Simulating Post-L F0 Bouncing by Modeling Articulatory Dynamics

Santitham Prom-on [1,2], *Yi Xu* [1], *Fang Liu* [3]

[1] Department of Speech, Hearing and Phonetic Sciences, University College London, UK
[2] Department of Computer Engineering, King Mongkut's University of Technology Thonburi, Thailand
[3] Center for the Study of Language and Information, Stanford University, Stanford, USA
santitham.prom-on@ucl.ac.uk, yi.xu@ucl.ac.uk, fliu4@stanford.edu

Abstract

Post-L F_0 bouncing (post-L bouncing for short) is a prosodic phenomenon whereby F_0 is temporarily raised following a very low pitch. The phenomenon is quite robust, but is not widely known, and it has never been computationally modeled. This paper presents the results of our simulation of the phenomenon by modeling articulatory dynamics. Using the quantitative Target Approximation (qTA) model, we were able to simulate the F_0 rise after the Mandarin L tone by adding an acceleration adjustment to the initial state of the first post-L Neutral tone. Furthermore, a linear relationship was found between the added acceleration and the amount of F_0 lowering in the L tone. We interpreted the results as evidence that post-L bouncing is directly related to the articulatory mechanism of producing a very low pitch.

Index Terms: speech prosody, post-L F_0 bouncing, articulatory dynamics, qTA model

1. Introduction

The variations in F_0 contours in speech have two very different sources [1]. The first consists of various communicative functions such as tone, stress and intonation that carry real information. The second is composed of various perturbations due to the F_0 production mechanisms such as consonantal perturbation [2-4], anticipatory dissimilation [5-8] and vowel intrinsic F_0 [9]. A much lesser known perturbation pattern is post-L F_0 raising or post-L bouncing, which involves the raising of F_0 contours following the Low (L) tone, especially when the following syllables are unstressed or having the Neutral (N) tone [10, 11]. An example in Mandarin is shown in Figure 1, where the third syllable carries four alternating tones as indicated by line color and pattern. In the case of the L tone, F_0 within the third syllable drops sharply within the L-tone syllable, but it starts to rise from the beginning of the following N tone, and continues to rise in the next N tone. By the second N tone, the F_0 height has surpassed that of all the other tone sequences.

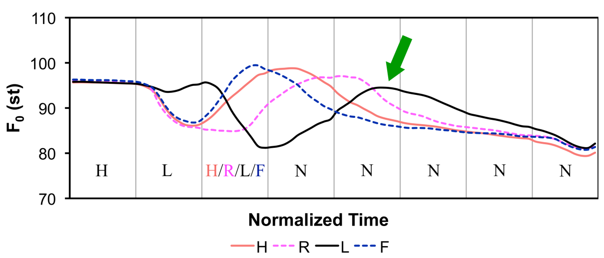

Figure 1: *Time-normalized mean F_0 contours, by 8 native Mandarin speakers, of the sentences "ta1 mai3 [ma1ma0/ye2ye0/nai3nai0/mei4mei0] men0 de0 le0 ma0", where 1, 2, 3, 4 and 0 represent the H, R, L, F and N tones, respectively.*

The behavior of the F_0 contour in the L-N tone sequence in Figure 1 is almost like that of a bouncing ball after hitting the ground, hence the name, post-L bouncing. Such pitch bouncing is different from the known carryover effects which are always assimilatory [5, 7]. The assimilatory effects have been successfully modeled by the quantitative Target Approximation (qTA) based on simple mechanical dynamics [12, 13]. But qTA in its current form cannot simulate post-L bouncing, as will become clear in the Result section. qTA is based on the understanding that normal F_0 variation is achieved by manipulating two antagonistic forces, generated by the intrinsic laryngeal muscles, that jointly control vocal fold tension. These forces are produced by the contraction of the cricothyroids to lengthen the vocal folds, and that of the thyroarytenoids and a number of other muscles to shorten the vocal folds [14, 15].

