The temporal organization between initial consonants and lexical tones

by

Alif Silpachai

First Screening Paper (MA thesis)

Submitted on April 1, 2014

Defended on April 14, 2014

Revised on May 3, 2014

Minor edits made (with permission from Khalil Iskarous) on November 5, 2020[1]

Committee:

Khalil Iskarous (chair)

Louis Goldstein

Karen Jesney

University of Southern California

Los Angeles, California

2014

[1] "Initial consonant" was changed from an indefinite mass noun to a count noun throughout the paper. These changes were the only edits.

1. Introduction

This study addresses the question of how speakers coordinate articulatory movements. In particular, it considers the temporal organization between initial consonants and tones within a syllable. Previous studies have focused on the organization between consonants and vowels in non-tonal languages, but little is known about the organization between segments and lexical tones. Therefore, the focus of this experiment is the investigation of the temporal organization between initial consonants (Cs) and lexical tones (T). It investigates the evidence of the so-called C-center effect of segmental and tonal sequences in the onset position of a syllable which has been hypothesized for Mandarin by Gao (2008, 2009). The exhibition of the C-center effect in her studies seems to contradict the findings from Morén & Zsiga (2006) for Thai, a language with more lexical tones and syllable types than Mandarin. However, it is unclear because previous studies did not directly investigate the temporal organization between segments and lexical tones. The present study therefore examines the temporal organization between initial Cs and Ts in Thai. The experiment described below uses acoustic aspects of C and Ts from Thai. This paper assumes the gestural theory which is grounded in Articulatory Phonology (Browman & Goldstein, 1986, 1988). It will be argued that Thai does not exhibit any C-center effect in the organization between segments and lexical tones. The present study conjectures that the observation that Mandarin exhibits this effect in previous studies may perhaps have been influenced by the peak delay of preceding[2] high tones and other factors.

1.1 C-center in a non-tonal language: evidence from American English

[2] "Preceding syllable" or "preceding tone" in the present study specifically refers the syllable or tone that immediately precedes target syllables.

The C-center effect was originally observed by Browman & Goldstein (1988, 2000) in the temporal organization between C and V(owel). The effect is basically that the midpoint of a C cluster (defined as "C-center") is produced synchronously with the V gesture. In other words, the beginning time of the V is aligned to the midpoint of a C cluster. This effect was first observed in American English (Browman & Goldstein, 1988) from X-ray microbeam data. The effect has since been the main subject of a number of later studies on the organization of consonant clusters. (e.g. Byrd, 1995; Goldstein, Chitoran, & Selkirk, 2007; Honorof & Brownman, 1995; Marin & Pouplier, 2008, 2014; Nam, Goldstein, & Saltzman, 2009; Shaw, Gafos, Hoole, & Zeroual, 2009) For American English, the production of sequences like #pV, #sV, #lV, #spV, #plV, and #splV, where # is a word boundary, were examined, and it was concluded that a single initial C gesture and sequences of initial Cs are timed globally with the following V gesture by phasing their C-center to the V gesture. The relative beginning times in #CV and #CCV are schematized in Figure 1a and b.

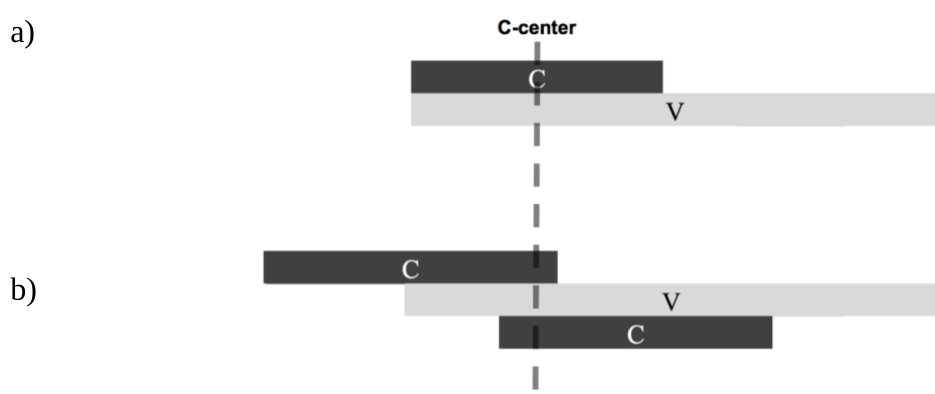


Fig. 1. (a) Hypothesized temporal organization between C gesture and V gesture in #CV. (b) Hypothesized temporal organization between CC and V in #CCV.

As can be seen in Fig. 1, when there is only one C gesture, it begins at the same time as the V gesture (a). However, as can be seen in (b), when there is another C gesture, i.e. when there is a consonant cluster, the two C gestures become spaced out with respect to the V gesture—C_1 begins earlier than the V gesture, and the V gesture begins earlier than C_2; the V gesture begins approximately in the midpoint of the two C gestures.

Browman & Goldstein (2000) proposed that the temporal organization of segmental sequences can be explained in terms of coupling relations under the coupled oscillator planning model of speech production, developed by Browman & Goldstein (2000), Goldstein *et al.* (2006), Goldstein *et al.* (2009), Nam *et al*., (2009), and Nam & Saltzman (2003) within the framework of Articulatory Phonology. Within this framework, speech can be decomposed into phonological units, or *gestures*, and the gestures are temporally coordinated with each another (Browman & Goldstein, 1989). In the planning model, gestures are associated with nonlinear planning oscillators (or *clocks*), and the clocks are coupled with each other in a pattern specified by a *coupling graph*. The C and the V gestures in #CV, as in words like *paid*, are assumed to be in in-phase mode since they begin at the same time, whereas the C gestures in #CC, as in words like *spayed*, are assumed to be in anti-phase mode since they do not begin at the same time. #CCV then can be understood as a consequence of phase relations competing. In other words, the observed C-center effect can be thought as the result of the competition between the coupling of the two C gestures and the coupling of the C and the V gestures, as shown in Fig. 2. The coupling graphs in Fig. 2 can account for why the two C gestures become spaced out from each other and why the V gesture begins at the midpoint of the two C gestures.

—— in-phase coupling mode
------ anti-phase coupling mode

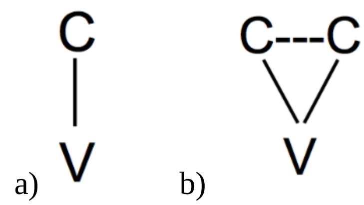


Fig. 2. Specified coupling relations by coupling graphs for #CV (a) and #CCV "competitive coupling" (b)

1.2 C-center in a tonal language: evidence from Mandarin

Gao (2008, 2009) examined the production of lexical tones in #CV and #CVC syllable types in Mandarin, where the coda C can only be a nasal consonant (N). She observed that the temporal organization between lexical tones and segments resembles the C-center effect. In her studies using kinematic data (for segments) and acoustic data (for tones), she observed that T gestures behaved like the C_2 in C_1C_2. Specifically, she observed that in #CV and #CVN, the onset of the V gesture was aligned to the midpoint of the C and the T gestures. With the assumption that lexical tones are composed of invariant gestures: either L(ow), H(igh), or a combination of both, she observed that there are two patterns among the four lexical tones of Mandarin: Pattern I for tone 1 (high level: H), tone 2 (mid rising: $\frac{H}{L}H$[3]), and tone 3 (low: L) and Pattern II for tone 4 (falling: HL). In particular, it was observed that for Pattern I, the V gesture began around the midpoint between the onset of the C gesture and the onsets of the T gestures, as schematized in Fig. 3a and b. For Pattern II on the other hand, the onset of the V gesture was observed to be start, not exactly, but roughly around the onset of the T1 gesture which is the midpoint of the #CT1T2 sequence.

[3]The $\frac{H}{L}$ in $\frac{H}{L}H$ (tone 3; mid-rising tone) is used in the present study to represent Gao's (2008, 2009) analysis under which the "mid" component of the tone is analyzed as a co-production between H and L tone gestures.

The T1 of HL tone was also observed to begin earlier with respect to the onset of the initial C gesture, than the other three tones. Pattern II is schematized in Fig. 3c.

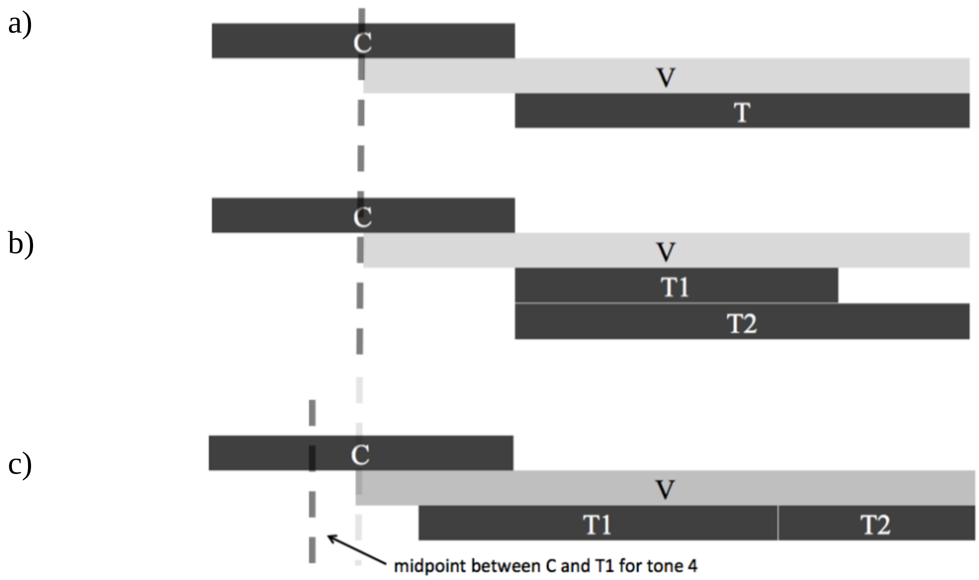


Fig. 3. Organization between segments and lexical tones in Mandarin. Tones 1 and 3 are represented in (a), tone 2 is represented in (b), and tone 4 is represented in (c).

She observed that the first pattern is similar to the temporal organization between the C and the V gestures in *spayed* in American English (as in Fig. 2b): the T gesture appeared to behave like the C_2 in #CCV, i.e. [p] in *spayed*. This led her to conclude that T gestures behave like a C gesture regardless of syllable type; T has a fixed organizational relationship with C and V gestures. She thus hypothesized that Mandarin exhibits C-center effect too but with lexical tones, and she proposed that at the planning level, the temporal organization observed in Pattern I can be represented by the coupling graph in Fig. 4a and the organization observed in Pattern II can be represented by the coupling graph in Fig. 4b.

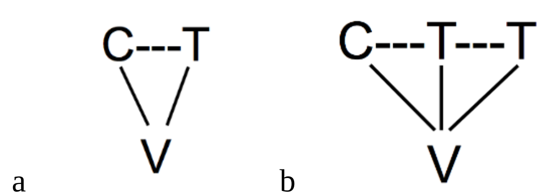


Fig. 4. Specified coupling relations for Pattern I (a) and Pattern II (b).

1.3 Thai

1.3.1 Tones and Syllable structures in Thai

Standard Thai, or Siamese, has five lexical tones which are traditionally described as high, mid, low, falling, and rising. (Abramson, 1962, 1975) Abramson (1962) categorized the tones into two groups: i) level, or "static", tones which include high, mid, and low tones, and ii) contour, or "dynamic", tones which include falling and rising tones. The "static" tones were observed to exhibit small pitch excursions when compared to "dynamic" tones which exhibit larger pitch excursions. A closer examination of Thai tones however revealed that the tones are more complex than the phonological labels. The actual phonetic shapes of each tone do not match "high", "mid", "low", "falling", and "rising" labels. As noted in previous studies (e.g. Abramson, 1962; Erickson, 1974; Gandour *et al.*, 1991; Morén & Zsiga, 2006), in citation forms, "high" tone is phonetically realized as mid-rising, "low" tone is phonetically mid-falling, "falling" tone is phonetically realized as rise-falling, and "mid" tone tends to fall slightly towards the end. Whether the low and the high tones should be analyzed as contour tones is beyond the scope of the present study. In the present study, high, mid, low, falling, and rising tones will be represented as H, M, L, HL, and LH respectively: only HL and LH are assumed to be contour tones in the present study. The five lexical tones are presented in (1) below.

(1)	high	H	[kʰáa]^{H}	'to trade'
	mid	M	[kʰāa]^{M}	'to be stuck'
	low	L	[kʰàa]^{L}	'galangal'
	falling	HL	[kʰâa]^{HL}	'slave; to kill'
	rising	LH	[kʰǎa]^{LH}	'leg'

Morén & Zsiga (2006) in following previous studies grounded in Autosegmental Phonology, (e.g. Gandour, 1974; Leben, 1971, 1973; Woo, 1969) proposed that Thai tones can be represented with H and L: *none* (mid), H (high), L (low), HL (falling), and LH (rising). (In contrast to their study, the present study assumes that the mid tone has a M(id) tone gesture; I will leave the issue of whether Thai mid tone is toneless or underspecified for future study.) Morén & Zsiga (2006) also argued that the tone-bearing unit (TBU) in Thai is the mora, as shown in (1). Follow up perception experiments by Zsiga & Nitisaroj (2007) further support that the TBU in Thai is the mora.

