Title: Pre-low raising in Japanese pitch accent

Albert Lee

Department of Linguistics, University of Hong Kong, Hong Kong.

Santitham Prom-on

Department of Computer Engineering, King Mongkut's University of Technology Thonburi, Thailand.

Yi Xu

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

Correspondence address:

Albert Lee, Department of Linguistics, Run Run Shaw Tower, University of Hong Kong, Pokfulam, Hong Kong. Tel: +852 39178603, Fax: +852 25467477, Email: albertlee@hku.hk.

Running title: Pre-low raising in Japanese pitch accent

Abstract

Japanese has been observed to have two versions of the H tone, the higher of which is associated with an accented mora. However, the distinction of these two versions only surfaces in context but not in isolation, leading to a long-standing debate over whether there is one H tone or two. This article reports evidence that the higher version may result from a pre-low raising mechanism rather than being inherently higher. The evidence is based on an analysis of F_0 of words that varied in length, accent condition and syllable structure, produced by native speakers of Japanese at two speech rates. The data indicate a clear separation between effects that are due to mora-level pre-planning and those that are mechanical. These results are discussed in terms of mechanisms of laryngeal control during tone production, and highlight the importance of articulation as a link between phonology and surface acoustics.

I.INTRODUCTION

In Japanese, an accented word is characterized by an initial rise in F_0 (unless it bears an initial accent or when its first syllable is heavy CVV or CVN, see Gussenhoven, 2004, pp. 188–189), followed by a sharp fall starting from the accented mora (as shown in the solid curve in Figure 1), which has been transcribed as *%L H*+L* in the Autosegmental-Metrical (AM) representation (Pierrehumbert & Beckman, 1988). In contrast, an unaccented word has an initial rise (again, unless when its first syllable is heavy CVV or CVN) but no sharp fall, namely the dashed curve in Figure 1, which has been transcribed as *%L H-* (Pierrehumbert & Beckman, 1988). The two distinct surface tones (H* and H-) are argued by some (e.g. Kindaichi, 2005, first published in 1947) to bear the same underlying phonological representation (for a comprehensive review see Sugiyama, 2012 and; Warner, 1997). Proponents of this hypothesis support their view by the lack of perceptual distinction between unaccented and final-accented words like *hashi* (%LH-) 'edge' vs. *hashi'* (%LH*, the accented mora marked by a following **'**) 'bridge' when said in isolation, and the fact that most speakers cannot produce this distinction in isolation (Sugito, 1968; Vance, 1995). For example, Vance (1995) reports in a production study that three out of four speakers made no reliable distinction between *hana* and *hana'*. Also, Warner (1997, p. 58) points out that 'many linguists, including many native speakers of Japanese, fail to detect this difference when working by ear'. Meanwhile, clear acoustic differences have been reported between the two accent conditions when produced in context (Poser, 1984). This has led to proposals to assign different representations (e.g. H* vs. H-) to accented and unaccented words (Pierrehumbert & Beckman, 1988; Sugiyama, 2012).

Though Beckman and Pierrehumbert (1986b, p. 177) explicitly stated that H* was 'generally higher than' H-, these two tonal categories are primarily a phonological distinction in AM. This is because H* and H- serve different linguistic functions and manifest different behaviors in the tonal phonology of Japanese. Specifically, H- serves to mark an Accentual Phrase whereas H* (along with a +L) marks the lexical pitch accent (Beckman & Pierrehumbert, 1986a; Pierrehumbert & Beckman, 1988). In forming compound words, H* is subject to phonological restrictions that govern its occurrence whereas the same is not applicable to H- (Kubozono, 1993). Thus H* and H- coexist due to their surface phonetic differences as well as the different roles they play in Japanese phonology.

A link thus seems to be missing between the perceptual indistinguishability (in isolation) and the phonological distinction of these two H tones. The present study is an attempt to bridge this gap by proposing that the surface distinction between H* and H- is due to a well attested articulatory mechanism, namely, pre-low raising of surface F_0, or PLR in short. Under this proposal, while there may be a phonological distinction between H* and H, at the level of articulatory planning there is only one H target for marking lexical contrasts, and it is the L tone associated with the pitch accent that gives rise to the boosted F_0 peak of H*. In turn, we reiterate the point that the mapping between underlying phonological representation and surface acoustics should not be direct, but should take into account the possible influence from such intermediate stages as articulation.

Also known as F_0 polarization (Hyman & Schuh, 1974), anticipatory dissimilation (Gandour, Potisuk, & Dechongkit, 1994; Xu, 1997), regressive H-raising or anticipatory raising (Connell & Ladd, 1990; Laniran, 1992; Xu, 1999), PLR is a local anticipatory tonal variation where the F_0 of a High tone becomes higher when preceding a Low tone. For example, the F_0 of first H in the sequence HLH would be higher than in HHH, ceteris paribus. PLR has been reported for other languages, including Bimoba (Snider, 1998), Cantonese (Gu & Lee, 2007), Gurma (Rialland, 1981), and Igbo (Laniran & Gerfen, 1997). And it has also been observed in singing, in which it is referred to as 'preparation' (Saitou, Unoki, & Akagi, 2005). Though widely observed, the underlying mechanism of PLR is still unclear despite some speculations (Gandour et al., 1994; Sugiyama, 2012; Warner, 1997; Xu, 1997). Moreover, the precise condition that triggers PLR is reported to vary from one language to another. For example, in Yoruba PLR is observed when a high tone is followed by a low tone (Laniran & Clements, 2003), in Cantonese it appears to occur mainly in rising tones (Gu & Lee, 2007), while in Mandarin the rising tone, the low tone and the falling tone can all trigger PLR in a preceding tone (Xu, 1997). What is common to all these cases is that the trigger contains a low pitch point, and the preceding tone has a high pitch point. The Japanese case seems to satisfy this condition, as can be seen in Figure 1. However, whereas in previous studies PLR is established by showing the surface F_0 realization of the High tone under different contexts, this approach is not applicable to Japanese. This is because the high F_0 points in the two curves come from two different accent conditions, attributing the higher F_0 of the solid line to PLR could be partially circular; showing the difference of two (arguably) different entities does not prove the existence of PLR for Japanese. One way to reduce the level of circularity is to show that the effect of PLR is gradient rather than all-or-none, because the gradience would be incompatible with a two-H-target hypothesis, but would be more compatible with a biomechanical account.

