in slow speech in general. The positive estimates for Tx − T1 (lexical tone on syllable 2) contrasts show that all these tones could give rise to PLR in syllable 1 that bore T1 or T2, except that the T2T2 sequence was not significantly higher than T2T1 ($\beta = 0.071$, s.e. $= 0.067$, $t = 1.058$). The significant interactions between speech rate and lexical tone show the change in magnitude of PLR in normal speech vs slow speech. For example, before a T4, mean T1 f_{o} was 20.9 Hz higher than the baseline in normal speech but 11.0 Hz higher in slow speech ($\beta = 0.741$, s.e. $= 0.130$, $t = 5.701$).

Figure 2 shows the averaged f_{o} contours of 30 repetitions from six speakers, with the second interval kept constant (T1 or T2 on syllable 1). Vertical lines represent syllable boundaries. Here the TxT1 sequences serve as the baseline. Any contour significantly higher than the baseline in syllable 1 would constitute a case of PLR. In the two upper panels, the T1T4 contours are significantly higher than T1T1 across the entire syllable 1, showing clear evidence of PLR. In the bottom panel, the T2T4 contour is also significantly higher than T2T1, though in only part of the second interval, while in the rest of the syllable the two conditions overlapped.

In other tonal contexts, PLR appeared to be dependent upon speech rate, i.e., present in faster speech but absent in slower speech. For example, for the T1T6 sequence in Fig. 2, PLR was observed only in normal speech and not in slow speech (i.e., the T1T6 contour is not higher than T1T1 in slow speech in syllable 1). The same was true for T1T2, T1T3, and T1T5, where PLR was only observed in faster speech. While slow speech has a lower global f_{o} register (global mean f_{o} in our data is 172 Hz for normal speech and 145 Hz for slow speech), the resultant lower f_{o} in syllable 2 did not give rise to more PLR; this suggests that a low syllable 2 is not the only factor underlying this phenomenon.

Finally, as Table III has shown, where syllable 1 was not a high tone (T1 or T2), PLR did not occur even if syllable 2 was low (T4). Refer to the supplemental data[1] for a complete set of SS ANOVA plots for all syllable 2 tone and speech rate conditions.

2. f_{o} *velocity*

Next, f_{o} velocity in syllable 2 (third interval) is considered. Recall that there was PLR in T1T6 (see Fig. 2) in normal speech but not in slow speech. Figure 3 shows the maximum falling velocity of all syllable 2 tone × speech rate conditions. In cases of PLR, the maximum falling velocity was much greater than otherwise. The same pattern was observed after visual inspection of the velocity profiles of other tone sequences (see supplemental data[1]). Judging from Fig. 3, it appears that all cases of PLR in this corpus had a maximum falling velocity in syllable 2 greater than 400 ST/s; similarly, in the slow condition, those without PLR all appear to have peak velocity values below 300 ST/s.

3. *Correlation analysis*

Finally, linear regression analysis was performed to verify the observations in Figs. 2 and 3. To calculate the correlation between mean syllable duration and PLR, we (i) first averaged all repetitions of the same speaker and then (ii) for each tone (T1 and T2) in syllable 1, measured the difference between each tone in syllable 2 (T2 − T6) and T1. For normal speech, mean syllable duration was inversely correlated

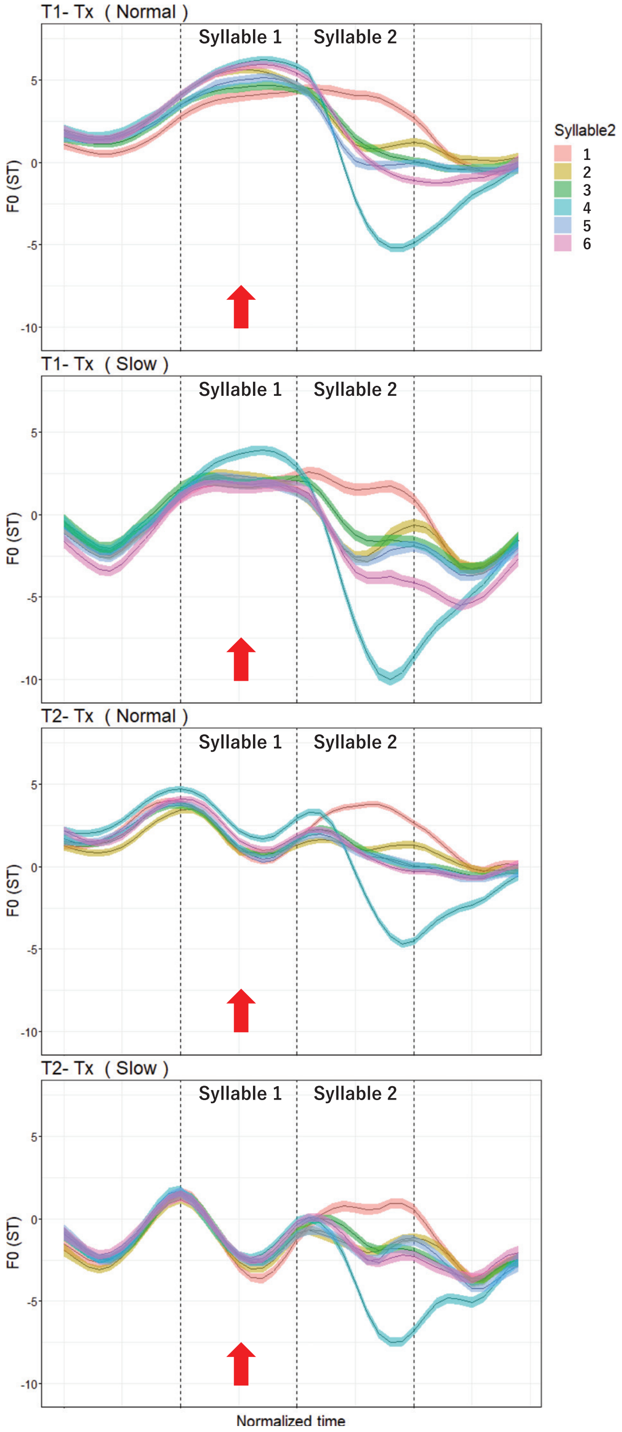

FIG. 2. (Color online) SS ANOVA plots showing mean f_o contours averaged across six Cantonese speakers.

with mean PLR, $r = -0.234$, $N = 60$, $p = 0.036$ (one-tailed); for slow speech, the same correlation was non-significant, $r = 0.026$, $N = 60$, $p = 0.423$.