The production of very low F_0, however, is known to involve the extrinsic laryngeal muscles such as the sternohyoids, sternothyroids, and omohyoids [16-18]. The main function of these muscles is to control the vertical position of the larynx, and it has been suggested that the lowering of the larynx drags the cricoid cartilages across a curvature of the cervical spine, which tilts the top of the cricoids forward and shortens the vocal folds by an extra amount [19]. Given these findings, it is possible that after producing a very low F_0, the extrinsic laryngeal muscles simply stop contracting, which temporarily tips the balance between the two antagonistic forces maintained by the intrinsic laryngeal muscles, resulting in a sudden increase of the vocal fold tension. If this is the case, post-L bouncing could be modeled by adding an extra F_0 raising force after F_0 is lowered beyond a certain threshold. The present study is an attempt to explore this possibility by modeling simulation using the qTA model [12, 13].

2. Method

2.1. Corpus

The corpus consists of utterances by 8 native Mandarin speakers recorded for a study of question intonation involving long sequences of consecutive neutral tones [20]. Each utterance consists of 8 syllables, and the tone of the third syllable varies across High (H, T1), Rising (R, T2), Low (L, T3) and Falling (F, T4) tones, as shown in Figure 1. The first syllable is always H and the second syllable always L. The fourth to sixth syllables are always the Neutral tone (N, T0), and the final two syllables are either both H or both N. Each utterance was also said as either a statement or a question, and with prosodic focus either on the second or third syllable. The utterance was repeated five times by each speaker. Table 1 shows the structure of the sentences in this corpus.

For this study, only those utterances with L tone on the third syllable were used. There are 320 utterances in total that fit this criterion. However, the results in 3.1 were based on the whole corpus.

Table 1. *Structure of the sentence.*

<table>
<tr><td rowspan="4">ta1 mai3
(他买)
H L</td><td>ma1 ma0 (妈妈) H N</td><td rowspan="4">men0 de0
(们的)
N N</td><td rowspan="2">le0 ma0
(了嘛)
N N</td></tr>
<tr><td>ye2 ye0 (爷爷) R N</td></tr>
<tr><td>nai3 nai0 (奶奶) L N</td><td rowspan="2">mao1 mi1
(猫咪)
H H</td></tr>
<tr><td>mei4 mei (妹妹) F N</td></tr>
</table>

2.2. Baseline simulation

The quantitative Target Approximation (qTA) was used to simulate the baseline F_0 contour. The qTA model simulates the production of tone and intonation as a process of syllable-synchronized sequential target approximation [12, 13]. It represents F_0 as the surface response of the target approximation process driven by underlying pitch targets. In qTA, an F_0 contour is composed of two parts: forced response and natural response. The forced response is the output of the process when it reaches the desired target and the natural response is the additive transient in transition from the current articulatory state to the desired target. F_0 realization in qTA model is computed by the following equation,

$$f_0(t) = (mt + b) + (c_1 + c_2 t + c_3 t^2)e^{-\lambda t} \quad (1)$$

The first term in parenthesis is the forced response which is the pitch target and the second term, the polynomial and the exponential, is the natural response. The model has three parameters, *m* and *b* which specify the pitch target slope and height, respectively, and λ which represents the rate or strength of target approximation. The coefficients c_1, c_2 and c_3 are jointly determined by the initial dynamic F_0 state of the syllable and its pitch target. The initial state consists of F_0 level, $f_0(0)$, velocity $f_0'(0)$, and acceleration, $f_0''(0)$. This state is transferred directly from the preceding syllable. The three transient coefficients are computed with the following formulae.

$$c_1 = f_0 - b \quad (2)$$

$$c_2 = f_0'(0) + c_1\lambda - m \quad (3)$$

$$c_3 = \left(f_0''(0) + 2c_2\lambda - c_1\lambda^2\right)/2 \quad (4)$$

For each syllable in an utterance except those carrying the N tone, the qTA parameters *m*, *b* and λ are simultaneously estimated by searching for the parameter combination with the lowest sum of square error. For the N tone, because their pitch target takes several consecutive N-tone syllables to reach, for each utterance, all the N tones were grouped together as one segment during parameter estimation. The results of this simulation formed the baseline in the study.

