The Articulatory TBU: Gestural Coordination of Lexical Tone in Thai

Robin Karlin[*]

1 Introduction

Recent literature in Articulatory Phonology (AP) has promoted the idea that tone is an articulatory gesture, similar to those that make up consonants and vowels (Prieto and Torreira 2007, Gao 2008). When treated as gestures, tones should be co-selected and coordinated like other articulatory gestures. The co-selection of T(one) gestures with C(onsonant) and V(owel) gestures is reminiscent of the autosegmental "association" of tone with tone-bearing unit (TBUs); however, work with tone in the framework of AP has not formalized the articulatory conceptualization of the TBU. In this paper, I argue that a TBU is an articulatory gesture that T gestures can coordinate with. I also argue that moras correspond to groups of coordinated articulatory gestures, and that these mora-sized sets of gestures are coordinated with each other. In some languages, tone is coordinated within the group of articulatory gestures. I present evidence from an articulatory and acoustic study on Thai, a tonal language that uses the mora as its TBU.

The rest of the paper is organized as follows: in Section 2, I provide the theoretical background for this study, as well as the hypotheses and predictions under the current framework. In Section 3, I describe the methodology used to collect data. In Section 4, I present the data from the study. In Section 5, I discuss the data and its implications for the treatment of tone in AP.

2 The Articulatory TBU

2.1 Tone-Bearing Units

Probably the most widely accepted view of lexical tone in generative phonology is the autosegmental approach. In this theory, lexical tone is represented as a series of tone levels that are associated with tone-bearing units (TBUs). This is in contrast to a theory where each TBU is stored in the phoneme inventory multiple times, once for each tone value: /á/ and /à/, for example. The autosegmental approach additionally proposes that, instead of a detailed representation of the phonetic values, tones are made up of one or more abstract tone levels, usually H(igh) or L(ow), but sometimes including M(id). Simple tones have just one tone level, while contour tones are represented as a sequence of two or more tones—for example, a falling tone is represented as an HL sequence, and a rising tone as a LH sequence, where each level tone is associated with some segment.

The realization of tones on segments comes from the "association" of tones with TBUs. The two most popular candidates for TBUs are the syllable and the mora. Different languages are held to have different TBUs; for example, the TBU in Mandarin Chinese is argued to be the syllable, while the mora is more relevant for Japanese pitch accent (Yip 2002). In Thai, the language of study in this paper, there has been a debate as to which unit provides a more accurate representation. Most recently, Morén and Zsiga (2006) have argued for the mora as the TBU in Thai. The argument holds that monomoraic words can only carry simple tones, while bimoraic words can carry both simple and contour tones. Thus, words with two moras can carry lexical tones with two elements, while those with one mora can only carry singleton lexical tones (see Fig. 1).

Morén and Zsiga further demonstrated that in bimoraic words of shape CVː, the "inflection point" of falling tones (HL), where F0 stopped rising and started to fall, occurred halfway through the duration of the long vowel. They thus argued that in addition to the mora being the TBU in Thai, tone levels more specifically align to the mora's right edge. This accounts for why the peak of the falling tone is not at the beginning of the word, as might be expected, but rather at the end of the first mora. However, describing a tone as associated to the "right edge" of a TBU is more descriptive than

[*] Special thanks to Sam Tilsen for feedback and guidance, as well as to Draga Zec, Abby Cohn, Hao Yi, and the members of Research Workshop. All remaining errors are my own.

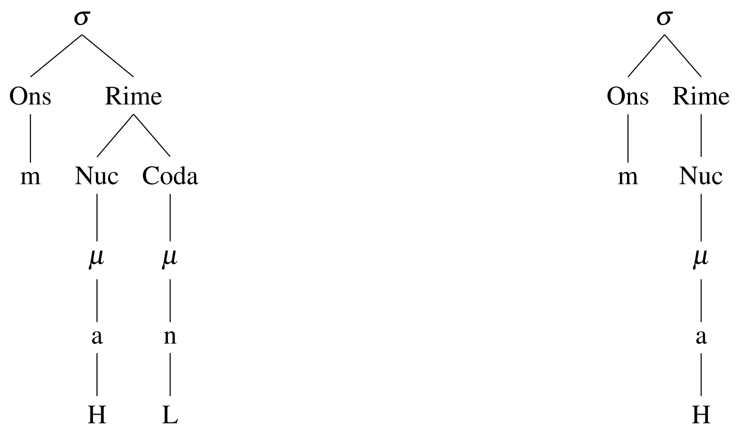


Figure 1: A basic illustration of two possible Thai syllables, with differing moraic weight. The syllable on the left, /man/, is bimoraic and can have a falling (HL) tone. The syllable on the right, /ma/, is monomoraic, and can only have a simple tone (H).

predictive; other than the F0 shapes that the account is based on, Morén and Zsiga (2006)'s analysis does not account for further data. On a broader theoretical scale, the autosegmental approach is problematic in that it does not concretely define what it means for a tone to "associate" with a TBU, and does not make a connection between abstract representation and implementation. The Articulatory Phonology (AP) framework, introduced in the following subsection, aims to link abstract representations with implementation, and thus makes explicit predictions regarding the relationships between tones and their TBUs.

2.2 Articulatory Phonology

In AP, the fundamental unit of information is the gesture (Browman and Goldstein 1989, 1992), where a gesture is the motor coordination of multiple articulators involved in achieving some specific task, such as closing the lips. In the specific gesture of bilabial closure, the jaw and upper and lower lips all coordinate to achieve closure. At one step higher in the coordinative hierarchy, gestures are coordinated to form a segment-like phonological unit. For example, the phoneme /m/ is composed of three gestures: lip constriction (bilabial closure), velar lowering (nasality), and glottal adduction (voicing). These so-called "constellations" of gestures (Browman and Goldstein 1989) can further coordinate with other constellations to form more even more complex phonological units, such as the mora or prosodic phrase. These layers of coordination point to a hierarchical organization of language production, from the gesture to the utterance, and reflects the hierarchical structure of phonology similar to those described in prosodic theory.

AP draws from general human motor control to describe the patterns observed in speech articulation. There are two major motor control regimes: coordination and (competitive) selection (Tilsen 2014, Browman and Goldstein 1992). Coordination is the more complex mode of control, and is acquired later in both speech and in other motor activity (Tilsen 2014, Jeannerod 1986). When gestures are coordinated, they are selected and activated within one selection event, forming a bundle of coordinated gestures called a "co-selection set" (Tilsen 2014). That is to say, feedback is not necessary to trigger the onset of a "second" gesture after a "first" gesture; the gestures within a co-selection set are activated before the controller receives sensory feedback of the achievement of any one gesture.

Within a co-selection set, gestures can be in-phase coordinated or anti-phase coordinated. Of the two types of coordination, in-phase is the more stable. For the purposes of this paper, in-phase coordination refers to the *onsets* of two gestures occurring at the same time. Anti-phase coordination, on the other hand, refers to the onsets of two gestures occurring 180° out of phase with each other; that is, for two gestures with the same oscillation period and in the absence of other coupling interactions, the onset of one gesture would be timed with the offset of the other, assuming that the onset and offset of a gesture are separated by a half-period of the gesture's virtual cycle. The classic example of in-phase coordination within AP is the coordination of a CV syllable. In this case, the co-selection set for the {C} and the co-selection set for the {V} are initiated at the same time, forming a larger co-selection set for the CV syllable. In contrast, the co-selection sets {V} and {C} of a VC syllable are anti-phase coordinated: the {C} co-selection set starts at the offset of the {V} co-selection set, again forming a larger co-selection set for the VC syllable. Anti-phase coordination can give the impression of sequential events, or competitive selection. However, the contrast lies in that {C} is not activated *as a result of* the {V} achieving its target—rather, the 180° phasing is determined prior to the activation of the {VC} co-selection set.

Gestural coordination lies in direct contrast with competitive selection, where feedback does trigger the activation of the next co-selection set. Competitive selection is the less complex mode of combination, and involves the sequential activation of co-selection sets. "Competitive" in this sense refers to the availability of all co-selection sets at that level, with one "winner" given priority and activated first. When the first co-selection set reaches a certain threshold—either achieving the task, or even the offset of the task—the next co-selection set is fully activated. One example is the implicit assumption of competitive selection-like sequencing of co-selection sets for syllables in a word, and an explicit model was proposed by Tilsen (2013). Under this assumption, the word "mommy" [mami] would have two syllable-sized co-selection sets: the {CV} set that produces [ma], and the {CV} set that produces [mi]. The gestures within each set are in-phase coordinated, as described above; the two syllables in the word "mommy" are organized in a selectional sequence so that [ma] is produced before [mi], as illustrated in 2 below.

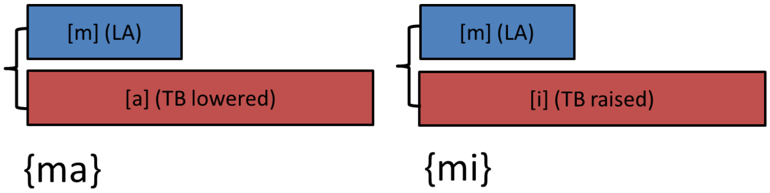


Figure 2: A gestural score of [mami], using lip aperture (LA) to represent [m] and tongue body height (TB) for vowels. The C and V of [ma] and [mi] are in-phase coordinated, while the syllables are competitively selected.