(2)	Mid	High	Low	Falling	Rising
		H	L	H L	L H
		|	|	| |	| |
	μ μ	μ μ	μ μ	μ μ	μ μ

(Morén & Zsiga, 2006: 114)

Tones in Thai are fully realized when they are stressed and are located before a phrasal boundary—they are not fully realized otherwise. Specifically, in non-final positions, the pitch excursion in the second half of the syllable is truncated. This is most apparent with contour tones: the HL tone stays high and does not fall, and the LH tone stays low and rarely rises. (Abramson, 1979; Gandour *et al.*, 1994; Potisuk *et al.*, 1997; Morén & Zsiga, 2006) Although, the second half of the tones is curtailed, these previous studies all agree that all five tones still remain distinctive in non-final positions: the HL tone, albeit phonetically high, is phonetically higher than phonemic H tone, the LH tone, albeit phonetically low, has a different contour from that of the phonemic L tone. Morén & Zsiga (2006) argued that that truncation of the second half of the tones in this context indicates that the second mora is dropped, and as a result, the T2s of contour tones are truncated.

1.3.2 Restricted Distribution of Tones in Thai

Not all syllables can bear all lexical tones in Thai. (Abramson, 1962; Gandour, 1974a, 1974b, 1977; Yip, 1982; Zhang, 2002; Morén & Zsiga, 2006) Open syllables and syllables ending in a sonorant (S), which are traditionally called "live syllables", can bear all five tones, whereas in syllables ending in an obstruent (O) or a glottal stop, which are traditionally called "dead syllables," can only bear certain tones. In particular, while CVV, CVS, and CVVS can bear all five tones, CVO and CVVO can only bear two tones: CVO can bear H and L tones, and CVVO can bear HL and L tones. The distribution of tones is shown in (3).

(3)		CVV		CVS		CVVS	
a)	H	[kʰáa]	'to trade'	[kʰáw]	'he/she/they'	[kʰáan]	'to contradict'
	M	[kʰāa]	'to be stuck'	[kʰān]	'to itch'	[kʰāaw]	'fishy smell'
	L	[kʰàa]	'galangal'	[kʰàw]	'knee'	[kʰàaw]	'news'
	HL	[kʰâa]	'slave; to kill'	[kʰâw]	'to enter'	[kʰâaw]	'rice'
	LH	[kʰǎa]	'leg'	[kʰǎw]	'mountain'	[kʰǎaw]	'white'

b)		CVO		CVVO	
	H	[sák]	'to wash clothes'	-	
	M	-		-	
	L	[sàk]	'to tattoo'	[sàak]	'pestle'
	HL	-		[sâak]	'remains; corpse'
	LH	-		-	

There are exceptions, however. In English loanwords, CVO can also bear HL tone, and CVVO can also bear H tone. (Gandour, 1979; Kenstowicz & Suchato, 2006) These syllable types were not included in the experiment.

1.4 C-center in Thai?

Previous findings from Thai suggest that there is a possibility that the language does not exhibit C-center-like effect in the temporal organization between initial consonants and lexical tones. The findings from previous studies on Thai suggest that syllable length may affect the temporal organization between initial Cs and Ts. Morén & Zsiga's (2006) proposed that Thai words in citation form are bimoraic: in CVS, V is aligned to the first mora, and S is aligned to the second mora, but in CV_1V_2S on the other hand, V_1 is aligned to the first mora, and V_2 and S share the second mora. In their study, for CVS and CVVS in citation form, the turning point of f0 of T2 was observed to be aligned to the midpoint of the rhyme in CVS (defined as the boundary between V and S), but the turning point was observed to be aligned to the midpoint of the rhyme in CVVS. However, it is unclear whether the boundary between the V and the S in CVS is temporally aligned to the midpoint point of the rhyme in CVVS. Under the assumption that V_2 and S in CV_1V_2S share the second mora (Morén & Zsiga, 2006) and that the middle of V_2 is the midpoint of the rhyme, it is possible that the boundary between the V and the S in CVS and the midpoint of the rhyme in CVVS are not be temporally aligned. This is illustrated in Fig. 5.

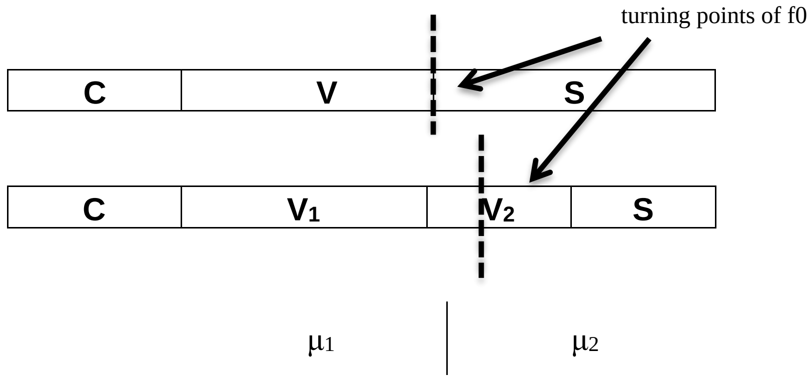


Fig. 5. The turning points of f0 of contour tones in CVS and in CVVS under the assumption the V_2 in CV_1V_2S is the midpoint of the rhyme and that V_2 and S share the second mora.

However, it is not transparent if their study truly contradicts the C-center effect in the temporal organization between segments and tones as hypothesized by Gao (2008, 2009). Because their research was on a different topic, Morén & Zsiga (2006) did not directly compare the turning points of f0 across syllable types. Crucially, they did not directly examine the relationship between the onsets of initial Cs and the onset of tones: they did not measure the distance from the initial consonant to the turning point of f0. Furthermore, the turning points of f0 used in their study different from those in Gao (2008): while they examined the turning points of f0 across all five tones, the turning points of f0 that they examined were in the middle of the syllable, i.e. the turning points occurring *within* a single lexical tone, unlike Gao (2008, 2009) who examined the turning points of f0 occurring *between* two different tones, i.e. the f0 from a preceding syllable in a carrier sentence to target syllables. Thus, Morén & Zsiga's (2006) representation of H and L tones in (2) do not represent the onset of H tone and the onset of L tone the same way Gao (2008, 2009) defined the onsets of tone; instead the autosegmental representation of lexical high and low tones in Morén & Zsiga (2006) study would represent the tonal *targets* in Gao's view.

The present study focuses on the organization between initial C and T gestures in Thai—a language with five tones and complex syllable structure, in an attempt to expand Gao's (2008, 2009) effort which only considered two types of syllable structure and four types of tone. The limited types of syllables and tones in her studies may have influenced her findings. Specifically, this experiment evaluates the predictions of Gao (2008, 2009) regarding the C-center effect in the temporal organization between segments and lexical tones. By doing so, the present study will also have filled in the gap in knowledge from previous studies on Thai which did not directly examine the temporal organization between the onsets of initial Cs and the onset of tones.

2. Methodology

2.1 Stimuli

The stimuli consisted of 14 target words which are presented in Table I. The segmental material was controlled to the greatest extent possible. All target syllables contained the same V(s): /i/ for CVC and /ia/ for CVVC, the same initial C: /l/, the same sonorant coda: /ŋ/ for CVS and CVVS, and the same obstruent coda: /k/ for CVO and CVVO.

Syllable type	Tone allowed	Target word
CVS	H	líŋ
	HL	lîŋ
	L	lìŋ
	LH	lǐŋ
	M	līŋ
CVVS	H	líaŋ
	HL	lîaŋ
	L	lìaŋ
	LH	lǐaŋ
	M	līaŋ
CVO	H	lík
	L	lìk
CVVO	HL	lîak
	L	lìak

Table 1. Unrandomized stimuli consisting of 14 target words

Since the segmental material was controlled in such a strict fashion, some uncommon words were chosen as target syllables. To make each target syllable sound as natural as possible, they were used as proper names (nicknames), which, according to my Thai consultants, are all possible nicknames in Thai.

Each speaker was asked a read the stimuli which were presented to them in a list of sentences. The stimuli were written on five sheets of paper. All sentences were written in Thai script.

It was crucial that the target words were placed in sentence-final position, so that the onset of T2 could be apparent when the data were analyzed. Thus, each target word was placed in sentence-final position.

Each target word in Table I was placed in one of the following two contexts:

1. PRECEDING H: preceding H tone;
   /kʰáw^{H} t͡ɕʰûɯ^{HL} tʰáaw^{H} ___________ /
   he name TITLE ___________
   His name is Lord ___________.

2. PRECEDING L: preceding L tone;
   /kʰáw^{H} t͡ɕʰûɯ^{HL} bàaw^{L} ___________ /
   he name TITLE ___________
   His name is Servant ___________.

The onset of the T gesture that is adjacent to the initial C gesture was defined as the turning point between the f0 of the preceding word and the f0 of the T gesture in target syllables. To ensure that such turning point of f0 occurred, the T gesture of the preceding syllable had to be different from the T or T1 gesture of the target syllables. Thus, L, LH, and M tones were placed in PRECEDING H context, and H, HL, and M tones were placed in PRECEDING L context. Note that M tone was placed in both contexts. The syllables that immediately preceded the target syllables were long (CVVS). This was done to maximally minimize any potential peak delay effects that could affect the onset time of T and T1. (Xu, 2001) As a result, words somewhat archaic which denote a person's rank—roughly translated as "Lord" and "Servant"—were chosen as preceding words.

The material was randomized in 6 blocks. There were a total of 16 sentences per each block; each sentence was repeated twice; thus, 12 repetitions for each sentence were recorded. In sum, there were a total number of 192 tokens recorded for each speaker: 16 sentences X 12 repetitions (from 6 blocks X 2 repetitions of each sentence in each block).

The recordings were digitized and analyzed by in Praat (Boersma, 2001).

2.2 Participants

Five native speakers of Thai, who at the time of the experiment were undergraduate and graduate students of USC, participated in the experiment; they were paid at a standard compensation rate. These included four men and one woman, ages from 19-27. It was important that the speakers were around this age group because a previous study by Thepboriruk (2009) observed that people ages from 18-24[4] produced tones slightly differently from groups ages 50 and up. Specifically, the T2s of HL and LH for the younger groups were observed to be more delayed than the oldest group.

Before participating in this experiment, they had lived in Los Angeles for no longer than two years. All participants were from Bangkok, except for one male who was from Lampang, a northern province of Thailand. All speakers reported no speech or hearing pathology. Speakers henceforth will be referred to as M1, M2, M3, M4, and W1 (M = male, W = female).

A head-mounted microphone was used to record each speaker. During the recording session, each was seated in a quiet room located at the Linguistics Department of USC.

2.3 Measurements

---

[4] Our oldest participant who was 27 at the time of the experiment would have been around 22 at the time of the experiment by Thepboriruk (2009).

To investigate tone-to-consonant alignment patterns, a series of onset-to-onset lags were measured. Onset-to-onset lags were defined as the temporal distance between the beginning time of the initial C gesture and the beginning times of the T gesture. All lags were measured in milliseconds.

2.3.1 How onsets were defined

Three onset landmarks had to be located before the lags were computed. These include:

- The onset of the initial C gesture
- The onset of T gesture (for H, M, and L tones) or T1 gesture (for HL and LH tones)
- T2 (for HL and LH tones)

The onset of the initial C gesture was defined as the formant transition of F2 between the preceding segment and the initial C gesture of the target segment. Since the preceding words ended in [w] while the target words began with [l], a clear formant transition of F2 going from [w] (low F2) to [l] (high F2) could be observed in the spectrogram, as shown in Fig. 6. Thus, the lowest F2 value, i.e. the F2 “valley”, was used to define as the onset of the [l].

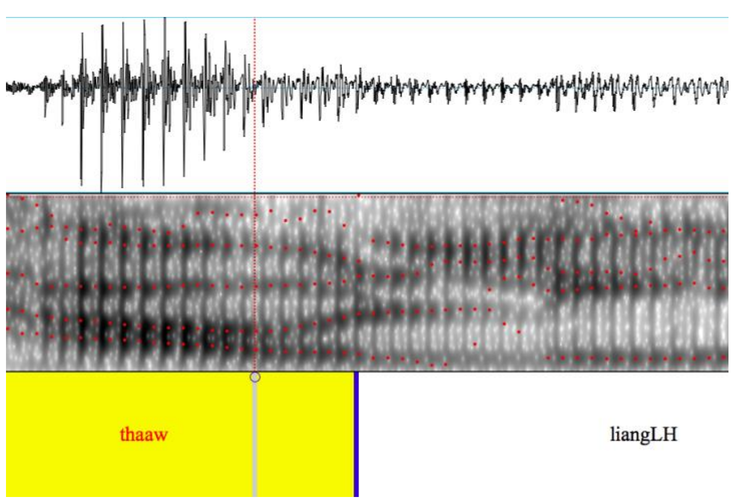


Fig. 6. The transition of F2 in [w#l]. F2 in [aaw] is low, whereas F2 in [l] is high. The onset of [l] is defined as the "valley" or the "dip" in F2 shown in the figure.