2.3. Post-L F_0 bouncing simulation

To implement the extra force as hypothesized for post-L bouncing, we added an adjustment of either the acceleration or velocity value to the initial state of the first N tone after the L tone. Such an adjustment would not cause a sudden change in F_0, but would influence the F_0 trajectories of the tone following the L tone. To determine the amount of adjustment needed, we modified the qTA parameter estimation algorithm implemented in PENTAtrainer [21] to iteratively search for the optimal extra values of either velocity or acceleration that resulted in the lowest sum of square errors. This was done only at the beginning of the fourth syllable which was immediately after the L tone.

2.4. Developing a generalized post-L bouncing rule

The initial-state adjustments in 2.3 were trained on individual utterances and reapplied to the same utterances. To develop a generalized rule, we analyzed the relation between the estimated extra force and the amount of F_0 lowering in the preceding L tone in terms of the difference between the final F_0 and the initial F_0 of the utterance. A 5-point median filter was applied to eliminate outliers in the resulting pairs of adjustment and the F_0 lowering values. The post-L bouncing rule was then established in the form of a linear relation between the amount of F_0 lowering in the L tone and the estimated adjustment of the initial state of the first post-L N tone.

We then tested the generalizability of the post-L bouncing rule by applying it in the qTA resynthesis of the F_0 contours. The synthesized F_0 contour was then compared to the original F_0 contours of each utterance in terms of root-mean-square error (RMSE) and Pearson's correlation coefficient. RMSE indicates point-by-point discrepancies between the synthesized and original F_0 contours, while correlation represents the degree of linear relationship between the two.

3. Results

3.1. L tone baseline compared to other tones

The necessity for modeling post-L bouncing is illustrated in Table 2, which shows synthesis accuracies of utterances having each of the four Mandarin full tones in the third syllable *without applying any adjustments for post-L bouncing*. The utterances with the L tone show both higher error and lower correlation compared to other tones.

Table 2. *Means and standard errors of utterances having each full tone in the third syllable.*

Tone	RMSE	Correlation
H	1.23 ± 0.12	0.964 ± 0.003
R	1.39 ± 0.11	0.950 ± 0.006
L	1.85 ± 0.25	0.906 ± 0.013
F	1.05 ± 0.10	0.972 ± 0.003

3.2. Effectiveness of adjusting initial acceleration

Table 3 shows accuracies of F_0 synthesis when the learned post-L bouncing rule in terms of either velocity or acceleration was applied. These results are based on only those utterances in the corpus that have the L tone on the third syllable. As can be seen, implementing the acceleration adjustments yielded greater improvements than implementing the velocity adjustments, which is intuitively reasonable since acceleration is directly proportional to force according to Newton's second law of motion. Implementing both acceleration and velocity adjustments yielded only marginal additional improvements over implementing either alone. This is presumably because velocity variations are the direct result of acceleration variations and so the two are not independent of each other. In the following discussion, we will therefore only report the results of implementing the post-L bouncing adjustments in terms of acceleration. Figure 2 shows examples of qTA-synthesized F_0 contours with and without implementing the acceleration adjustment. Without the additional acceleration added to the initial state of the first post-L N tone, F_0 simply continues the low final F_0 of the L tone, resulting in a contour that is very different from the original one. With the acceleration adjustment, the F_0 contours seem to resemble the original one much better.

Table 3. *Means and standard errors of different post-L F_0 raising simulation strategies.*

Strategy	RMSE (st)	Correlation
Baseline	1.85 ± 0.25	0.906 ± 0.013
Baseline + V	1.12 ± 0.11	0.967 ± 0.004
Baseline + A	1.01 ± 0.10	0.974 ± 0.003
Baseline + V + A	0.97 ± 0.10	0.975 ± 0.003

Comparing Table 3 to Table 2, we can also see that, with the implementation of the acceleration adjustments, the overall errors and correlations of the utterances with the L tone on the third syllable are now comparable to those of other full tones.