One such account, based mainly on the physics of motion, is that PLR is an anticipatory action to increase the peak velocity of F_0 movement in order to reach a low F_0, given that the production of a low F_0 is harder than that of a high F_0 since it involves external laryngeal muscles (Atkinson, 1978; Erickson, 1976). Other things being equal, reaching a target quickly requires a high velocity, but peak velocity is positively related to movement magnitude, whether the movement is articulatorily (Ostry & Munhall, 1985) or acoustically measured (Cheng & Xu, 2013; Xu & Sun, 2002). The present data seem to fit this account. In Xu and Sun (2002), the average speaker could lower F_0 as fast as 61~67 semitones per second; in the present data, mean mora duration is 118 ms for normal rate, while mean excursion size of accentual fall is 9.35 semitones. Thus for cases where the accent peak is adjacent to the right edge of the target word (e.g. final accented words), the speaker would be intending to drop F_0 by 9.35 semitones in some 118 ms, or 79 semitones per second, which exceeds Xu and Sun's (2002) reported maximum speed of F_0 change. By implication, when producing an accentual fall Japanese speakers may well be at their maximum speed. As such, it would be helpful to pre-raise F_0 to increase the total movement distance in order to achieve the peak velocity needed to reach a low F_0. Such a pre-movement can be also seen in the act of throwing an object or striking a tennis ball with a racket: the harder the throwing or the more striking force required, the farther the arm needs to first pull back in the opposite direction.

Another account is based on more specific physiological mechanisms of F_0 production. F_0 is determined by the tension of the vocal folds, which is controlled by a combination of the intrinsic laryngeal muscles, mainly the cricothyroids (CT), and the extrinsic laryngeal muscles —mainly the sternohyoids (SH) and thyrohyoids (TH) (Atkinson, 1978; Erickson, 1976). When F_0 changes from either a high or low level across the mid range, there is a 'switch-over point' in the mid-F_0 range around which there is a slight overlap of CT and SH/TH activities. When F_0 changes from a low to a non-low range, a post-low bouncing effect is observed in Mandarin (Chen & Xu, 2006) and Cantonese (Gu & Lee, 2007), and possibly in English (Pierrehumbert, 1980, as interpreted by Chen & Xu, 2006). For example, in a LM (i.e. Low-Neutral) sequence in Mandarin, if the L is low enough and the following M is not given enough articulatory strength, the M is realized with a much higher F_0. This bouncing effect was argued to be due to a temporal loss of balance between the CT and SH/TH control over the vocal folds during the 'switch over' (Prom-on, Liu, & Xu, 2012). Thus it is possible that the driving force of PLR is a preemptive CT activity to pre-balance the SH/TH activity in anticipation of a very low F_0.

Although these two accounts are quite different, positing that the raised H is to either facilitate the downward F_0 movement, or counteract the contraction of the external laryngeal muscles, both would predict (i) a gradient negative correlation between the F_0 values of H and L, as opposed to a categorical dual-height division. Furthermore, assuming that articulatory planning cannot be fully precise, it is also predicted that (ii) the negative correlation is weaker than the highly linear negative correlation observed in post-low bouncing (Prom-on et al., 2012), a related articulatory phenomenon where F_0 is boosted to a high level after a Low target. The result of testing these two predictions will therefore either support or reject the biomechanical accounts, thus providing evidence either for or against the single-H hypothesis for Japanese pitch accents. If supported, the single-H-target hypothesis for Japanese word prosody would provide yet another piece of evidence for PLR as a general articulatory mechanism applicable to languages in general, whenever a low tonal target occurs.

II.METHODOLOGY

As shown in Table I, a total of 33 Japanese words were chosen as stimuli. The target words varied in length (1~4 morae), accent condition (unaccented/initial-/medial-/penultimate-/final-accent), and syllable structure (CVCV, CVN, CVV). In Table I H stands for the accent target, and L stands for both the low tone after a pitch accent and the low pitch at the beginning of a non-initial-accented word. The tokens were presented in the unaccented carrier sentence *Jiten-ni ___-mo nottemasu* 'The word ___ too is found in the dictionary'.

A number of factors were taken into consideration when designing the stimuli. First, past studies of Japanese word prosody often used only bimoraic target words (e.g. Sugiyama, 2012) or words that are incomparable to one another in terms of vowel height, consonant manner, etc. (e.g. Warner, 1997; cf. microprosody), leaving the possibility that the results could have been different with longer words and words that have similar segments. Stimuli used in the present study cover a wider range of phonological contexts (length, accent condition, and syllable structure). Also, the use of only nasals as initial consonants avoids most of the distortions from segmental perturbation of F_0. Second, as fast speech can lead to F_0 target undershoot (Cheng & Xu, 2014), we recorded each target sentence at two speech rates to control for the effect of speed of articulation. Third, though less directly relevant to the present research question, introducing three types of syllable structure into our stimuli allowed us to gain further insights into the shape of F_0 contours under different conditions, which may be relevant in future studies on prosody modeling.