D. Interim discussion

This experiment set out to test two hypotheses: (H1) PLR can occur in T1, and (H2a) more PLR would be observed in slower speech/(H2b) in faster speech. We found that PLR occurred in T1 as well as in T2 and that there was more PLR in fast speech than in slow speech. These results clearly refuted (H2a), while supporting (H1) and (H2b).

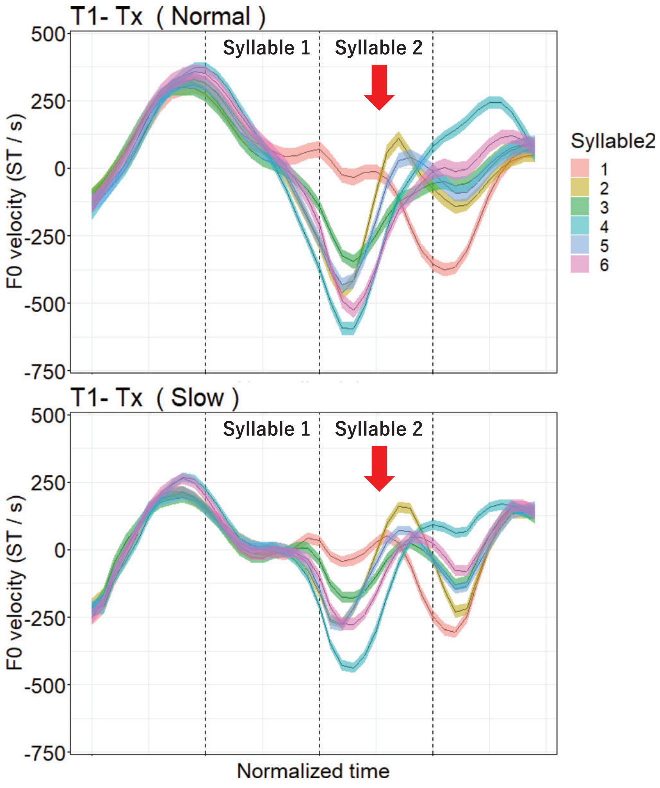

FIG. 3. (Color online) SS ANOVA plots showing mean f_o velocity contours averaged across six Cantonese speakers. In the left panel, PLR occurred in all tone pairs; in the right panel, PLR was observed only in T1T4 (turquoise).

That PLR could occur in T1 in our data is not surprising, as PLR commonly occurs in the high tone in many languages. PLR in rising T2 in our data was also consistent with Gu and Lee (2009), in which the raising appeared not to span entire syllable 1-equivalent either. What is more mysterious is why PLR was not observed in T1 in Gu and Lee (2009). Conceivable reasons for this discrepancy include task effect (i.e., focus vs speech rate, different target syllables) and precision in speakers' articulation [use of professional news readers in Gu and Lee (2009)].

However, it was interesting that H2a was not supported. In Lee *et al.* (2017), we observed a higher H* before a lower following L and attributed this to PLR. The gradient effect observed in Japanese could not be applied to Cantonese, likely because of the difference in lexical prosody of the two languages—the L target in the Japanese data of Lee *et al.* (2017) was probably way lower than any non-T4 Cantonese tones even at its highest phonetic realization. Taking together the results of Lee *et al.* (2017) and the present data, it seems that whether PLR occurs may be binary and conditional upon a low enough syllable 2; then in cases where PLR does occur, the exact amount of raising is gradient and determined by the lowness of the following target.

The durational effect found in this experiment is novel and requires further verification. As we have seen how speech rate affects PLR, a natural extension would be to see whether the phonological use of duration (i.e., vowel quantity) has the same effect. To this end, we chose Thai for our follow-up experiment, to be described below.

III. EXPERIMENT 2: THAI

A. Introduction

In Sec. II, we reported the effect of duration on PLR realization in Cantonese. As PLR is assumed to be an articulatory, in turn universal, phenomenon, it is important to understand its nature by comparing any proposed effect across different languages. In this section, we explore PLR in Thai, which provides a suitable testing ground for the effect of duration in the abstract sense (i.e., vowel quantity). While speech rate is concerned with syllable duration at a global level (i.e., utterance or longer), it would be interesting to see whether durational contrasts at the syllable level would affect the realization of PLR in a similar way.

Thai has five lexical tones, which contrast in height and contour, namely, mid, low, fall, high, and rise [Tingsabadh and Abramson (1993); see also Table IV and Fig. 4]. Vowels contrast in quantity, with duration being the primary cue (Potisuk *et al.*, 1998), though in specific stress conditions, the durational contrast can be lost (Potisuk *et al.*, 1998).