The onset of T gesture (or T1 for contour tones) was defined as the turning point between the f0 of the preceding T gesture in the preceding word and the f0 of the T gesture (or T1) in target syllables. Thus, in PRECEDING H context, the highest point of f0—the "peak"— was used to define as the onset of L, LH, and M tones, and similarly in PRECEDING L context, the lowest point of f0—the "valley"— was used to define as the onset of H, HL, and M tones. This is illustrated in Fig. 7.

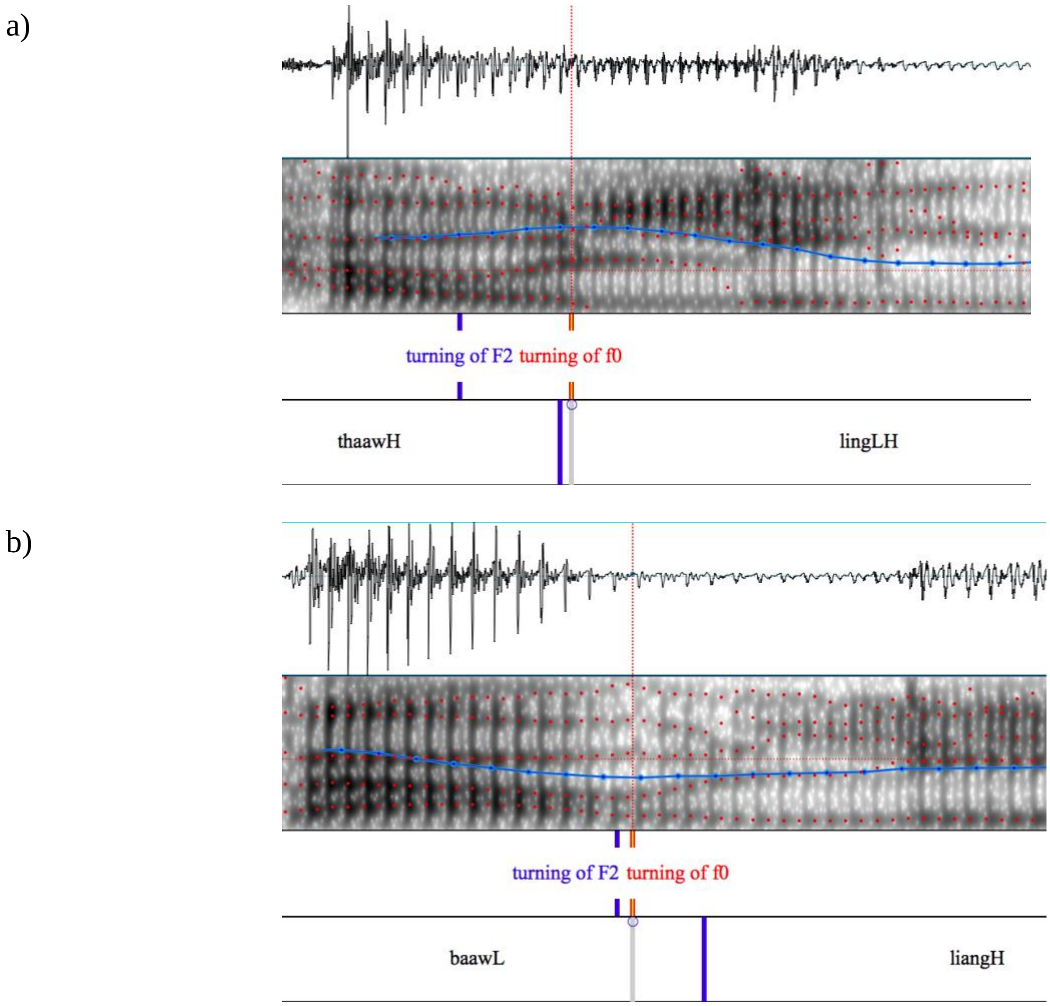


Fig. 7. Onset of T1 in LH tone, defined as the turning of f0—the f0 peak in [tʰáaw^{H}#lǐŋ^{LH}] (a), and onset of H tone, defined as the turning point of f0—the f0 valley in [bàaw^{L}#líaŋ^{H}] (b).

The onset of T2 was defined as the turning point of f0 going from T1 gesture to T2 gesture. Thus, for HL, the onset of T2 gesture was defined as the peak of f0, and similarly, the onset of T2 of LH was defined as the valley of f0.

2.3.2 Onset-to-onset lags

The following are the three types of onset-to-onset lag used in this experiment:

- C-to-T lag: temporal distance between C gesture and T gesture.
    a. Onset of C gesture – onset of T gesture.

b. This lag was measured for H, L, and M tones.

- C-to-T1 lag: temporal distance between C gesture and T1 gesture

  a. Onset of C gesture – onset of T1 gesture.

  b. This lag was measured for HL and LH tones.

- C-to-T2 lag: temporal distance between C gesture and T2 gesture

  a. Onset of C gesture – onset of T2 in complex tones.

  b. This lag was measured for HL and LH tones.

For C-to-T and C-to-T1 lags, positive values would indicate that the C gesture precedes the T gesture, while negative values would indicate that the T gesture precedes the C gesture. For all lags, the higher the value, the more delayed the T gesture is with respect to the initial C gesture. The lags are illustrated in Fig. 8. T(1) represents either any level tones or the T1 gesture of contour tones.

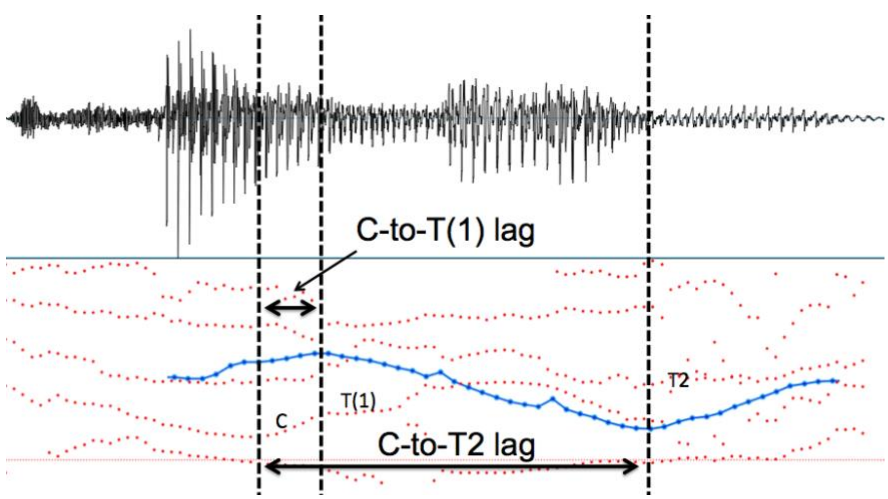


Fig. 8. C-to-T1 lag and C-to-T2 lag in [tʰáaw^{H}#lǐaŋ^{LH}].

All statistical analyses were carried out with the software SPSS^{TM}.

3. Hypotheses

Hypothesis A: If the initial C gesture and T gesture (or T1 gesture for contour tones) are in a competitive coupling mode, then they should not begin synchronously. Thus, it was predicted that the T gesture would begin after the C gesture regardless of syllable type (CVO, CVS, CVVO, and CVVS), and that the T gesture would not begin before the C gesture. In particular, it was predicted that the formant transition of the initial C gesture would always occur before the turning point of f0 of the T gesture that is adjacent to the initial C gesture. This is schematized in Fig. 9. The lags were expected to be around 60-100 ms long because the lags observed in Gao (2008, 2009) with the range from 60-100 ms.

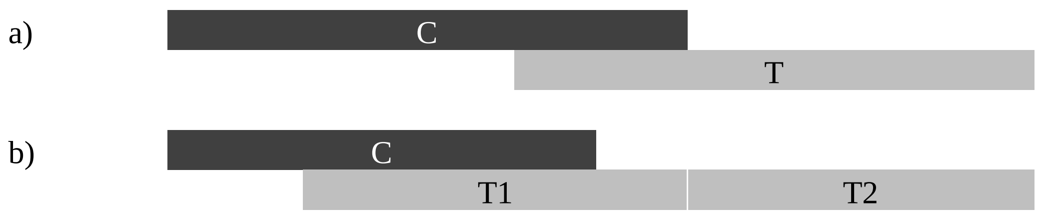


Fig. 9. Hypothesized timing relation between the onsets of C gesture and T gestures for H, M, and L tones (a) and HL and LH tones (b).

Hypothesis B: If the T2 gesture of contour tones (HL and LH) is coupled with respect to the initial C gesture, then the organization between the initial C gesture and the T2 gesture of contour tones (HL and LH) within a syllable should not be affected by syllable type. Thus, it was predicted that the T2 gesture of HL and LH would begin at the same time in both CVC and CVVC. In other words, the distance between the formant transitions and the turning points of f0 of T2 gesture should be the same across all syllable types. In particular, it was predicted that for HL tone, the turning points of f0 of the T2 gesture in CVS, CVVO, and CVVS would begin at the same time with respect to the formant transitions of the initial C gesture. Similarly, for the LH tone, the turning points of f0 of the T2 gesture in CVS and CVVS were also predicted to begin at the same time in both syllable types.

Hypothesis C: If the T1 gesture of a contour tone is in a competitive coupling mode with the initial C gesture and T2 gesture, while the T gesture of a level tone is only in a competitive coupling mode with the initial C gesture, then the T1 gesture of a contour tone should begin before the T gesture of a level tone. Thus, it was predicted that the T1 gesture of a contour tone would begin before level tones. This is illustrated in Fig. 9 where the onset of the T1 gesture (b) occurs before the onset of the T gesture (a). In particular, it was predicted that the turning points of f0 of the T1 gesture in HL and LH would begin before the turning points of f0 of level tones. However, for simplicity this study only made two comparisons: i) the distance between the onset of H tone and the onset of the T1 gesture of HL and ii) the distance between the onset of L tone and the onset of the T1 gesture of LH. The turning point of f0 of the T1 gesture of HL tone was predicted to begin before the turning point of f0 of H tone, and the turning point of f0 of the T1 gesture of LH was predicted to begin before the turning point of f0 of L tone. Hypotheses A, B, and C come from Gao (2008, 2009).

Hypothesis D: If the T2 gesture of a contour tone is not in a competitive coupling mode with the initial C gesture, then its beginning time should be affected by syllable type (contra Hypothesis B). Thus, it was predicted that the organization between the initial C gesture and the T2 gesture would be affected by syllable length: the T2 gesture in HL and LH in CVVC were predicted to begin after the T2 gesture in HL and LH in CVC. This is schematized in Fig. 10. The present study made comparisons within the same tone. For HL tone, the T2 gesture of HL in CVS was compared to the T2 gesture in HL in CVVO and

CVVS; for LH tone, the T2 gesture of LH in CVS was compared to the T2 gesture in HL in CVVS. This hypothesis comes from my interpretation of the findings by Morén & Zsiga (2006). It must be noted that in the original study, the alignments between T2 and the S in CVS and the alignment between T2 and the V_2S in CV_1V_2S are based on syllables produced in isolation and in citation speech.

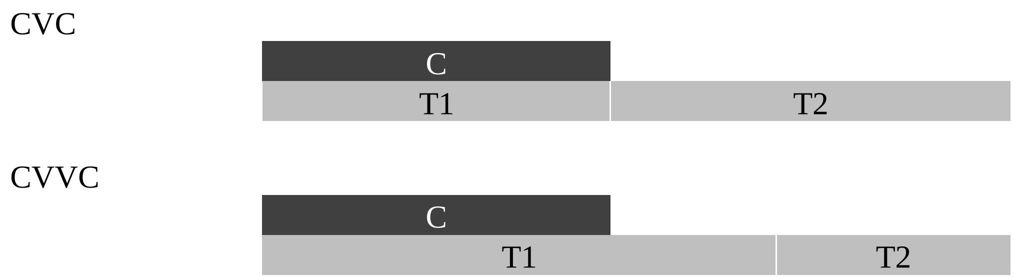


Fig. 10. Hypothesized timing relation between T2 in CVS and T2 in CVVS.

4. Results

4.1 L tone

4.1.1 C-to-T lag

Fig. 11 shows the C-to-T lags for L tone in milliseconds across different syllable types from individual subjects. Error bars represent standard deviations. Table 2 shows the comparison of lag duration for L tone across different syllable types.

