A GESTURAL ACCOUNT OF MANDARIN TONE 3 VARIATION

Fang-Ying Hsieh

University of Southern California, USA
fangyinh@usc.edu

ABSTRACT

The current study provides a simple and unified account for the Tone 3 variations in Mandarin Chinese by making reference to the organization of articulatory gestures in the framework of Articulatory Phonology. It is proposed that the varying pitch contours of Tone 3 result from changing the coordination among tone gestures in different positions. Two invariant tone gestures: H(igh) and L(ow) are hypothesized to be the basic components of Tone 3 and these two gestures may be produced sequentially or synchronously, thus giving rise to the low-falling-rising, low-falling, and rising pitch contours of Tone 3.

This hypothesis is supported by results from two experiments: (1) in the TADA modeling, the coupling relations among the gestures are specified according to the proposal and the simulated output shows resemblance to the real speech data in many aspects; (2) the hypothesis is also tested in a speech production experiment, in which speakers repeated phrases that include a Tone 3 with the L and H gestures in a sequential mode. Production rate was gradually increased because it has been shown that increasing speech rates could induce gestural synchrony. Results indicate that the H gesture becomes more synchronous with the L gesture as has been predicted by the hypothesis, suggesting the rising contour variant of Tone 3 could have resulted from mode shift of gesture coordination.

Keywords: Articulatory Phonology, gesture coordination, Mandarin tone sandhi

1. INTRODUCTION

Mandarin Chinese has four lexical tones: a high level tone (Tone 1), a mid rising tone (Tone 2), a low falling tone (Tone 3), and a high falling tone (Tone 4). While the other three tones are generally produced with the same pitch contour across contexts, Tone 3 shows several variations depending on its position: Tone 3 with a low falling contour (*low Tone 3*) occurs in non-final position when followed by another tone; but in pre-pausal position it has a low falling-rising pitch contour (*full Tone 3*); and when preceding another Tone 3 in the same domain, it becomes a derived tone 2 rising contour (*sandhi Tone 3*).

This study proposes a simple and unified account for these variations by referring to the control structure of speech, i.e. the organization of articulatory gestures, in the framework of Articulatory Phonology (henceforth AP) [1] and the coupled oscillator model of gestural timing control and syllable structure [4, 7]. It is hypothesized that the alternates of full Tone 3 result from (i) qualitative shift in coupling of tone gestures (sandhi Tone 3) and (ii) overlap of identical tone gestures (low Tone 3).

The first account of Mandarin tones in terms of tone gestures is provided by Gao [3]. She proposed that two invariant tone gestures, targeted as H(igh) and L(ow) pitch respectively, can be used to represent the four Mandarin tones. The tone gestures were viewed as dynamic events that can couple (coordinate) with other gestures, such as vowel and consonant gestures. Based on the alignment patterns among the articulatory gestures of segments and tonal contours, Gao proposed that the tone gestures behave like onset consonants: they are synchronous (in-phase coupled) with the vowels. Her account of Mandarin Tone 3 only concerns the non-final low Tone 3. The current study expands her view on Chinese tones in terms of tone gestures and accounts for all the three variants of Tone 3 as below:

1.1. Full Tone 3

The current study hypothesizes that Tone 3 is composed of sequential L and H tone gestures, which can model the full Tone 3. Unlike the other tonal gestures in Mandarin, the H gesture of Tone 3 is modeled as a coda consonant, i.e. the H gesture anti-phase coupled with the vowel, illustrated in Fig.1, instead of in-phase coupled the way an onset would be (e.g. Tone 4 in Fig. 2, in which both H and L are onsets); and therefore the onset of the H tone is predicted to occur much later with respect to the vowel than, e.g., the onset of the L tone gesture is in Tone 4. This analysis is consistent with the findings that the rising contour is realized much later in the syllable than the falling contour (L gesture) of Tone 4, and that a Tone 3 syllable in isolation is much longer than other tones [9].

Figure 1: Proposed gestural coordination in Tone 3 syllable.

