ANTICIPATORY TONAL COARTICULATION: HOW, WHEN AND WHY IT OCCURS

Yan Sun & Chilin Shih

University of Illinois at Urbana-Champaign, USA
yansun5@illinois.edu; cls@illinois.edu

ABSTRACT

This study examines anticipatory tonal coarticulation exhibited on a string of neutral tones in Mandarin. In addition to the well-documented dissimilatory anticipation triggered by the Low tone, the results show that the Falling tone (with high onset) consistently exerts assimilatory anticipatory effect which could extend over three preceding neutral-tone syllables; the High tone, on the other hand, doesn't exert such effect. 2D density plots of the surrounding full tones revealed that dissimilatory anticipation was strongest when the neutral tones were both preceded and followed by a low pitch target, whereas assimilatory anticipation tended to occur when the following tone had an early high pitch target. This finding was interpreted as the result of speech planning and articulatory constraint.

Keywords: tonal coarticulation, assimilation, dissimilation, speech planning, articulatory constraint

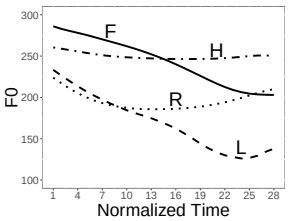
1. INTRODUCTION

In continuous speech, F0 contours of lexical tones are constantly affected by the preceding tone (i.e. carryover tonal coarticulation) and/or the following tone (i.e., anticipatory tonal coarticulation, or ATC). Previous studies often focused on the directionality, nature (in terms of assimilation vs. dissimilation) and magnitude of tonal coarticulation (see [6] for a review). In general, carryover effect has been found to be stronger than anticipatory effect (cf. [4,6]) and to be mostly assimilatory in nature (cf. [4,36]). The nature of ATC, on the other hand, is language- or even tone-specific: in Vietnamese, ATC is reported to be entirely assimilatory [2, 3, 13]; by contrast, in Thai [1, 9–11, 16], Cantonese [28], and Tianjin Chinese [14, 36], ATC is totally dissimilatory; in languages like Mandarin [20, 21, 29, 30], Taiwanese [15], Nanjing Chinese [6], and Malaysian Hokkein [4], both assimilatory and dissimilatory ATC have been reported. This complexity raises questions about why ATC occurs and how it is specific to different languages and tones. To investigate the possible underlying mechanism for ATC, it is necessary to first examine when it occurs or does not occur. Since Mandarin exhibits complex patterns of ATC, the present study takes Mandarin as an example to investigate how, when and why ATC occurs.

1.1. Tonal system in Mandarin

Mandarin distinguishes four lexical tones phonologically described as High (H), Rising (R), Low (L) and Falling (F). Fig. 1 (adapted from Figure 1 in [22]) displayed the averaged F0 contours of the four tones. The data include all possible Mandarin syllables spoken in a sentence frame. Note that the rising slope of R is shallow for its high target is often delayed to the next syllable in continuous speech [21, 32], and that L has a rising tail when spoken in isolation.

Figure 1: F0 contours of Mandarin lexical tones.



In addition to the four full tones, Mandarin has another tonal category called the neutral tone (N), which is phonologically targetless [5]. N only occurs in prosodically weak positions and the surface F0 contour associated with N changes dramatically depending on its tonal context.

1.2. ATC in Mandarin

In studies on tonal coarticulation in Mandarin, anticipatory dissimilation and assimilation have both been attested.

Anticipatory dissimilation refers to the raising of a tone when it is followed by a low pitch target. This dissimilatory ATC is widely reported in studies of Mandarin [20, 21, 29–31]. For example, Xu [30] examined tonal coarticulation using bi-tonal nonsense

sequences and found that when the first syllable was followed by a low-onset tone (i.e., R and L), its F0 contour was higher than when it was followed by a high-onset tone (i.e., H and F), and the strongest effect was triggered by L in the second syllable.

Compared to the well-documented anticipatory dissimilation, assimilatory anticipation is only sporadically reported in previous studies [20,22,24,26]. For example, a recent study [26] examined the F0 contour of consecutive Ns and found that when a string of Ns and the following full tone (L or F) were in the same noun phrase, the F0 contours of Ns were higher when followed by F than by L, except when L preceded the sequence of Ns. Such assimilatory anticipation could extend over two N syllables.