Eight native speakers (four of each gender, mean age 28.5, s.d. 4.72) of Tokyo Japanese from the Greater Tokyo Area (Tokyo, Saitama, Kanagawa, and Chiba) served as subjects. They were living in London at the time of recording. All speakers had moved to the country for less than half a year, except for one speaker who had been in London for 1.5 years, and another for five years. No atypical F_0 behavior was observed in either. None of the speakers reported any history of speech, language, or hearing impairment.

The recording took place in a quiet room in University College London, using a RØDE NT1-A microphone. Subjects were seated in front of a computer screen, on which stimuli were displayed one by one in random order. They produced each sentence first at normal speed, then immediately followed by a slow production. Though speech rate was not stipulated in actual terms, subjects were instructed to speak obviously more slowly in the second production. When an undesired emphasis was placed on the particle *–mo*, in which case *–mo* sees an abrupt F_0 rise, the subject was asked to repeat the utterance without any emphasis. From each subject a total of 33 sentences × 5 repetitions × 2 speech rates = 330 tokens were collected. The sampling rate was 44.1 kHz.

Because only nasal stops were used in syllable-initial positions, some low frequency words had to be included. In light of this, subjects were given time to rehearse and familiarize themselves with the experiment material until they felt comfortable enough to start recording. No F_0 patterns peculiar to the less familiar stimuli were observed in subsequent analyses.

Sound files were then annotated using ProsodyPro (Xu, 2013), a Praat (Boersma & Weenink, 2012) script for prosody analysis. Each sound file was labeled, and markings of vocal pulses were manually rectified (to correct apparent errors in the vocal pulse markings generated by the autocorrelation algorithm in Praat that would have led to octave jumps). Segmentation was done by the 'mora', such that a light syllable (CV) counts as one mora while a heavy syllable (CVN or CVV) counts as two. In the latter case, two labeled intervals equal in duration were assigned. Apart from the target word itself, the mora before (*-ni*) as well as the one after (*no*) were also labeled during annotation, in order to capture any carryover effect extended from or into the target word. Other parts of the carrier sentence were not analyzed in the present study. The script then generated all the acoustical measurements from individual files, as well as ensemble files containing data ready for graphical and statistical analysis.

For each utterance, the following measurements were taken (see Figure 1): (i) MaxF0 —maximum F_0 (in semitones, same for variables ii-vi) in the host mora of a pitch accent and the following mora, wherever it occurs; (ii) MinF0A—minimum F_0 of the final mora of the target prosodic word, i.e. the final particle *–mo* in the carrier sentence; (iii) MinF0B—minimum F_0 value of the mora immediately after the target word, i.e. *no*- in the carrier sentence, with the last 30 ms of the mora excluded (to avoid the effect of segmental perturbation on F_0 from the following geminate consonant /*tt*/); (iv) RiseSize—the difference between MaxF0 and minimum F_0 of the initial rise;

(v) FallSizeA—the difference between MaxF0 and MinF0A; (vi) FallSizeB—MaxF0 less MinF0B; (vii) VMaxRise—maximum velocity of initial rise; (viii) VMaxFall—maximum velocity of accentual fall; and (ix) PeakDelay—the difference between accent peak time and the beginning of the mora that hosts the pitch accent. Measurements (i)~(vi) are illustrated in Figure 1. Pearson's correlations were calculated (as shown in Table II) between all these measurements to examine the relationship between accent peak and the following low tone.

III.RESULTS

Our main hypothesis is that the height of H is a function of the following L, as opposed to previous report of H as a function of word/phrase length (Selkirk, Shinya, & Kawahara, 2004). An examination of the properties of unaccented words in our data is thus necessary before proceeding to examine the behavior of H. Figure 2 shows that contrary to the observation of Selkirk et al (2004), it is only for some, but not all, speakers that a longer unaccented word has a higher maximum F_0. Although a one-way ANOVA shows a significant main effect of word length on maximum F_0 (in semitones based on initial F_0 value), $F(3,156) = 7.270$, $p < 0.01$, most contrasts do not reach statistical significance in post-hoc pairwise comparisons. For example, three-mora words are not significantly different from words of any other lengths. This echoes with our observation in Lee et al (2014) that adding word length as extra predictors did not yield noticeable improvement in modeling accuracy – if word length systematically influences F_0 in our data, having it as an additional predictor should have allowed the model to capture more variations, which was not the case. Having established the inconsistency of word length effect on the height of unaccented F_0 peak across speakers in our data, next we proceed to investigate if H is gradient. Figure 3 displays time-normalized F_0 contours averaged across five repetitions by all subjects. We can see that peak-to-end distance (i.e. number of morae from the accented mora) is positively related to accent peak height, but inversely related to the F_0 of the right edge of target word. That is, other things being equal, the earlier the pitch accent in a word, the higher its peak F_0 and the lower the F_0 at word end. Figure 4 shows how peak-to-end distance affects MaxF0. The four bars on the left represent words that bear initial accent, and differ from one another in terms of peak-to-end distance. For example, the leftmost bar represents initial accented-words of which accent peak is one mora away from word end, i.e. it is one mora-long. We can see that **MaxF0** is higher when peak-to-end distance is greater (e.g. the fourth bar from the left). To verify these observations, we performed repeated measures ANOVAs on the data of accented words. Peak-to-end distance significantly affects MaxF0, F(3,21) = 30.297, $p < 0.001$. A post-hoc pairwise comparison (see **Table III**) confirms that a word has higher MaxF0 when its accent is further away from word end ($p < 0.05$), except that pitch accents that are 3 or 4 morae away do not have significantly different MaxF0. Note that this cannot be, at least not solely, be attributed to the well-known phonetic phenomenon of declination (Cohen, Collier, & ’t Hart, 1982; Ladd, 1984), or else words of the same accent condition but different lengths would have the same peak height. In fact, Figures 3 and 4 show that among initial-accented words, a longer word tends to have a higher peak, in line with Selkirk et al. (2004).

