INTERSPEECH 2016
September 8–12, 2016, San Francisco, USA

Interaction between lexical tone and intonation: an EMA study

Hao Yi, Sam Tilsen

Department of Linguistics
Cornell University, Ithaca, New York

hy433@cornell.edu, tilsen@cornell.edu

Abstract

This paper aims to examine the interaction of intonation and lexical tone within the framework of Articulatory Phonology, by investigating the timing relationship between oral articulatory gestures and tone-related/intonation-related F0 dynamics. Specifically, we compared the consonant-vowel-F0 (C-V-T) coordinative patterns at phrase-final position and at phrase-medial position. We found that the C-V-T coordination was altered by the presence of boundary tones, which is in line with the sequential model in which tone and intonation are conceptualized as events that interact at the phonological level before the phonetic implementation. However, the effect of boundary tone on the C-V-T coordination seemed to be tone-specific. Moreover, the presence of pitch accents also influenced the intra-syllabic C-V-T coordinative patterns. By presenting evidence from the coordinative patterns between articulatory gestures and F0 dynamics, the current study lent support to the sequential model of the interaction between intonation and lexical tone from a gestural perspective.

Index Terms: interaction, lexical tone, intonation, boundary tone, pitch accent, gestural coordination, electromagnetic articulography

1. Introduction

Both intonation and lexical tone make use of F0 variation contrastively: intonation is used for the expression of discourse meaning and for marking phrases; lexical tone is used contrastively for lexical or grammatical meaning [1]. Two categories of models have been proposed to account for the interaction between lexical tone and intonation: the overlay model and the sequential model [2, 3]. The overlay model characterizes the overall F0 contour as the result of local perturbations for lexical tone superimposed onto the global intonation, while the sequential model treats intonation as interacting locally with lexical tone (the targets can be a mixture of lexical tones and prosodic tones). The overlay model supports the notion that intonation and lexical tones are encoded and implemented in a parallel fashion, consistent with the lexical/post-lexical distinction, while the sequential model holds that intonation and lexical tone interact at the same phonological level before the sequential phonological output is interpreted to generate acoustic output. The current study aims to investigate the interaction between intonation and lexical tone from the perspective of gestural coordination between oral articulatory gestures and F0 dynamics.

Previous research has shown that components of intonation, e.g., pitch accents, phrase accents, and boundary tones, are consistently timed with acoustic landmarks of consonants and vowels [4, 5, 6]. The alignment between lexical tones and acoustic landmarks of segments has also been shown to be consistent [7, 8]. Importantly, acoustic-based findings like these, which associate the offsets of the F0 movement with the acoustic landmarks of segments, can be interpreted differently from articulatory-based findings, which often associate the onsets of the F0 movement with the articulatory landmarks of segments.

Conceptualizing the control of F0 to reach a target as a tone gesture (both prosodic and lexical), Articulatory Phonology analyzes the temporal coordinative patterns between tone gestures (H and L gestures) and constriction gestures (C and V gestures) using a coupled-oscillator planning model of speech timing [9, 10, 11].

Articulatory-based alignment between tonal gestures and constriction gestures appears to more stable and synchronous than acoustic-based alignment. For example, it was found that in Catalan, the H gesture of a bitonal pitch accent, L+H, is in-phase coupled with the accented V gesture, and thus the onset of H gesture coincides with the onsets of constriction gestures for the consonant and the vowel [12]. It was suggested that the addition of tonal gestures coupled only to the V gesture produces no effect on the C-V coordination, in which the C gesture and the V gesture begin synchronously. Similarly, boundary tone gestures do not alter the C-V coordination. For example, it was found that in Greek, the boundary tone gesture is anti-phase coupled with the V gesture of the phrase-final syllable, and that the C-V coordination remains constant regardless of the presence of boundary tones [13].

In Mandarin, lexical tone gestures behave as onset C gestures in that they are in-phase coupled with the V gesture of the tone-bearing syllable and anti-phase coupled with the onset C gesture, resulting the C-center effect (lexical tone gesture is a C gesture): in an H-tone-bearing syllable, the V gesture is initiated halfway between the C gesture and the H gesture [14, 15]. Mandarin tones thus contrast with Catalan pitch accents and Greek boundary tones in the sense that lexical tone gestures appear to interact with the within-syllable oral articulatory gestures, whereas pitch accent gestures and boundary tone gestures do not. It was argued that lexical tone gestures are fully integrated into the syllabic coupling network, whereas pitch accent gestures and boundary tone gestures are not coupled with the consonant gestures, and are therefore less tightly integrated into the syllabic coupling network [12, 13].

