



Effect of speech rate on pre-low raising in Cantonese

Albert Lee¹, Yi Xu²

¹ Department of Linguistics, University of Hong Kong, Hong Kong

² Dept. of Speech, Hearing and Phonetic Sciences, University College London, United Kingdom

albertlee@hku.hk, yi.xu@ucl.ac.uk

Abstract

This paper investigates the effect of speech rate on pre-low raising in Cantonese. Pre-low raising is an anticipatory tonal process where a high tone is raised when followed by a low tone (i.e. the trigger). Six native speakers of Cantonese were recorded saying a disyllable in 36 tone combinations (6 tones × 6 tones) at two speech rates (normal and slow). It was found that in slow speech pre-low raising only occurred when the trigger was extremely low in pitch, whereas at normal speech rate it was observed in more tonal contexts. It is argued that pre-low raising is a result of enhanced cricothyroid activity in preparation for an upcoming low pitch target.

Index Terms: tone, production, Cantonese

1. Introduction

Pre-low raising (PLR) refers to the raising of a High tone when followed by a low tone. It is a well-known phenomenon in contextual tonal variation widely reported across languages. Also known as F0 polarisation [1], anticipatory dissimilation [2], [3], regressive H-raising or anticipatory raising [4]–[6], PLR is a local anticipatory tonal variation where the F0 of a High tone (H₁) is higher in a H₁L sequence than in H₁H₂. Consider Figure 1, where the first peak of the HL sequence is higher than the all-H sequence. Despite extensive reports on the tonal context in which PLR occurs, little is known about its underlying mechanism. This poses a problem when there is a suspected case of PLR, because 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.

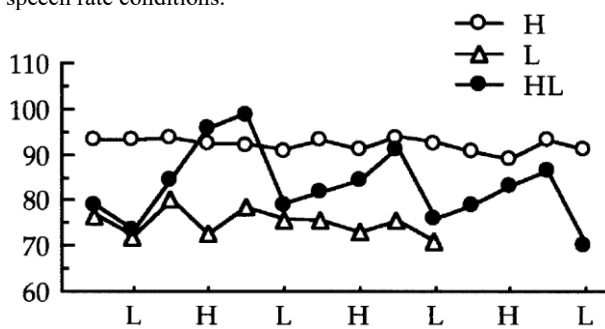


Figure 1. Example of PLR from Yoruba [7]

Cantonese was chosen in this study because of its rich tonal inventory. Figure 2 shows representative F0 contours of the six contrastive tones (T1~T6) in Hong Kong Cantonese. The highest tones are T1 and T2, while T4 is the lowest tone. Presumably, T1 and T2 would be the ideal hosts for PLR, whereas T4 would be an ideal trigger (but see [8] who report that T2 and T5 are the main hosts instead). While PLR would be expected in a T1T4 sequence, we are also interested in whether the less-low tones can also trigger and if the less-high tones can also host PLR. Against this backdrop, this exploratory

study has three goals: (i) verify the findings of [8] that only rising tones can serve as PLR hosts; (ii) examine if speech rate has an effect on PLR (e.g. whether a lower general F0 register associated with slow speech would provide a better trigger for PLR); and (iii) offer an account on the cause of PLR.

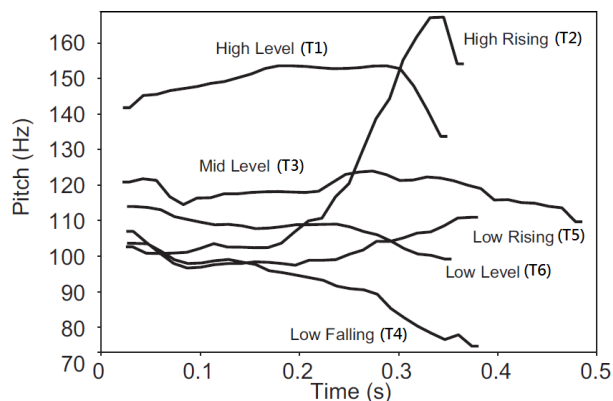


Figure 2: Cantonese tones (adapted from [9])

2. Methods

2.1. Participants

Six native speakers (3 male, including the first author) of Hong Kong Cantonese were recruited in London, aged 22–30 (S.D. 4.49) for a production experiment. None reported any (history of) speech and hearing impairment. All participants were briefed about the experiment and granted their written consent before the recording began. Five of the speakers were remunerated a small sum for their time.