1.3. Effects of prosodic strength and prosodic structure on ATC

Before moving on to investigate tone-specific effect on ATC, it should be noticed that other factors, such as prosodic strength and prosodic structure, could also play a role. For example, Xu [31] found that when a tone occurred in a focused word, it exerted greater influence on adjacent tones and sometimes also on non-adjacent tones; Shih and Kochanski [22] demonstrated that tones of prosodically weak syllables were more likely to accommodate the shapes of neighboring strong tones. As for the effect of prosodic structure, studies [26, 37] showed that different levels of prosodic phrase boundaries weaken tonal coarticulation to different extent; Scholz and Chen [19] also demonstrated that tones in prosodic head position were more likely to resist coarticulation and maintain the canonical shapes.

2. EXPERIMENTAL DESIGN

2.1. Stimuli

The experiment is designed to examine how the F0 contour of three consecutive Ns is realized in different tonal context. N is chosen because it is phonologically targetless and always occurs in weak position; as a result, the effect of ATC should be the clearest on N. (1) shows an example of the stimuli:

(1) *ta shuo (ma-ma men de mao) zai shui-jiao.*
H H H-N N N H F F-F
he say mother PL POSS cat PROG sleep
"He said that the mothers' cat was sleeping."

The five syllables in parentheses is the region to be examined. The underlined words/tones were manipulated to provide different tonal context for the sequence of Ns. The syllable preceding and following the Ns took one of the four full tones (H, R, L and F), resulting in a total of 16 (4×4) tonal combinations. To control the effect of prosodic strength, focus was always on the last word (i.e. *shui-jiao* "sleep") of the sentence so the region under examination was always pre-focus. Prosodic structure was controlled in a way that the parenthesized region formed a noun phrase and the full-tone syllable following Ns (e.g., *mao* "cat" in (1)) was the head of the phrase.

2.2. Subjects

Twenty native speakers (13 females and 7 males) of Mandarin participated as subjects. They ranged in age from 19 to 28 years old at the time of the recording and were all born and raised in Beijing.

2.3. Procedures

The recording was conducted in a sound-treated booth in the Speech Acquisition and Intelligent Technology Lab at Beijing Language and Culture University. In order to make the production more natural, experimental sentences were elicited using questions and pictures. In each trial, the subject first saw a written question (e.g., *What did he say that the mothers' cat was doing?* in Mandarin), then a picture (e.g., a cat that is sleeping) was shown on the screen and the subject was instructed to answer the question based on the picture. Each question-picture pair was repeated 3 times during the recording and all trials were automatically randomized.

3. ANALYSIS AND RESULTS

The audio files were first auto-segmented using a forced aligner program, then segmentation errors were manually corrected. Time-normalized (20 points/syllable) F0 values of each sentence were generated using ProsodyPro [33], and the erroneous vocal cycle marks were also corrected by hand. The extracted F0 values in Hz were then converted to semitones following the procedure used in [23].

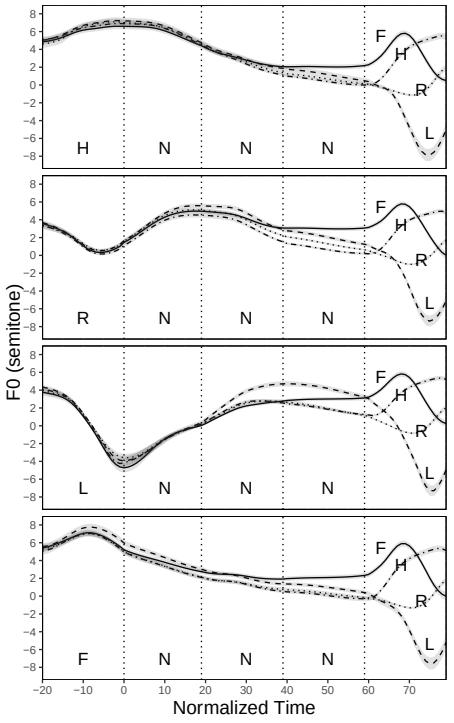
3.1. Graphical comparison of mean F0 contours

Fig. 2 displays the mean F0 contours (with standard errors indicated by the shaded bands) of the 5-syllable region across speakers and repetitions. The dotted vertical lines indicate syllable boundaries. In each panel, the tone of the first syllable is held constant while the tone of the last syllable is varied.

The graph shows that, in general, the F0 contours of the three Ns are the lowest when followed by H (dash-dotted lines), regardless of the preceding full tone. When the following full tone is R (dot-

ted lines), the contours of Ns generally overlap with those followed by H, except when the preceding full tone is R, in which case the contour of Ns followed by R is slightly higher, suggesting a weak dissimilatory effect triggered by the following R. When the following full tone is L (dashed lines), it exerts much stronger dissimilatory anticipation in that the F0 contours of Ns are higher when followed by L than by H or R, and the effect is strongest when the preceding full tone is L. Finally, when the following full tone is F (solid lines), the F0 contours of Ns are consistently lifted up, and the contours of the last N are the highest (except when the preceding full tone is L, in which case the contour of Ns followed by L is higher). This final observation suggests strong assimilatory anticipation triggered by F.