Research in the AP literature has largely focused on the coordinative structure of consonants and vowels. In the next subsection, I discuss the addition of lexical tone to the inventory of articulatory gestures.

2.3 Tone as a Gesture

The AP representation of segmental organization is not yet fully developed; however, the organization of tones is even less so. Recent work (Mücke et al. 2011, Prieto and Torreira 2007, Prieto et al. 2007) in AP has promoted the idea of tones as gestures, similar to those involved in consonants and vowels. Prieto and Torreira (2007) examined the timing of pitch accent peaks in Spanish, and found that the timing relationships between articulatory landmarks and F0 trajectories were more robust than those between acoustic landmarks and F0. However, Gao (2008) was the first to run a full articulatory study on lexical tone, and treated tone as a gesture with an F0 target. Gao bases her analysis of Mandarin Chinese on the assumption that the gestures for tone are very similar to those proposed in autosegmental theory—that is, H(igh) and L(ow)[1]. The key finding in her study was

[1] In this paper I do not include M(id), as it is currently unnecessary; however, a M gesture could very well be necessary in another language.

that tone does demonstrate a consistent timing relationship with consonant and vowel gestures; in particular, tone behaves as an additional consonant gesture.

The evidence for this claim lies in the coupling relationships between the onset, nucleus, and tone: instead of the tone gesture being simply "laid over" the TBU, tone coordinates with the {C} and {V} co-selection sets of the CV syllable, resulting in a pattern similar to that found in C_1C_2V syllables. As previously described, {C} and {V} are in-phase coordinated with each other to make a CV syllable. However, when there is a complex (cluster) onset, the phasing relationships are more complex. Both {C} gestures are attracted to the more stable in-phase relationship with {V}; however, they are repelled from each other and in anti-phase coordination, possibly to enhance the perceptual recoverability of each segment (Marin and Pouplier 2010). The result of these competing forces is a pattern called the C-center effect, where the order of onsets is C_1, V, C_2 (see Fig. 3). This arrangement reflects the minimization of the potential forces that describe pairwise coupling relationships between the consonant and vowel gestures.

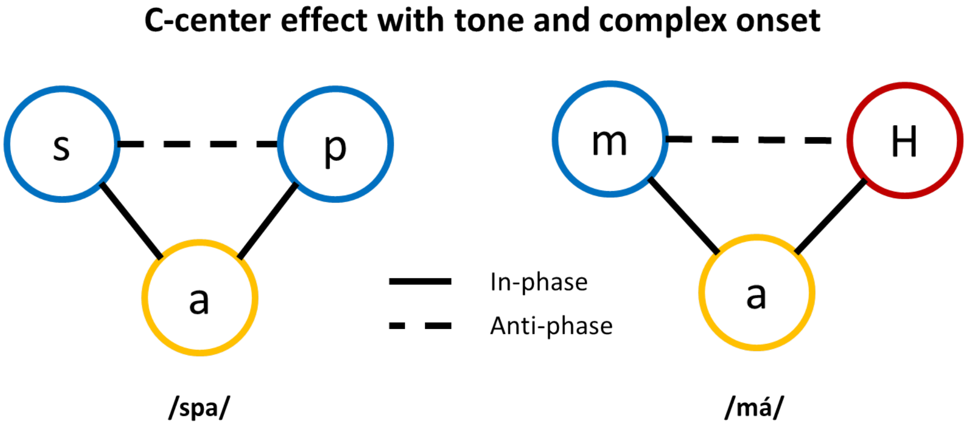


Figure 3: Coupling graphs of a CCV syllable and a CV syllable with tone. Consonants and tones are in-phase coupled with vowels, and anti-phase coupled with each other

Gao (2008) found that the {C}, {V}, and {T} co-selection sets exhibit a C-center effect—that is, the onset of the {V} gesture is in the center of the onsets of the {C} and {T} gestures, with {T} acting as C_2. The clearest example of the C-center effect is tone 1, or the level high tone, illustrated in Fig. 4.

The more complicated relationships are the contour tones, or tones 2 (rising) and 4 (falling)[2]. For tone 2, Gao proposed that the L and H gestures are activated at the same time within one co-selection set (see Fig. 5). However, she argues that tone 4 is composed of an H and L gesture anti-phase coordinated with each other (see Fig. 6). Gao does not make a distinction between sequential selection and coordination, instead conflating anti-phase coordination and sequential selection.

Gao further attempts to extend her analysis to Thai tones, in which the falling tone is also composed of an H and L in anti-phase coordination, or what she describes as a "sequential" arrangement. Gao assumes a syllable-sized domain for tone gesture coordination in Thai, the same as in her analysis of Mandarin; she makes no reference to the mora despite very good evidence that moras are greatly relevant to tone in Thai. However, AP in its current state does not have a well-developed notion of what the mora is. In the following section, I present the argument for articulatory moraic structure proposed by Tilsen (2014), as it applies to Thai in particular.

[2] Gao does not examine the third tone (dipping) in non-final position. Thus instead of a falling-rising contour, she only examines tone 3 as a low tone, in which the only F0 gesture is a L gesture. Tone 3 exhibits the same coordinative patterns as tone 1; that is, tone 3 demonstrates a C-center effect

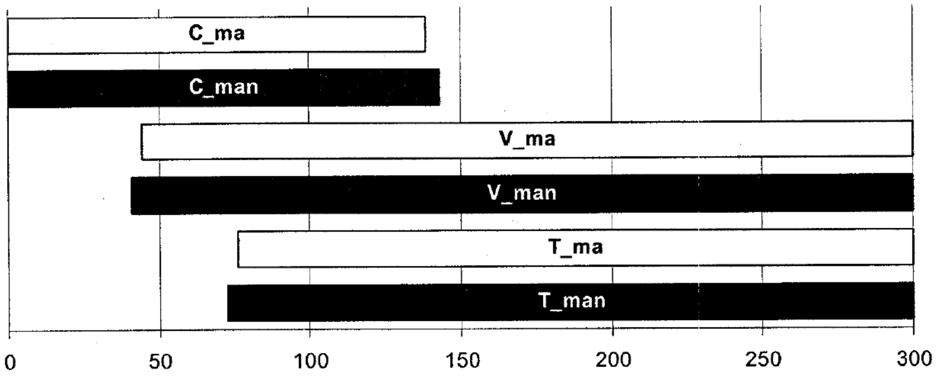


Figure 4: Gestural scores of the $\{CVT\}$ co-selection set of /ma/ and /man/ with tone 1 (Gao 2008). For both syllable types, the onset of the *V* gesture is between the onsets of the *C* and *T* gestures.

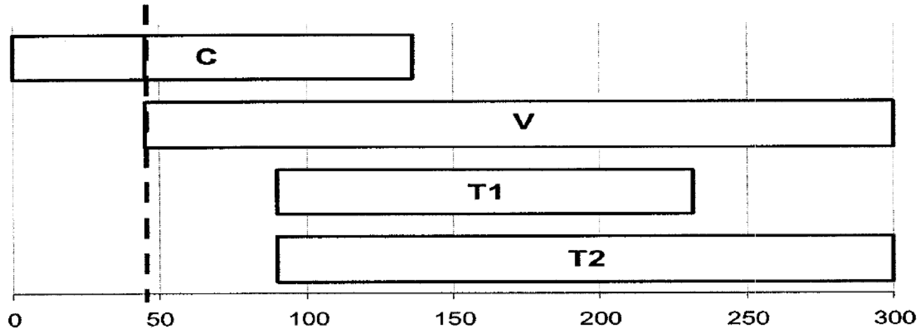


Figure 5: Gestural scores of /ma/ with tone 2 (Gao 2008). For this tone, Gao proposes that T1 and T2 are in-phase coupled with each other.

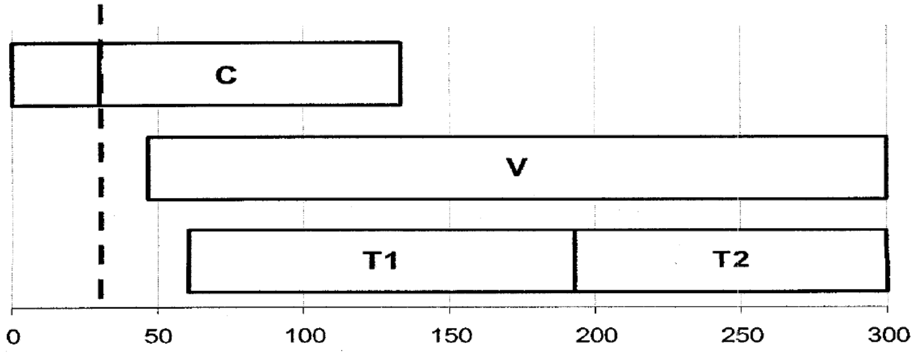


Figure 6: Gestural scores of /ma/ with tone 4 (Gao 2008). For this tone, Gao proposes that T1 and T2 are anti-phase ("sequentially") coupled with each other.