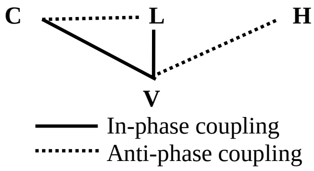
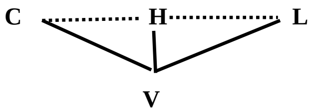


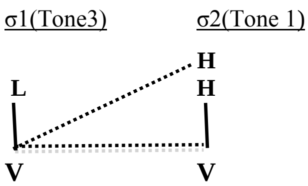
Figure 2: Gestural coordination in Tone 4 syllable following Gao [3].



1.2. Low Tone 3

When Tone 3 precedes another (non-tone 3) tone in non-final position, the coupling relationship of the L and H gestures is hypothesized to remain invariant but the H now overlaps with the H gesture of the following syllable, exemplified in Fig. 3. As the H does not occur within the Tone 3 syllable in such cases, only the low falling contour (low Tone 3) resulting from the L gesture can be perceived. This hypothesis is supported by the neutral tone sandhi rule [2]: when the neutral (targetless) tone syllable follows a Tone 3, its pitch is relatively high, while it is relatively low when preceded by Tone 1, Tone 2, or Tone 4. This high pitch is explained by the proposed analysis: it is the coda H tone, which in non-final position is synchronous with the following syllable.

Figure 3: Coupling graph for [Tone 3 Tone 1].



1.3. Sandhi Tone 3

As Tone 3 is the only tone that does not have an H in the onset, c.f. [3], when Tone 3 precedes another Tone 3, the H gesture of the preceding Tone 3 does not have an immediately following H to overlap with and instead, it is hypothesized to undergo a qualitative shift to the more stable in-phase coupling with the vowel, resulting in a pitch contour similar to Tone 2, which is composed of synchronous L and H in Gao's account. The derived tone 2 therefore results from shifting the sequential relationship between L and H to be synchronous. The proposed coupling graph for the sandhi Tone 3 thus will be similar to Fig. 1, but with an additional in-phase coupling relation between the L and H. Such shifts in coupling to the most stable in-phase mode have been shown to underlie the aspiration of stops following [h] in Western Andalusian Spanish [8] and the re-syllabification of coda consonants to onsets in VCV contexts [6].

To test the proposed account, two types of experiments are conducted in this study: (1) TADA (TAsk-Dynamic Application) modeling, in which Tone 3 bearing syllables were modeled by specifying in the input the inter-gestural coordination according to the proposed account. (2) A speech production experiment was further conducted to test the proposal of sandhi Tone 3, in which the production rate was gradually increased. Such a rate control has been shown to potentially destabilize a grammatical anti-phase mode and shift the system to an in-phase mode, as shown by [8]. The proposed hypothesis accordingly predicts that as the speech rate increases, the anti-phase coupled H in a full Tone 3 will shift to be in-phase coupled with the L gesture in the syllable, resulting in sandhi Tone 3.

2. MODELING

In this computational simulation experiment, artificial speech outputs synthesized from TADA and HLsynTM [5] are compared with real speech data. The TADA takes as input the coupling relations among the planning oscillators (clocks) that trigger the articulatory gestures and HLsyn generates acoustic signals using constriction and articulator time functions from TADA.

2.1. Method

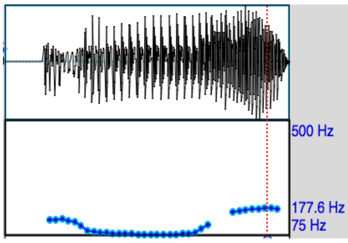
The three variants of Tone 3 were simulated on [ma] syllables. The gestural scores and coupling graphs for [ma] were automatically generated by the system based on generalizations of gestural control regimes and Syllable-based Coupling Principles accumulated from the theoretical and empirical studies in AP. But the tone gestures and the coupling of their oscillators were manually specified in the input for simulation since this aspect has not been specified in the present system. Following Gao [3], tone gestures are assumed to have the same natural frequency as the consonants. The H and L tone gestures share the same task variable: "F0," but they differ in their targets.