Figure 2: Anticipatory effect of the following full tone on F0 contours of neutral-tone sequences.



3.2. Statistical analysis

The whole F0 contours of the sequence of Ns are modeled using generalized additive mixed models (GAMMs) due to the non-linear shape [25, 27]. In the models, the contours of Ns followed by H were selected to be the *reference smooth* and the other contours were modelled as *difference smooths*. In model summary, GAMMs give both a set of *parametric* terms, indicating the overall height difference between two contours, and a set of *smooth* terms, indicating the shape difference between two contours. Table 1 and Table 2 summarized the *p*-values of the parametric terms and the smooth terms, respectively, reported by GAMMs. When the following tone is R, neither the overall height nor the shapes of the F0 contours of Ns significantly differ from the reference contour (an alpha level of .05 was used for all statistical tests). When the following tone is L and the Ns are preceded by R or F, only the overall height difference reached significance level; when the following tone is L and the preceding tone is also L, the overall height and the shape difference are both significant, suggesting a strong anticipatory effect. When the following tone is F, the effect is also strong: a following F consistently exerts significant effects on the shape (and sometimes the overall height as well) of Ns regardless of what the preceding full tone is.

Table 1: *p*-values of the parametric terms reported by GAMMs

	pre:H	pre:R	pre:L	pre:F
(fol:H)	<0.001***	<0.001***	<0.001***	< 0.001***
fol: R	0.858	0.210	0.731	0.798
fol: L	0.443	0.004**	0.033*	0.022*
fol: F	0.284	<0.001***	0.695	0.003**

Table 2: *p*-values of the smooth terms reported by GAMMs

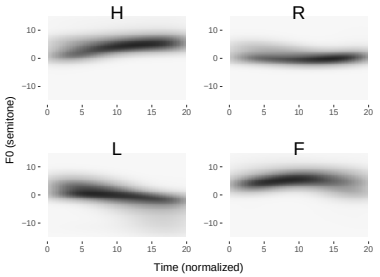
	pre:H	pre:R	pre:L	pre:F
(fol:H)	<0.001***	<0.001***	<0.001***	< 0.001***
fol: R	0.962	0.754	0.190	0.758
fol: L	0.089 .	0.068 .	<0.001***	0.074 .
fol: F	<0.001***	<0.001***	<0.001***	<0.001***

In order to see how far the anticipatory effects triggered by L and F can reach, the F0 contour of each N and the preceding full tone were separately modelled by GAMMs. Results suggest that the anticipatory effect of F could extend to as far as the first N when the preceding tone is F (p=.007 for the smooth term), while the effect exerted by L could extend to the second N when the preceding tone is R (p=.009 for the parametric term) or L (p=.002 for the parametric term and p=.006 for the smooth term).

In order to explain why the anticipatory effects exerted by the four tones differ in their nature and/or magnitude, 2D Kernel density plots of the full tones (preceding and following tones pooled together) were drawn and shown in Fig. 3. Higher density (i.e., darker colour) indicates smaller F0 variations across individual observations. Therefore, the high-density region of each tone can be interpreted as the common F0 target across speakers and tokens [35]. Under this interpretation, a high target can be identified at the late portion of H and early portion of F, and a low target can be identified at the late portion of R (the high target of R is delayed to the following

syllable, see [32] and references therein) and early portion of L[1]. In the discussion section, the height and position of these targets are argued to contribute to the different ATC patterns observed in this study.

Figure 3: Density plots of the full tones.



4. DISCUSSION

The F0 patterns of consecutive Ns examined in this study revealed interesting information about ATC in Mandarin. Results suggested that L and F exerted strong yet different anticipatory effects on preceding Ns: while the effect of L was dissimilatory in nature, the effect of F was assimilatory. Further examination showed that the effect of L was strongest when the preceding tone was L or R, whereas the effect of F was strongest when the preceding tone was F. H and R, on the other hand, didn't trigger ATC. The F0 target of each tone identified in the density plots seems to contribute to the discrepancies: the two ATC triggers, L and F, both have an early F0 target, whereas the pitch target of the other two tones, H and R, occurs late in the syllable. This finding raises the next question: why an early target triggered ATC in the current study, but a late target didn't?