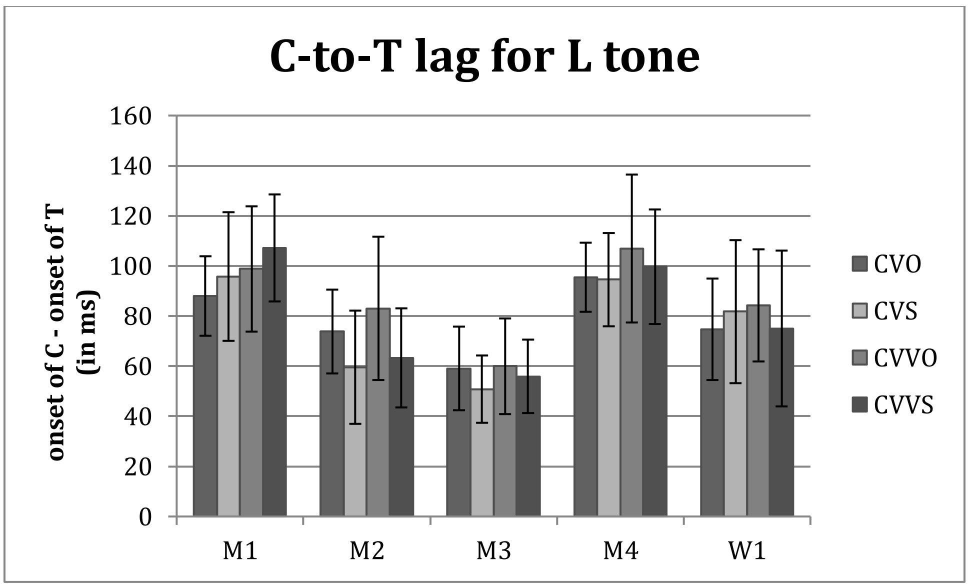


Fig. 11. C-to-T lags across different syllable types from individual subjects (L tone) (ms).

	CVO		CVS		CVVO		CVVS		
speaker	M	SD	M	SD	M	SD	M	SD	
M1	87.99	15.87	95.76	25.68	98.77	25.01	107.18	21.36	*n.s.*
M2	73.81	16.71	59.55	22.59	83.05	28.59	63.30	19.77	*n.s.*
M3	59.09	16.68	50.82	13.44	59.95	19.10	55.92	14.66	*n.s.*
M4	95.46	13.79	94.51	18.60	106.94	29.51	99.68	22.86	*n.s.*
W1	74.70	20.24	81.74	28.54	84.25	22.38	75.02	31.08	*n.s.*

Table 2. Comparison of lag duration across different syllable types (L tone)

According to Hypothesis A, the onset of the C gesture must start before the onset of the T gesture regardless of syllable type. The results show that all lags had positive values. This suggests that the onset of L tone never precedes the onset of the C gesture. A one-way ANOVA was used to test for C-to-T lag differences for L tone among four syllable types for each subject. C-to-T lags did not differed significantly across the four syllable types for all subjects. The results suggest that syllable type has no effects on this temporal organization. Therefore, the results support Hypothesis A

4.2 H tone

4.2.1 C-to-T lag

Fig. 12 shows the C-to-T lags for H tone in milliseconds across different syllable types from individual subjects. Table 3 shows the comparison of lag duration for H tone across different syllable types.

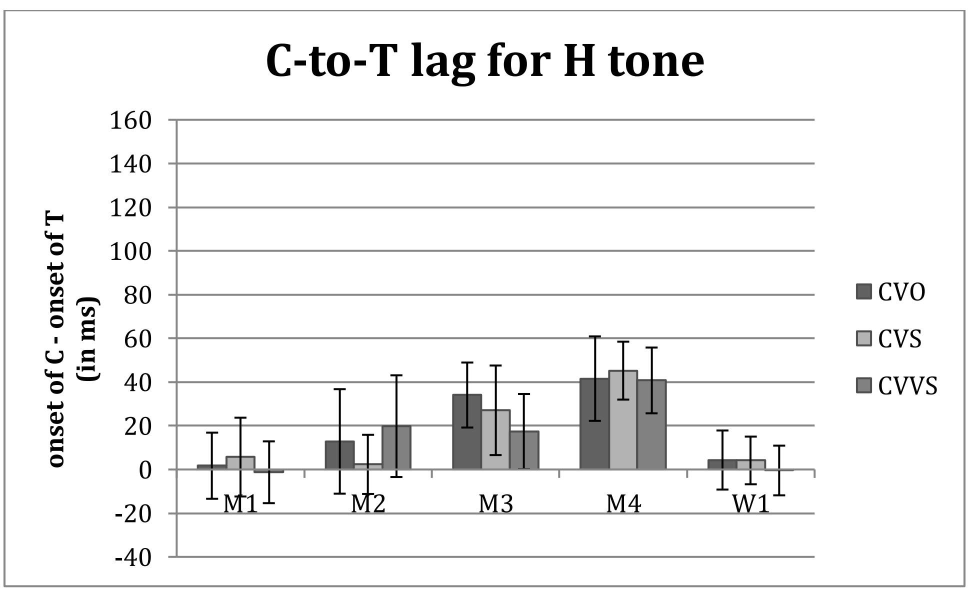


Fig. 12. C-to-T lags across different syllable types from individual subjects (H tone) (ms).

	CVO		CVS		CVVS		
speaker	M	SD	M	SD	M	SD	
M1	1.71	15.13	5.66	18.02	-1.26	14.14	*n.s.*
M2	12.84	23.92	2.32	13.53	19.82	23.28	*n.s.*
M3	34.06	14.88	27.08	20.48	17.35	17.18	*n.s.*
M4	41.57	19.33	45.22	13.26	40.75	15.04	*n.s.*
W1	4.32	13.52	4.13	10.90	-0.46	11.36	*n.s.*

Table 3. Comparison of lag duration across different syllable types (H tone)

The results for C-to-T lags for H tone show that the lags sometimes have negative values, except for M3 and M4. This suggests that the onset of H tone sometimes precedes the onset of the C gesture for all subjects except M3 and M4. A one-way ANOVA was used to test for C-to-T lag differences for H tone among three syllable types for each subject. C-to-T lags did not differ significantly across the three syllable types for all subjects. The results suggest that syllable type has no effects on this temporal organization. However, when compared to the temporal distance between the onset of the initial C gesture and the onset of L tone, results suggest that the onset of H tone is more synchronous to the initial C gesture than the onset of L tone. The lags are overall smaller (with means ranging between -0.26 – 45.22 ms), whereas the lags for L tone are larger (with means ranging between 50.82 – 107.18 ms). Moreover, although results from M3 and M4 do not show that the onset of H tone precedes the initial C gesture, their C-to-T lags for H tone were found to be significantly smaller than their C-to-T lags for L tone; a paired-samples t-test was conducted to compare the C-to-T lags for H tone and the C-to-T lags for L tone for each of the two subjects (all syllable types pooled together for each tone), and it was found that the C-to-T lags for H tone are significantly smaller than the C-to-T lags for L tone, $t(35) = -4.34$, $p < .001$ for M3 and $t(35) = -11.93$, $p < .001$ for M4. The results suggest that the onset of H tone for M3 and M4 is much more synchronous to the initial C than the onset of L. Therefore, the results do not support Hypothesis A.

4.3 M tone

4.3.1 C-to-T lag

Results for C-to-T lags for M tone were divided into two groups: one in PRECEDING H context, and one in PRECEDING L context, due to my observation that the M tone in both contexts do not pattern each other well. Fig. 13 shows the C-to-T lags for M tone in PRECEDING H context in milliseconds across different syllable types from individual subjects, and Fig. 14 shows the C-to-T lags for M tone in PRECEDING L context in milliseconds across different syllable types from individual subjects. Table 4 shows the comparison of lag duration for M tone in PRECEDING H context across different syllable types, and Table 5 shows the comparison of lag duration for M tone in PRECEDING L context across different syllable types.

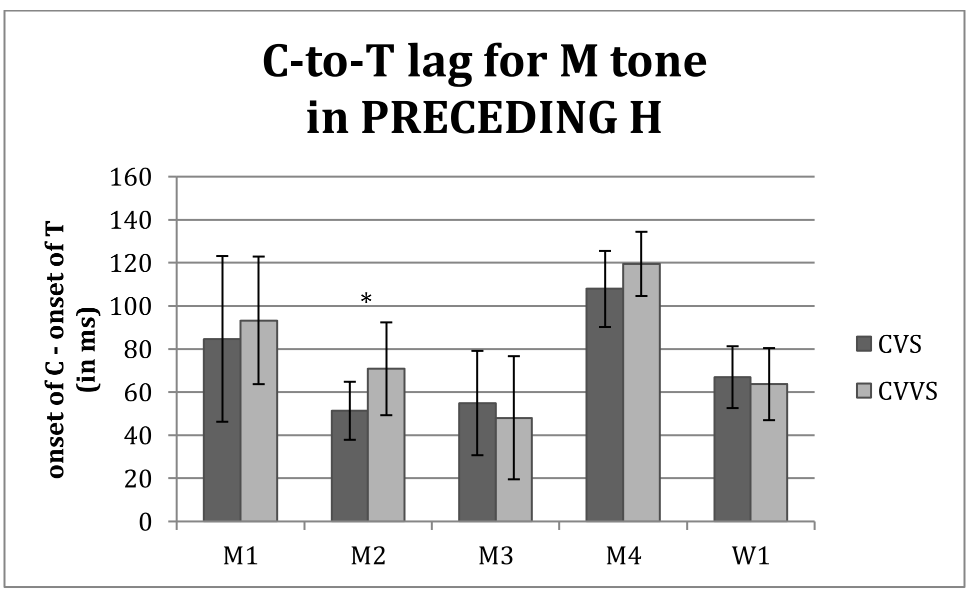


Fig. 13. C-to-T lags across different syllable types from individual subjects (M tone in PRECEDING H context) (ms).

	CVS		CVVS		
speaker	M	SD	M	SD	
M1	84.66	38.38	93.27	29.64	*n.s.*
M2	51.37	13.46	70.78	21.54	p<0.05
M3	54.95	24.25	48.06	28.54	*n.s.*
M4	107.91	17.65	119.54	14.90	*n.s.*
W1	66.92	14.30	63.67	16.71	*n.s.*

Table 4. Comparison of lag duration across different syllable types (M tone in PRECEDING H context)

A paired-samples t-test was conducted to compare the C-to-T lags for M tone in PRECEDING H context in CVS and the C-to-T lags in CVVS for each subject. No significant differences were found for all subjects except M2. For M2, C-to-T lags differed significantly between the two syllable types, $t(11) = 3.01$, $p < .05$, with the lags in CVVS being longer than the lags in CVS. The results from M2 might be due to idiosyncrasy in speech. For the rest of the subjects the lags always contained positive values. This suggests that the onset of the M in this context never precedes the initial C gesture. The results overall are parallel to the results for C-to-T lags for L tone

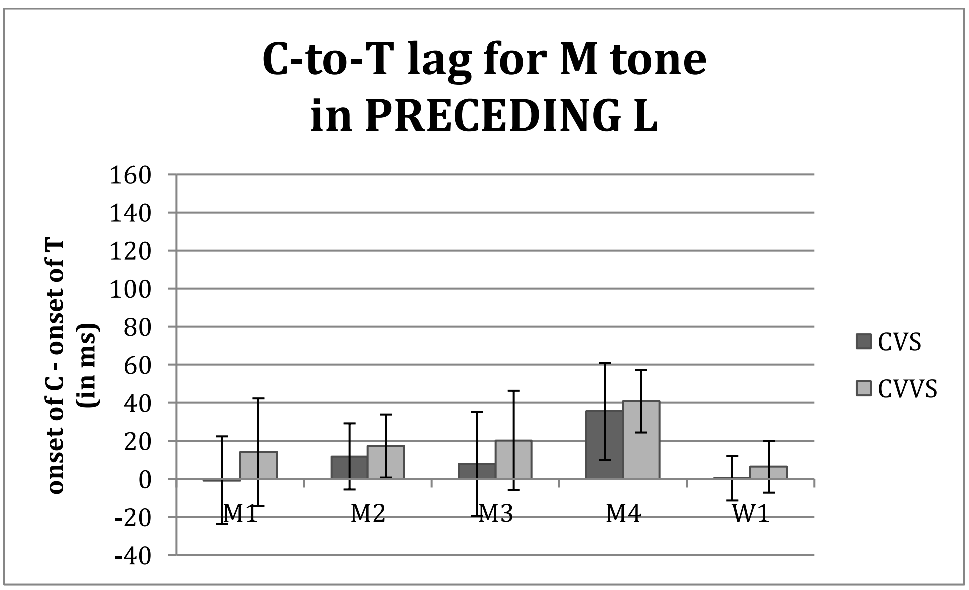


Fig. 14. C-to-T lags across different syllable types from individual subjects (M tone in PRECEDING L context) (ms).

	CVS		CVVS		
speaker	M	SD	M	SD	
M1	-0.67	23.07	14.11	28.28	*n.s.*
M2	11.87	17.34	17.32	16.53	*n.s.*
M3	7.89	27.31	20.34	26.06	*n.s.*
M4	35.50	25.47	40.78	16.36	*n.s.*
W1	0.47	11.72	6.47	13.60	*n.s.*

Table 5. Comparison of lag duration across different syllable types (M tone in PRECEDING L context)

A paired-samples t-test was conducted to compare the C-to-T lags for M tone in PRECEDING L context in CVS and the C-to-T lags in CVVS for each subject. No significant differences were found for individual subjects. However, unlike the C-to-T lags for M tone in PRECEDING H context, it was found that the C-to-T lags were smaller. This suggests that the M tone in PRECEDING L context is synchronous to the initial C. In contrast to the M tone in PRECEDING H context, the results are parallel to the results for C-to-T lags for H tone, and this is not predicted by Hypothesis A.