The distinction between the the overlay model and the sequential model can be investigated by drawing evidence from gestural coordination between oral articulatory gestures and tone-related and intonation-related F0 gestures. The current study approaches this issue using two intonation types, i.e., STATEMENT and QUESTION, to elicit boundary tones at phrase-final position. A STATEMENT elicitation renders an L boundary tone (L%) and a QUESTION elicitation renders an H boundary

http://dx.doi.org/10.21437/Interspeech.2016-662

tone (H%), respectively. We further measure the alignment pattern (C-V-T) between the F0 gestures and articulatory gestures. If the overlay model holds, it will predict that the presence of boundary tones will not alter the intra-syllabic C-V-T coordinative patterns; if the sequential model holds, it will predict that the C-V-T coordinative patterns at phrase-final position and at phrase-medial position differ due to the presence of boundary tones at phrase-final position.

Hypothesis 1 (Overlay model) Intonation and lexical tone are encoded and implemented in a parallel fashion.

Prediction 1 The C-V-T coordination at phrase-final position does not differ from that at phrase-medial position.

Hypothesis 2 (Sequential model) Intonation and lexical tone interacts at the same phonological level before the phonetic implementation.

Prediction 2 The C-V-T coordination at phrase-final position differs from that at phrase-medial position.

2. Methods

2.1. Participants

Three female participants who are native speakers of Beijing Mandarin participated in this experiment. They were born and raised in Beijing, and were graduate students at Cornell University at the time of recording. The participants were naïve to the purpose of the study. They gave informed consent and received financial compensation for their participation. The experiment took place in Cornell Phonetics Lab in the Department of Linguistics at Cornell University.

2.2. Stimuli

The target syllable [ma] had either a rising tone (Tone2) or a falling tone (Tone4). It was immediately preceded by a syllable bearing the same tone. The target syllable [ma] was further embedded in three carriers: phrase-medial (MEDIAL), de-accented phrase-final (FINAL), and accented phrase-final (ACCENTED). Each carrier can end either with a question mark, indicating a QUESTION, or with a period, indicating a STATEMENT. A QUESTION elicitation and a STATEMENT elicitation only differed in the punctuation at the end of the sentence. High boundary tones (H%) and low boundary tones (L%) were elicited with QUESTION and STATEMENT, respectively. Boundary tones were only elicited in phrase-final carriers (FINAL and ACCENTED).

The target sentences were presented on a monitor approximately 1.5 m away from the participant. The experiment was organized into blocks of 16 unique trials in random order. Each speaker completed approximately 16 blocks.

2.3. Data processing and analysis

Articulatory data were collected by an NDI WAVE Electromagnetic Articulography (EMA). Acoustic data were simultaneously collected at a sampling rate of 22.5 kHz. Kinematic and F0 trajectories were extracted using MATLAB. The articulatory gestures that were involved in the target syllable [ma] were a bilabial closure of [m] and tongue root retraction of [a]. The bilabial closure was determined according to lip aperture (LA), the vertical distance between upper lip and lower lip; the tongue root retraction was determined according to tongue body height (TBy), the vertical displacement of tongue body. The F0 trajectories were extracted using the normalized cross-correlation pitch tracking algorithm developed in [16].

Carrier Target	(Background) Stimuli
MEDIAL	(–) (–)
Tone2	lu^2 yan^2 yi^2 ma^2 yi^2 de hen^3 kuai^4? ‘Lu Yan moves ma^2 very fast?’
FINAL	(bu^2 shi^4 luo^2 yan^2? bu^2 shi^4 li^3 yan^2?) ‘(Not Luo Yan? Not Li Yan?)’
Tone2+H%	**lu**^{2} yan^2 yao^4 yi^2 ma^2? ‘**Lu** Yan will move ma^2?’
ACCENTED	(bu^2 shi^4 ma^3? bu^2 shi^4 ma^4?) ‘(Not ma^3? Not ma^4?)
Tone2+H%	lu^2 yan^2 yao^4 yi^2 **ma**^{2}? ‘Lu Yan will move **ma**^{2}?’