We believe that dissimilatory and assimilatory ATC are both the result of active speech planning with the aim of accommodating articulatory constraints. Electromyographic (EMG) studies [7, 8, 12, 18] have demonstrated that F0 raising mainly requires contraction of the cricothyroid (CT) muscle, whereas F0 lowering involves the relaxation of CT as well as the activation of strap muscles and is consequently more effortful. Therefore, one could hypothesize that, to reach a low pitch target, a speaker should either initiate the lowering earlier or increase the velocity of the movement. The first strategy requires longer duration, while the latter requires a higher starting point. Since the low target of a following R occurs in the late portion of the tone-bearing syllable, it allows more time for the initiation of F0 lowering. By contrast, to plan for the early low target (and also the even lower target towards the end) in L, a speaker often needs to intentionally raise a preceding peak (for H, R and F, this is their intrinsic high pitch target; for L, this is the peak resulting from post-low bouncing, see [17] and references therein) by enhancing CT muscle activation, resulting in the anticipatory dissimilation. Moreover, since the preceding peak occurs the latest when the preceding full tone is L, given the shortest time for pitch lowering, the raising effect in this case is the strongest. This hypothesis could also explain why R triggered dissimilatory ATC in previous studies, e.g., [30], but not in the current one: in this study, the string of targetless Ns provides longer time for F0 lowering.

As for assimilatory anticipation, research on the maximum speed of pitch change [34] found that direction shift (either from falling to rising or from rising to falling) took time and that pitch raising was slower than pitch lowering. Therefore, it is expected that if a high pitch target is anticipated to follow a falling F0 contour, a speaker will decelerate the falling earlier in order to get prepared for the rising and thus makes the falling much flatter. This is exactly what happens in the current study. Because N is targetless, it is natural for its F0 to approach the so-called "neutral" level [12], and for a string of Ns like those examined here, it means to end with a falling contour. However, when the string of Ns is followed by an early high pitch target, like the one in F, the F0 contour needs to change its direction from falling to rising. In preparation for that, the falling is decelerated and the CT muscle is activated earlier, resulting in the raising of the contour. On the other hand, since the high target in H occurs late, the rising could be done inside the syllable itself, hence no need to raise the falling contour of Ns.

To further test our hypothesis that the height and position of pitch target in a tone would contribute to the various language- and tone-specific ATC patterns reported in the literature, it is necessary to look at other tonal languages and their tonal coarticulation patterns. Ultimately, it will also be necessary to test the hypothesis using articulatory methods.

5. CONCLUSION

In this paper we have shown that in Mandarin, tones with early low and high pitch target are better triggers of dissimilatory and assimilatory ATC, respectively, than tones with late target. Based on these observations, we have argued that, ATC, whether it is assimilatory or dissimilatory in nature, is a result of enhanced CT muscle activation in preparation for an upcoming early pitch target.

6. REFERENCES

[1] Abramson, A. S. 1979. The coarticulation of tones: An acoustic study of Thai. In: Thongkum, T., Panupong, V., Kullavanijava, P., Kalaya Tingsabadh, M., (eds), *Studies in Tai and Mon-Khmer phonetics and phonology in honour of Eugenie JA Henderson*. Chulalongkorn University Press: Bangkok 1–9.

[2] Brunelle, M. 2003. Tonal coarticulation in Northern Vietnamese. *Proc. 15th ICPhS* 2673–2676.

[3] Brunelle, M. 2009. Northern and Southern Vietnamese tone coarticulation: A comparative case study. *Journal of Southeast Asian Linguistics* 1(1), 49–62.

[4] Chang, Y.-C., Hsieh, F.-F. 2012. Tonal coarticulation in Malaysian Hokkien: A typological anomaly? *The Linguistic Review* 29, 37–73.

[5] Chao, Y. R. 1933. Tone and intonation in Chinese. *Bulletin of the Institute of History and Philology* 4, 121–134.

[6] Chen, S., Wiltshire, C., Bin, L. 2017. An updated typology of tonal coarticulation properties. *Taiwan Journal of Linguistics* 79–114.

[7] Erickson, D., Baer, T., Harris, K. S. 1983. The role of the strap muscles in pitch lowering. *Vocal Fold Physology. College-Hill press, San Diego*.

[8] Erickson, D., Honda, K., Hirai, H., Beckman, M. E. 1995. The production of low tones in English intonation. *Journal of Phonetics* 23(1-2), 179–188.

[9] Gandour, J. 1994. Tonal coarticulation in Thai. *Journal of Phonetics* 22, 477–492.