Gandour *et al.* (1994) have reported clear evidence of PLR in Thai, though vowel quantity was not investigated in that study. They found that both the rising and low tones could lead to PLR in the preceding syllable (mid, rising, or high). This echoes their remark that, of the five Thai tones, "low and rising tones had low f_o onsets, falling and high tones high f_o onsets, and mid tone intermediate onsets" (Gandour *et al.*, 1994, p. 483). They also noted that raised f_o due to PLR spanned only a portion of the duration of syllable 1 (e.g., the last 30% of a high tone, unlike in Fig. 2, where PLR effects in Cantonese spanned the entire syllable 1). To better understand the findings in the Cantonese experiment above, here we reanalyzed the production data from Xu and Prom-on (2014) on contextual tonal variation, which are highly comparable with our Cantonese data in terms of design and elicitation method. Xu and Prom-on (2014) pointed out PLR as one source of residual errors in their f_o synthesis but did not provide further acoustic details. Although this set of data was originally designed for a different purpose (i.e., f_o modeling), it would also be an ideal corpus for examining PLR in Thai in greater detail than before.

Based on the Cantonese results reported above and in Gandour *et al.* (1994), here we tested the hypotheses that there are (H3) always PLR in high-low, high-rise, rise-low, and rise-rise sequences and (H4) more cases of PLR in short syllables (comparable to fast speech) than long syllables. H3 is based on the observation by Gandour *et al.* (1994) that the low and rising tones have low f_o onsets, whereas the offsets of high and rising are high. The resultant long falling excursion would thus be a likely environment for PLR regardless of vowel quantity. H4 assumes that short vowels are comparable to the faster speech rate in Cantonese and would thus permit PLR in contexts otherwise not possible for PLR in the long vowel conditions. Furthermore, we are also interested in whether the apparent 400 ST/s threshold in the Cantonese data also holds for Thai.

B. Methods

1. Corpus

The speech material was recorded by five native speakers (two females) of Standard Thai (Xu and Prom-on, 2014). They were undergraduate students age 20–25 yr, studying at King Mongkut's University of Technology Thonburi, Bangkok, Thailand. The dataset consists of four-syllable sentences in which the tones of the two middle syllables vary across all five Thai tones [mid (T0), low (T1), falling (T2), high (T3), and rising (T4)] and two vowel lengths (short and long), cf. Table V. The first and the last syllables were always the mid tone to minimize carryover and anticipatory influences on the two middle syllables.

Altogether, there were 100 tone × vowel length combinations. Each utterance was produced five times by each speaker, and the recording was done at the sample rate of 22.05 kHz and 16-bit resolution. Participants were recorded at the normal speaking rate. Altogether, there were 5 speakers × 4 quantity conditions × 25 tone combinations × 5 = 2500 utterances. Six utterances (0.24%) were excluded from subsequent analysis due to misproduction. In the subset of corpus of interest (high or rising on syllable 1, $N = 994$), mean syllable 1 duration was 305 ms (SD = 31 ms)

TABLE IV. Thai words contrasting five lexical tones on open syllable /k^ha:/.

Tone	Lexical item	Tone contour
T0	คา "stick"	Mid
T1	ข่า "galangal"	Low
T2	ค่า "value"	Falling
T3	ค้า "to trade"	High
T4	ขา "leg"	Rising

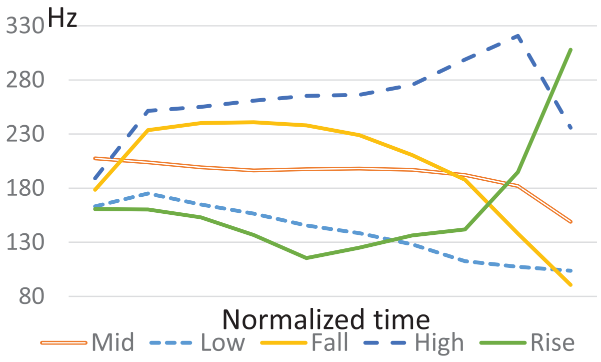

FIG. 4. (Color online) Time-normalized f_o contours of the five lexical tones of Thai (carrier syllable /ga/) produced by a female native speaker.

TABLE V. Target sentences of the Thai corpus [first reported in Xu and Prom-on (2014)].

1st	2nd	3rd	4th
k^hun0 “คุณ”	ʔa:0/nim0 “อา/นิม” no:j1/mam1 “หน่อย/หม่ำ” mɛ:2/nim2 “แม่/นิ่ม” na:3/miŋ3 “น้า/มิ้ง” la:n4/jiŋ4 “หลาน/หญิง”	la:0/loŋ0 “ลา/หลง” a:n1/man1 “อ่าน/หมั่น” wa:ŋ2/maj2 “ว่าง/ไม่” ne:n3/lom3 “เน้น/ล้ม ha:4/loŋ4 “ลา/หลง”	ŋa:n0 or ma:0 “งาน/มา”

for long vowels and 288 ms (SD = 33 ms) for short vowels. One-tailed paired samples *t* test confirmed that the difference was significant [$t(9) = 4.151, p < 0.001$].

2. Data analysis

Data extraction and analysis procedures were the same as in the Cantonese analysis above. For the growth curve models, the fixed factor of speech rate was replaced by quantity. In the model for the rising tone (see Table VI), by-speaker random slopes were not included due to non-convergence of the model. Like for the Cantonese data, velocity values greater than ±1000 ST/s were discarded, accounting for 0.19% ($N = 97$) of the velocity data.

C. Results

1. f_o contours

This experiment set out to test whether duration in terms of phonological quantity influences the occurrence of PLR in Thai. Our hypotheses were that there is (H3) always PLR in high-low, high-rise, rise-low, and rise-rise sequences and (H4) more cases of PLR in short syllables.