2.2. Stimuli

The disyllable *lau man* was chosen for this study. There is a 6-way tonal contrast for these two syllables, which yields all 36 (6 × 6) possible tonal combinations. Also, with sonorant initial consonants these two syllables ensure that continuous F0 contours can be tracked. Target words were framed in the carrier *zoi3 gong2 ___ go2 deoi3 zi6* 再講___個對字 ‘Say the disyllable ___ again’.

2.3. Recording procedures

Recording took place in a quiet room at UCL, using a RØDE NT1-A microphone. Sampling rate was 44,100 Hz. Speakers were seated in front of a computer screen, which displayed the stimuli in a randomised 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 obviously more slowly in the second production. Fast speech, which tends to cause target undershoot, was not included in the present study. Altogether 6 speakers × 2 speech rates × 36 tone combinations × 5

repetitions = 2160 utterances were elicited. Seven utterances were subsequently discarded due to mispronunciation.

2.4. Data analysis

Sound files were then annotated using ProsodyPro [10]. Each sound file was labeled, and markings of vocal pulses were manually checked and rectified. Segmentation was done by the syllable. Apart from the target word itself, the syllable before (*gong2*) as well as the one after (*go2*) were also labeled during annotation (i.e. analyzed), in case any carryover effect 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 from individual files, as well as ensemble files containing data ready for graphical and statistical analyses.

3. Results

3.1. Non-speech rate-dependent cases

First, tonal contexts in which PLR always occurs are considered, namely T1T4 and T2T4. In our data, mean syllable duration is 336 ms for slow speech and 202 ms for normal speech. Figure 3 shows the averaged F0 contours of 40 repetitions from six speakers, with the second interval kept constant (T1). Vertical lines represent syllable boundaries. Here the solid black contour (T1T1 sequence) serves as the baseline. Any contour higher than T1T1 in the second interval would constitute a possible case of PLR.

As expected, PLR occurs in T1T4 (dotted red) given a low trigger. It is 20.91 Hz higher (see Figure 4) than the baseline in the second interval in normal speech and 11.10 Hz higher in slow speech. T2T4 also shows evidence of PLR at both speech rates, although the magnitude is much smaller (10.11 Hz and 3.17 Hz for normal and slow speech respectively). Paired samples T-tests confirm that all contrasts are statistically significant ($p < 0.05$).

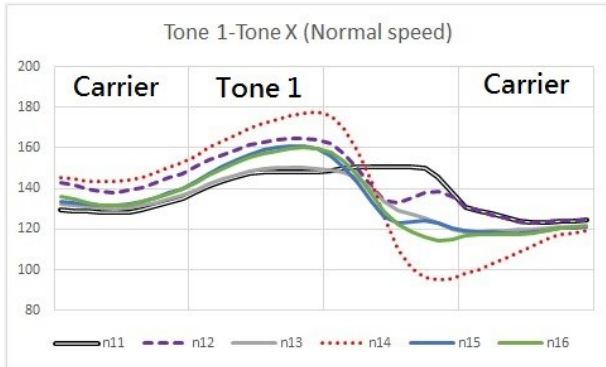


Figure 3. *F0* contours of T1-T(x) at normal speed. X-axis represents normalised time, while Y-axis is *F0* in Hz.