The results from M tone suggest that the onset of the T gesture is dependent on the T gesture of the preceding tone; therefore, the results overall do not support Hypothesis A. The results show that the onset of T is delayed with respect to the initial C gesture when the preceding tone is H, whereas the onset of the T gesture is much more synchronous to the initial C gesture when the preceding tone is L. A follow-up paired-samples t-test was conducted to compare the C-to-T lags for M tone in PRECEDING H context and the C-to-T lags in PRECEDING L context (all subjects and syllable types pooled together for each preceding tone context), and it was found that the C-to-T lags for M tone in PRECEDING H context are significantly larger than the C-to-T lags for M tone in PRECEDING L context, $t(119) = 19.81$, $p < .001$.

4.4 HL tone

4.4.1 C-to-T1 lag

Fig. 15 shows the C-to-T1 lags for HL tone in milliseconds across different syllable types from individual subjects. Table 6 shows the comparison of lag duration for HL tone across different syllable types.

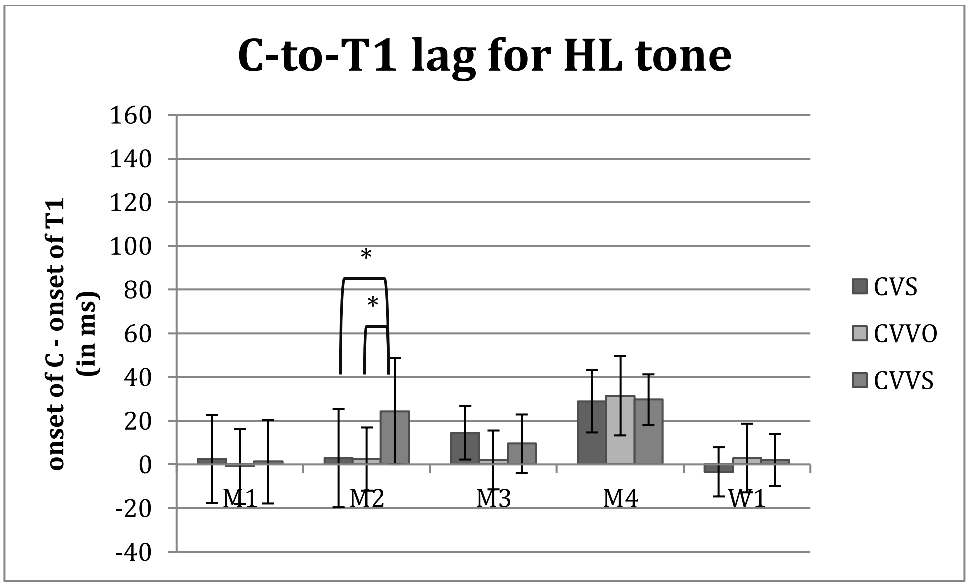


Fig. 15. C-to-T1 lags across different syllable types from individual subjects (HL tone) (ms).

	CVS		CVVO		CVVS			
speaker	M	SD	M	SD	M	SD		
M1	2.49	20.06	-0.86	17.15	1.28	19.15	*n.s.*	
M2	2.84	22.46	2.45	14.45	24.38	24.34	One-way ANOVA: Post-hoc Tukey tests CVS vs CVVO: CVS vs CVVS: CVVO vs CVVS:	$p < .05$ *n.s.* $p < .05$ $p < .05$
M3	14.54	12.31	2.05	13.47	9.50	13.34	*n.s.*	
M4	28.96	14.33	31.39	18.12	29.60	11.61	*n.s*	
W1	-3.42	11.25	2.92	15.71	2.05	11.97	*n.s*	

Table 6. Comparison of lag duration across different syllable types (HL tone)

The results from C-to-T1 lags for HL tone show that the onset of T1 of HL sometimes precedes the initial C gesture, except M4. However, the lags from M4 are small, with means ranging between 28.96 – 31.39 ms, which is below 60 ms. The results here are similar to the results from the C-to-T lags for H and M tone in PRECEDING L context. A one-way ANOVA was conducted to assess the effect of syllable type on the lag, and the results were negative for M1, M3, M4, and W1. Syllable type has no effects on the C-to-T1 lags for HL tone for these subjects. For M2, however, C-to-T1 lags differed significantly across the three syllable types, $F(2, 33) = 4.34$, $p < .05$. Post hoc comparisons using the Tukey HSD test indicated that the mean lag in CVVS is significantly longer than the mean lags in CVS ($p < .05$) and CVVO ($p < .05$). As with the C-to-T lag for M tone in PRECEDING H context, the results from M2 might be due to idiosyncrasy in speech. The overall results here do not support Hypothesis A because the T1 gesture of HL and the initial C gesture are synchronous.

4.4.2 C-to-T2 lag

Fig. 16 shows the C-to-T2 lags for HL tone in milliseconds across different syllable types from individual subjects. Table 7 shows the comparison of lag duration for HL tone across different syllable types.

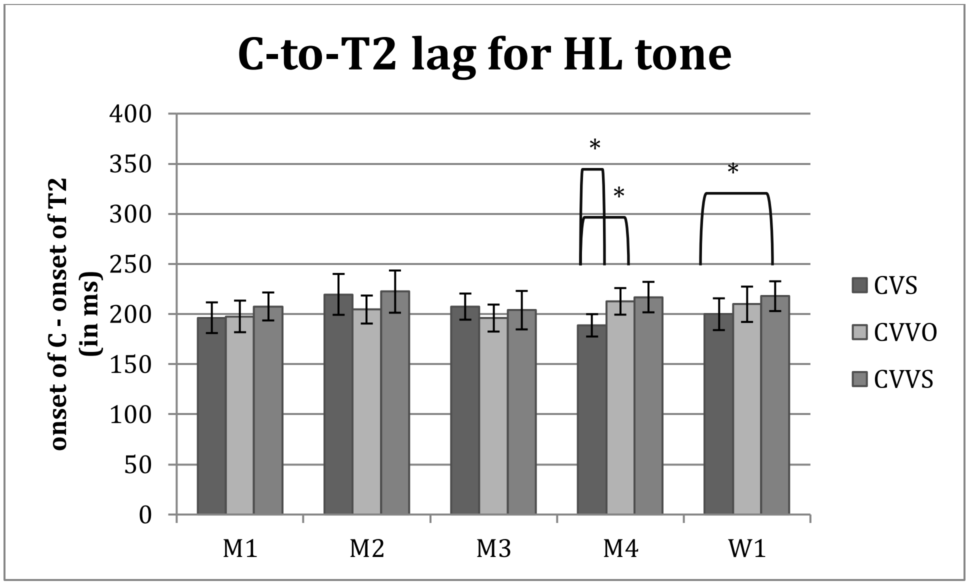


Fig. 16. C-to-T2 lags across different syllable types from individual subjects (HL tone) (ms).

	CVS		CVVO		CVVS			
speaker	M	SD	M	SD	M	SD		
M1	196.35	15.32	197.66	15.77	207.59	13.95	*n.s.*	
M2	219.64	20.39	204.51	13.96	222.41	21.07	*n.s.*	
M3	207.45	13.04	196.01	13.46	203.91	19.18	*n.s.*	
M4	188.73	11.16	212.67	13.27	216.93	15.19	One-way ANOVA: Post-hoc Tukey tests CVS vs CVVO: CVS vs CVVS: CVVO vs CVVS:	p < .001 p < .05 p < .05 *n.s.*
W1	199.86	15.86	209.76	17.63	217.86	14.86	One-way ANOVA: Post-hoc Tukey tests CVS vs CVVO: CVS vs CVVS: CVVO vs CVVS:	p < .05 *n.s.* p < .05 *n.s.*

Table 7. Comparison of lag duration across different syllable types (HL tone)

According to Hypothesis B, the distance between the onset of the initial C gesture and the onset of the T2 gesture will not be affected by syllable length, but according to Hypothesis D, the distance between the onset of the initial C gesture and the onset of the T2 gesture *will* be affected by syllable length. The results show that M1, M2, and M3 did not produce the lags the same way as M4 and W1. A one-way ANOVA was conducted to assess the effect of syllable type on the lag, and the results were negative for M1, M2, and M3; it was found that syllable type had an effect on the lags for M4 and W1. C-to-T2 lags differed significantly across the three syllable types, $F (2, 33) = 15.65$, $p < .001$ for M4 and $F (2, 33) = 3.81$, $p < .05$ for W1. Post hoc comparisons using the Tukey HSD test indicated that for M4 the mean lag in CVS is significantly shorter than the mean lags in CVVO ($p < .05$) and CVVS ($p < .05$), and that for W1 the mean lag in CVS is significantly shorter than the mean lag in CVVS ($p < .05$). Hence, the results from M1, M2, and M3 suggest that Hypothesis D must be rejected, while the results from M4 and W1 suggest that Hypothesis B must be rejected. It is puzzling as to why there are two patterns among the speakers.

4.5 LH tone

4.5.1 C-to-T1 lag

Fig. 17 shows the C-to-T1 lags for LH tone in milliseconds across different syllable types from individual subjects. Table 8 shows the comparison of lag duration for LH tone across different syllable types.

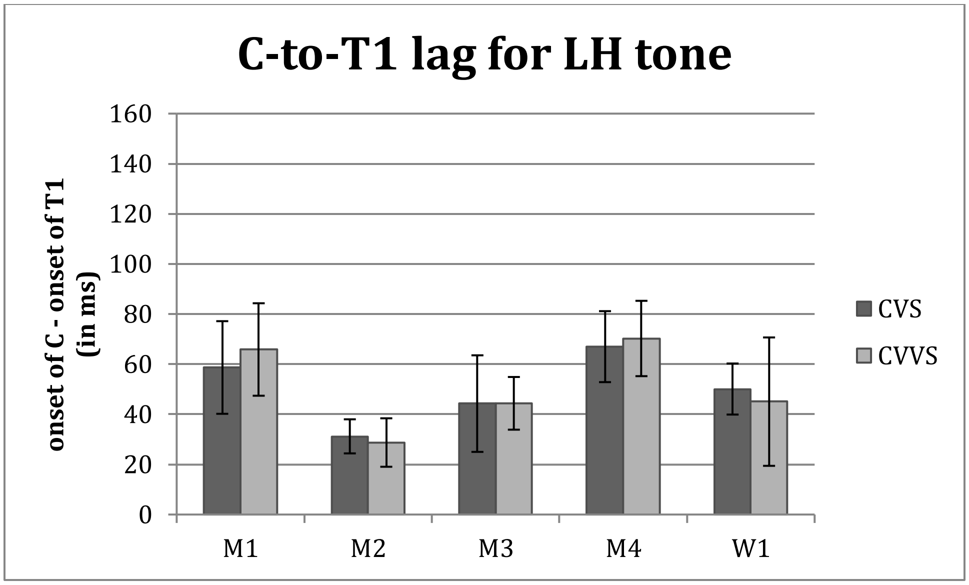


Fig. 17. C-to-T1 lags across different syllable types from individual subjects (LH tone) (ms).

	CVS		CVVS		
speaker	M	SD	M	SD	
M1	58.68	18.49	65.86	18.45	*n.s.*
M2	31.20	6.80	28.73	9.68	*n.s.*
M3	44.27	19.28	44.38	10.51	*n.s.*
M4	67.00	14.17	70.25	15.03	*n.s.*
W1	50.08	10.19	45.05	25.62	*n.s.*

Table 8. Comparison of lag duration across different syllable types (LH tone)

A paired-samples t-test was conducted to compare the C-to-T1 lag for LH tone in CVS and the C-to-T lag in CVVS for each subject. No significant differences were found for individual subjects. This suggests that syllable type has no effects on the distance between the onsets of the initial C gesture and the T1 gesture of LH. The results also show that the lags always have high values in both syllable types. This suggests that the onset of the T1 gesture does not begin at the same time as the initial C gesture. In contrast to the C-to-T1 lag for HL tone, the results here support Hypothesis A.

4.5.2 C-to-T2 lag

Fig. 18 shows the C-to-T2 lags for LH tone in milliseconds across different syllable types from individual subjects. Table 9 shows the comparison of lag duration for LH tone across different syllable types.