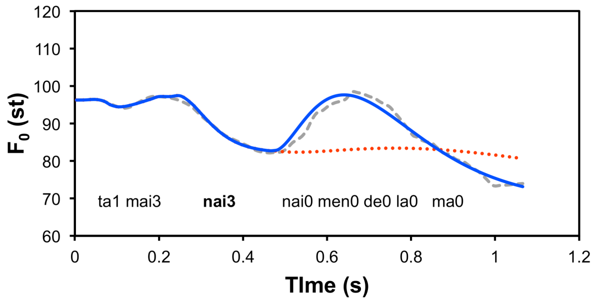

Figure 2: *Examples of F_0 contours synthesized with (solid blue) and without (dotted red) post-L bouncing simulation, together with the original F_0 contour (dashed gray).*

3.3. Relation of acceleration adjustment and amount of F_0 lowering

Figure 2 shows the scatter plot the simulated acceleration adjustment as a function of the amount of F_0 lowering measured in terms of the difference between the final F_0 value of the L tone of the third syllable and the utterance-initial F_0. A strong linear relation can be seen, with a coefficient of determination (R^2) of 0.80. The linear trend in the scatter plot can be expressed by the follow equation:

$$Acceleration = 414.6 \times F_0 Lowering - 1787.4 \quad (5)$$

This linear equation indicates that the reduction of F_0 level of the L tone by 1 *st* would increase the initial acceleration of the following N tone by 414.6 st/s^2. Since the negative value of acceleration would mean a lowering rather than raising of F_0, equation (5) is valid only when the size of F_0 lowering is greater than 4.32 *st*, i.e., at the intercept of the regression line with the x-axis in Figure 2. For the F_0 displacement lower than 4.32 *st*, the acceleration adjustment should be set to 0.

3.4. Effectiveness of a generalized post-L bouncing rule

The linearity of the scatter plot in Figure 3 means that equation (5) can be used as a generalized rule that can be applied to any N tone that immediately follows the L tone. When we applied this rule, as it turned out, the synthesis accuracy did not decrease from those shown in Table 3 (RMSE: 1.02 ± 0.09, Correlation: 0.973 ± 0.002). This is a clear indication of the effectiveness of the generalized post-L bouncing rule as expressed by equation (5).

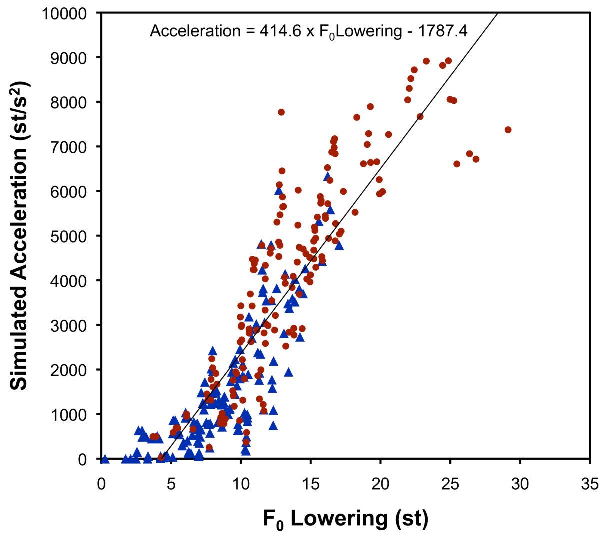

Figure 3: *Relationship between simulated acceleration adjustment and amount of F_0 lowering in the L tone. Red circles are cases where the L tone is focused, while blue triangles are cases where the L tone is post-focus.*

To further test how the post-L bouncing rule would interact with the communicative functions included in the current corpus, we conducted function-specific modeling in the same way as done in [12], i.e., averaging qTA parameters in the same functional category and using them as the representation of that function. The post-L bouncing rule was applied only after the third syllable. Figure 4 shows the synthesis accuracy of each imposed functional combination with and without the post-L bouncing rule. As is apparent, for each and every functional combination, the qTA-generated F_0 contours show better precision in terms of both RMSE and correlation when the post-L bouncing rule was applied. Moreover, the improvements with increased number of functions included in the modeling can be seen only when the post-L bouncing rule was implemented. Without the implementation, there was no improvement in RMSE at all. This is likely because the raising of F_0 after the focused L tone can be simulated only when the post-L rule is applied, and the corpus used in this study contains many sequences of consecutive N tones after focus which are highly susceptible to post-L bouncing.