Table VI shows the summary of growth curve analysis on f_o realization of Thai high tone and rising tone on syllable 1 (see results for other tones in the supplemental data[1]). All of the mid ($\beta = 0.606$, s.e. = 0.073, $t = 8.332$), low ($\beta = 1.205$, s.e. = 0.073, $t = 16.567$), falling ($\beta = 0.285$, s.e. = 0.073, $t = 3.921$), and rising ($\beta = 0.645$, s.e. = 0.073, $t = 8.867$) tones on syllable 2 led to significantly higher realization of the high tone in syllable 1, compared to the baseline condition (high tone on syllable 2). Compared to the short-short quantity condition, in all of the long-long ($\beta = -0.899$, s.e. = 0.103, $t = -8.733$), long-short ($\beta = -0.587$, s.e. = 0.103, $t = -5.702$), and short-long ($\beta = -0.498$, s.e. = 0.103, $t = -4.840$) conditions, the low tone on syllable 2 led to significantly less increase in f_o in preceding high tone, i.e., more PLR in short-short. Similarly, when syllable 1 bore the rising tone, all of the

TABLE VI. Growth curve analysis on f_o realization of Thai high tone and rising tone on syllable 1. Notation of the baseline level for tone (high) is omitted in the interaction terms. Significant effects are in bold ($t < 2.0$).

	High on syllable 1				Rising on syllable 1			
	Fixed			Random	Fixed			Random
	β	s.e.	*t*	by-speaker SD	β	s.e.	*t*	by-speaker SD
(Intercept)	1.501	0.338	**4.439**	0.747	−1.899	0.212	**−8.943**	0.458
Time (linear)	−176.647	8.106	**−21.792**		−647.900	8.999	**−72.001**	
Time (quadratic)	239.900		**29.595**		474.000		**52.669**	
Quantity (LL − SS)	−1.416	0.443	**−3.197**	0.977	−0.149	0.080	−1.855	
Quantity (LS − SS)	−0.991	0.234	**−4.232**	0.498	−0.218		**−2.718**	
Quantity (SL − SS)	−0.182	0.236	−0.773	0.501	0.335		**4.178**	
Mid − high	0.606	0.073	**8.332**		0.358		**4.455**	
Low − high	1.205		**16.567**		0.335		**4.175**	
Falling − high	0.285		**3.921**		0.169		**2.099**	
Rising − high	0.645		**8.867**		0.319		**3.971**	
Quantity (LL − SS) × mid	−0.192	0.103	−1.863		−0.223	0.114	−1.965	
Quantity (LS − SS) × mid	−0.416		**−4.038**		−0.285		**−2.506**	
Quantity (SL − SS) × mid	−0.028		−0.277		−0.364		**−3.208**	
Quantity (LL − SS) × low	−0.899		**−8.733**		0.056		−0.494	
Quantity (LS − SS) × low	−0.587		**−5.702**		−0.120		−1.052	
Quantity (SL − SS) × low	−0.498		**−4.840**		−0.005	0.117	−0.041	
Quantity (LL − SS) × falling	−0.299		**−2.908**		−0.147	0.114	−1.297	
Quantity (LS − SS) × falling	−0.191		−1.856		−0.329		**−2.893**	
Quantity (SL − SS) × falling	−0.349		**−3.393**		−0.583		**−5.130**	
Quantity (LL − SS) × rising	0.214		**2.077**		−0.022		−0.196	
Quantity (LS − SS) × rising	−0.018		−0.179		−0.317		**−2.794**	
Quantity (SL − SS) × rising	−0.034		−0.329		0.009		−0.081	

mid ($\beta = 0.358$, s.e. $= 0.080$, $t = 4.455$), low ($\beta = 0.335$, s.e. $= 0.080$, $t = 4.175$), falling ($\beta = 0.169$, s.e. $= 0.080$, $t = 2.099$), and rising ($\beta = 0.319$, s.e. $= 0.080$, $t = 3.971$) tones on syllable 2 led to significantly higher realization in the preceding syllable. Both the high and the rising tones on syllable 1 were significantly higher in f_o in the short-short condition than in the long-short condition ($\beta = 0.169$, s.e. $= 0.080$, $t = 2.099$, $\beta = 0.169$, s.e. $= 0.080$, $t = 2.099$).

Figure 5 shows the f_o contours of high-x and rise-x sequences in short-short and long-long contexts. For high-x sequences, in both quantity conditions, there was clear PLR in high-mid, high-low, and high-rise, but not in high-fall, all compared with the high-high baseline. In the short-short context, high-low manifested the greatest PLR effect; in the long-long context, high-rise showed the most PLR instead. Moreover, in the short-short context, the PLR contours all diverged from the high-high baseline in the first half of the first syllable, whereas in the long-long context, this divergence mostly began at 50% into the first syllable. Where syllable 1 was the rising tone, the mid tone on syllable 2 did not seem to incur PLR in the preceding syllable. The low, falling, and rising tones led to significantly higher realization of preceding rising tone, but this raising effect spanned only the last 30% of syllable 1. Refer to the supplemental data[1] for a complete set of SS ANOVA plots for all syllable 2 tone and speech rate conditions.

2. f_o velocity

For f_o velocity, we were interested in whether the 400 ST/s dividing line in Cantonese would also apply to Thai. Figure 6 shows that although all PLR cases had a greater maximum falling f_o velocity than the baseline, only some of them exceeded 400 ST/s, namely, high-low and high-rise in the short-short context and high-rise in the long-long context. Refer to the supplemental data[1] for a complete set of SS ANOVA plots for all syllable 2 tone and speech rate conditions.

3. Correlation analysis

Finally, linear regression showed that for the short-short condition, mean syllable duration was positively correlated with mean PLR, $r = 0.169$, $N = 100$, $p = 0.046$ (one-tailed). No significant correlation between syllable duration and PLR was observed in any other quantity conditions.