[10] Gandour, J., Potisuk, S., Dechongkit, S., Ponglorpisit, S. 1992. Anticipatory tonal coartlculation in Thai noun compounds. *Linguistics of the Tibeto-Burman Area* 15(1), 111–124.

[11] Gandour, J., Potisuk, S., Dechongkit, S., Ponglorpisit, S. 1992. Tonal coarticulation in Thai disyllabic utterances: a preliminary study. *Linguistics of the Tibeto-Burman Area* 15(1), 93–110.

[12] Hallé, P. A. 1994. Evidence for tone-specific activity of the sternohyoid muscle in modern standard Chinese. *Language and Speech* 37(2), 103–123.

[13] Han, M. S., Kim, K.-O. 1974. Phonetic variation of Vietnamese tones in disyllabic utterances. *J. Phonet.* 2, 223–232.

[14] Li, Q., Chen, Y. 2016. An acoustic study of contextual tonal variation in Tianjin Mandarin. *Journal of Phonetics* 54, 123–150.

[15] Peng, S.-h. 1997. Production and perception of Taiwanese tones in different tonal and prosodic contexts. *Journal of Phonetics* 25(3), 371–400.

[16] Potisuk, S., Gandour, J., Harper, M. P. 1997. Contextual variations in trisyllabic sequences of Thai tones. *Phonetica* 54(1), 22–42.

[17] Prom-on, S., Liu, F., Xu, Y. 2012. Post-low bouncing in Mandarin Chinese: Acoustic analysis and computational modeling. *J. Acoust. Soc. Am.* 132(1), 421–432.

[18] Sagart, L., Hallé, P., de Boysson-Bardies, B., Arabia-Guidet, C. 1986. Tone production in modern Standard Chinese: An electromyographic investigation. *Cahiers De Linguistique-Asie Orientale* 15(2), 205–221.

[19] Scholz, F., Chen, Y. 2014. The independent effects of prosodic structure and information status on tonal coarticulation. *Above and beyond the segments: Experimental linguistics and phonetics* 275.

[20] Shen, X. S. 1990. Tonal coarticulation in Mandarin. *J. Phonet.* 18, 281–295.

[21] Shih, C. 1988. Tone and intonation in Mandarin. *Working Papers, Cornell Phonetics Laboratory* 3, 83–109.

[22] Shih, C., Kochanski, G. P. 2000. Chinese tone modeling with Stem-ML. *Sixth International Conference on Spoken Language Processing*.

[23] Shih, C., Lu, H.-Y. D. 2015. Effects of talker-to-listener distance on tone. *Journal of Phonetics* 51, 6–35.

[24] Shih, C., Sproat, R. 1992. Variations of the Mandarin rising tone. *Proc. IRCS workshop on prosody in natural speech* volume 92 193–200.

[25] Sóskuthy, M. 2017. Generalised additive mixed models for dynamic analysis in linguistics: a practical introduction. *arXiv preprint arXiv:1703.05339*.

[26] Sun, Y., Shih, C. 2018. A reexamination of tonal coarticulation: how boundary and tone-specific effects influence anticipatory tonal coarticulation. Manuscript submitted for publication.

[27] Wieling, M. 2018. Analyzing dynamic phonetic data using generalized additive mixed modeling: a tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics* 70, 86–116.

[28] Wong, Y. W. 2006. Contextual tonal variations and pitch targets in Cantonese. *Proc. Speech Prosody* 317–320.

[29] Xu, Y. 1994. Production and perception of coarticulated tones. *J. Acoust. Soc. Am.* 95(4), 2240–2253.

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

[31] Xu, Y. 1999. Effects of tone and focus on the formation and alignment of f0 contours. *Journal of phonetics* 27(1), 55–105.

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

[33] Xu, Y. 2013. ProsodyPro—a tool for large-scale systematic prosody analysis. Laboratoire Parole et Langage, France.

[34] Xu, Y., Sun, X. 2002. Maximum speed of pitch change and how it may relate to speech. *J. Acoust. Soc. Am.* 111(3), 1399–1413.

[35] Zhang, H. 2018. Analyzing Thai tone distribution through functional data analysis. *Proc. Interspeech 2018* 2137–2141.

[36] Zhang, J., Liu, J. 2011. Tone sandhi and tonal coarticulation in Tianjin Chinese. *Phonetica* 68(3), 161–191.

[37] Zhang, J.-S., Kawanami, H. 1999. Modeling carryover and anticipation effects for Chinese tone recognition. *Sixth European Conference on Speech Communication and Technology*.

[1] L in Mandarin is often produced with creaky voice, thus extremely low F0. However, not every subject has this feature so the density at the late portion of L is not as high as that at the early portion.