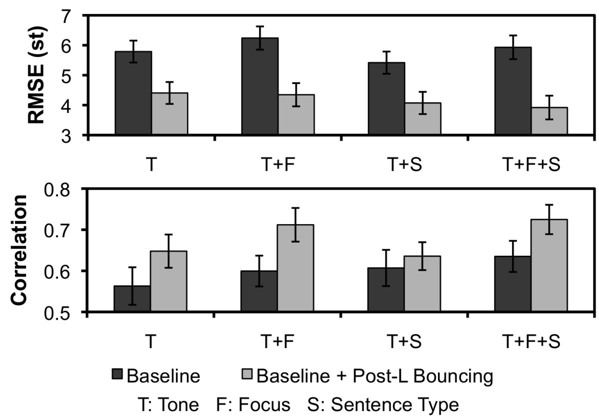

Figure 4: *Means and standard errors of RMSE and correlation of each imposed functional combination compared between the baseline (dark gray) and baseline plus post-L simulation (light gray).*

4. Discussion

Post-L F_0 bouncing is a prosodic phenomenon that is specifically recognized only recently [10]. But its effect on the N tone in Mandarin seems to be strong enough that we cannot afford to ignore it in either computational modeling or theoretical understanding of tone and intonation. Theoretically, the behavior of the N tone after the L tone has long been a puzzle in both phonetics and phonology research ever since its early description by Chao [22]. The bouncing effect may also be responsible for the constant valley-to-peak interval reported for English [11], which has been viewed as evidence for the existence of underlying phonological tonal units in the language. The importance and benefit of simulating post-L bouncing in computational modeling of tone and intonation have been clearly shown by the examples and results in this paper.

To achieve a better understanding of post-L F_0 bouncing, we tested a hypothesis based on previous reports and proposals in regard to the articulatory mechanisms of producing very low F_0 in speech [16, 17, 19]. That is, post-L bouncing is due to an added articulatory force introduced by the sudden cessation of the contraction of the external laryngeal muscles after producing a very low F_0, which temporarily tips the balance in the antagonistic control of the vocal fold tension by the intrinsic laryngeal muscles. We tested the hypothesis by training the qTA model with an additional force in term of either velocity or acceleration added to the initial state of the first post-L N tone. The results showed clear supports for the hypothesis in several ways. The first is the highly linear relationship between the amount of F_0 lowering in the L tone and the simulated acceleration adjustment as shown in Figure 3. This linear relation can be conceptualized if we link Hooke's law of elasticity with Newton's second law of motion. In such a relation, the acceleration as part of the restoring force is linearly proportional to displacement of one end of a spring from its equilibrium. Further support was seen in the fact that the application of the generalized post-L bouncing rule based on equation (5) lead to virtually identical performance as that of applying utterance-specific acceleration adjustments shown in Table 3. Finally, the post-L bouncing rule was shown to be able to be implemented in tandem with the functional modeling strategy developed in our previous research [12] and to significantly improve the performance of the modeling simulation.

5. Conclusions

In this study we tested the hypothesis that post-L F_0 bouncing is directly related to the articulatory mechanism of producing very low pitch. We simulated the bouncing effect by adding an extra force in the form of acceleration adjustment to the initial state of the first post-L N tone using the qTA model. The results of simulating a Mandarin corpus containing many cases of consecutive N tones following the L tone, where the effect is the strongest, indicate a clear success of the strategy, with dramatic performance improvement in terms of both RMSE and correlation relative to the baseline. A highly linear relation was also found between the amount of F_0 lowering in the L tone and the degree of acceleration adjustment, which provides further support for the extra-force hypothesis. The greater generality of the hypothesis can be tested in future studies on tones other than the Neutral tone and in languages other than Mandarin.