Table 1: QUESTION elicitations of Tone2-bearing target syllables in three different carriers: phrase-medial (MEDIAL), de-accented phrase-final (FINAL), and accented phrase-final (ACCENTED). Target syllables are underlined. Accented syllables are in bold. Background information are provided for FINAL and ACCENTED elicitations. Out of the four combinations (Tone2 with QUESTION, Tone2 with STATEMENT, Tone2 with QUESTION, Tone4 with STATEMENT), only Tone2 with QUESTION elicitations are exemplified here due to space limitations. A QUESTION elicitation and a STATEMENT elicitation only differ in the punctuation at the end of the sentence.

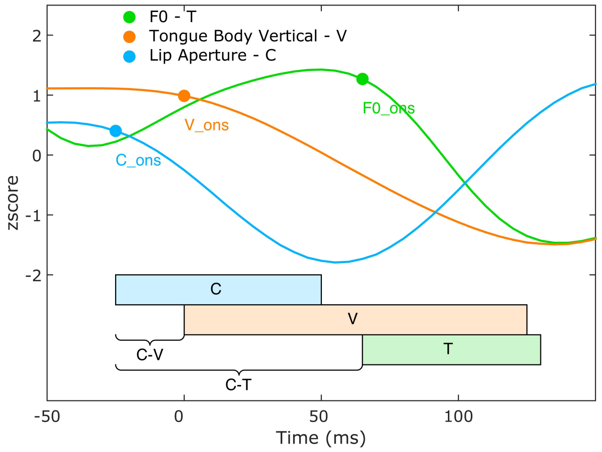

Figure 1: Normalized trajectories of lip aperture (LA), tongue body height (TBy), and F0 for a STATEMENT elicitation of a Tone2-bearing syllable in MEDIAL. The target syllable is preceded by a Tone2-bearing syllable. The C, V, and T gestures are determined according to tract variables LA (blue), TBy (orange), and F0 (green), respectively. The LA$_{\text{ons}}$, TBy$_{\text{ons}}$, and F0$_{\text{ons}}$ respectively mark the onset of the C, V, and T gestures (horizontal bars at the bottom). Onset lags between C and V, and between C and T are recorded.

As shown in Figure 1, for each articulatory trajectory of interest (i.e., LA and TBy), a corresponding velocity profile was computed to determine the articulatory landmarks: minimum velocity, onset, and peak velocity. Specifically, the onset was defined as the point when 30% of the velocity range between the minimum velocity and the peak velocity had passed. The F0 landmarks of were consistently defined in the same way as kinematic landmarks. Importantly, the onset of a high F0 gesture was defined with reference to the preceding F0 minimum, and the onset of a low F0 gesture was defined with reference to the preceding F0 maximum.

Temporal lags between onsets were calculated from the landmarks, as shown at the bottom of Figure 1. The C-V lag is defined as the temporal lag between the onset of the C gesture and the onset of the V gesture; the C-T lag is defined as the temporal lag between the onset of the C gesture and the onset of the T gesture. The phase of the V gesture relative to the C-V-T lag was further computed for each trial (CV% = C-V lag / C-V-T lag × 100%). The smaller the CV%, the closer the C gesture initiation and the V gesture initiation are; the larger the CV%, the closer the V gesture initiation and the T gesture initiation are.

3. Results

Figure 2 shows the mean F0 contours for the target syllables (bearing either Tone2 or Tone4) in different environments. Unsurprisingly, F0 contours in Question were higher than those in Statement. The difference in F0 between Question contours and Statement contours was of greater magnitude at phrase-final position than at the phrase-medial position, and was the most pronounced in Final. Interestingly, such difference was also present in Medial, even though the target syllable was four syllables away from the end of the carrier sentence. Moreover, this difference was also more pronounced for Tone4 than for Tone2.