Paired samples t-tests					
	Trigger	t	df	p	Mean diff.
Normal	T1T1 - T1T2	-7.813	29	< .001	-13.381
	T1T1 - T1T3	-3.334	29	0.001	-7.405
	T1T1 - T1T4	-9.557	29	< .001	-20.914
	T1T1 - T1T5	-4.602	29	< .001	-10.756
	T1T1 - T1T6	-3.927	28	< .001	-16.049
	Slow	T1T1 - T1T2	-0.856	29	0.199
T1T1 - T1T3		-1.015	29	0.159	-2.755
T1T1 - T1T4		-6.484	29	< .001	-11.104
T1T1 - T1T5		-0.453	29	0.327	-1.136
T1T1 - T1T6		0.315	28	0.623	1.176

Normal	T2T1-T2T2	0.569	29	0.713	1.459
	T2T1-T2T3	-0.243	29	0.405	-0.342
	T2T1-T2T4	-8.104	29	< .001	-10.111
	T2T1-T2T5	0.005	29	0.502	0.009
	T2T1-T2T6	-3.573	29	< .001	-3.83
	Slow	T2T1-T2T2	-0.224	29	0.412
T2T1-T2T3		-2.625	29	0.007	-3.333
T2T1-T2T4		-2.222	29	0.017	-3.17
T2T1-T2T5		-1.533	29	0.068	-3.026
T2T1-T2T6		-1.829	28	0.039	-3.259

Table 1. One-tailed T-tests comparing Mean *F0* in the PLR host

3.2. Speech rate-dependent cases

In other tonal contexts, PLR appears to be dependent upon speech rate, i.e. present at faster rate but absent at slower rate. For example, Figure 4 shows that for the T1T6 (solid curves) sequence, PLR is observed only in normal speech but not in slow speech. The same is true for T1T2, T1T3, and T1T5 (see Table 1), where PLR is only observed in faster speech.

While slow speech has a lower global *F0* register (global mean *F0* in our data is 172 Hz for normal speech, and 145 Hz for slow speech), the resultant lower trigger does not give rise to more PLR; this shows that a low trigger is not the only factor underlying this phenomenon.

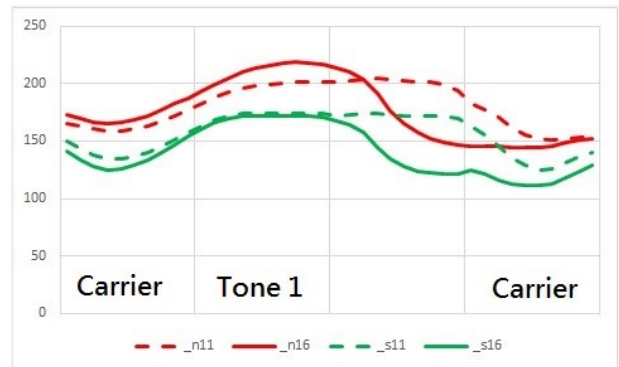


Figure 4. *F0* contours of T1T1 vs. T1T6 in normal (red) vs. slow (green) speech.

Next, *F0* velocity in the trigger (third interval) is considered. Recall that there is PLR in T1T6 (see Figure 4) in normal speech but not in slow speech. Figure 5 shows that where there is PLR (solid red), the falling velocity is much greater than otherwise. The same pattern was observed after visual inspection of the velocity profiles of other tone sequences.

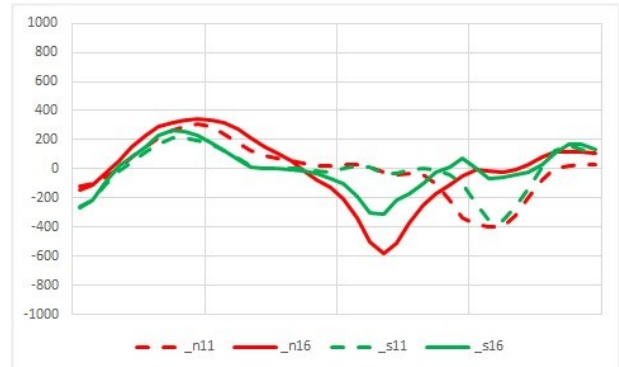


Figure 5. Averaged *F0* velocity of T1T1 vs. T1T6 in normal and slow speech.