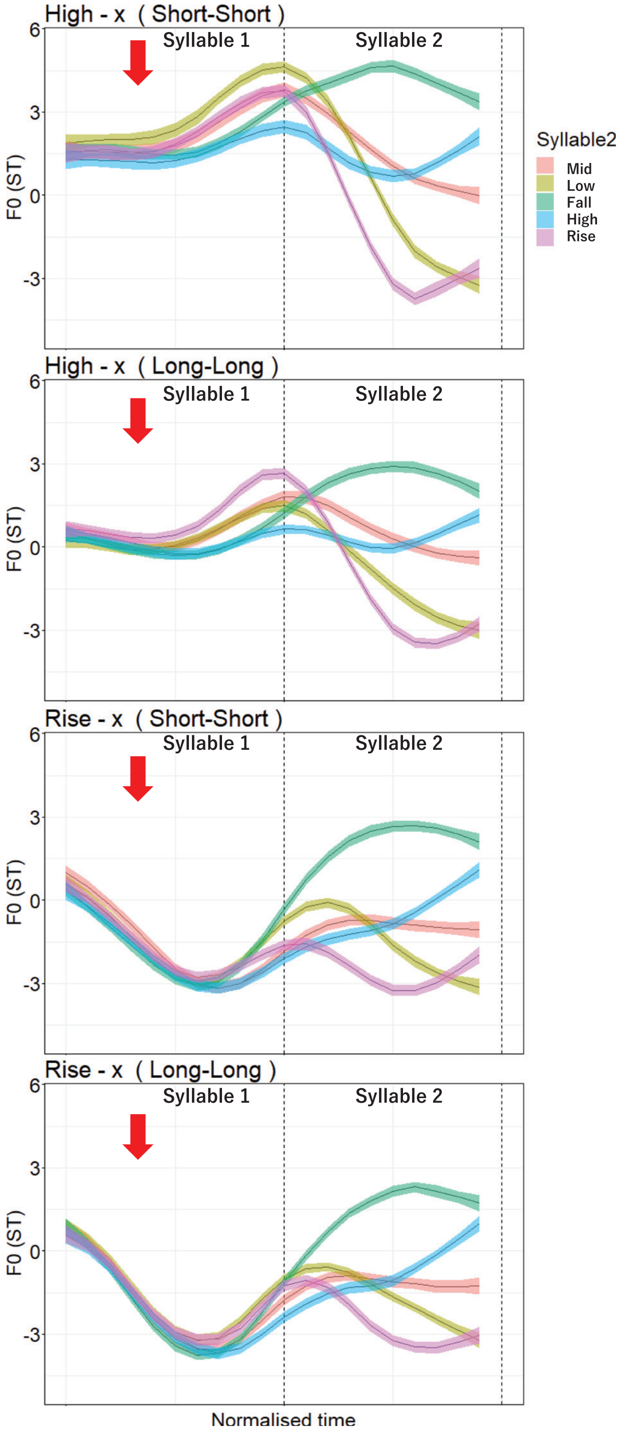

FIG. 5. (Color online) SS ANOVA plots showing mean f_o contours averaged across five Thai speakers.

IV. GENERAL DISCUSSION

A. Summary of findings

1. Cantonese

This paper set out to extend previous work by Gu and Lee (2009) and explored the underlying mechanism of PLR. We observed PLR when the falling excursion is large (T1T4 and T2T4) or when the fall is fast (T1Tx in faster speech). We also found that for any PLR to occur, syllable 1 must be high, as syllable 1 low in f_o did not have PLR. Although one might assume that a low syllable 2 is the key to PLR, the results suggest that a high syllable 1 and a fast fall are at least as important if not more.

These findings are compatible with Gu and Lee (2009) in general, though there are also differences. In Gu and Lee (2009), where the effect of focus was examined, PLR was mainly observed in T2 and T5 on syllable 1. On the other hand, in the present study, we looked at the effect of speech rate and found instead that PLR consistently occurred in T1 and T2. Taken together, these two studies suggest that PLR

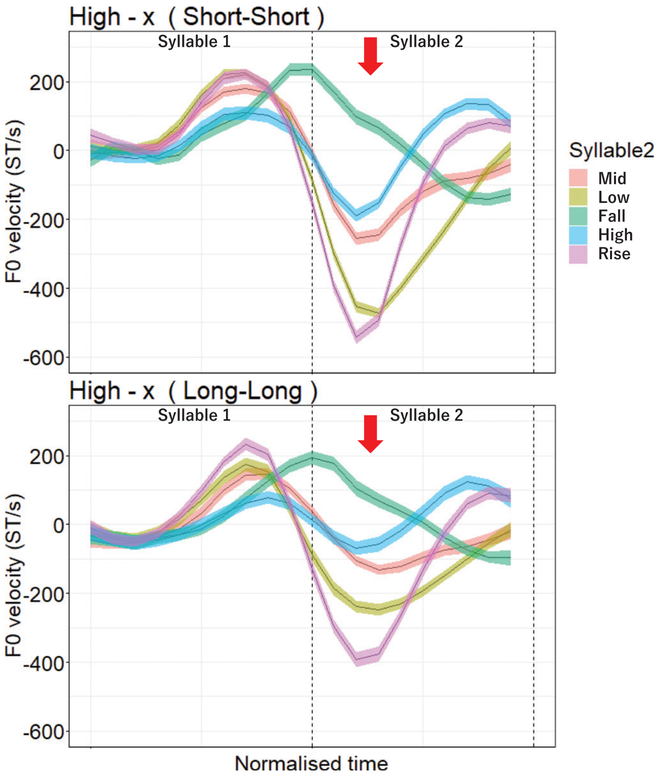

FIG. 6. (Color online) SS ANOVA plots showing mean f_o velocity contours averaged across five Thai speakers.

in Cantonese is subject to factors including f_o of syllable 1, f_o of syllable 2, speech rate, and focus.