2.4 The Articulatory Mora

This paper does not aim to discard the mora; rather, this interpretation provides a conceptualization of the mora that is motorically grounded. Tilsen (2014) proposes that children first employ the easier mode of control, competitive selection, in order to build larger structures, such as the syllable, or a phrase. However, with additional practice, these competitively selected gestures are consolidated into coordinated co-selection sets. These co-selection sets are built from the bottom up; there is a co-selection set that corresponds with a segment, as well as co-selection sets that correspond with moras, phonological words, and other units in the phonological hierarchy. I argue that there are two co-selection sets in bimoraic words, which correspond to previously proposed groupings of moraic and non-moraic segments.

When examining Thai phonological patterning, it is easy to see what a mora is not. There is not, for example, a one-to-one correspondence from segmental level co-selection sets to moras. Additionally, the correspondence obviously does not stem from the syllable level, which is predicted by previous accounts of hierarchical prosodic structure. The difficulty lies in establishing how the different levels of the prosodic hierarchy are reflected in coordination and competitive selection. As only the nucleus and coda of a syllable are candidates for the mora, as a starting point, I consider the patterning for that would result in a bimoraic structure for CVC syllables: {CV(V)}-{C}, where the onset and nucleus are coordinated as one co-selection set, and the coda C is the sole member of a second co-selection set. In Fig. 1 below, I show the results of a regular application of this rule to various syllable types in Thai:

Syllable Type	**Example**	**Moras**	**Co-selection sets**
CV	{ma}	1	1
CVO	{ma}-{t}	2	2
CVN	{ma}-{n}	2	2
CVV	{maa}	2	1
CVVO	{maa}-{t}	2	2
CVVN	{maa}-{n}	2	2

Table 1: Regular application of competitive selection between onset/nucleus co-selection set and coda co-selection set

The major problem treating nuclei and codas as the moras is evidenced by the long vowels. There are two logical possibilities: (1) a long vowel is one gesture that is extended in time, or (2) a long vowel is composed of the same gesture, selected twice. In Thai, the additional consideration of diphthongs indicates that (2) is the correct solution, as diphthongs pattern phonologically the same way as long vowels. As diphthongs have two distinct vowel qualities, they thus obligatorily have two vocalic gestures; both diphthong and long monophthong nuclei are bimoraic and allow contour tones. As Thai syllables are maximally bimoraic, there must also be a restriction on the moraicity of codas; specifically, codas following diphthongs and long vowels are not moraic. Gesturally speaking, consonantal coda gestures following a second vowel gesture are incorporated into the co-selection set headed by the second vocalic gesture. The previous division of Thai syllables is amended to {CV}-{(V)(C)} (see Fig. 2.), where phonological words always have two moraic co-selection sets.

There is still one complication: although *mat* is bimoraic, as evidenced by its status as a phonological word, it cannot hold a contour tone, such as falling (HL) or rising (LH). This restriction is not due solely to the voicelessness of the coda, as *maat* can hold a falling tone, nor to the singleton consonant acting as a mora, as *man* can carry both a falling and a rising tone. I will discuss the phonological patterning further in Section 5. In defining the articulatory moraic TBU, then, I argue that the gesture that corresponds to the moraic segment is what the tone gesture, along with other non-moraic segments, is first coordinated with. This coordination to a segmental gesture is the concrete implementation of tonal "association." This conceptualization of the TBU and of association makes a number of predictions, described in the following section.

Syllable Type	Example	Moras	Co-selection sets
CV	{ma}	1	1
CVO	{ma}-{t}	2	2
CVN	{ma}-{n}	2	2
CVV	{ma}-{a}	2	2
CVD	{mu}-{a}	2	2
CVVO	{ma}-{at}	2	2
CVVN	{ma}-{an}	2	2
CVDO	{mu}-{at}	2	2
CVDN	{mu}-{at}	2	2

Table 2: Final organization of moraic co-selectional events

2.5 Hypotheses and Predictions

The current study is designed to test timing relationships between consonant, vowel, and tone gestures in bimoraic words. Specifically, the study is designed to test for evidence of the proposed articulatory TBUs. The predictions are as follows:

1. **Verification of F0 as an articulatory gesture**. The initial problem for this study is verifying if F0 is indeed an articulatory gesture. If F0 is an articulatory gesture, similar to consonants and vowels, there will be consistent timing relationships between F0 and consonant and vowel gestures. For the relationship to be considered robust or stable, it must hold in the face of variables such as time pressure (speed), word shape, phrasal position, and environmental differences. In this study, I manipulate time pressure, word shape, and environment.

2. **T gestures will behave similarly to C gestures**. The first T gesture (*T1*) will pattern like an onset consonant, and there will be a C-center effect with *m*, *T1*, and the first V gesture (*V1*). The second T gesture (*T2*), in contrast, will pattern like a coda consonant. In line with Marin and Pouplier (2010)'s findings, *T2* will be in anti-phase coordination with the second TBU gesture; the patterning will be most visible in words with diphthong nuclei and *mân*, where the second TBU gesture is distinct from the first—*V2* and *n*, respectively. When another non-moraic coda is present, *T2* will be delayed with respect to the second TBU gesture.

3. **The existence of an articulatory TBU**.

   (a) **There is an articulatory TBU**. An articulatory TBU would be some gesture with which tone gestures consistently coordinate. In the case of Thai, the TBU gesture will be the gesture that corresponds to moraic segments—i.e., vowels and moraic consonants. In this particular study, the gestures will be organized into two mora-sized co-selection sets, and the gesture that corresponds to the moraic segment will act as the articulatory TBU. For diphthongs, they will be *V1* and *V2*; for long monophthongs, *V2* will not be visible in contrast to *V1*, but the same patterning as in diphthongs will be present. Further, the moraic *n* in *mân* will act as an articulatory TBU, while non-moraic codas will be coordinated to the mora gestures. This hypothesis does not include any immediate coordinative relationships between the T gestures. The proposed coupling relationships and predicted scores are in Fig. 7.

   (b) **T gestures are coordinated first with each other**. An alternative hypothesis is that T gestures are not primarily coordinated with C and V gestures, but rather with other T gestures. The coordinated T gestures could either additionally be coordinated with each articulatory TBU, or possibly at the syllable level, with T2 not coordinated with an articulatory TBU at all. In this case, the timing relationship between T gestures would be consistent across target words. Two possible coupling schemata for *mâan*, *mâat*, *mûan*, and *mûat* are illustrated in Fig. 8 below.