The observation about word-end F_0 was also statistically confirmed. Repeated measures ANOVAs show that **MinF0A** is significantly lower when peak-to-end distance is greater (F(3,21) = 23.255, $p < 0.001$). Likewise, post-hoc pairwise comparison confirms that different peak-to-end distance conditions have significantly different **MinF0A**, except for those that are 3 morae and 4 morae away from word end, in which case they are not significantly different, as is the case for **MaxF0**.

We then compared the measurements introduced above for possible correlations (N = 2640). The F_0 data were first converted to semitones using the utterance-initial F_0 of each utterance as reference. The reason for so doing, rather than using other reference like speaker mean F_0, was to avoid distortion from the lower F_0 register in slow speech rate observed across all speakers. MaxF0 and MinF_0A were inversely correlated, $r = -0.354$ (two-tailed, $p < 0.001$), suggesting that a lower word-end F_0 is associated with a higher accent peak. However, part of the negative correlation comes from the bimodal distribution of accented and unaccented words, where accented words naturally have a higher MaxF0 and lower MinF0A. Therefore, to assure that raising, if there is any, is gradient within accented words, we repeated the same correlation analysis with unaccented words excluded, and the results are shown in Table II and Figure 6. It can be seen that **MaxF0** is positively correlated with **PeakDelay** (r = 0.415). That is, in an accented word, when peak occurs later in or after the accented mora, it tends also to be higher. Meanwhile, **PeakDelay** is also inversely related to **MinF0A**, r = -0.201 (or r = -0.250 when normalizing data with word-initial F_0 value instead). That is, the lower the word-end F_0, the later the F_0 peak, which in turn is related to peak height. Similarly, a lower value of **MinF0A** also gives rise to a larger initial rise: for **RiseSize~MinF0B**, r = -0.198 (see also Figure 6).

Given the design of the stimuli, which contrast accent condition and word length at the same time, it is possible that parts of the effect could be confounded. Hence, a logical extension of our analysis would be to further divide the data into subsets according to four peak-to-end distance and word length conditions (see Table IV). Here a gradient pattern emerges – **RiseSize~MinF0A** was r = -0.317 when accent was 4 morae away from word end, r = -0.205 when 3 morae away, r = -0.190 when 2 morae away, and r = -0.142 when 1 mora away. Note that this is not to be mistaken for the word length effect of F_0, because when data was grouped by word length the same gradient pattern became much weaker, with 1-mora words losing negative correlation between **RiseSize~MinF0A** altogether. Meanwhile, for unaccented words, **MaxF0~MinF0A** r = 0.639, suggesting a completely opposite behavior to accented words. Finally, grouping data by speech rates reveals that, for **PeakDelay~MinF0A**, r = -0.229 in normal speech and r = -0.194 in slow speech.

IV.DISCUSSION

The above analysis has yielded support for PLR as the mechanism of raising F_0 of the accented H in Japanese. First, there is evidence of PLR in the PeakDelay~MaxF0 and PeakDelay~MinF0A correlations. On the one hand, other things being equal, a higher MaxF0 should take longer to achieve, hence a greater peak delay. This is confirmed by the positive PeakDelay~MaxF0 correlation. On the other hand, a greater PeakDelay relative to the accent host mora would leave less time for the movement toward the low F_0 at word end, resulting in an undershoot of the low F_0. But this is contradicted by the negative PeakDelay~MinF0A correlation, which shows that a greater peak delay is associated with a lower F_0. Thus a lower MinF0A has led to higher MaxF0, which in turn led to greater peak delay. This is consistent with previous findings about PLR in other languages, except that no measurement of peak delay was taken in the earlier studies.

The second piece of evidence is that these correlations become stronger as the accent is further away from word end. Given the relatively fast tempo in the present data (mean mora duration 117 ms for normal speech), time pressure may have masked part of the PLR effect in the correlations. As peak-to-end distance is reduced, carryover assimilation to the preceding H- due to inertia is more likely to obscure any raising effect, and this also explains why when the subsets in Table IV are collapsed the general correlations were quite weak. The fact that the negative correlation in RiseSize~MinF0A is the strongest when accent is four morae away from word end indicates that a lower F_0 indeed gives rise to greater initial rise. Meanwhile, *r* is the smallest in RiseSize~MinF0A when accent is adjacent to word end, in which case PeakDelay is the smallest because there is a lack of time to reach the low tone, which in turn has led to relatively low MaxF0, thus also a weaker negative correlation. Note that the effect of peak-to-end distance is not to be confused with gradient measurements like PeakDelay and speech rate. Recall that for PeakDelay~MinF0A, r = -0.208 at normal speech rate, but r = -0.146 in slow speech. This is because, when given more time, the low tone is better reached and so less variable, leading to a weaker correlation. Note also that although PeakDelay per se is not an indicator of PLR, it serves to confirm that accent peak height is the result of articulatory processes, which in turn supports the view that the raised peak results from an adjustment of muscle activities, as outlined above.