2. Thai

The Thai experiment served as a cross-linguistic verification and extension of the findings of experiment 1. Growth curve analysis (Table VI) suggests that all of the four tones could lead to some raising in the preceding syllable in Thai compared to the high baseline, thus supporting H3. Furthermore, the significant interaction between quantity and tone on syllable 2 shows that there was greater PLR in short-short than in any other quantity conditions, thus supporting H4. These two observations bring the Thai data in line with Cantonese in terms of the behavior of PLR.

However, there were also notable differences between Thai and Cantonese. First, upon careful inspection of SS ANOVA plots, we noticed that the raising effect of PLR was largely restricted to the final portion (approximately 30%) of syllable 1 for Thai. Duration in terms of phonemic quantity appears to mainly affect the relative timing of the divergence of the baseline and the PLR condition. This is in contrast to Cantonese, where PLR effects often span the entire syllable 1. This could potentially be attributed to the longer mean syllable duration in the Thai corpus [mean syllable 1 duration with high or rising, 296.1 ms (SD = 32.9 ms)] than in the Cantonese corpus [mean syllable 1 duration with T1 or T2, 247.6 ms (SD = 86.6 ms)]. Second, in cases where PLR was large in Thai (e.g., syllable 2 = rising), maximum falling velocity exceeded −400 ST/s, like in Cantonese. But in other PLR cases, it was ∼−200 ST/s (Fig. 6). Thus, the difference in PLR between Cantonese and Thai lay not only in how far they spanned in syllable 1, but also in their relationship with the corresponding maximum falling velocity, which in turn is associated with articulatory strength. A third difference is that unlike Cantonese, the Thai rising tone does not seem to allow as much PLR as the high tone does. A closer inspection of the SS ANOVA plots reveals that the Thai rising tone occupies a much lower f_o range than the high tone. In fact, to produce the Thai rising tone, speakers first dip toward their pitch floor before rising again—likely involving a completely different set of laryngeal muscles (i.e., pitch-lowering extrinsic laryngeal muscles) than the Thai high tone. Thus, the smaller PLR effect here seems to lend further support to the physiological account, which will be explained further.

B. PLR to increase maximum velocity

The results of this study are consistent with the velocity account of PLR. That is, by raising pitch in the preceding syllable, the distance of the downward movement toward the low tone is increased, which would help generate a high downward velocity to push toward the pitch floor, which is known to be hard to reach. The speech rate effect in the Cantonese data fits into this account, because faster speech (where PLR occurs) requires a high maximum velocity; thus, a higher starting point would be required for acceleration. A non-low syllable 2 (e.g., T1T3) spoken slowly involves no fast movement or large excursion and thus yields no PLR.

The smaller PLR effect on the Thai rising tone, meanwhile, is likely attributable to another property of CT—allowing quick changes in f_o. While CT would not otherwise be very active in one's lower f_o range, here some PLR is still observed because the rising tone followed by other tones requires very rapid f_o movements—the specialty of the *pars recta* belly (Mu and Sanders, 2009), which will be explained further below.

C. A perceptual account for PLR?

PLR may enhance tonal contrasts to aid comprehension. Researchers have shown that Cantonese is undergoing tone-merger (Mok *et al.*, 2013) and that some native speakers are becoming less able to perceive the difference between certain similar tones; the magnitude of PLR can help distinguish between, for example, T4 and T6 in syllable 2. That said, while PLR may possibly facilitate tonal identification to some extent, this benefit cannot explain the occurrence of PLR *per se*. This is because PLR occurs only at the upper end of the tonal space, where tonal contrasts are hardly ambiguous; the fact that PLR is absent in non-high syllable 1, where tonal contrasts are ambiguous, renders this hypothesis rather unlikely. More importantly, PLR does not only occur in languages with many tones, but also in languages with fewer tones {e.g., three tones in Yoruba [see Laniran and Clements (2003)] and in Bimoba [see Snider (1998)]}, where contrast enhancement is not necessary. A contrast enhancement account, therefore, cannot be taken as the underlying mechanism of this phenomenon.

D. An anatomical account for PLR?

Yet another possible account for PLR comes from the innervation patterns of intrinsic laryngeal muscles. Here CT is hypothesized to be the direct cause of PLR. If PLR was not actively planned, it may be the result of physical constraints (nature of CT in relation to other laryngeal muscles). Recall that PLR depends on the excursion size as well as the speed of f_o fall, both of which are closely related to the properties of CT. The former, in particular the fact that PLR is absent when the fall starts from a non-high tone, echoes the fact that CT is active in one's upper pitch range; when the fall starts from the middle of one's pitch range, there may be little CT activity to begin with, thus no PLR. The latter point ties in well with the fact that CT activity is not responsible for a f_o fall that is steady and gradual (Collier, 1975). It is also consistent with a part of CT that is capable of very fast f_o movements, namely, the *pars recta* belly (Mu and Sanders, 2009). Hence, even when the falling excursion is small, PLR would still occur before a steep fall, as CT is required for fast f_o movement.

Laryngeal muscles work together to maintain balance in vocal fold tension, and some are antagonistic to one another. Normally, the contraction of different laryngeal muscles is timed to ensure precise f_o control. However, if we assume that some intrinsic laryngeal muscles (i.e., CT) are faster than others, then the slower ones may not catch up in fast speech as well as CT; and if it is the ones antagonistic to CT that do not catch up, then the effect of CT contraction would stand out unchecked, resulting in PLR.