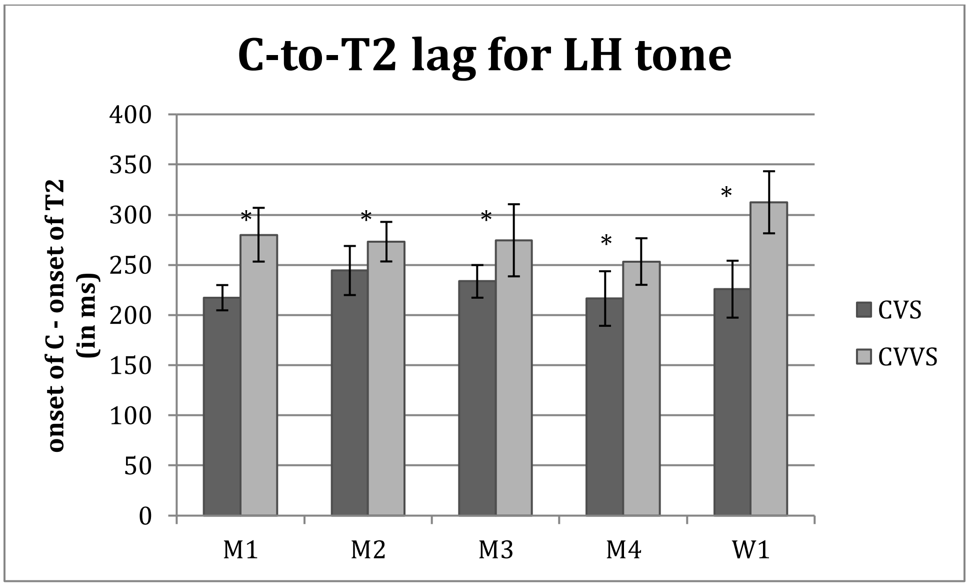


Fig. 18. C-to-T2 lags across different syllable types from individual subjects (LH tone) (ms).

	CVS		CVVS		
speaker	M	SD	M	SD	
M1	217.25	12.57	280.05	26.79	$p < .001$
M2	244.35	24.45	273.10	19.72	$p = .001$
M3	233.53	16.27	274.48	35.95	$p = .001$
M4	216.43	27.20	253.29	23.22	$p < .01$
W1	225.75	28.37	312.37	30.94	$p < .001$

Table 9. Comparison of lag duration across different syllable types (LH tone)

A paired-samples t-test was conducted to compare the C-to-T2 lag for LH tone in CVS and the C-to-T lag in CVVS for each subject. It was found that for each subject the C-to-T2 lag in CVS is significantly shorter than the lag in CVVS for all subjects: t(11) = -

7.11, p < .001 for M1, t(11) = -4.36, p = .001 for M2, t(11) = -4.42, p = .001 for M3, t(11) = -3.77, p < .01 for M4, and t(11) = -8.53, p < .001 for W1. This suggests that syllable structure has an effect on the C-to-T2 lag for LH. The results therefore support Hypothesis D, not Hypothesis B.

4.6 H and HL tones

4.6.1 C-to-T lag and C-to-T1 lag

Fig. 19 shows the C-to-T lags for H tone and the C-to-T1 lags for HL tone in milliseconds from individual subjects (CVS and CVVS pooled together). Table 10 shows the comparison of lag duration for H and HL tones across different syllable types.

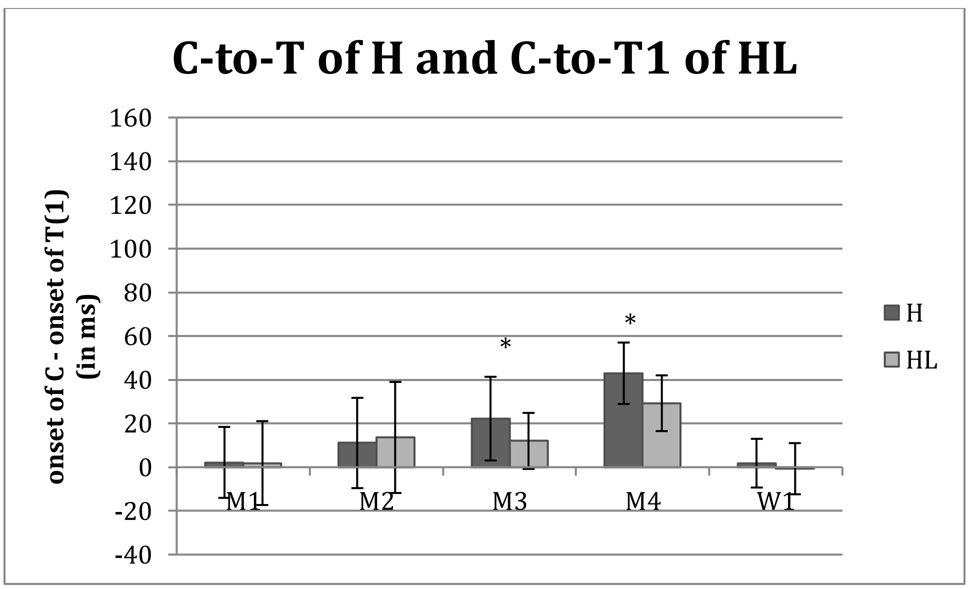


Fig. 19. C-to-T lags of H tone and C-to-T1 lags of HL across different syllable types from individual subjects (ms).

	H in CVS and CVVS		HL in CVS and CVVS		
speaker	M	SD	M	SD	
M1	2.20	16.23	1.88	19.19	*n.s.*
M2	11.07	20.66	13.61	25.41	*n.s.*
M3	22.22	19.14	12.02	12.82	$p < .05$
M4	42.99	14.05	29.28	12.76	$p < .01$
W1	1.84	11.14	-0.69	11.70	*n.s.*

Table 10. Comparison of lag duration across different syllable types (H and HL tones)

According to Hypothesis C, the T1 gesture of contour tones must begin before the T of level tone. The prediction was that the C-to-T lag of H would be longer than the C-to-T1 lag of HL. This would suggest that the H gesture in HL starts earlier than phonemic H tone. A paired-samples t-test was conducted to compare the C-to-T lag of H tone in CVS and CVVS pooled together and the C-to-T1 lag of HL in CVS and CVVS pooled together for each subject. No significant differences were found except for M3 and M4; it was found that for M3, the mean C-to-T lag of H tone is significantly longer than the mean C-to-T1 lag of HL, $t(23) = 2.39$, $p < .05$, and that for M4 the mean C-to-T lag of H tone is significantly longer than the mean C-to-T1 lag of HL, $t(23) = 3.41$, $p < .01$. Thus, results from M3 and M4 support Hypothesis C, while results from other subjects do not.

4.7 L and LH tones

4.7.1 C-to-T lag and C-to-T1 lag

Fig. 20 shows the C-to-T lags for L tone and the C-to-T1 lags for LH tone in milliseconds from individual subjects (CVS and CVVS pooled together). Table 11 shows the comparison of lag duration for H and HL tones across different syllable types.

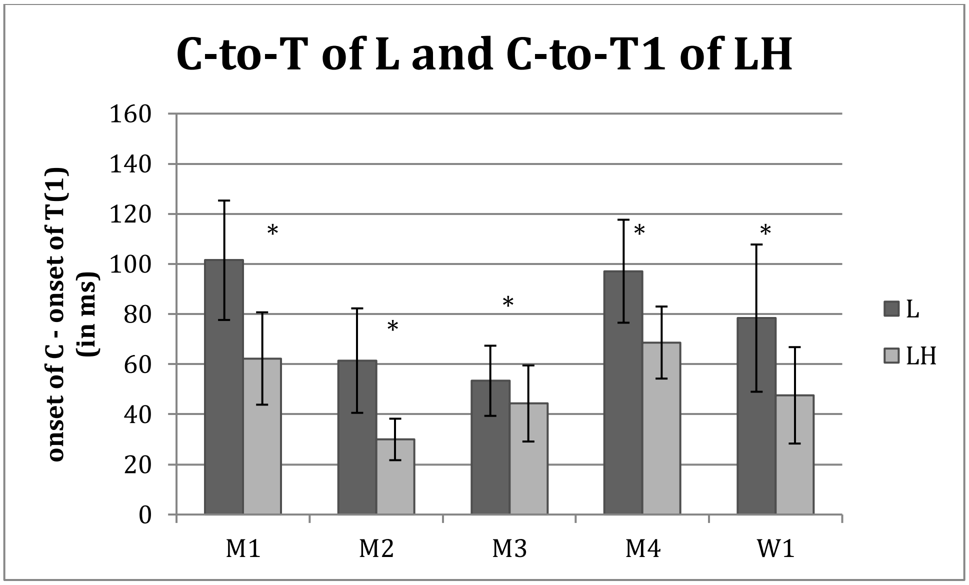


Fig. 20. C-to-T lags of L tone and C-to-T1 lags of LH across different syllable types from individual subjects (ms).

	L in CVS and CVVS		LH in CVS and CVVS		
speaker	M	SD	M	SD	
M1	101.47	23.82	62.27	18.43	p < .001
M2	61.43	20.85	29.97	8.28	p < .001
M3	53.37	14.00	44.32	15.19	p < .05
M4	97.10	20.55	68.63	14.38	p < .001
W1	78.38	29.38	47.57	19.24	p = .001

Table 11. Comparison of lag duration across different syllable types (L and LH tones)

Unlike the results from the comparison the C-to-T lag of H tone and C-to-T1 lag of HL in the previous section, the comparison between C-to-T lag of L and C-to-T1 lag of LH shows more uniformity among speakers. For all subjects, it was found that the tone type has an effect on the lags: $t(23) = 6.71$, $p < .001$ for M1, $t(23) = 8.06$, $p = .001$ for M2, $t(23) = 2.33$, $p = .05$ for M3, $t(23) = 5.55$, $p < .001$ for M4, and $t(23) = 4.01$, $p = .001$ for W1. The results suggest that the T1 of LH starts before the L tone, thus support Hypothesis C.

4.8 Effects of preceding tone contexts

A follow-up paired-samples t-test was conducted to compare the mean C-to-T(1) lags in both preceding T contexts, (subjects and syllable types were pooled together for each preceding T context) for each subject, and it was found that the mean C-to-T(1) lag in PRECEDING H context is significantly larger than the mean C-to-T(1) lag in PRECEDING L context for individual subjects: $t(95) = 23.28$, $p < .001$ for M1, $t(95) = 13.14$, $p < .001$ for M2, $t(95) = 12.09$, $p < .001$ for M3, $t(95) = 18.62$, $p < .001$ for M4, and $t(95) = 22.60$, $p < .001$ for W1.

Fig. 21 shows the C-to-T(1) lags in PRECEDING H context and the C-to-T(1) lags in PRECEDING L context in milliseconds from individual subjects (all tones and syllable types pooled together for each preceding T context). Table 12 shows the comparison of lag duration for all tones and syllable types across different preceding T contexts.

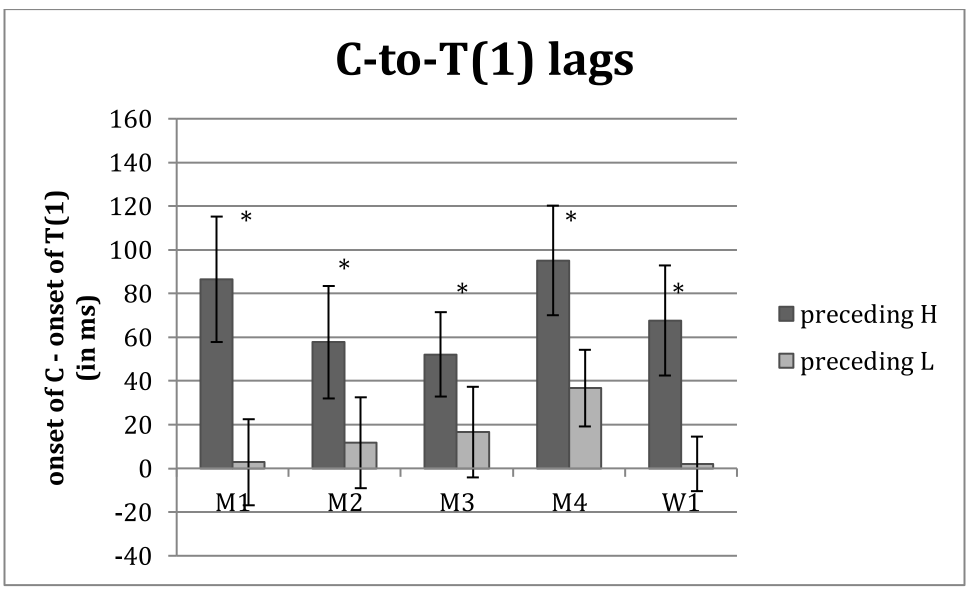


Fig. 21. C-to-T(1) lags of all tones and syllable types across different preceding T contexts from individual subjects (ms).

	PRECEDING H		PRECEDING L		
speaker	M	SD	M	SD	

M1	86.52	28.71	2.80	19.69	$p < .001$
M2	57.73	25.73	11.73	20.80	$p < .001$
M3	52.18	19.31	16.60	20.74	$p < .001$
M4	95.16	25.08	36.72	17.53	$p < .001$
W1	67.68	25.18	2.06	12.48	$p < .001$

Table 12. Comparison of lag duration across different syllable types (all tones and syllable types)

5. Discussion

The purpose of the present study was to examine the temporal organization between initial C gestures and T gestures in Thai. In particular, this experiment evaluates the predictions of the C-center effect in the temporal organization between segments and lexical tones hypothesized by Gao (2008, 2009).

5.1 Descriptive generalizations of results

Some generalizations can be drawn from the results. First, results suggest that within a syllable, the initial C gesture begins before the T gesture when the preceding T gesture is H, but when the preceding T gesture is L, the C gesture and the T gesture begin at the same time. This is schematized in Fig. 22.