These two pieces of evidence are in support of the hypothesis that variation of F_0 height associated with an accent is a function of the height of the following low F_0; the lower the following low, the higher the preceding high. But there is also a previously unreported interaction between categorical and gradient effects. At the word level, there seems to be an effect of gross pre-planning, based on the number of morae available to the speaker for achieving the upcoming low tone, which the speaker can deduce from lexical knowledge. Within a mora, the exact amount of PLR is dependent on how well the low tone is actually achieved, better at the slow speech rate but worse when accent is adjacent to the targeted low. The height of acoustical landmarks in Japanese word prosody thus appears to be shaped by both mora-sized planning and mechanistic articulatory inertia, working in opposite directions.

The absence of a consistent effect of word length on the height of unaccented words is both important and interesting. If such an effect was observed in our data, it would have been impossible to tease apart PLR which is sensitive to peak-to-end distance on one hand and pre-planning as based on word length on the other. Unlike in Selkirk et al. (2004), the effect of word length on F_0 scaling in unaccented words is at best limited. Post-hoc Bonferroni tests showed that although words that are several morae different in length (i.e. 1-mora vs. 4-mora) had significantly different peak F_0, most other contrasts did not reach statistical significance, e.g. 3-morae words were not significantly different from words of any other word length. In Selkirk et al (2004), word length ranged from 3 to 7 syllables/morae, whereas in the present study words were 1 to 4 morae long. Another difference lies in how height was measured — maximum F_0 of the entire word in the present study, whereas in Selkirk et al (2004) it was 'the F_0 at the start of the vocalic nucleus of the initial syllable… and the F_0 at the peak of the initial rise (if there was one) or the F_0 at the beginning of the second syllable (if there was not)'. The differences in choice of word length and measurement could both be the sources of discrepancy between our results and theirs.

Logically, the existence of PLR only enhances the surface difference between the two versions of observed H in Japanese, and does not entail that they are the same. Further evidence of their identity comes from their perceptual and acoustic indistinguishability in contexts where PLR does not apply, i.e. utterance-finally or when in isolation (Sugito, 1968; Vance, 1995). Combined together, there seems to be sufficient evidence for us to conclude that at the articulatory level there is only a single H target in Japanese word prosody.

At this point one may argue that the data presented above could also be taken to support an upstep account for the Japanese pitch accent. In English (Pierrehumbert, 1980), a boundary tone H% is said to undergo upstep when it is preceded by a H phrase accent. This is because without upstep two consecutive H tones would have resulted in a sustained high, instead of the observed final rise. The Japanese pitch accent appears to be comparable, as its H* always follows a H- at least within the AM framework. However this possibility is ruled out because when the L tone of a pitch accent (which follows the H*) is not realized (i.e. in isolation or utterance-final) H- and H* are not distinguishable from each other, thus the boosted F_0 peak of H* must have been due to the following L and not to a neighboring H tone, or upstep in other words.

The finding of gradient nature of PLR is consistent with the biomechanical accounts mentioned in the Introduction. Unlike post-low bouncing (Prom-on et al., 2012), however, PLR is a result of local pre-planning in anticipation of an imminent low tone. The current data indicate that the amount of F_0 increase depends on the predicted amount of forthcoming lowering based on the number of post-accent morae. Interestingly, the present results also suggest that pre-planning can be done only at the level of the smallest unit of individual movement, which is likely the mora in the case of Japanese. That is, speakers seem to anticipate that a low tone will be better reached as the amount of time available is increased based on the count of number of morae, as shown by the positive relation between MaxF0 and peak-to-end distance. Meanwhile, as seen above, the within-mora effect (mechanical in nature and more gradient) interacts with pre-planning at the word level (mora-by-mora and more discrete).

An alternative to PLR is Truckenbrodt's (2004) downstep account. He reports in German that a H tone before downstep is higher than a H tone not followed by downstep, and attributes the raised H to the following downstepped accent. He then further suggests that downstep also applies to Tokyo Japanese, where the raised H is triggered by a downstepped L. While this account appears to share some similarity to our PLR account here, there are two fundamental differences. The first is that our current data have established that the raised H in Japanese is directly triggered by the following L, whereas downstep, as used by both Liberman and Pierrehumbert (1984) and Truckenbrodt (2004), refers to the lowering of any tone relative to its neighbor of the same tone category. Furthermore, what is also established by the present data is that the raising of the pre-L H is contingent upon the L being very low in F_0. Therefore, PLR is much more specific than downstep. Secondly, the raising of H in the current data was established with the non-raised H as the reference. In contrast, the downstepped L proposed by Truckenbrodt is not established by using a non-downstepped L as a reference. In the present data, although the L tone was realized with different F_0, there is no good reason to argue that the lower ones should be considered as downstepped from the other cases. We therefore believe that the PLR account is a more precise, mechanism-specific and coherent account

The present finding has implications for models of tone and intonation, especially those that simulate articulatory dynamics of F_0 production, such as the Fujisaki model (Fujisaki, Wang, Ohno, & Gu, 2005) and the PENTA model (Prom-on, Xu, & Thipakorn, 2009; Xu, 2005). Both models assume that surface F_0 contours result from laryngeal responses to muscle activities, which can be simulated by a spring-mass system. PENTA, in particular, assumes that the most basic laryngeal movements are unidirectional toward underlying pitch targets that are specified by communicative functions such as lexical (encoded through tone, pitch accent and word stress), focal and sentential contrasts. Although it has been demonstrated that computationally simulating these unidirectional movements can generate F_0 contours that are close to those of natural speech in English, Mandarin and Thai (Liu, Xu, Prom-on, & Yu, 2013; Prom-on et al., 2009; Xu & Prom-on, 2014), there is residual variability that is not yet fully modeled. Prom-on et al. (2012) showed that part of the residual variability comes from post-low bouncing, which can be modeled by adding an F_0-raising force at the end of a low-approaching movement. The present results show another source of residual variability not yet accounted for by PENTA or any other computational model. The present version of PENTA could capture this variability as a context-specific target change, akin to Tone