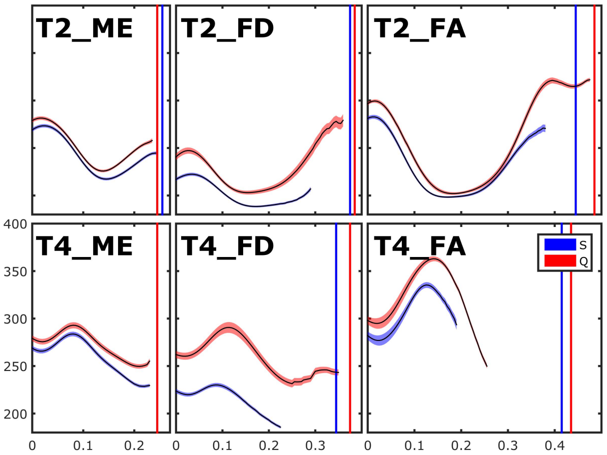

Figure 2: Mean F0 contours $+/-$ 1.0 standard error for the target syllables. Tone2 is shown in the upper panel and Tone4 the lower panel. Medial (ME), Final (FD), and Accented (FA) carriers are shown in left, middle, and right column, respectively. In each subfigure, Statement and Question are coded in red and blue, respectively. Time 0 corresponds the acoustic onset of the target syllable. Vertical blue lines indicate the acoustic offset for each target syllable in Statement, and red each target syllable in Question.

Figure 3 shows the CV% for the target syllables in 12 environments (two Tones × three Carriers × two Intonations).

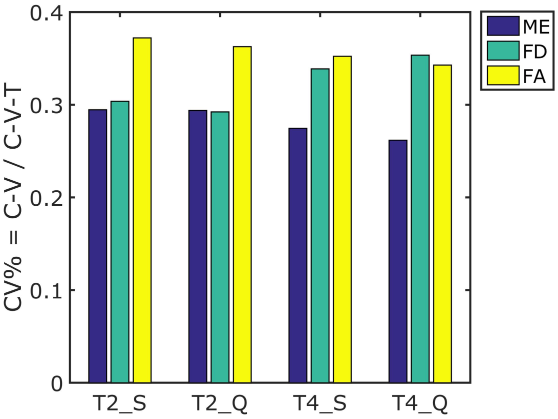

Figure 3: The CV% for the target syllables bearing Tone2 and Tone4 elicited with Statement and Question in Medial (ME), Final (FD), and Accented (FA).

The CV% was submitted to a three-way ANOVA having two levels of Tone (Tone2 and Tone4), three levels of Carrier (Medial, Final, and Accented), and two levels of Intonation (Statement and Question). Interaction between the main effect variables were also included (Table 2).

There was a significant main effect for Carrier, $F(1, 643) = 0.28$, $p < 0.0001$, indicating the C-V-T coordinative patterns in the three carriers were significantly different. A post-hoc Tukey HSD test further showed that the CV% difference between any two different carriers was significantly different. Specifically, the V onset was closer to the C onset in Medial than in Final than in Accented.

Source	d.f.	F	p-value
Tone	1	0.01	0.9130
Intonation	1	0.28	0.5980
Carrier	**1**	**24.87**	**0.0000*** **
Tone×Intonation	1	0,06	0.8099
Tone×Carrier	**2**	**7.25**	**0.0008*** **
Intonation×Carrier	2	0.15	0.8648
Error	634		

Table 2: Three-way ANOVA on the CV% showed a significant main effect for Carrier and a significant interaction effect between Tone and Carrier.

The interaction effect between Tone and Carrier was significant, $F(2, 643) = 0.09$, $p < 0.001$. There was no significant main effect for Intonation. For Tone2, regardless of the Intonation, the CV% in Final patterned with that in Medial, differing from that in Accented, while for Tone4, regardless of the Intonation, the CV% in Final patterned with that in Medial, differing from that in Medial. For both Tone2 and Tone4, the difference in CV% between Medial and Accented was significant.

Two three-way ANOVAs were conducted separately on the C-V and the V-T onset lags. The results showed that for both the C-V lag and the V-T lag, there was a significant main effect for Carrier and a significant interaction effect for Tone × Carrier, consistent with the ANOVA on the CV%. A series of post-hoc Tukey HSD tests showed that for Tone2, the C-V lag in Final was not different from that in Medial, but was significantly shorter than that in ACCENTED; for Tone4, the C-V lag in FINAL was not different from that in ACCENTED, but was significantly longer than that in MEDIAL. In terms of the V-T lag, for Tone2, MEDIAL had a significant longer lag than ACCENTED, but not FINAL; for Tone4, MEDIAL had a significantly shorter lag than both FINAL and ACCENTED.