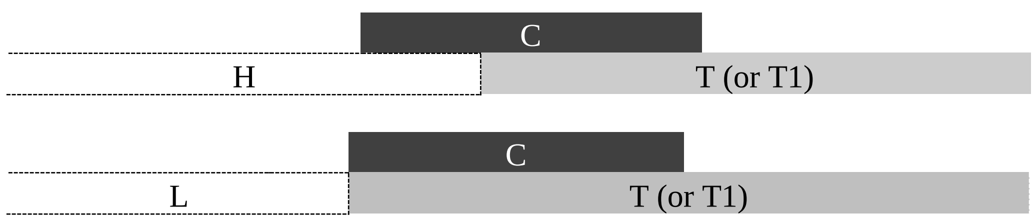


Fig. 22. The distance between the onset of the C gesture and the onset of the T or T1 gesture is large when the preceding tone gesture is H.

Second, in a long syllable, the T2 gesture of LH begins later with respect to the initial C gesture in long syllables when compared to the T2 gesture of LH in a short syllable. This is schematized in Fig. 22a. However, the T2 gesture of HL in long and short syllables begin at the same time with respect to the initial C gesture for three out of five subjects, as illustrated in Fig. 22b pattern A, for the other two subjects, the T2 gesture of HL in long syllables begins after the T2 gesture of HL in a short syllable (For M4, the T2 gesture in CVS begins later than the T2 gesture in CVVO and CVVS. For W1, the T2 gesture in CVS begins later than the T2 gesture in CVVS not CVVO.) This is schematized in Fig. 23b pattern B.

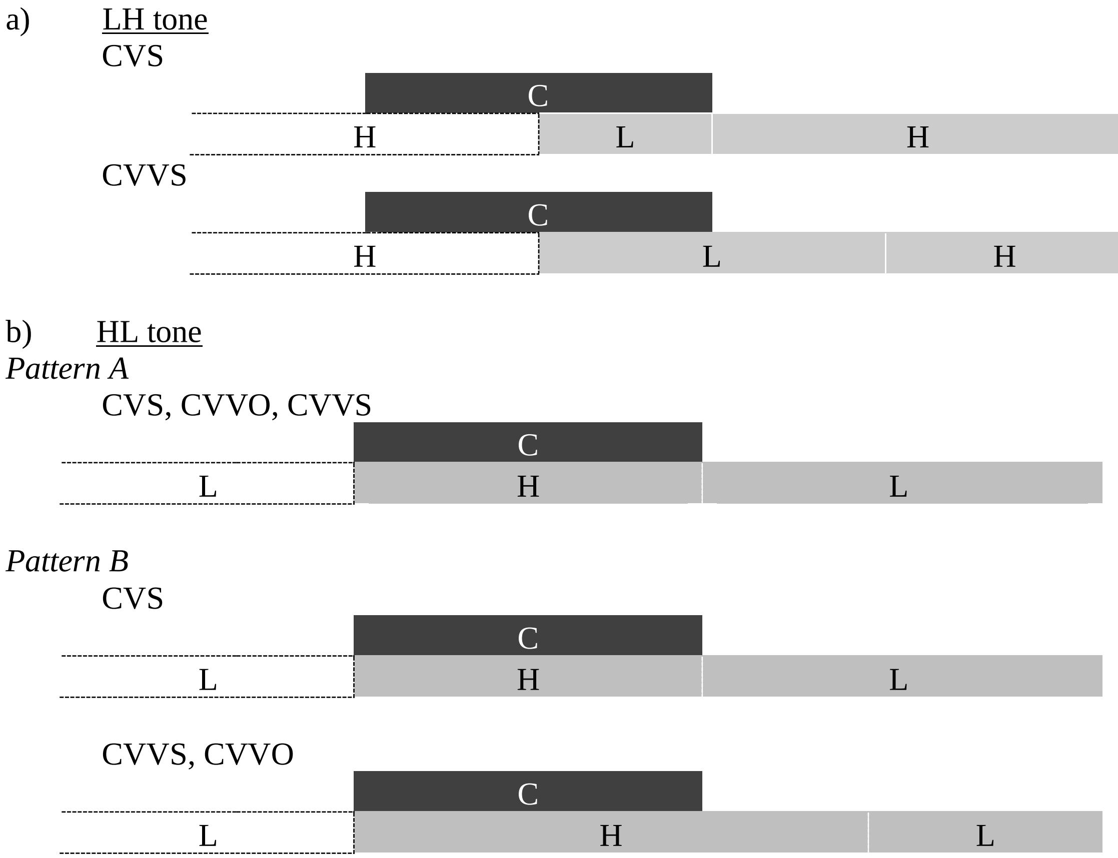


Fig. 23. The onset of the T2 gesture in LH is delayed with respect to the C gesture in long syllable (a). For HL tone however there are two patterns (b). For three out of five subjects, the T2 gesture of HL tone begins at the same with respect to the C gesture across three syllable types (pattern A). For the other two subjects, the T2 gesture of HL tone is delayed with respect to the C gesture in long syllable (pattern B).

Third, the T1 gesture in LH begins before L tone with respect to the initial C gesture. This is schematized in Fig. 24a. However, the T1 gesture in HL begins at the same time as H tone with respect to the initial C gesture for three subjects out of five (Fig. 24b pattern A); for the other two subjects, the T1 gesture of HL begins before H tone (Fig. 24b pattern B).

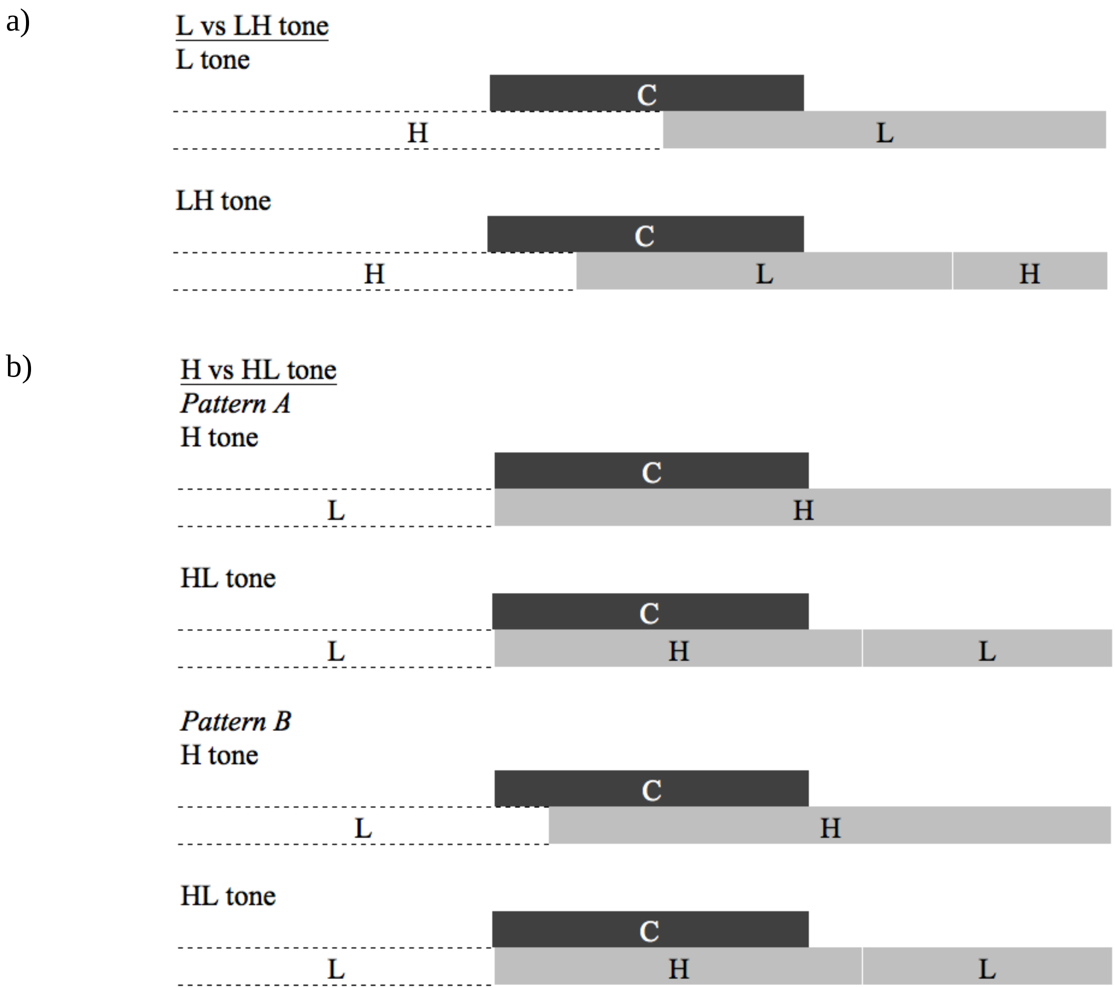


Fig. 24. The T2 gesture of LH begins later than L tone with respect to the C gesture (a). For HL tone however two patterns were observed among the subjects (b). For three out of five subjects, the T2 gesture of HL tone and H tone begin at the same time with respect to the C gesture (pattern A); for the other two subjects, the H tone begins later than the T2 gesture of HL with respect to the C gesture. (pattern B)

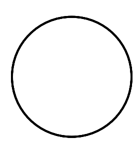
5.2 No C-center effect

If Thai exhibits C-center effect between the initial C gesture and T gesture(s), then the organization between the initial C gesture and T gesture(s) should always be the same regardless of syllable type. However, results show that the temporal organization between the initial C gesture and T gesture(s) can sometimes be affected by syllable type. Thus, this suggests that there is no C-center-like effect in the temporal organization between segments and tones in Thai. The temporal organization appears to be influenced by two factors. The first factor is the preceding T context. The initial C and T gestures do not always exhibit the same temporal organization: depending on the preceding tone context, the T gesture can either be realized at the same time as the C gesture or after the C gesture. Xu (1997) also has similar observation in his study on Mandarin where it was observed that the beginning time of tones is dependent on the offset of the preceding tone. In addition, results suggest that the unfixed organization is not influenced by the T gestures in target syllables *per se* but is rather influenced by the *interaction* between preceding T gestures and the T gestures of target syllables. This is evident in M tone which was placed in both preceding T contexts. The M tone in PRECEDING H context begins later than the M tone in PRECEDING L context. Other tones that were placed in PRECEDING H context (L and LH) behave similarly to the M tone in PRECEDING H context, while the tones that were placed in PRECEDING L context (H and HL) behave similarly to the M tone in PRECEDING L context.

The second factor that may have influenced the unfixed organization between the initial C gesture and T gestures is syllable structure. Although the results suggest that syllable structure does not have an effect on the organization between the C gesture and T(1) gesture, it was found that syllable structure has an effect on the organization between the initial C gesture and *T2* gesture. For LH tone, the T2 gesture in long syllables begins later than the T2 gesture in short syllables with respect to the initial C gesture. Thus, the T2 gesture in contour tones does not exhibit a fixed relationship with C and T1.

The observation that the preceding H tone has an effect on the onset of the T gesture in target syllables can be explained by the peak delay phenomenon. In both lexical tone and intonational languages, the phenomenon is when f0 peak of a high tone is realized after the syllable with which it associates. (Xu, 2001) This usually occurs when the syllable that carries a high tone is short in duration and thus does not have enough time for the tone to be realized within its boundaries; as a result, the f0 peak of the tone finishes its realization on the following syllable. (Xu, 2001) Although the preceding T gestures in the present study occurred on a long syllable (CVVC), the H tone nevertheless was observed to be delayed, possibly because of the position they were in. The syllables that preceded the target syllables were all phonetically shortened because they were all unstressed. (Recall that unstressed syllables in Thai occur in non-final positions.)

It is possible that the peak delay does not truly affect the beginning times of the T gestures in the target words. Recall that the present study assumes that the turning point of f0 in the acoustic data indicates the beginning time of each T gesture. This assumption implies that tones are realized in sequence and therefore do not overlap articulatorically. This might be an incorrect assumption, however. The turning point of f0 may not completely correspond to the activation time of laryngeal muscles which are responsible for tonal production. It is therefore possible that all tonal gestures in the target words do in fact begin at the same time; the preceding H tone may be overlapped with the tonal onsets in the target words because H tone requires more time to reach its tonal target—it takes longer time to realize than preceding L tone. This is illustrated in Fig. 25.



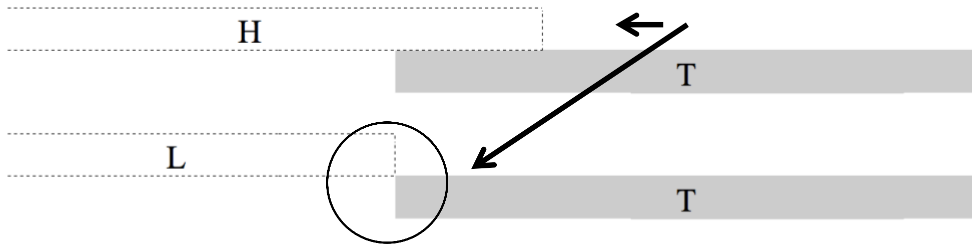


Fig. 25. It is possible that the T gestures in the target words begin at the same time, but the preceding H is partially overlapped with the onsets of the tones. The turning points of f0 might not reflect the activation time of tones articulatorically.