3 sandhi in Mandarin (Yip, 2002), but such an account would bear no relevance to the underlying mechanism being proposed, and achieves nothing more than introducing an ad hoc predictor in the model. Given its planning nature, and relatively weak correlation with low F_0, the modeling simulation of PLR would be rather different from that of post-low bouncing, which needs to be explored in future studies.

All in all, our PLR account for Japanese pitch accent has three implications for phonology. Firstly, the gradient nature of pre-low raising means that this type of surface difference in F_0 is unlikely due to a phonemic or tonomic contrast. Put in another way, what are proposed as different tonal categories (H- vs. H*) do not have to be distinct in terms of articulatory planning. At the phonological level, H- and H* were proposed in AM to serve different functions; at the articulatory planning level, our proposal is that H- and H* have the same target, given their indistinguishability in isolation. A (phonological) tone can be realized by multiple articulatory targets (in turn surface acoustics), just as different tones can share the same target. Taking morphology for example, the former would be reminiscent of the allophones of the English plural (-s, -z-, -ɨz), and the latter would be /s/ in English serving to mark both plurality and possession.

Secondly, the pervasiveness of PLR as seen in both Japanese and many other languages suggests that the kind of downstep that involves PLR is unlikely a contrastive phonological process. As such, it should not be grouped together with cases of downsteped H that do not involve L as a trigger, for example, final lowering (contra Truckenbrodt, 2004), vocative chant (Frota & Prieto, 2015), or boundary tone (Ladd, 2008).

Thirdly, the identification of PLR as an independent process as well as its possible articulatory mechanism, together with the identification of other articulatory mechanisms, such as post-low bouncing, inertia-triggered extensive F_0 transitions (Gandour et al., 1994; Xu, 1997, 1999), undershoot (Xu & Wang, 2009) and peak delay (Xu, 2001) suggests that surface F_0 patterns alone cannot be used as direct basis for positing underlying phonological tones or tonal processes such as scaling, downstep, and spreading. Instead, for each observed F_0 pattern, it is imperative to determine not only the potential contrastive function behind it, but also the likely articulatory mechanisms involved. That articulation is an intermediate stage between surface acoustics and phonological representation echoes classical concepts like the speech chain, as well as more recent theories like the DIVA Model (Guenther, 1994) and the PENTA Model (Xu, 2005).

Needless to say, the generalizability of the present finding needs further empirical support from other languages. One of our next steps will be to investigate whether the same correlations can be observed in languages where PLR is well established, e.g. Thai, Cantonese and Mandarin. Moreover, although we have found evidence of mora-level categorical pre-planning, it is also possible that the syllable is the real tone bearing unit of the language, as is the case for languages like Mandarin and English. This issue will also be examined in future studies.

V.CONCLUSION

In this paper, we have used a quantitative approach to show that in Japanese the F_0 peak associated with a pitch accent varies with its following low tone. We have found evidence that the variable F_0 peak height is the result of pre-low raising. Pearson's *r* reveals an inverse relation between accent peak and the following low tone, and that such relation becomes more pronounced when the peak is further away from the low tone. That the effect of PLR is masked by the proximity between F_0 peak and its following low tone may explain the absence of similar findings in the literature. These results suggest that in Japanese a low tone raises its preceding high tone, which is consistent with our current understanding of the physiology of vocal fold tension control in F_0 production.

References

Atkinson, J. E. (1978). Correlation analysis of the physiological factors controlling fundamental voice frequency. Journal of the Acoustical Society of America 63(1):211–222.

Beckman, M. E., & Pierrehumbert, J. B. (1986a). Intonational structure in Japanese and English. Phonology Yearbook, 3, 255–309.

Beckman, M. E.; Pierrehumbert, J. B. (1986b). Japanese prosodic phrasing and intonation synthesis. In Proceedings of the 24th Annual Meeting on Association for Computational Linguistics (ACL1986) (pp. 173–180). New York, NY.

Boersma, P. P. G.; Weenink, D. J. M. (2012). Praat: Doing phonetics by computer. Retrieved from http://www.praat.org/ on 2012/12/1.

Chen, Y.; Xu, Y. (2006). Production of weak elements in speech: Evidence from F0 patterns of neutral tone in Standard Chinese. Phonetica 63(1):47–75.

Cheng, C.; Xu, Y. (2013). Articulatory limit and extreme segmental reduction in Taiwan Mandarin. Journal of the Acoustical Society of America 134(6):4481–4495.

Cheng, C., & Xu, Y. (2014). Mechanism of disyllabic tonal reduction in Taiwan Mandarin. Language and Speech, 1–34.

Cohen, A., Collier, R.; ’t Hart, J. (1982). Declination: Construct or intrinsic feature of speech pitch? Phonetica 39:254–273.

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

Erickson, D. M. (1976). A physiological analysis of the tones of Thai. PhD Thesis. University of Connecticut, Storrs, CT.