4. Discussion

In this section, we will argue that the evidence from the C-V-T coordinative patterns is in line with the sequential model which allows for the local interaction between intonation and lexical tone.

The results can be summarized as:

- There was no significant main effect for Intonation or significant interaction effect between Intonation and Tone on the CV%.
- There was a significant main effect of Carrier. The CV% in MEDIAL was significantly smaller than that in ACCENTED.
- There was also a significant interaction effect between Tone and Carrier on the CV%.

The observation that the intonation type (STATEMENT and QUESTION) has no effect on the C-V-T coordinative patterns indicates that either both H% and L% are present or neither H% nor L% is present. Taking into consideration the evidence that for Tone4, the C-V-T coordinative patterns in MEDIAL and in FINAL were significantly different, we argue that both H% and L% are present. The presence of boundary tones is also supported by the interdependence between boundary tones and boundary lengthening: it was found that boundary tones are triggered when the boundary lengthening gesture, i.e., π-gestures reach a specific high-level of activation in Greek [13]. Moreover, there was no significant tone-specific Intonation effect, which calls for a unified boundary tone gesture for both H% and L% boundary tones, which is consistent with the observation that the boundary tone gesture coordination patterns do not vary across boundary tone types (L%, H%, and !H%) in Greek [13].

Taking the presence of boundary tones at phrase-final position as the point of departure, we further argue that consistent with the sequential model, boundary tone, at least in Mandarin, interacts with lexical tone locally before the phonetic implementation takes effect.

The argument seems counterfactual given that for Tone2, no statistical difference in the CV% was detected between MEDIAL and FINAL. (This is in line the overlay model, because it predicts that intonation events, due to their post-lexical status, have no effect on the C-V-T coordinative patterns.) However, for Tone2, the significant difference in the C-V-T coordination between MEDIAL and ACCENTED suggests the opposite. In ACCENTED, the presence of the pitch accents altered the C-V-T coordinative patterns by drawing the V gesture onset closer to the T gesture onset. This is true for both Tone2 and Tone4. Like boundary tone, pitch accent is also post-lexical. Therefore, the notion that intonation is encoded in parallel with lexical tone on the basis of the lexical/post-lexical distinction could not be entertained. Moreover, for Tone4, the evidence that the CV% in MEDIAL was significantly smaller than that in FINAL further suggests that the C-V-T coordination was affected by the presence of boundary tones, regardless of the intonation type. Therefore, we argue for the sequential model to account for the intonation-lexical tone interaction. However, the current study cannot provide a concrete explanation for the tone-specific Carrier effect in the sense that the boundary tone gestures seem to interact with Tone2 differently than with Tone4.

To account for the observation that the CV% in ACCENTED was significantly higher than that in MEDIAL, we argue that the pitch-accent gesture, i.e., μ-gesture is triggered by the focus-introduced pitch accent. The μ-gesture, behaving like a lexical tone gesture, is in-phase coupled to the V gesture and anti-phase coupled to the onset C gesture. This renders a stronger coordination between the V gesture and the lexical tone gesture (and the μ-gesture), attracting the V gesture towards the tone gestures, accounting for the higher CV% in ACCENTED (Figure 4). Importantly, as argued earlier, this corroborates the claim that intonation interacts with lexical tone at the phonological level.

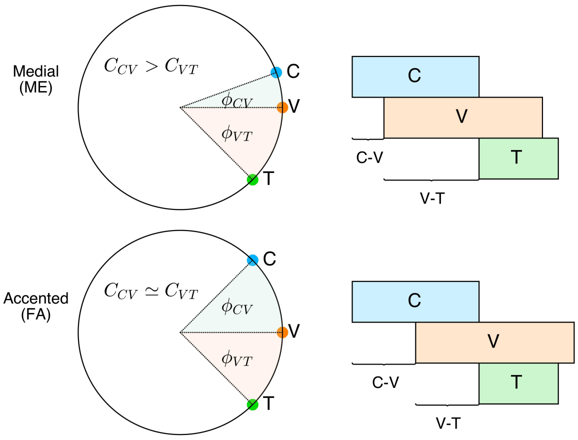

Figure 4: Coupled oscillator graphs (left) and gestural scores (right) in MEDIAL (top) and ACCENTED (bottom). In ACCENTED, the presence of the pitch accent increases the coupling strength between the V gesture and the T gestures, drawing the V gesture closer to the T gestures, thereby increasing the CV% in ACCENTED.