Although it is difficult to measure the laryngeal muscles which are responsible for tonal production (Gao 2008), more research on gestural overlapping in tones is needed. I will leave this issue for future research.

The present study conjectures that the observed C-center effect in Gao (2008, 2009) may have been influenced by the preceding tone context. In her experiments, tone 1 (H) and tone 4 (HL) were used as preceding tones. These two tones occurred on short syllables: C[i]. All target syllables and preceding syllables in her study were uttered with no focus. Xu (2001) reports that H tone in Mandarin can exhibit peak delay especially in fast speech. Although he reports that HL tone does not exhibit peak delay, it is however possible that the delay in Gao (2008, 2009) for HL tone may have been influenced by other factors. The delay of HL tone may have been motivated by the focus condition it was in. The fact that the preceding and the target syllables were produced under no focus may have phonetically shortened the syllable and thus triggers the delay.

There are possible explanations for why the T2 gesture of HL and the T2 gesture of LH do not behave the same way. First, the T2 gesture of each contour tone might be aligned to different locations within a syllable. Specifically, the T2 gesture of LH might be aligned to the coda C gesture or the syllable offset in general, whereas the T2 gesture of f0 of HL is not. In coupling terms, perhaps one can assume that the T2 gesture of LH shares a coupling relation with the coda C gesture, whereas the T2 gesture of HL does not—perhaps the latter shares a coupling relation with the initial C gesture. Xu (2001) observed that tone 2 (rising tone) in Mandarin, which he assumes to be a single unit (R), is aligned to the syllable offset because the tone moves rightwards when the syllable becomes lengthened. It is therefore possible that the T2 gesture of LH tone is aligned to the syllable offset in Thai.[5] Whether it is aligned to the coda C gesture or syllable offset in general, I will leave this for future research. Second, it has been observed by previous studies (Morén & Zsiga, 2006; Sundberg, 1973, 1979; Xu, 1999; Zhang, 2002) that a pitch rise takes more time to be implemented than a pitch fall. Thus, it is possible that in longer syllables the LH requires more time to transition from T1 gesture to T2 gesture than does HL. Third, the difference between LH and HL in this context may be affected by the fact that they were located in a sentence-final position. Lexical tones have been reported to behave differently when they are located at the end of phrases because lexical tones interact with boundary tones. (Ladd, 2008) Other factors such as final lengthening may have an effect on both tones in different ways as well. I will leave this for future research.

5.2 T and early T1

It is unclear why the T1 gesture of HL behaves differently from the T1 gesture of LH. Specifically, the present study cannot explain why the T1 gesture of LH is realized earlier than L tone with respect to the initial C gesture across all subjects, while the T1 gesture of HL is realized earlier than H tone for only two subjects (M3 and M4); the T1 gesture of HL for other subjects is realized at the same time as H tone. There might be a

[5] Note that Gao (2008, 2009) assumes that the dip in tone 2 in Mandarin is a consequence of the L gesture terminating. In other words, she did not assume that the dip is where T2 begins, as can be seen in Fig 3b.

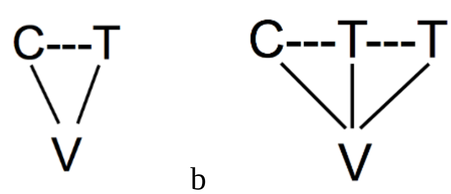
correlation between the early T1 gesture and the delay of preceding T gesture. In other words, it might be possible that there is an unknown variable that influences both the delay in peak delay and the early realization of T1 gesture of HL. Although the data from M3 and M4 show that they produced early T1 for HL, it was also observed that their overall C-to-T(1) lags are larger than the lags from other subjects. This is merely a speculation, however. No systematic comparisons were made; therefore, it cannot be concluded that there is a correlation between tonal delay and early T1. Further study on early T1 is needed.

5.5 How does V behave with respect to the organization between initial Cs and Ts?

The present study cannot explain why the onset of the V gesture begins in between the onset of the C gesture and the onset of the T gesture in Gao's (2008, 2009) experiments. Because the onset of the V gesture is rather difficult to locate in the acoustic, given that it is overlapped with the C gesture, this study did not measure the onset of the V gesture. Further investigation is needed for future research.

5.6 Implications for the coupled oscillator planning model of speech production

The results from the present study suggest that the models proposed by Gao (2008, 2009) in Fig. 4, repeated below as Fig. 26, cannot be extended to Thai.



a b

Fig. 26. Specified coupling relations by coupling graphs for Pattern I (a) and Pattern II (b) in Mandarin.

The proposed models predict that the onset of the C gesture begins before the onset of T(1) gesture. The models also imply that this organization is stable. However, results in the present study show that the onset of the C gesture does not always begin before the onset of the T(1) gesture; the onset of the C gesture can sometimes begin at the same time as the onset of the T(1) gesture. The models thus would wrongly predict the Thai data. An adequate planning model for Thai is needed. I will leave this for future research.

6. Conclusion

In sum, the present study investigated the organization between initial consonants and tones within a syllable in Thai and did not find the C-center effect in the temporal organization between the consonant and the tones. The findings suggest that preceding tones and syllable type affect the organization. A unified account of the temporal organization between segments and tones is needed. Hence, more research—not only on the temporal organization among initial consonants, vowels, and tones, but also the organization among coda consonants, vowels, and tones—is needed.

References

Abramson, A. (1962). The vowels and tones of Standard Thai: Acoustical Measurements and Experiments. *International Journal of American Linguistics*, *28*(2), part II.

Abramson, A. (1975). The Tones of Central Thai: Some Perceptual Experiments. In J. G. Harris & J. R. Chamberlain (Eds.), *Studies in Thai Linguistics in Honor of William J.Gedney* (pp. 1–16). Bangkok: Central Institute of English Language, Office of State Universities.

Abramson, A. (1979). The Coarticulation of Tones: An Acoustic Study of Thai. In T. L. Thongkum, V. Panupong, P. Kullavanijaya, & M.R.K. Tingsabadh (Eds.), *Studies in Tai and Mon-Khmer Phonetics and Phonology in honor of Euge ́nie J. A. Henderson* (pp. 1–9). Bankok: Chulalongkorn University Press.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International*, *5*(9/10), 341–345.

Browman, C. P., & Goldstein, L. (1986). Towards an articulatory phonology. *Phonology Yearbook*, *3*, 219–252.

Browman, C. P., & Goldstein, L. (1988). Some Notes on Syllable Structure in Articulatory Phonology. *Phonetica*, *45*(140-155).

Browman, C. P., & Goldstein, L. (1989). Articulatory Gestures as Phonological Units. *Phonology*, *6*(201-251).

Browman, C. P., & Goldstein, L. (2000). Competing Constraints on Intergestural Coordination and Self-organization of Phonological Structures. *Bulletin de La Communication Parle*, *5*(25-34).

Byrd, D. (1995). C-Centers revisited. *Phonetica*, *52*, 263–282.

Erickson, D. (1974). Fundamental Frequency Contours of the Tones of Stan- dard Thai. *Pasaa*, *4*, 1–25.

Gandour, J. (1974a). On the Representation of Tone in Siamese. In J. G. Harris & J. R. Chamberlain (Eds.), *Studies in Tai Linguistics in Honor of William J. Gedney* (pp. 170–195). Bangkok: Central Institute of English Language.

Gandour, J. (1974b). The Glottal Sop in Siamese: Predictability and Phono- logical Description. *UCLA Working Papers in Phonetics*, *27*, 84–91.

Gandour, J. (1977). On the Interaction Between Tone and Vowel Length: Evidence from Thai Dialects. *Phonetica*, *34*, 54–67.

Gandour, J. (1979). Tonal rules for English loanwords in Thai. In T. L. Thongkum, P. Kullavanijaya, P. V. Panupong, & K. Tingsabadh (Eds.), *Studies in Tai and Mon-Khmer Phonetics and Phonology in Honour of Eugénie J. A. Henderson* (pp. 94–105). Bangkok: Chulalongkorn University Press.

Gandour, J., Potisuk, S., & Dechongkit, S. (1994). Tonal Coarticulation in Thai. *Journal of Phonetics*, *22*, 474–492.

Gandour, J., Potisuk, S., Ponglorpisit, S., & Dechongkit, S. (1991). Inter- and Intraspeaker Variability in Fundamental Frequency of Thai Tones. *Speech Communication*, *10*, 355–372.

Gao, M. (2008). *Tonal Alignment in Mandarin Chinese: An Articulatory Phonology Account*. *Yale University unpublished dissertation*. Yale University unpublished dissertation.

Gao, M. (2009). Gestural Coordination among Vowel, Consonant and Tone Gestures in Mandarin Chinese. *Chinese Journal of Phonetics*, Beijing: Commercial Press.

Goldstein, L., Byrd, D., & Saltzman, E. (2006). The role of vocal tract gestural action units in understanding the evolution of phonology. In M. Arbib (Ed.), *From action to language: The mirror neuron system* (pp. 215–249). Cambridge: Cambridge University Press.

Goldstein, L., Chitoran, I., & Selkirk, E. (2007). Syllable structure as coupled oscillator modes: evidence from Georgian vs. Tashlhiyt Berber. In *Proceedings of the XVI International Congress of Phonetic Sciences* (pp. Saarbrücken, Germany, 241–244).

Goldstein, L., Nam, H., Saltzman, E., & Chitoran, I. (2009). Coupled oscillator planning model of speech timing and syllable structure. In G. Fant, H. Fujisaki, & J. Shen (Eds.), *Frontiers in Phonetics and Speech Science* (pp. 239–250). Beijing: The Commercial Press.

Honorof, D. N., & Brownman, C. P. (1995). The center of edge: how are consonant clusters organised with respect to the vowel? In K. Elenius & P. Branderud (Eds.), *Proceedings of the XIIIth International Congress of Phonetic Sciences* (pp. 3, 552–555). Stockholm, Sweden: Congress Organisers at KTH and Stockholm University.

Kenstowicz, M., & Suchato, A. (2006). Issues in loanword adaptation: A case study from Thai. *Lingua*, *116*(7), 921–949.

Ladd, D. R. (2008). *Intonational phonology* (2nd ed.). Cambridge: Cambridge University Press.

Leben, W. (1971). On the Segmental Nature of Tone in Thai. *Quarterly Progress Report, Research Laboratory of Electronics*, *101*, 221–224.

Leben, W. (1973). *Suprasegmental Phonology*. MIT.

Marin, S., & Pouplier, M. (2008). Organization of Complex Onset and Codas in American English: Evidence for a Competitive Coupling Model. In *Proceedings of the 8th ISSP* (pp. Strasbourg, France, 437–440).

Marin, S., & Pouplier, M. (2014). Articulatory synergies in the temporal organization of liquid clusters in Romanian. *Journal of Phonetics*, *42*, 24–36.

Morén, B., & Zsiga, E. (2006). The Lexical and Post-lexical Phonology of Thai Tones. *Natural Language and Linguistic Theory*, *24(1)*(113-178).

Nam, H., Goldstein, L., & Saltzman, E. (2009). Self-organization of syllable structure: A coupled oscillator model. *Approaches to Phonological Complexity*, 299–328.

Nam, H., & Saltzman, E. (2003). A competitive, coupled oscillator model of syllable structure. In *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 2253–2256).

Potisuk, S., Gandour, J., & Harper, M. (1997). Contextural Variations in Trisyllablic Sequences of Thai Tones. *Phonetica*, *54*, 22–42.

Shaw, J. A., Gafos, A., Hoole, P., & Zeroual, C. (2009). Syllabification in Moroccan Arabic: evidence from patterns of temporal stability in articulation. *Phonology*, *26*, 187–215.

Sundberg, J. (1973). Data on maximum speed of pitch changes. In *Speech Transmission Laboratory Quarterly Progress and Status Report* (Vol. 4, pp. 39–47). Stockholm, Sweden: Royal Institute of Technology.

Sundberg, J. (1979). Maximum speed of pitch changes in singers and untrained subjects. *Journal of Phonetics*, *7*, 71–79.

Thepboriruk, K. (2009). Bangkok Thai tones revisited. *JSEALS*, *3*, 86–105.

Woo, N. (1969). *Prosody and Phonology*. MlT.

Xu, Y. (1997). Contexual tonal variations in mandarin. *Journal of Phonetics*, (25), 61–83.

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

Xu, Y. (2001). Fundamental Frequency Peak Delay in Mandarin. *Phonetica*, *58*, 26–52.

Yip, M. (1982). Against a Segmental Analysis of Zahao and Thai: A Laryngeal Tier Proposal. *Linguistic Analysis*, *9,* 79–94.

Zhang, J. (2002). *The effects of duration and sonority on contour tone distribution: a typological survey and formal analysis*. New York: Routledge.

Zsiga, E., & Nitisaroj, R. (2007). Tone features, tone perception, and peak alignment in Thai. *Language and Speech*, *0*(3), 343–383.