Frota, S.; Prieto, P. (2015). Intonation in Romance : Systemic similarities and differences. In S. Frota & P. Prieto (Eds.), Intonation in Romance (pp. 392–418). Oxford: Oxford University Press.

Fujisaki, H.; Wang, C.; Ohno, S.; Gu, W. (2005). Analysis and synthesis of fundamental frequency contours of Standard Chinese using the command–response model. Speech Communication 47(1-2):59–70.

Gandour, J. T.; Potisuk, S.; Dechongkit, S. (1994). Tonal coarticulation in Thai. Journal of Phonetics 22:477–492.

Gu, W.; Lee, T. (2007). Effects of tonal context and focus on Cantonese F0. In Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS 2007) (pp. 1033–1036). Saarbrücken, Germany.

Guenther, F. H. (1994). A neural network model of speech acquisition and motor equivalent speech production. Biological Cybernetics, 72(1), 43–53.

Gussenhoven, C. H. M. (2004). The phonology of tone and intonation. New York, NY: Cambridge University Press.

Haraguchi, S. (2002). Accent. In N. Tsujimura (Ed.), The Handbook of Japanese Linguistics (pp. 1–30). Malden, MA: Blackwell Publishers.

Hyman, L. M.; Schuh, R. G. (1974). Universals of tone rules: Evidence from West Africa. Linguistic Inquiry 5(1):81–115.

Kindaichi, H. (2005). Toukyougo-ni okeru hana to hana no kubetsu: Toukyou akusento shin-nidankan kyouchouron. In Kindaichi Haruhiko Chosakushuu Vol. 6 (pp. 303‐326). Tokyo: Tamagawa University.

Kubozono, H. (1993). The organization of Japanese prosody. Tokyo: Kurosio.

Ladd, D. R. (1984). Declination: A review of some hypotheses. Phonology Yearbook 1:53–74.

Laniran, Y. O. (1992). Intonation in tone languages: The phonetic implementation of tones in Yoruba. PhD Thesis. Cornell University.

Laniran, Y. O.; Clements, G. N. (2003). Downstep and high raising: Interacting factors in Yoruba tone production. Journal of Phonetics 31(2):203–250.

Laniran, Y. O.; Gerfen, C. (1997). High raising, downstep and downdrift in Igbo. In Paper presented at the 71st annual meeting of the Linguistic Society of Americ. Chicago, IL.

Lee, A. (2015). The dynamics of Japanese prosody. PhD Thesis. University College London.

Lee, A.; Xu, Y.; Prom-on, S. (2014). Modeling Japanese F0 contours using the PENTAtrainers and AMtrainer. In Proceedings of the 4th International Symposium on Tonal Aspects of Languages (TAL 2014) (pp. 164–167). Nijmegen.

Lee, A., & Xu, Y. (2016). Effect of speech rate on pre-low raising in Cantonese. In Proceedings of the 5th International Symposium on Tonal Aspects of Languages (TAL 2016). Buffalo, NY.

Liu, F.; Xu, Y. (2005). Parallel encoding of focus and interrogative meaning in Mandarin intonation. Phonetica 62:70–87.

Liu, F.; Xu, Y.; Prom-on, S.; Yu, A. C. L. (2013). Morpheme-like prosodic functions: Evidence from acoustic analysis and computational modeling. Journal of Speech Sciences 3(1):85–140.

Ostry, D. J.; Munhall, K. G. (1985). Control of rate and duration of speech movements. Journal of the Acoustical Society of America 77(2):640–648.

Pierrehumbert, J. B. (1980). The phonology and phonetics of English intonation. PhD Thesis. Massachusetts Institute of Technology, Cambridge, MA.

Pierrehumbert, J. B.; Beckman, M. E. (1988). Japanese Tone Structure. Cambridge, MA: Massachusetts Institute of Technology.

Poser, W. J. (1984). The phonetics and phonology of tone and intonation in Japanese. PhD Thesis. Massachusetts Institute of Technology, Cambridge, MA.

Prom-on, S.; Liu, F.; Xu, Y. (2012). Post-low bouncing in Mandarin Chinese: Acoustic analysis and computational modeling. Journal of the Acoustical Society of America 132(1):421–432.

Prom-on, S.; Xu, Y.; Thipakorn, B. (2009). Modeling tone and intonation in Mandarin and English as a process of target approximation. Journal of the Acoustical Society of America 125(1):405–424.

Rialland, A. (1981). Le système tonal du gurma (langue gur de Haute-Volta). Journal of African Languages and Linguistics 3:39–64.

Saitou, T.; Unoki, M.; Akagi, M. (2005). Development of an F0 control model based on F0 dynamic characteristics for singing-voice synthesis. Speech Communication 46(3-4):405–417.

Selkirk, E. O.; Shinya, T.; Kawahara, S. (2004). Phonological and phonetic effects of Minor Phrase length on F0 in Japanese. In Proceedings of the 2nd International Conference on Speech Prosody (SP2004) (pp. 183–186). Nara, Japan.

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

Sugito, M. (1968). Comparison between “high-final” & “level” tone in Tokyo accent. Bulletin of the Phonetic Society of Japan 129:1–4.

Sugiyama, Y. (2012). The production and perception of Japanese pitch accent. Newcastle upon Tyne, England: Cambridge Scholars Publishing.

Truckenbrodt, H. (2004). Final lowering in non-final position. Journal of Phonetics 32(3):313–348.

Vance, T. J. (1995). Final accent vs. no accent: Utterance-final neutralization in Tokyo Japanese. Journal of Phonetics 23(4):487–499.

Warner, N. (1997). Japanese final-accented and unaccented phrases. Journal of Phonetics 25(1):43–60.