On the whole, there seems to be a dividing line for PLR vs. no PLR based on maximum falling velocity in the trigger

syllable. Figure 5 shows the maximum falling velocity of all Trigger \times Speech rate conditions, with y-axis representing maximum F0 velocity. It appears to be possible to identify cases of PLR (based on Figure 3 and Figure 4) by drawing a line at 400 semitones / second (red line).

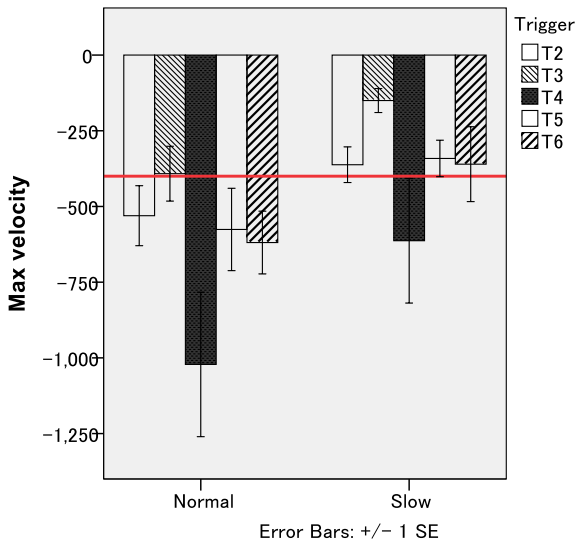


Figure 6. Max falling velocity in the PLR triggering syllable (Tone \times Speech rate) where the host is T1

3.3. Cases without PLR

Finally, where the host is not a high tone, PLR does not occur even if the trigger is low. T4T4 sequences are not considered as they are beyond the scope of PLR. Table 2 shows that except the T5T4 sequence in normal speech, the difference between all pairs of tone sequences is non-significant. Even T4T4 that appears to undergo PLR (cf. [8]), the magnitude of raising is small (3.75 Hz). In fact, in all these cases, maximum falling F0 velocity never exceeds 400 semitones / second in the trigger syllable.

Paired samples t-tests

	Trigger	t	df	p	Mean diff.
Normal	T3T1 - T3T4	-0.452	29	0.655	-1.573
	T5T1 - T5T4	-2.634	29	0.013	-3.749
	T6T1 - T6T4	-0.072	29	0.943	-0.237
Slow	T3T1 - T3T4	0.494	29	0.625	0.916
	T5T1 - T5T4	-0.474	29	0.639	-0.571
	T6T1 - T6T4	-1.357	29	0.185	-3.077

Table 2. Two-tailed T-tests comparing Mean F0 in the PLR host while trigger is T4

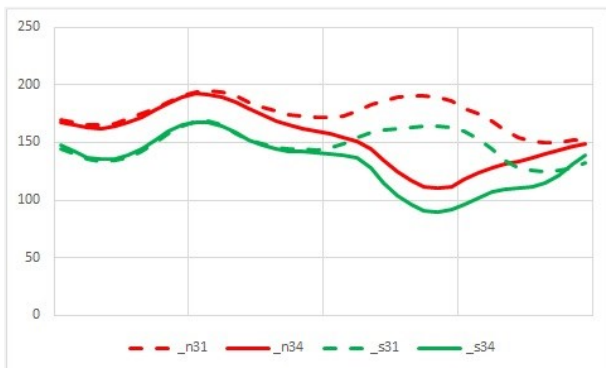


Figure 7. F0 contours of T3T1 vs. T3T4 in normal (red) vs. slow (green) speech.

In sum, although PLR does not require the lowest trigger provided fast enough speech rate, a high host is on the other hand necessary.