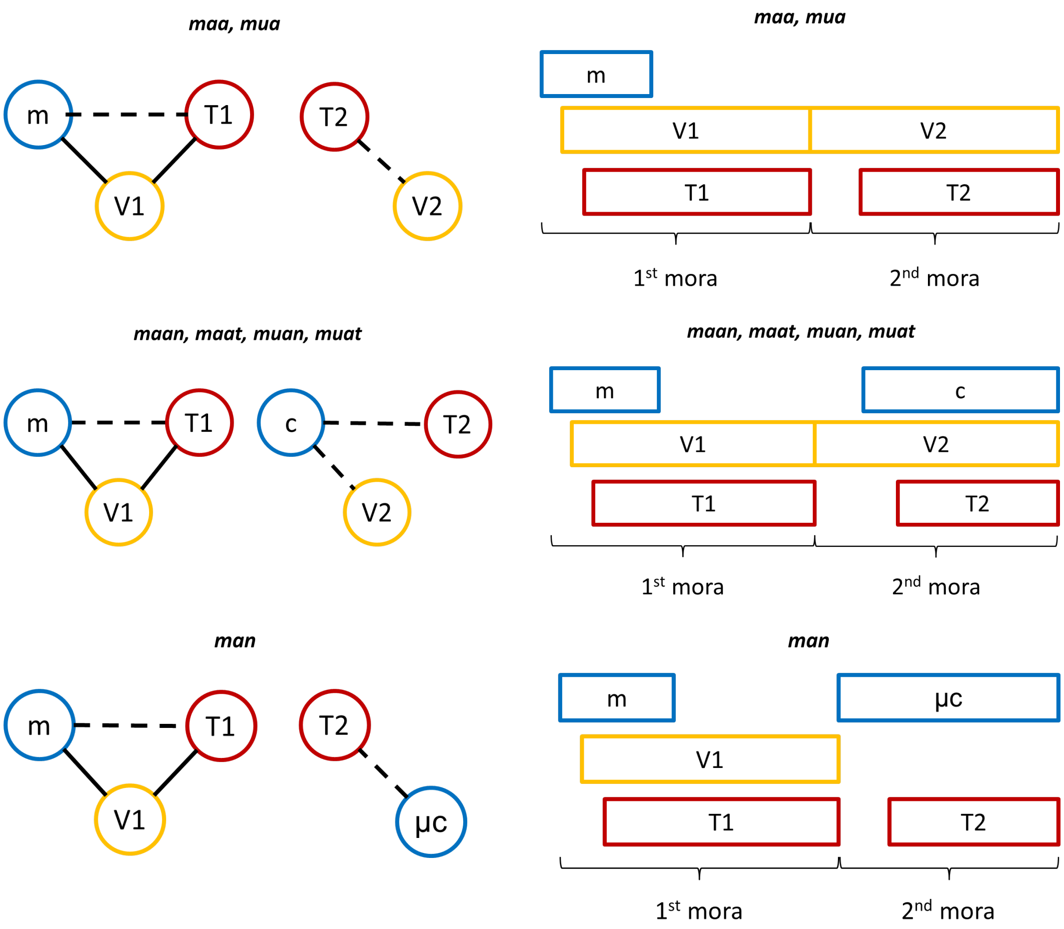


Figure 7: The proposed coupling relationships and corresponding predicted scores for each token type. In all cases, *m*, *T1*, and *V1* are in a C-center relationship and form the first moraic co-selection set, with *m* and *T1* in-phase coupled with *V1*, and anti-phase coupled with each other. When there is no coda, the second moraic co-selection set sees *T2* anti-phase coupled with *V2*, and the onset of *V2* is before the onset of *T2*. When there is a non-moraic coda, *coda* is anti-phase coupled with *V2*, and *T2* is anti-phase coordinated with *coda*; the onset of *coda* is after the onset of *V2*, and the onset of *T2* follows the onset of *coda*. When the coda is the second mora, *T2* is anti-phase coupled with μ*coda*, and the onset of *T2* is after the onset of μ*coda*.

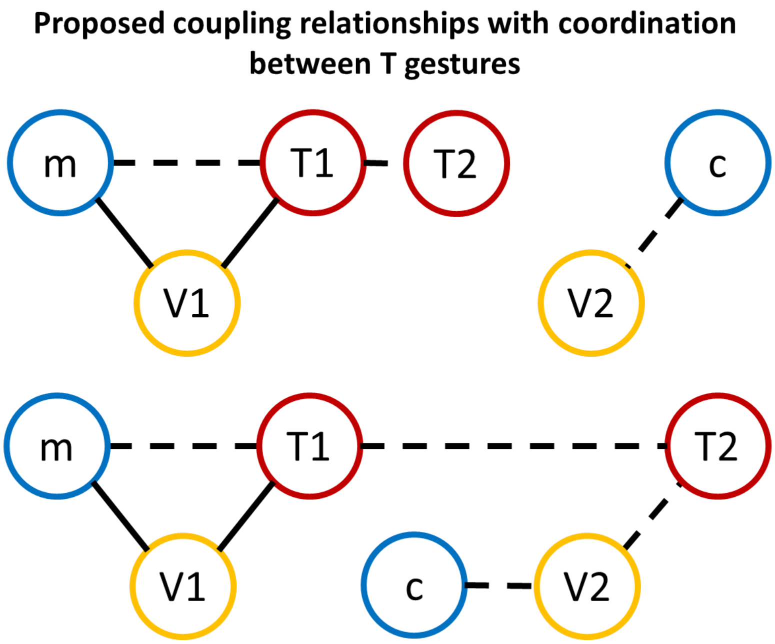

Proposed coupling relationships with coordination between T gestures
m
T1
T2
c
V1
V2
m
T1
T2
V1
c
V2

Figure 8: Two possible coupling relationships with coordination between T gestures for *mâan*, *mâat*, *mûan*, and *mûat*. In the first, T2 is anti-phase coordinated with T1, but not with V2; in the second, T2 is also coordinated with a TBU gesture.

3 Methods

3.1 Stimuli

In order to examine the coordination of *T* gestures with articulatory TBU gestures, the target word list was restricted to only those that can carry falling tones. Using a contour tone instead of a level tone ensures that there will be a unique *T* gesture for each TBU, which in turn provides the opportunity to examine *T* coordination with two TBUs. Target words had three nucleus types: diphthongs, in order to have two distinct vowel gestures for each TBU; long monophthongs, to compare the effect of identical vowel gestures for each TBU; and a short monophthong with a moraic coda, to compare vocalic TBUs to consonantal TBUs. In order to minimize disturbances on F0, target words were additionally restricted to sonorant onset; the bilabial /m/ was used in order to minimize mechanical interactions between the consonantal and vocalic gestures of the tongue. In order to test for effects of coda type on *T* coordination, the diphthong and long monophthong sets included words with a sonorant coda (/n/) and a non-sonorant coda (/t/). In order to meet the specifications, target words were a mix of real and nonce words; all were presented in Thai script.

Coda	Monophthong	Diphthong
none	maa	mua
/t/	maat	muat
/n/	maan	muan
	man	

Table 3: Target words used in the study.

Carrier phrases had either high or mid tones around the target word. In order to maximize the vertical tongue body displacement between adjacent vowels and make the extrema of the tongue body contour more visible to the landmarking script, the monophthong and diphthong target words had different preceding vowels: before a low vowel (monophthongs), the carrier phrase had a high vowel, and before a high vowel (diphthongs), the carrier phrase had a low vowel (Fig. 4).

Environment	**Monophthong**	**Diphthong**
High	phî: lí: __ lɛ́:w	phî: lá: __ lɛ́:w
	Lii already __	Laa already __
Mid	phî: li: __ wan ní:	phî: la: __ wan ní:
	Lii __ today	Laa __ today

Table 4: Carrier phrases used in the study.

3.2 Procedure and Data Collection

Four monolingual native speakers of Thai (three female, one male) between the ages of 18 and 45 participated in the study. Prior to recording, the experimenter told the participants that some of the words they would see were not real, but that they should read the sentences as naturally as possible. Data from one participant was excluded due to consistent phonological misproduction of the tone of both real and nonce words.

A full experiment included 20 blocks, or 60 tokens of each target word, with an upper limit of 2 hours on experiment duration. Only Participant 2 completed the full experiment; participants 1, 3, and 4 completed 12, 19, and 14 blocks, respectively. Before each block, participants saw a carrier phrase with a blank for the target words, presented in the Thai script. The number of trials per block varied depending on the nucleus type; diphthong blocks had 18 trials, and monophthong blocks had 24 trials (6 tokens of each target word for both cases). In each trial, participants saw the target word, presented in Thai script, as well as a red rectangle that moved from the bottom to the top of the screen. The rectangle moved at three different speeds; the speed for the trial was chosen pseudo-randomly. The experimenter instructed participants to wait until after the rectangle had moved off the screen, and then to say the full sentence at a speed that was analogous to the speed of the rectangle. For example, if the rectangle moved quickly, then they should speak quickly. Stimuli remained on the screen for 6 seconds, with 1 second between each trial.

Acoustic data was collected with a shotgun microphone approximately 3 feet from the participant. Articulatory data was collected with an NDI-WAVE articulograph, at a sampling rate of 100 Hz, which was up-sampled to 200 Hz for data processing. Reference sensors were affixed to the nasion and left and right mastoid processes. Five articulator sensors were affixed in the mid-saggital plane: the upper lip (UL); the lower lip (LL); the gumline below the lower front incisors (JAW); at the tongue tip (TT), approximately 1 cm behind the apex of the tongue; and on the tongue body (TB), approximately 4 cm behind the TT sensor. Prior to recording, a biteplate was used to determine the angle of the occipital plane relative to the sensors. The biteplate is a thin piece of plastic with three sensors affixed in a triangle. To collect the data, participants inserted the biteplate into their mouth and bit down lightly for five seconds. The recording provides for a "basis" position for the reference sensors; the most anterior sensor of the biteplate is treated as the origin of the coordinate system. For each time sample, the sensors are rotated and translated such that the reference sensors are aligned with the basis position. This both corrects for head movement and creates a standard coordinate system across participants, with axes that are parallel and perpendicular to the occipital plane.

3.3 Data Analysis

Data analysis was conducted in Matlab, using scripts developed by the Cornell Phonetics Lab, as well as the VOICEBOX toolkit (Brookes et al. 1997). For each trial, time-aligned acoustic and articulatory data was extracted for use in kinematic and acoustic landmarking. In order to extract the kinematic landmarks, the target word from each trial was first hand-segmented in Praat. Then, the relevant point of maximum velocity was located, followed by the extrema preceding and following the velocity extremum. Onsets, targets, and releases were identified at a 20% threshold of the related maximum velocity. In order to ensure accurate landmarking, additional time boundaries for each landmark were set. Landmarks were also inspected by hand, and corrected or discarded when necessary. The landmarks identified by the Matlab script are presented in a schematic in Fig. 9.

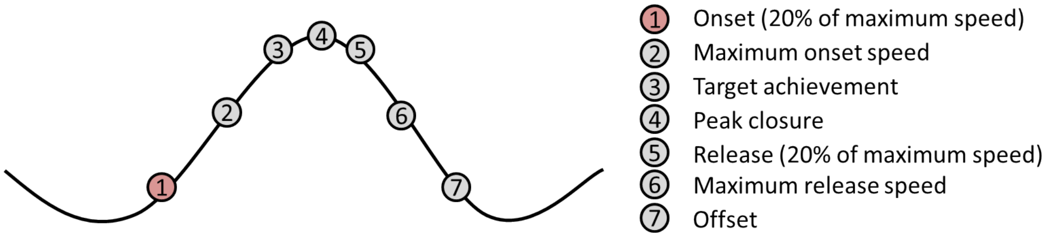


Figure 9: A schematic of the trajectory landmarking used in data processing.