Xu, Y. (1997). Contextual tonal variations in Mandarin. Journal of Phonetics 25(1):61–83.

Xu, Y. (1999). Effects of tone and focus on the formation and alignment of F0 contours. Journal of Phonetics 27(1):55–105.

Xu, Y. (2001). Fundamental frequency peak delay in Mandarin. Phonetica 58(1-2):26–52.

Xu, Y. (2005). Speech melody as articulatorily implemented communicative functions. Speech Communication 46(3-4):220–251.

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) (pp. 7–10). Aix-en-Provence, France.

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

Xu, Y.; Sun, X. (2002). Maximum speed of pitch change and how it may relate to speech. Journal of the Acoustical Society of America 111(3):1399–1413.

Xu, Y.; Wang, M. (2009). Organizing syllables into groups: Evidence from F0 and duration patterns in Mandarin. Journal of Phonetics 37(4):502–520.

Xu, Y.; Wang, Q. E. (2001). Pitch targets and their realization: Evidence from Mandarin Chinese. Speech Communication, 33(4):319–337.

Yip, M. J. W. (2002). Tone. New York, NY: Cambridge University Press.

Tables

1-mora	CV		
Unaccented	ne (price)		
1 (H-L)	ne’ (root)		
2-mora	CV	CVV	CVN
Unaccented	mane (imitate)	mai (dance)	
1 (HL-L)	me’mo (memo)	me’i (May)	me’n (face)
2 (LH-L)	mune’ (aim)		
3-mora	CV	CVV	CVN
Unaccented	mimono (ornamental plant)	mimei (dawn), neimo (shoot)	momen (cotton)
1 (HLL-L)	me’nami (small wave)	me’imu (fog), ni’mei (two people)	ni’nmu (mission)
2 (LHL-L)	nana’me (oblique)	mema’i (dizzy)	nima’n (20,000)
3 (LHH-L)	mimono’ (attraction)	nuime’ (seam)	
4-mora	CVCV	CVV	CVN
Unaccented	monomane (mimicry)	meimei (naming)	nennen (annually)
1 (HLLL-L)		mu’umin (Moomin)	na’nnen (what year)
2 (LHLL-L)	mina’mina (everyone)		
3 (LHHL-L)	namana’ma (lively)	meime’i (individual)	menme’n (everyone)
4 (LHHH-L)	anomama’ (as it is)	nimaime’ (second piece)	ninenme’ (second year)

Table I. A list of stimuli used and corresponding English gloss in brackets. For simplicity, tonal representation used in the first column of this Table follows Haraguchi (2002), which comprises only H and L.

N = 1840		VMaxFall	RiseSize	FallSizeA	FallSizeB	PeakDelay
MaxF0	*r*		.694	.318	.267	.415
	p		<.001	<.001	<.001	<.001
MinF0A	*r*	.057	-.198	-.955	-.745	-.201
	p	.015	<.001	<.001	<.001	<.001
MinF0B	*r*		-.214	-.732	-.967	-.103
	p		<.001	<.001	<.001	<.001
RiseSize	*r*			.394	.382	.229
	p			<.001	<.001	<.001
FallSizeA	*r*				.786	.314
	p				<.001	<.001

Table II: Pearson's correlations of normalized data (converted into semitones using utterance-initial F_0 value). Non-significant correlations are not displayed. Data of unaccented words have been removed.

Pairwise Comparisons

Measure: **MaxF0 (semitones)**

(I) Peak-to-end distance (mora)		Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
1	2	-.284*	.037	.001	-.420	-.149
	3	-.658*	.099	.002	-1.017	-.299
	4	-.975*	.128	.001	-1.441	-.508
2	3	-.374*	.075	.009	-.646	-.101
	4	-.691*	.117	.004	-1.116	-.265
3	4	-.317	.158	.514	-.893	.259

Table III: Post-hoc Bonferroni test comparing MaxF0 under different peak-to-end distance conditions (α = 0.05).

Peak-to-end distance	**Pearson's *r* RiseSize~ MinF0A**	**Word length**	**Pearson's *r* RiseSize~ MinF0A**
4 morae	**-0.317**	**4 morae**	**-0.225**
3 morae	**-0.205**	**3 morae**	**-0.210**
2 morae	**-0.190**	**2 morae**	**-0.202**
1 mora	**-0.142**	**1 mora**	**0.306**
Combined	**-0.198**	**Combined**	**-0.198**

Table IV: Pearson's *r* of **RiseSize~ MinF0A** (converted into semitones using utterance-initial F_0 value). Data were subsetted into four groups according to peak-to-end distance and word length.

Collected figure captions

Figure 1: F_0 contours of a 4-mora unaccented word *monomane* 'mimickry' and a 4-mora accented word *mina'mina* 'everyone' averaged across 40 repetitions from eight speakers (color online). X-axis shows normalised time, whereas y-axis represents F_0. The first interval from the left and the last two on the right are parts of the carrier sentence.

Figure 2: Mean Max F0 of unaccented words by word length (morae).

Figure 3. Time-normalized average F_0 contour of four 4-mora words. The solid vertical lines show target word boundaries, while the dashed vertical line marks the end of the particle *–mo*.

Figure 4. The effect of peak-to-end distance (number of morae between the accented mora and the mora *no-*) on MaxF0 in semitones (y-axis). Bar colors represent peak-to-end distance.

Figure 5. Averaged F_0 contours of four initial-accented words of different lengths. X-axis shows actual time (color online).

Figure 6. Scatterplot of MaxF0~MinF0A (N=2640) (color online).