4. Discussion

4.1. The present study vs. Gu & Lee (2007)

This paper set out to complement previous work by [8] 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, the first syllable (i.e. the host) must be high; hosts that are low in F0 would not have PLR. Although one might assume that a low trigger is the key to PLR, our results suggest that a high host and a fast fall are at least as important if not more.

Our findings are compatible with [8] in general, though there are also differences. In [8], where the effect of focus was examined, T2 and T5 were found to be good PLR hosts. On the other hand, in the present study, we looked at the effect of speech rate, and found instead that T1 and T2 were reliable PLR hosts. Taken together, these two studies suggest that PLR in Cantonese is subject to factors including F0 of the host, F0 of the trigger, speech rate, and focus.

4.2. A perceptual account for PLR?

This leaves us with the last question, namely why PLR occurs in the first place. Given the results, several explanations are conceivable. The first is that PLR may enhance tonal contrasts to aid comprehension. Researchers have shown that Cantonese is undergoing tone-merger [11], 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 the trigger position. 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 hosts, 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 [7] and Bimoba [12]) where contrast enhancement is not necessary. A contrast enhancement account, therefore, cannot be taken as the underlying mechanism of this phenomenon.

4.3. PLR to increase maximum velocity

A likelier account, in our opinion, is that PLR occurs to allow a larger excursion to achieve a high maximum F0 velocity (cf. [13]). Acceleration takes time, hence a higher starting point (raised peak) would be required to achieve a very low target. This is reminiscent of hitting a tennis ball with a racket: the harder one hits, the further s/he needs to first pull back his/her arm. The speech rate effect fits in this account too, because faster speech (where PLR occurs) also requires a high maximum velocity, a higher starting point would still be required for acceleration. A non-low trigger (e.g. T1T3) spoken slowly involves no fast movement or large excursion, and thus yields no PLR as would be predicted by this account.

As for the physiological mechanism behind, it is likely that during PLR the trigger (a low tone) activates pitch-lowering strap muscles (e.g. sternohyoid, or SH), while in the preceding syllable (PLR site) the contraction of pitch-raising cricothyroid (CT) muscle is **enhanced** to allow more distance for

acceleration. The end result is a raised F0 peak. Such an account sees PLR as active planning, like pulling one's arm back for a hard tennis serve. By implication, one may predict rather categorical pattern for PLR given it is based on the speaker's knowledge of the utterance.

4.4. 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 F0 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 F0 fall that is steady and gradual [14]. It is also consistent with a part of CT that is capable of very fast F0 movements, namely the pars recta belly [15]. Hence even when the fall excursion is small, PLR would still occur before a steep fall as CT is required for fast F0 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 F0 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 F0 control. Two pieces of evidence appear to be supportive. Firstly, CT is innervated by the external superior laryngeal nerve (ESLN), whereas all other intrinsic laryngeal muscles are supplied by the recurrent laryngeal nerve. In mammals, ESLN 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 [16]. Moreover, the rectus belly of CT that is responsible for fast F0 changes is supplied by 3~7 branches of ESLN [15], lending further support to this account.

Secondly, factors which raise F0 usually raise intensity as well. Where F0 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 F0 rises [17]. 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 F0 is due to CT contraction alone, like in [17].

4.5. Further implications and future directions

These results are also in line with PLR accounts for Japanese. It is argued [18] that the extra high F0 associated with the Japanese pitch accent is 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 F0 excursion and velocity at various speech rate conditions.

To further test our hypotheses one could look at languages with a complex tone system and which is generally spoken slowly. One suitable candidate would be Thai. Our analysis of Thai production data is now under way, and will hopefully shed more light on our quest for the mechanism underlying PLR. Ultimately, it will also be necessary to test these hypotheses using articulatory methods such as (surface) electromyography (for muscle action potentials), ultrasound imaging, and electromagnetic articulography (for larynx movement).

5. Conclusions

In this paper we have shown that PLR occurs when the falling excursion is large or when the falling velocity is high (i.e. steep fall). In other words, the surface realisation of a high tone is raised if it is followed by a very low tone (T4 in Cantonese), or when it is followed by a steep fall (due either to fast speech rate or steeply declining underlying target). Based on these observations, we have argued that PLR is a result of enhanced CT muscle activation in preparation for an upcoming low target.