For this hypothesis to be true, it is necessary to establish that CT is a much faster muscle than other intrinsic laryngeal muscles that are involved in f_o control. Two pieces of evidence appear to be supportive. First, CT is innervated by the external superior laryngeal nerve, whereas all other intrinsic laryngeal muscles are supplied by the recurrent laryngeal nerve. In mammals, the external superior laryngeal nerve is much shorter in length than the recurrent laryngeal nerve, meaning that motor commands go through a much shorter course to reach CT than they do to reach other muscles. One study looking at laryngeal muscle potentials under auditory stimulation found that CT had a shorter latency than lateral cricoarytenoid (Udaka *et al.*, 1988). Moreover, the rectus belly of CT that is responsible for fast f_o changes is supplied by 3–7 branches of the external superior laryngeal nerve (Mu and Sanders, 2009), lending further support to this account.

Second, factors that raise f_o usually raise intensity as well. Where f_o is deliberately held constant and intensity left to vary (e.g., production of swelltone), CT activity is found to decrease with increasing intensity, so as to suppress involuntary f_o rises (Hirano *et al.*, 1970). Although a full acoustical analysis would be beyond the scope of this paper, our intensity results show that cases with PLR do not also see higher intensity, suggesting that the raised f_o is due to CT contraction alone, like in Hirano *et al.* (1970). Needless to say, any speculation on the cause of PLR related to muscle coordination must be verified with articulatory measurements such as electromyography.

E. Suggestions for future research

The most direct implication of our findings is that we could test suspected cases of PLR in the future based on our new understanding of this phenomenon. For example, the present results are in line with the Japanese pitch accent, a case argued to be due to PLR (Lee *et al.*, 2017). The extra high f_o associated with the Japanese pitch accent is argued to be the result of PLR (i.e., derived), instead of being an underlying articulatory target in its own right. As an accented word ends in a steep fall, our data explain why "PLR" occurs even in slow speech in Japanese. Previously, it has been difficult to motivate this account due to theory-internal reasons regarding Japanese phonology. With a slightly better understanding of PLR, it is now possible to diagnose ambiguous cases like Japanese based on such acoustic properties as f_o excursion and velocity at various speech rate conditions.

Another interesting observation from the data that was beyond the scope of this study was that the T4T4 sequence in Cantonese was always realized significantly higher than any other T*x*T4 sequence, with the difference being much larger in slow speech. Similarly, though to a much lesser extent, the low-low sequence in Thai was also realized significantly higher than some other tonal contexts. It is unclear whether this is idiosyncratic or another articulatory phenomenon pertaining to continuous low targets. The reader is referred to the supplemental data[1] for details.

Third, it would be beneficial to verify the present findings with additional manipulation of speech rate of Cantonese and Thai or of other languages.[2] With more data, we may be able to predict when exactly PLR may occur in different conditions (e.g., speech rate, pitch excursion). In turn, this would contribute to the accuracy of f_o synthesis, among other applications.

Finally, while this paper has explored PLR from the perspective of speech production, currently little is known about the relationship between this phenomenon and perception, with exceptions such as Wong and Diehl (2003). How much PLR contributes to tonal perception in languages with many tones, e.g., Thai and Cantonese, warrants more detailed investigation.

V. CONCLUSION

In this study, we found that for Cantonese, there was PLR either when falling excursion was large or when speech was fast; Thai showed a similar behavior to Cantonese in that there was more PLR in short vowels. Cases with large PLR effects often coincided with great maximum falling velocity values, e.g., <-400 ST/s. Given our findings, we argue that PLR serves to allow more room for acceleration in preparation for an upcoming falling excursion.

---

[1] See supplementary material at https://www.scitation.org/doi/suppl/10.1121/10.0002976 for complete SS ANOVA f_o plots for Cantonese, complete SS ANOVA f_o velocity plots for Cantonese, complete SS ANOVA f_o plots for Thai, model summaries for Cantonese, and model summaries for Thai.

[2] We owe this suggestion to Professor Benjamin Tucker.

Bates, D. M., Mächler, M., Bolker, B. M., and Walker, S. C. (**2015**). "Fitting linear mixed-effects models using {lme4}," J. Stat. Softw. **67**, 1–48.

Bauer, R. S., and Benedict, P. K. (**1997**). *Modern Cantonese Phonology* (Mouton de Gruyter, Berlin).

Chao, Y.-R. (**1930**). "A system of tone-letters," Le Maître Phonétique **45**, 24–27.

Cho, T., and Ladefoged, P. N. (**1999**). "Variation and universals in VOT: Evidence from 18 languages," J. Phon. **27**, 207–229.

Collier, R. (**1975**). "Physiological correlates of intonation patterns," J. Acoust. Soc. Am. **58**, 249–256.

Connell, B., and Ladd, D. R. (**1990**). "Aspects of pitch realisation in Yoruba," Phonology **7**, 1–29.

Davidson, L. S. (**2006**). "Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance," J. Acoust. Soc. Am. **120**, 407–415.

Franich, K. (**2015**). "The effect of cognitive load on tonal coarticulation," *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS 2015)*, Glasgow, Scotland (August 10–14).

Gandour, J. T., Potisuk, S., and Dechongkit, S. (**1994**). "Tonal coarticulation in Thai," J. Phon. **22**, 477–492.

Gelman, A., and Hill, J. (**2007**). *Data Analysis Using Regression and Multilevel/Hierarchical Models* (Cambridge University Press, Cambridge, UK).

Gu, C. (**2014**). "Smoothing spline ANOVA models: R package gss," J. Stat. Softw. **58**, 1–25.

Gu, W., and Lee, T. (**2009**). "Effects of tone and emphatic focus on F0 contours of Cantonese speech: A comparison with Standard Chinese," Chin. J. Phon. **2**, 133–147.