Temporal lags were calculated between the onsets of each gesture. In accordance with AP, onsets (as opposed to targets or releases) were used to calculate lags; the assumptions of AP entail that only onsets and releases can be coordinated, while the timing of targets follows only from the timing of onsets. Analyses were conducted on a by-participant basis, and are presented in Section 4 below.

4 Results

The results presented below are from Participant 2, who completed all 20 blocks of the experiment. Only the first trial was preemptively excluded, due to a production error.

4.1 General Acoustic Results

In order to test the efficacy of the speed manipulation, a one-way ANOVA was conducted comparing the mean durations of each speed condition; results are significant [F(2,417) = 54.54, $p < 0.0001$, fast $<$ med $<$ slow].

	Word Duration	
	Mean	**St. Dev.**
Slow	315.9 ms	32.0 ms
Medium	294.7 ms	38.2 ms
Fast	271.7 ms	35.6 ms

Table 5: Acoustic duration of words in each speed condition.

Word duration was also compared across nucleus types to test for major differences between monophthongs and diphthongs. A one-way ANOVA shows that the total duration of long monophthong targets and diphthong targets are only marginally significantly different [F(1,358) = 4.64, p = 0.032]. Contrary to expectation, monophthongal words are actually longer than diphthongal words, but there is a small effect size [η^2 = 0.0121]. The duration of *mân* is significantly different from the duration of *mâan* [F(1,118) = 36.69, $p < 0.0001$, η^2 = 0.2372, mân $<$ mâan]. When comparing just the nucleus durations, the patterns are the same (see Fig. 6).

Overall, the tone trajectories are acoustically similar across speed conditions; as Nitisaroj (2006) found, speech rate does not greatly affect the overall shape of the tone. The time- and z-score normalized trajectories, separated by speed condition, are presented in Fig. 10. The onset of *T1* is the start of the rise, and the onset of *T2* is after the peak, at the start of the fall.

	Word Duration		**Nucleus Duration**	
	Mean	**St. Dev.**	**Mean**	**St. Dev.**
mân	277.5 ms	31.7 ms	81.9 ms	8.9 ms
long monophthongs	301.4 ms	40.8 ms	156.0 ms	32.6 ms
diphthongs	292.3 ms	39.2 ms	148.0 ms	30.4 ms

Table 6: A table of mean total and nucleus durations of target word types.

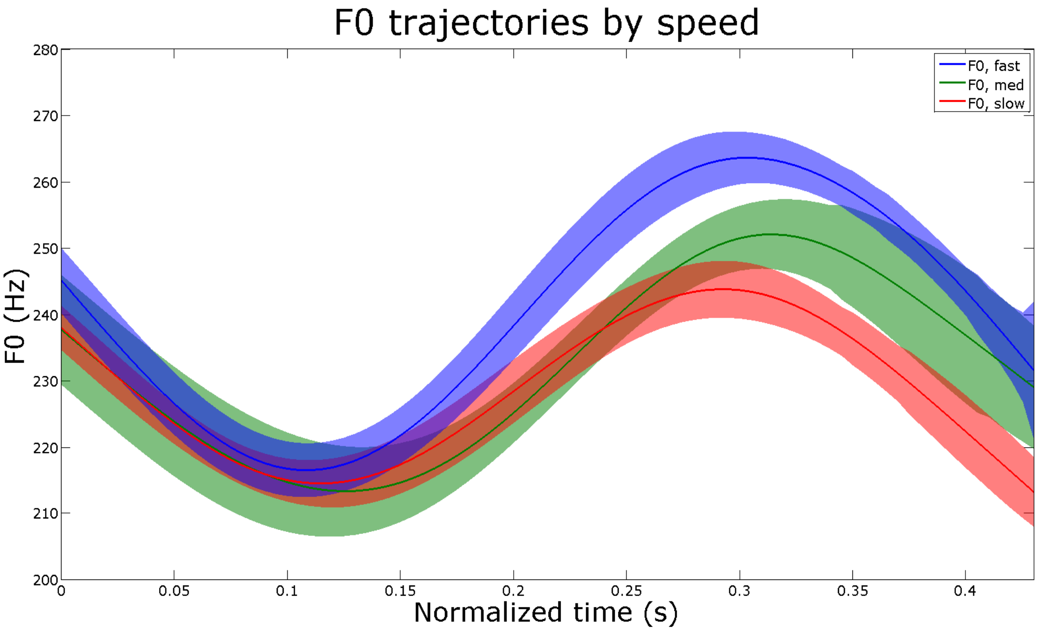


Figure 10: Time-normalized F0 trajectories, by speed.

In the sections that follow, I present the evidence for F0 as an articulatory gesture, as well as for patterns of coordination with gestures that correspond to moraic segments. I first address *T1* and its relationship with *m* and *V1*, and then continue to *T2* and its relationship both with *V2*, when available, as well as codas.

4.2 Hypothesis 1: F0 Is an Articulatory Gesture

Participant 2 demonstrated consistent timing of T gestures with C and V gestures. In order to consider T as in a coordinative relationship with another gesture, the relationship between the two gestures should be consistent within condition, as well as exhibit stability across different pressures, such as speed or environment. In other words, the standard deviation of a "lag" (difference in time between two landmarks) should be low, and the differences between groups should be minimal.

In both AP and autosegmental views of tone, the most straightforward relationship between T and a segment is that between *V1* and *T1*. As *V1* is supposed to be the first TBU, its relationship with *T1* should be both consistent and stable. A one-way ANOVA shows that the mean *acoustic T1 - V1* lag is significantly different across speed conditions [F(2,413) = 17.05, p $<$ 0.0001, slow $>$ med $>$ fast]. In contrast, the relationship between *articulatory T1 - V1* is both consistent as well as stable across speed conditions [F(2, 413) = 0.45, p = 0.6356]. Thus, the data supports the conclusion that F0 can be treated as an articulatory gesture that is coordinated with other articulatory gestures.

Similarly, the acoustic relationship between T2 and V2 is not stable across speed conditions (Fig. 8). There was a marginal effect of speed on the acoustic relationship between T2 and V2, [F(2,143) = 4.06, p = 0.0194]. However, the effect of speed was not significant on the articulatory

	Acoustic: T1 - V1		Articulatory: T1 - V1	
	$p < 0.0001$		p = 0.6356	
	Mean	**St. Dev.**	**Mean**	**St. Dev.**
Fast	-71.6 ms	25.3 ms	47.0 ms	23.8 ms
Medium	-81.5 ms	27.5 ms	43.8 ms	28.6 ms
Slow	-91.8 ms	33.3 ms	44.8 ms	33.6 ms

Table 7: Acoustic and articulatory lags between *T1* and *V1*

relationship between T2 and V2, [F(2,143) = 1.59, p = 0.2082]. In both cases, the data support the conclusion that tone is an articulatory gesture that is timed with other articulatory gestures.

	Acoustic: T2 - V2		Articulatory: T2 - V2	
	p = 0.0194		p = 0.2082	
	Mean	**St. Dev.**	**Mean**	**St. Dev.**
Fast	90.4 ms	21.3 ms	76.4 ms	19.4 ms
Medium	102.1 ms	20.1 ms	80.2 ms	24.0 ms
Slow	98.7 ms	21.7 ms	71.1 ms	31.2 ms

Table 8: Acoustic and articulatory lags between *T2* and *V2*

4.3 Hypothesis 2: T Gestures Behave like C Gestures

4.3.1 C-center Effect of T1

In line with Hypothesis 2, Participant 2 demonstrated a consistent C-center pattern with *m*, *T1*, and *V1* (see Fig. 11). Thus, *T1* behaves as the second member of a complex cluster, which backs up Gao (2008)'s finding that T gestures behave as C gestures. For a perfect C-center effect, the midpoint of *m* and *T1* would be the same as the onset of *V1*. The difference from C-center was calculated as $((m\ onset + T1\ onset) / 2) - V1\ onset$, or as the difference between *V1* and the midpoint of *m* and *T1*. For Participant 2, the mean difference between *V1* and the *m-T1* midpoint is 5 ms, with a standard deviation of 17 ms.

There was no effect of speed on the C-center [F(2,413) = 1.04, p = 0.353]. While there was a significant effect of nucleus type on the C-center [F(1,414) = 18.88, $p < 0.001$] (see Fig. 9), the magnitude of this difference is relatively small (7.6 ms), and may originate from a vowel-specific differences in landmark estimation. The C-center effect is demonstrated graphically in the gestural scores in Fig. 14 and Fig. 15. The relevant gestures are the *m* (blue), *V1* (yellow), and *T1* (red). Although the onset of *V1* is not precisely at the midpoint of the onset of *m* and *T1*, there is clearly a C-center arrangement. The data support the hypothesis that T is an articulatory gesture that behaves like C gestures.