6. Acknowledgements

The authors would like to thank the Royal Society and the Royal Academy of Engineering for the financial support through the Newton International Fellowship Scheme, and the Thai Research Fund through the TRF-CHE Research Grant for New Scholar with grant number MRG5380038. This work was supported in part by the Economic and Social Research Council with grant number PTA-026-27-2480.

7. References

[1] Xu, Y., "Speech melody as articulatorily implemented communicative functions," *Speech Communication,* vol. 46, pp. 220-251, Jul 2005.

[2] Hombert, J.-M., "Consonant types, vowel quality, and tone," in *Tone: A linguistic survey*, V. A. Fromkin, Ed., New York: Academic Press, 1974, pp. 77-111.

[3] Lehiste, I., "Some Basic Considerations in the Analysis of Intonation," *The Journal of the Acoustical Society of America,* vol. 33, p. 419, 1961.

[4] Xu, C. X. and Xu, Y., "Effects of Consonant Aspiration on Mandarin Tones," *Journal of the International Phonetic Association,* vol. 33, pp. 165-181, 2003.

[5] Gandour, J.*, et al.*, "Tonal Coarticulation in Thai," *Journal of Phonetics,* vol. 22, pp. 477-492, Oct 1994.

[6] Laniran, Y. O. and Clements, G. N., "Downstep and high raising: interacting factors in Yoruba tone production," *Journal of Phonetics,* vol. 31, pp. 203-250, Apr 2003.

[7] Xu, Y., "Contextual tonal variations in Mandarin," *Journal of Phonetics,* vol. 25, pp. 61-83, Jan 1997.

[8] Xu, Y., "Effects of tone and focus on the formation and alignment of F0 contours," *Journal of Phonetics,* vol. 27, pp. 55-105, 1999.

[9] Whalen, D. H. and Levitt, A. G., "The universality of intrinsic F0 of vowels," *Journal of Phonetics,* vol. 23, pp. 349-366, 1995.

[10] Chen, Y. Y. and Xu, Y., "Production of weak elements in speech - Evidence from F-o patterns of neutral tone in standard Chinese," *Phonetica,* vol. 63, pp. 47-75, 2006.

[11] Pierrehumbert, J., "The phonology and phonetics of English intonation," Ph.D., MIT, Cambridge, MA, 1980.

[12] Prom-on, S.*, et al.*, "Modeling tone and intonation in Mandarin and English as a process of target approximation," *J Acoust Soc Am,* vol. 125, pp. 405-24, Jan 2009.

[13] Xu, Y., "Pitch targets and their realization: Evidence from Mandarin Chinese," *Speech Communication,* vol. 33, pp. 319-337, 2001.

[14] Titze, I. R., "On the relation between subglottal pressure and fundamental frequency in phonation," *J Acoust Soc Am,* vol. 85, pp. 901-6, Feb 1989.

[15] Zemlin, W. R., *Speech and hearing science : anatomy and physiology*, 4th ed. Boston: Allyn and Bacon, 1998.

[16] Erickson, D. M., "A physiological analysis of the tones of Thai," Ph.D., University of Connecticut, 1976.

[17] Erickson, D.*, et al.*, "The production of low tones in English intonation," *Journal of Phonetics,* vol. 23, pp. 179-188, 1995.

[18] Halle, P. A., "Evidence for Tone-Specific Activity of the Sternohyoid Muscle in Modern Standard Chinese," *Language and Speech,* vol. 37, pp. 103-123, Apr-Jun 1994.

[19] Honda, K.*, et al.*, "Role of vertical larynx movement and cervical lordosis in F0 control," *Lang Speech,* vol. 42 (Pt 4), pp. 401-11, Oct-Dec 1999.

[20] Liu, F. and Xu, Y., "The Neutral Tone in Question Intonation in Mandarin," *Interspeech 2007: 8th Annual Conference of the International Speech Communication Association, Vols 1-4,* pp. 521-524, 2007.

[21] Xu, Y. and Prom-on, S. (2010-2011). *PENTAtrainer.praat*. Available: http://www.phon.ucl.ac.uk/home/yi/PENTAtrainer/

[22] Chao, Y. R., *A grammar of spoken Chinese*. Berkeley: University of California Press, 1968.