Last but not least, the tone-specific CV% difference in FINAL between Tone2 and Tone4 was neutralized in ACCENTED. The neutralization can be attributed to: 1) the coordination between the μ gesture and the V gesture is stronger than the coordination between the boundary tone gesture and the V gesture, and 2) the boundary tone gesture is integrated into the μ-gesture in ACCENTED by virtue of the saturation effect.

5. Conclusion

The current study investigated the interaction between intonation and lexical tone from a gestural perspective. We found that the C-V-T coordination was affected by the presence of both boundary tones and pitch accents. We argue that this piece of articulatory evidence regarding the relative timing of gestural activation is in line with the sequential model in which intonation interacts with lexical tone locally. Future studies should look into the tone-specific interaction between boundary tones and lexical tones that rendered the coordinative difference between Tone2 and Tone4 at de-accented phrase-final position.

6. References

[1] C. Gussenhoven, *The Phonology of Tone and Intonation*. Cambridge University Press, 2004.

[2] D. R. Ladd, *Intonational Phonology*, 2nd ed. Cambridge University Press, 2008.

[3] M. Gibson, "Lexical tone, intonation, and their interaction: a scopal theory of tune association," Ph.D. dissertation, Cornell University, 2013.

[4] P. Prieto, J. van Santen, and J. Hirschberg, "Tonal alignment patterns in Spanish," *Journal of Phonetics*, 1995.

[5] P. Prieto, M. D'Imperio, and B. G. Fivela, "Pitch accent alignment in Romance: primary and secondary associations with metrical structure," *Language and Speech*, vol. 48, no. 4, pp. 359–396, 2005.

[6] A. Arvaniti, D. R. Ladd, and I. Mennen, "What is a starred tone? Evidence from Greek," in *Papers in Laboratory Phonology V: Acquisition and the Lexicon*, M. Broe and J. Pierrehumbert, Eds. Cambridge: Cambridge University Press, 2000, pp. 119–131.

[7] Y. Xu, "Consistency of tone-syllable alignment across different syllable structures and speaking rates," *Phonetica*, vol. 55, pp. 179–203, 1998.

[8] B. Morén and E. Zsiga, "The lexical and post-lexical phonology of Thai tones," *Natural Language & Linguistics Theory*, vol. 24, pp. 113–178, 2006.

[9] C. P. Browman and L. Goldstein, "Articulatory gestures as phonological units," *Phonology*, vol. 6, pp. 201–251, 1989.

[10] ——, "Articulatory phonology: An overview," *Phonetica*, vol. 49, pp. 155–180, 1992.

[11] H. Nam and E. Saltzman, "A competitive, coupled oscillator model of syllable structure," in *Proceedings of the 15th International Congress of Phonetic Sciences*, 2003.

[12] D. Mücke, H. Nam, A. Hermes, and L. Goldstein, "Coupling of tone and constriction gestures in pitch accents," in *Consonat clusters and structural complexity*, P. Hoole, L. Bombien, M. Pouplier, C. Moonshammer, and B. Kühnert, Eds. Berlin: Mouton de Gruyter, 2012, pp. 205–230.

[13] A. Katsika, J. Krivopapić, C. Moonshammer, M. Tiede, and L. Goldstein, "The coordination of boundary tones and its interaction with prominence," *Journal of Phonetics*, vol. 44, pp. 62–82, 2014.

[14] M. Gao, "Mandarin tones: An articulatory phonology account," Ph.D. dissertation, Yale University, 2008.

[15] H. Yi and S. Tilsen, "Gestural timing in Mandarin tone sandhi," in *Proceedings of Meetings on Acoustics*, vol. 22, 2015.

[16] D. Talkin, "A robust algorithm for pitch tracking (RAPT)," *Speech conding and synthesis*, vol. 495, 1995.