	Mean Difference	Standard Dev.
Monophthongs	9.1 ms	19.6 ms
Diphthongs	1.5 ms	14 ms

Table 9: A table that shows the mean difference between the onset of *V1* and the midpoint of *m* and *T1*, separated by nucleus type.

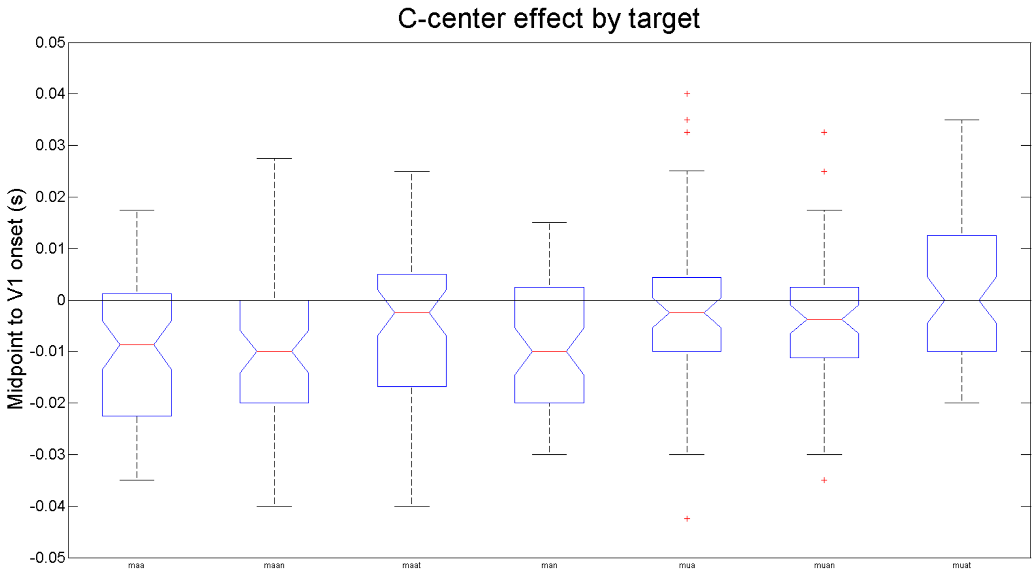


Figure 11: Boxplot of C-center effect by target word, comparing the midpoint of *m* and *T1* to the onset of *V1*

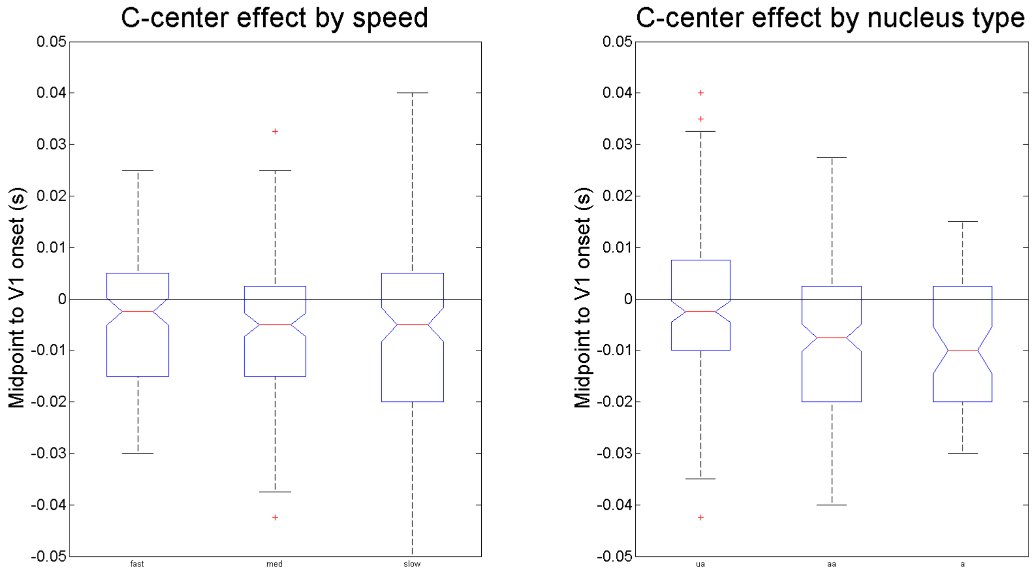


Figure 12: Boxplot of C-center effect by speed and by nucleus type, comparing the midpoint of *m* and *T1* to the onset of *V1*

4.3.2 T2 Behaves as a Consonantal Coda

In addition to *T1* behaving as a consonant when in onset position, *T2* also patterns like a consonantal coda. There are three different situations, which result in three coordinative patterns. For all target words, the onset of *T2* occurs after the onset of the gesture that corresponds to the second mora, which indicates that *T2* is anti-phase coordinated with a TBU gesture. For words without codas, this pattern is most readily visible with a diphthong, as the second mora gesture is distinct from the first (see Fig. 14 and 15 for gestural scores of all target words).

The second case is when there is a non-moraic coda. As predicted, the onset of *T2* is after the onset of *coda*, indicating that *T2* is anti-phase coordinated with *coda*. As can be seen when *V2* is distinct from *V1*, the onset of *coda* is after the onset of *V2*. The presence of the non-moraic coda increases the *T2*-*V2* lag [F(2,143) = 17.78, $p < 0.0001$, $\eta^2 = 0.1991$, mua < muan = muat; see Fig. 13, 10]. This suggests that *coda* is anti-phase coupled with *V2*, and then *T2* is anti-phase coupled with *coda*, which is in line with Marin and Pouplier (2010)'s observations of complex codas, as well as the predictions from Sec. 2.

	T2 - V2 (mean)	**Standard Dev.**
mua	62.8 ms	14.3 ms
muan	87.3 ms	27.7 ms
muat	80.0 ms	26.2 ms

Table 10: A table of mean *T2*-*V2* lags in diphthong target words.