2.2. Results and discussion

2.2.1. Modeling full Tone 3

In the current proposal, full Tone 3 has an L gesture in the onset and an H gesture in the coda. Following the coupling specifications for onset and coda consonants in the TADA system, the relative phasing between the onset of the vowel gesture and the onset of the L gesture was specified as in-phase coupling, and that between the onset of vowel gesture and the onset of the H gesture was specified as anti-phase (c.f. Fig. 1). The resulting synthetic output (Fig. 4) exhibits a rising contour around the middle of the syllable, similar to the real speech data. Besides, the synthesized full Tone 3 has a much longer duration (384 ms) than other tones (254-285 ms) that were synthesized according to Gao's proposal, which also accords with the real speech data.

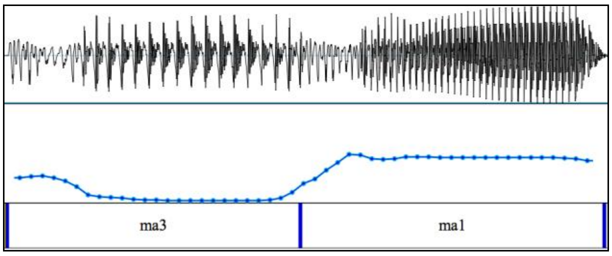
Figure 4: Simulated full Tone 3 [ma].



2.2.2. Modeling low Tone 3

The low Tone 3 was simulated on a disyllabic sequence [ma3.ma1] since its H gesture has been proposed to realize on the following syllable. The tonal gestures of Tone 3 and Tone 1 were specified for the two syllables respectively following the parameters used for monosyllabic full Tone 3 and monosyllabic Tone 1, with the relative phrasing between the two vowels in the two syllables being anti-phase (c.f. Fig. 3). As shown in Fig. 5, the Tone 3 syllable only contains a low tone in the synthesized output.

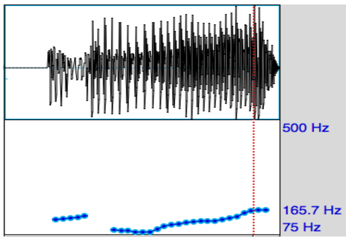
Figure 5: Low Tone 3 [ma] followed by Tone 1 [ma].



2.2.3. Modeling sandhi Tone 3

The sandhi Tone 3 was modeled by adding an in-phase coupling relationship between the L and H gestures to the coupling specifications for the full Tone 3, as proposed in 1.3. The result is illustrated in Fig. 6, which shows a rising contour, similar to an underlying Tone 2.

Figure 6: Simulated sandhi Tone 3 [ma].



As discussed above, the simulated results show a great amount of similarity to the real speech data. Thus results from this computational study support the proposed account to be an adequate model of Mandarin Tone 3 variations. To directly observe how this model could give rise to the variations in real speech, a production experiment is conducted focusing on the formation of sandhi Tone 3, which has been hypothesized to involve addition of an in-phase coupling relationship between L and H. Since the H in full Tone 3 is only anti-phase coupled to the vowel, it is predicted that under rate increase, the sequential L and H in full Tone 3 will shift to the more stable in-phase mode.

3. PRODUCTION EXPERIMENT

3.1. Method

Four native Mandarin speakers from China repeated pairs of disyllabic phrases that include a pre-pausal Tone 3 followed by another Tone 3 across an IP boundary (T3 # T3, e.g. *yan1yu3* "misty rain" # *yan3lian2* "eyelids"). The sandhi rule does not apply across the two Tone 3's in this environment. To determine whether Tone 3 in this environment is shifting in the direction to become a derived Tone 2, pairs of phrases containing a pre-pausal Tone 2 followed by a Tone 3 across the boundary were also included (T2 # T3). A pre-pausal Tone 3 followed by a Tone 2 (T3 # T2) is another type of test stimuli that serves as a comparison. The hypothesis predicts that the H gesture in such a Tone 3 will overlap with the H gesture in the following syllable, leaving the L gesture in the Tone 3 unaffected. These pairs of phrases were produced in five blocks, across which speech rate was gradually increased: Block 1 the slowest, Block 5 the fastest.