6. Acknowledgements

Part of the present data was presented at HKU Tone Conference 2015 at University of Hong Kong. Thanks go to Prof. Donna Erickson and anonymous reviewers for invaluable comments on an earlier version of this paper. All errors and inadequacies remain our own.

7. References

- [1] L. M. Hyman and R. G. Schuh, “Universals of tone rules: Evidence from West Africa,” *Linguist. Inq.*, vol. 5, no. 1, pp. 81–115, 1974.
- [2] J. T. Gandour, S. Potisuk, and S. Dechongkit, “Tonal coarticulation in Thai,” *J. Phon.*, vol. 22, pp. 477–492, 1994.
- [3] Y. Xu, “Contextual tonal variations in Mandarin,” *J. Phon.*, vol. 25, no. 1, pp. 61–83, 1997.
- [4] B. Connell and D. R. J. Ladd, “Aspects of pitch realisation in Yoruba,” *Phonology*, vol. 7, no. 1, pp. 1–29, Oct. 1990.
- [5] Y. O. Laniran, “Intonation in tone languages: The phonetic implementation of tones in Yoruba,” Cornell University, 1992.
- [6] Y. Xu, “Effects of tone and focus on the formation and alignment of F0 contours,” *J. Phon.*, vol. 27, no. 1, pp. 55–105, 1999.
- [7] Y. O. Laniran and G. N. Clements, “Downstep and high raising: Interacting factors in Yoruba tone production,” *J. Phon.*, vol. 31, no. 2, pp. 203–250, Apr. 2003.
- [8] W. Gu and T. Lee, “Effects of tonal context and focus on Cantonese F0,” in *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS 2007)*, 2007, pp. 1033–1036.
- [9] A. L. Francis, V. Ciocca, L. Ma, and K. Fenn, “Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers,” *J. Phon.*, vol. 36, no. 2, pp. 268–294, 2008.
- [10] Y. Xu, “ProsodyPro: A tool for large-scale systematic prosody analysis,” in *Proceedings of Tools and Resources for the Analysis of Speech Prosody (TRASP 2013)*, 2013, pp. 7–10.
- [11] P. K. P. Mok, D. Zuo, and W. Y. P. Wong, “Production and perception of a sound change in progress: Tone merging in Hong Kong Cantonese,” *Lang. Var. Change*, vol. 25, no. 3, pp. 341–370, Sep. 2013.
- [12] K. L. Snider, “Phonetic realisation of downstep in Bimoba,” *Phonology*, vol. 15, pp. 77–101, 1998.
- [13] H. Truckenbrodt, “Final lowering in non-final position,” *J. Phon.*, vol. 32, no. 3, pp. 313–348, Jul. 2004.
- [14] R. Collier, “Physiological correlates of intonation patterns,” *J. Acoust. Soc. Am.*, vol. 58, no. 1, pp. 249–256, 1975.
- [15] L. Mu and I. Sanders, “The human cricothyroid muscle: Three muscle bellies and their innervation patterns,” *J. Voice*, vol. 23, pp. 21–28, 2009.
- [16] J. Udaka, H. Kanetake, H. Kihara, and Y. Koike, “聴覚喉頭反射をめぐって [Human auditory-laryngeal reflex],” *Japan J. Logop. Phoniatr. [音声言語医学]*, vol. 30, pp. 360–367, 1989.
- [17] M. Hirano, W. Vennard, and J. J. Ohala, “Regulation of register, pitch and intensity of voice,” *Folia Phoniatr. Logop.*, vol. 22, no. 1, pp. 1–20, 1970.
- [18] A. Lee, Y. Xu, and S. Prom-on, “Mora-based pre-low raising in Japanese pitch accent,” in *Presentation at Interspeech 2013*, 2013, pp. 3532–3536.