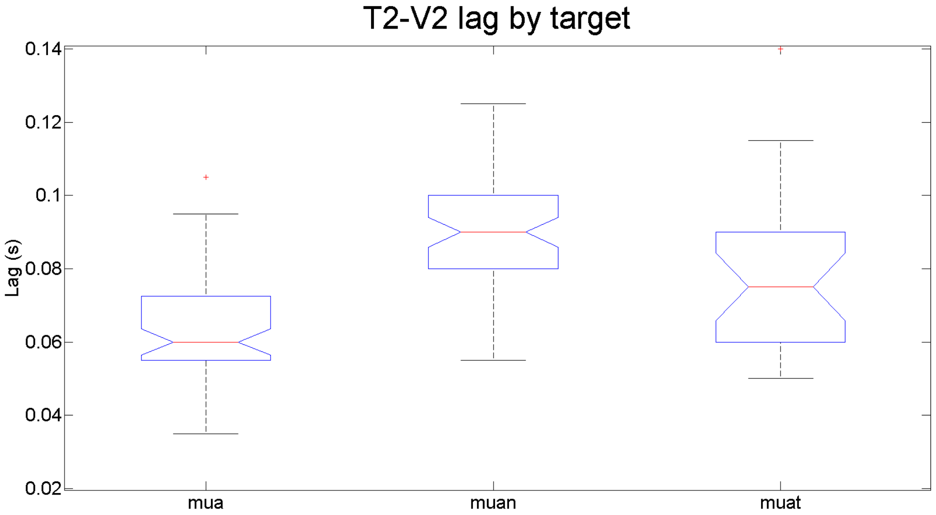


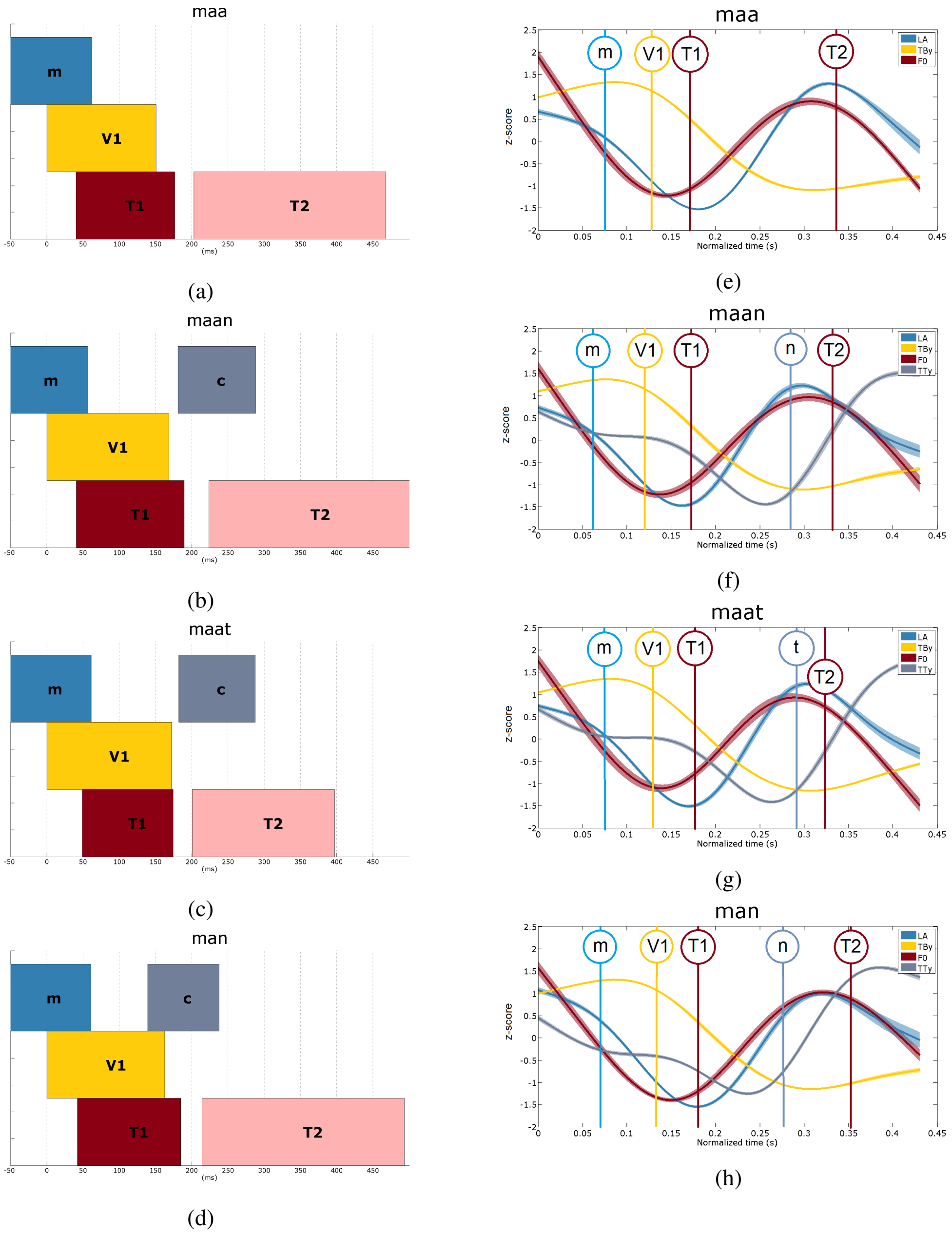
Figure 13: A one-way ANOVA shows that the *T2*-*V2* lag is shorter for *mûa* than for *mûan* and *mûat*.

Interestingly, the lag between *T2* and *coda* onsets is the same between non-moraic codas, even though [t] is voiceless and unable to carry tone—that is, regardless of the phonetic value of the coda, the onset of *T2* is always after the onset of *coda*. This phasing relationship does not sacrifice perceptibility of the tone, as the onset of *T2* would occur well before the onset of the glottal abduction gesture. The *coda*-*V2* lag is consistent across targets, as well as the *T2*-*coda* lag (see Fig. 11).

	T2 - coda		**coda - V2**	
	Mean	**St. Dev.**	**Mean**	**St. Dev.**
muan	35.8 ms	31.2 ms	51.3 ms	17.7 ms
muat	31.6 ms	45.8 ms	51.6 ms	26.0 ms

Table 11: *T2*-*coda* and *coda*-*V2* lags for diphthong targets.

However, the *T2*-*coda* lag is not the same when the coda is moraic, which is the third case. When there is a moraic coda, the *T2*-*coda* is much greater than it is for either the non-moraic coda *t* gesture or the non-moraic coda *n* gesture. When collapsing across nucleus type and comparing by coda type (see Fig. 16, 12), the *T2*-*coda* lag is significantly different between non-moraic codas as

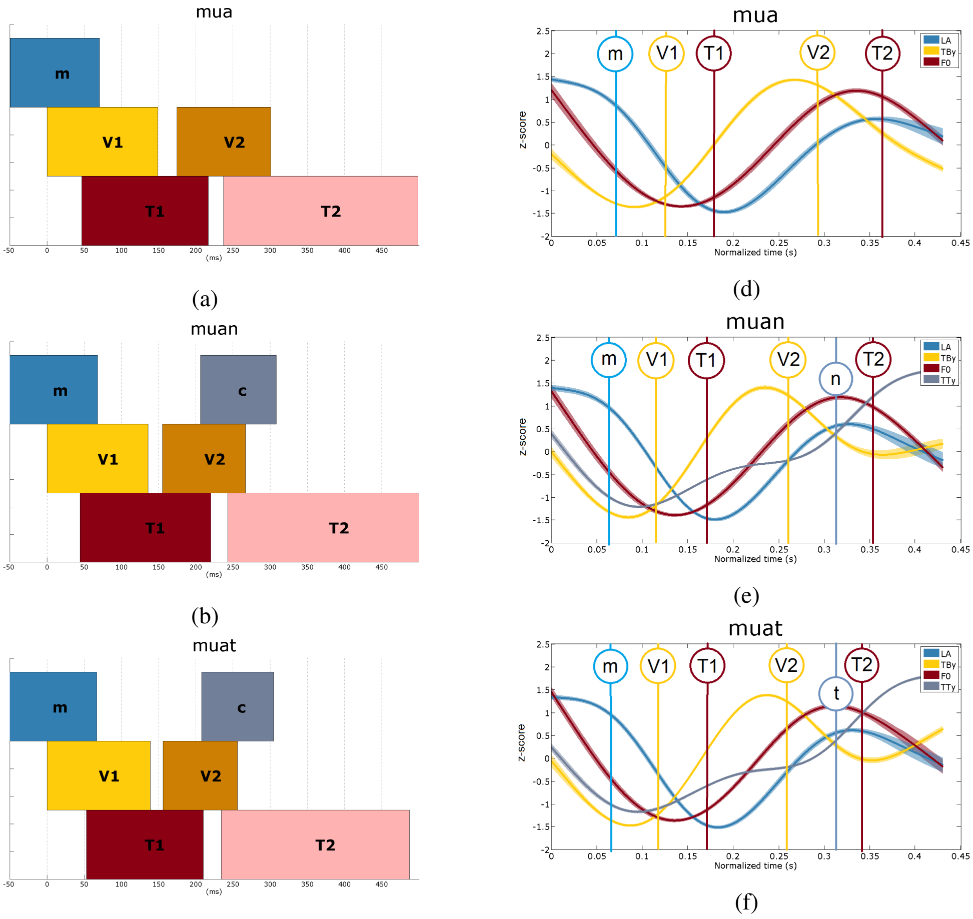


(a) (b) (c) (d) (e) (f) (g) (h)

Figure 14: Gestural scores and landmarked trajectories of monophthong target words, collapsed across speed and carrier conditions. The left edge of each rectangle is the onset of the gesture; the right edge is the target. Trajectories are z-scores of the raw trajectories; landmarks on the trajectories indicate onsets.

well [F(2,241) = 33.39, $p < 0.0001$, moraic n > n > t], though the magnitude of the effect is much smaller between non-moraic consonants (13.4 ms) than between moraic n and non-moraic n (35.8 ms).

It is also possible that the early portion of the glottal abduction gesture, or the start of the voicelessness of the /t/, had an effect on F0, which would in turn affect the landmark estimation for the F0 trajectory. For target words with /t/ as the coda, the *T2-coda* lag is smaller than for non-moraic n codas, but the timing is variable, as shown in Fig. 12. At the very least, the F0 trajectory would



(a)

(b)

(c)

(d)

(e)

(f)

Figure 15: Gestural scores and landmarked trajectories of diphthong target words, collapsed across speed and carrier conditions. The left edge of each rectangle is the onset of the gesture; the right edge is the target. Trajectories are z-scores of the raw trajectories; landmarks on the trajectories indicate onsets.

	T2 - coda (mean)	**Standard Dev.**
n	39.2 ms	26.8 ms
t	25.8 ms	52.1 ms
moraic n	75.0 ms	23.1 ms

Table 12: Mean *T2-coda* lags and standard deviations, by coda type.

stop at some point when the glottis was sufficiently opened, which could influence the landmarking of the F0 trajectory, ranging from a pitch perturbation that causes an early change in F0 to losing the F0 trajectory entirely, and thus not having a T2 landmark. However, the available landmarks indicate that *T2* is anti-phase coordinated with *coda*, regardless of the moraic status of *coda*.