To measure how the L gesture is affected in the pre-pausal tone syllable (the target syllable), the word preceding the target was designed to end in an H gesture in the stimuli. Given that both Tone 2 and Tone 3 of the target syllable have an L gesture in the onset (though the L in Tone 2 is synchronous with an H), F0 will fall from the H in the preceding syllable into the target syllable. The difference between the F0 peak in the preceding syllable and the F0 valley in the target syllable (F0 fall) is measured, which will be greater for Tone 3 than for Tone 2 when the L gesture in Tone 3 does not overlap with an H gesture and thus can achieve its target. But if the coda H gesture advances with the speech rate, the Tone 3 F0 fall should be reduced in the T3 # T3 condition. While other tone combinations (T2 # T3, T3 # T2) could also show a reduced F0 fall, due to undershoot of the L in the target syllable as a function of rate, the effect should not be as large as T3 # T3, where phase shift is also predicted to be going on. Thus the correlation between speech rate and F0 fall and the regression slope should be greater for T3 # T3 than for other tone combinations.

3.2. Results and discussion

The correlations and regression slopes for these three tone combinations are shown in Table 1. The results support the hypothesis that the relation between F0 fall and speech rate has a higher correlation (R) and regression slope (b) than other tone combinations. To test whether Tone 3 (# T3) becomes more similar to Tone 2 (# T3) in F0 fall as speech rate increases, the collected data for these two tone combinations were divided into three speech rate groups (slow, medium, and fast) and a two-way ANOVA F test was conducted. The results suggest that there is an interaction between the tone combination type and speech rate ($p < 0.05$). The differences between Tone 3 (# T3) and Tone 2 (# T3) in F0 fall are reduced at fast rate, compared to slow, indicating Tone 3 (# T3) is changing in the direction to become similar to Tone 2 as predicted. Nevertheless, comparing the F0 fall in Tone 2 and Tone 3 at the fast speech rate using a one-way ANOVA test shows that Tone 2 and Tone 3 still remains significantly different from each other in the current study ($p < 0.05$), implying there is no qualitative shift from Tone 3 to Tone 2 in this experiment.

Table 1: Correlations and slopes.

	T2 # T3	T3 # T2	T3 # T3
R	0.42	0.53	0.66
b	0.31	0.50	0.63

4. CONCLUSION

As shown above, predictions of the proposed account are generally borne out in the results from both TADA modeling and the speech production experiment. But the result from the production experiment indicates that there is no qualitative shift: Tone 3 remains to be distinct from Tone 2. Therefore, results of this experiment provide suggestive evidence that Tone 3 sandhi could have had its historical origins in a destabilization of the anti-phase mode at fast rates.

5. REFERENCES

[1] Browman, C.P., Goldstein, L. 1992. Articulatory Phonology: An overview. *Phonetica* 49, 155-180.

[2] Chao, Y.R. 1968. *A Grammar of Spoken Chinese.* Berkeley, CA: University of California Press.

[3] Gao, M. 2008. *Mandarin Tones: An Articulatory Phonology Account.* Doctoral Dissertation, Yale University.

[4] Goldstein, L.M., Byrd, D., Saltzman, E. 2006. The role of vocal tract gestural action units in understanding the evolution of phonology. In Arbib, M. (ed.), *From Action to Language: The Mirror Neuron System.* Cambridge: Cambridge University Press, 215-249.

[5] Hanson, H., Stevens, K.N. 2002. A quasi-articulatory approach to controlling acoustic source parameters in a Klatt-type formant synthesizer using HLSyn. *Journal of the Acoustical Society of America* 112, 1158-1182.

[6] de Jong, K.J., Okamura, K., Lim, B.J. 2003. The phonetics of resyllabification in English and Arabic speech. In Sole, M.J., Recasens, D., Romero, J. (eds.), *Proc. 15th ICPhS* Barcelona, 2621-2624.

[7] Nam, H. 2007. Syllable-level intergestural timing model: Split-gesture dynamics focusing on positional asymmetry and moraic structure. In Cole, J., Hualde, J.I. (eds.), *Laboratory Phonology 9 (Phonology and Phonetics)*. Berlin, New York: Walter de Gruyter, 483-506.

[8] Parrell, B. Forthcoming. The role of gestural phasing in Western Andalusian Spanish aspiration. *Journal of Phonetics*.

[9] Xu, Y. 1997. Contextual tonal variations in Mandarin. *Journal of Phonetics* 25, 61-83.