Hazan, V. L., and Baker, R. E. (**2011**). "Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions," J. Acoust. Soc. Am. **130**, 2139–2152.

Hirano, M., Vennard, W., and Ohala, J. J. (**1970**). "Regulation of register, pitch and intensity of voice," Folia Phoniatr. Logop. **22**, 1–20.

Honda, K., Hirai, H., Masaki, S., and Shimada, Y. (**1999**). "Role of vertical larynx movement and cervical lordosis in F0 control," Lang. Speech **42**, 401–411.

Honorof, D. N., and Whalen, D. H. (**2005**). "Perception of pitch location within a speaker's F0 range," J. Acoust. Soc. Am. **117**, 2193–2200.

Laniran, Y. O., and Clements, G. N. (**2003**). "Downstep and high raising: Interacting factors in Yoruba tone production," J. Phon. **31**, 203–250.

Lee, A., and Mok, P. K. P. (**2021**). "Lexical tone," in *The Cambridge Handbook of Phonetics*, edited by J. Setter and R.-A. Knight (Cambridge University Press, Cambridge, UK), in press, see https://www.cambridge.org/core/books/cambridge-handbook-of-phonetics/4C35C9BE4CE7D93EB21C06E90BEE68A9.

Lee, A., Prom-on, S., and Xu, Y. (**2017**). "Pre-low raising in Japanese pitch accent," Phonetica **74**, 231–246.

Lindblom, B. E. F. (**1990**). "Explaining phonetic variation: A sketch of the H&H theory," in *Speech Production and Speech Modelling*, edited by W. J. Hardcastle and A. Marchal (Kluwer, Dordrecht, The Netherlands), pp. 403–439.

Mirman, D. (**2014**). *Growth Curve Analysis and Visualization Using R* (CRC Press, Boca Raton, FL).

Mok, P. K. P., Zuo, D., and Wong, W. Y. P. (**2013**). "Production and perception of a sound change in progress: Tone merging in Hong Kong Cantonese," Lang. Var. Change **25**, 341–370.

Mu, L., and Sanders, I. (**2009**). "The human cricothyroid muscle: Three muscle bellies and their innervation patterns," J. Voice **23**, 21–28.

Ohala, J. J. (**1978**). "Production of tone," in *Tone: A Linguistic Survey*, edited by V. A. Fromkin (Academic Press, New York), pp. 5–39.

Potisuk, S., Gandour, J. T., and Harper, M. P. (**1998**). "Vowel length and stress in Thai," Acta Linguist. Hafniensia **30**, 39–62.

Prades, J.-M., Dubois, M.-D., Dumollard, J.-M., Tordella, L., Rigail, J., Timoshenko, A. P., and Peoc'h, M. (**2012**). "Morphological and functional asymmetry of the human recurrent laryngeal nerve," Surg. Radiol. Anat. **34**, 903–908.

Sims, H. S., Yamashita, T., Rhew, K., and Ludlow, C. L. (**1996**). "Assessing the clinical utility of the magnetic stimulator for measuring response latencies in the laryngeal muscles," Otolaryngol. Head Neck Surg. **114**, 761–767.

Smiljanić, R., and Bradlow, A. R. (**2009**). "Speaking and hearing clearly: Talker and listener factors in speaking style changes," Linguist. Lang. Compass **3**, 236–264.

Snider, K. L. (**1998**). "Phonetic realisation of downstep in Bimoba," Phonology **15**, 77–101.

Tingsabadh, M. R. K., and Abramson, A. S. (**1993**). "Illustrations of the IPA: Thai," J. Int. Phon. Assoc. **23**, 24–28.

Udaka, J., Kanetake, H., Kihara, H., and Koike, Y. (**1988**). "Human laryngeal responses induced by sensory nerve stimuli," in *Vocal Fold Physiology: Voice Production, Mechanisms and Function*, edited by O. Fujimura (Raven Press, New York), pp. 67–74.

Wong, P. C. M., and Diehl, R. L. (**2003**). "Perceptual normalization for inter- and intratalker variation in Cantonese level tones," J. Speech, Lang. Hear. Res. **46**, 413–422.

Xu, Y. (**1997**). "Contextual tonal variations in Mandarin," J. Phon. **25**, 61–83.

Xu, Y. (**1999**). "Effects of tone and focus on the formation and alignment of F0 contours," J. Phon. **27**, 55–105.

Xu, Y. (**2013**). "ProsodyPro: A tool for large-scale systematic prosody analysis," in *Proceedings of Tools and Resources for the Analysis of Speech Prosody (TRASP 2013)*, Aix-en-Provence, France (August 30), pp. 7–10, see http://www2.lpl-aix.fr/~trasp/.

Xu, Y., and Lee, A. (**2021**). "Tonal processes defined as tonal coarticulation," in *The Cambridge Handbook of Chinese Linguistics*, edited by C.-R. Huang, Y.-H. Lin, and I.-H. Chen (Cambridge University Press, Cambridge, UK), in press.

Xu, Y., and Prom-on, S. (**2014**). "Toward invariant functional representations of variable surface fundamental frequency contours: Synthesizing speech melody via model-based stochastic learning," Speech Commun. **57**, 181–208.

Xu, Y., and Sun, X. (**2002**). "Maximum speed of pitch change and how it may relate to speech," J. Acoust. Soc. Am. **111**, 1399–1413.

Zemlin, W. R. (**1988**). *Speech and Hearing Science—Anatomy and Physiology* (Prentice Hall, Englewood Cliffs, NJ).

Zhang, C., Peng, G., and Wang, W. S.-Y. (**2012**). "Unequal effects of speech and nonspeech contexts on the perceptual normalization of Cantonese level tones," J. Acoust. Soc. Am. **132**, 1088–1099.