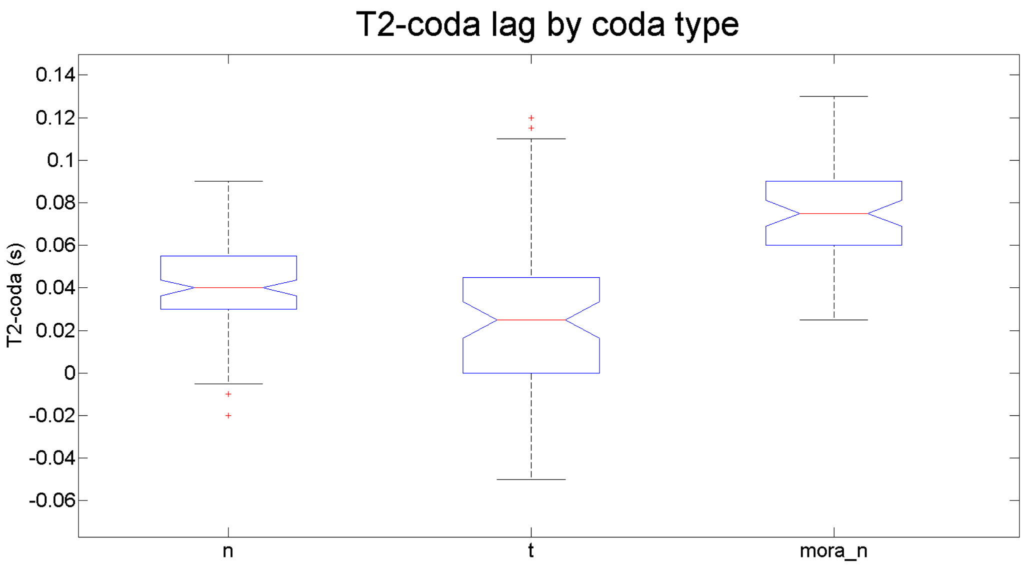


Figure 16: A boxplot of *T2-coda* lags, divided by coda type

4.4 Hypothesis 3: There Is an Articulatory Correlate of the Mora as a TBU

4.4.1 Coordination of Gestures Within and Across Moraic Co-selection Sets

Although there is consistency of timing relationships between all gestures in the target words, gestures that are generally considered part of one mora are more consistently timed with each other than with gestures that are in another mora. The consistent timing of gestures in different moraic co-selection sets can be attributed to the coordinative relationships between the mora gestures. For example, although the *mora 1-mora 2* lag is different for *mân* and *mûa* [F(1,118) = 56.40, p < 0.0001], the moraic gestures of each target word show a high degree of coordination with each other. The lags are fairly large, but the standard deviations are quite low, indicating that there is a consistent relationship between the two moraic gestures (see Fig. 13 for means and standard deviations); the standard deviations are comparable to those between *coda* and *V2* (see Fig. 16), which are in anti-phase coordination.

	mora 1 - mora 2 (mean)	**Standard Dev.**
mua (*V2-V1*)	174.6 ms	32.4 ms
man (*coda-V1*)	139.3 ms	16.8 ms

Table 13: Mean and standard deviation of mora 1-mora 2 lags for *mûa* and *mân*

However, the other diphthong target words have a different *V2-V1* lag, where the only significant difference is between *mûa* and the stimuli with a coda, illustrated in Fig. 17 [F(2,177) = 7.71, p = 0.0006, mua > muan = muat].

This difference is unexpected, as both vowels should serve as a mora/TBU gesture, to which the non-moraic gestures coordinate, and should be consistently timed relative to each other. However, in terms of absolute time lags, the most consistent relationship across moraic boundaries is between *T1* and *T2*, which is not significantly different for stimulus [F(2,142) = 0.98, p = 0.3792]. That is, the presence of a non-moraic coda does not delay the onset of *T2*, but rather pushes forward the onset of *V2*. This consistency in timing suggests that the T gestures are coordinated with each other, and not solely with the segmental gestures, a possibility that I discuss further in Section 5.

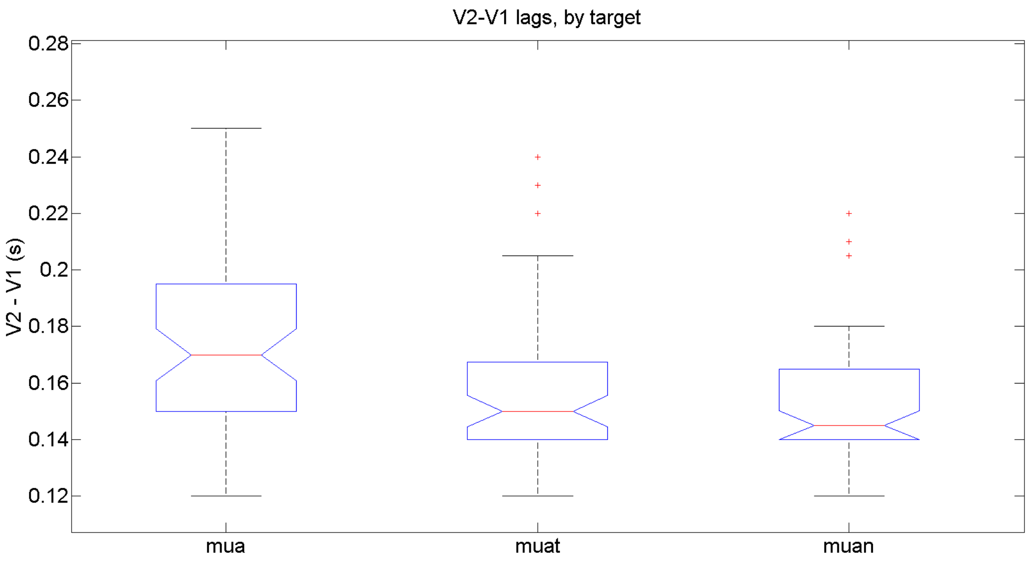


Figure 17: The *V2-V1* lag is greater in *mûa* than in *mûan* or *mûan*.

Overall, the time lags between gestures within a moraic co-selection set have lower standard deviations than the time lags between gestures across moraic co-selection sets. This indicates that there are mora-sized co-selection sets that can be coordinated with other mora-sized co-selection sets. Fig. 14 below compares the standard deviations of the time lags for *mûan* and *mûa*, arranged from lowest standard deviation to highest.

	mûan		mûa
	St. Dev.		**St. Dev.**
V1 - m	15.5 ms	*T2 - V2*	14.3 ms
V2 - n	17.7 ms	*V1 - m*	15.0 ms
T1 - V1	19.3 ms	*T1 - m*	22.5 ms
T1 - m	22.9 ms	*T1 - V1*	23.0 ms
T2 - V1	25.5 ms	*T2 - T1*	25.9 ms
T2 - V2	27.7 ms	*T2 - m*	28.6 ms
T2 - m	28.4 ms	*V2 - m*	30.0 ms
T2 - T1	29.4 ms	*T2 - V1*	32.0 ms
T2 - n	31.2 ms	*V2 - V1*	32.4 ms
V2 - V1	31.2 ms		
T1 - V2	33.2 ms		
V2 - m	35.3 ms		
T1 - n	42.1 ms		
n - m	42.6 ms		

Table 14: A table showing the (absolute value) standard deviation of each timing relationship, arranged from lowest standard deviation to highest. A blue background indicates that the gestures are in the same moraic co-selection set, a red background that the gestures are in different moraic co-selection sets, and a gold background that the gestures correspond to the moraic segments.

4.4.2 *T2-T1* coordination

The relationship between *T1* and *T2* is somewhat inconsistent; there are patterns that suggest that *T1* and *T2* may be directly coupled with each other, and others that indicate that their relative timing is determined by relationships within larger structures. For their part, the diphthong stimuli demonstrate a consistent *T2-T1* lag, both within stimulus and between stimulus; the means are not significantly different [F(2,142) = 0.98, p = 0.3792], and the standard deviations within stimulus are low (see Fig. 15).

	T2 - T1 lag	
	Mean	**St. Dev.**
mua	191.3 ms	25.9 ms
muan	198.2 ms	29.4 ms
muat	191.3 ms	36.9 ms

Table 15: T2-T1 lags for diphthong target words, divided by target.

In contrast, the long monophthong *T2-T1* lag is significantly shorter than the diphthong lag [F(2,302) = 42.18, $p < 0.0001$, $\eta^2 = 0.1225$]. There are also differences between stimulus word: the *T2-T1* lag for *mâan* patterns with the diphthongs and is significantly longer than the lags for *mâa* and *mâat*, though not *mân* (see Fig. 16).

	T2 - T1 lag	
	Mean	**St. Dev.**
maa	95.1 ms	16.9 ms
maan	182.9 ms	24.1 ms
maat	157.4 ms	56.3 ms
man	171.5 ms	36.9 ms

Table 16: T2-T1 lags for monophthong target words, divided by target.

Interestingly, the pattern of *T2-T1* lag is not the same as acoustic durations: overall, long monophthongs had a significantly (though only slightly) longer duration than diphthongs. Within monophthongs, *mâan* and *mâat* were significantly longer than *mâa* and *mân*. Thus, the differences in *T2-T1* lag between monophthong target words exhibit a pattern distinct from the acoustic durations, as well as a greater effect size. This suggests that *T2* and *T1* are not coordinated exclusively with each other, but rather that their timing relative to each other is determined at least partially through coordination with other gestures.

5 Discussion

This study tested the hypothesis of an articulatory TBU. Overall, the results point to some TBU-like organization, though with some unexpected coordinative patterns. In the sections that follow, I divide the discussion of the results by target word type: diphthongs, *mân*, and long monophthongs. I also consider the possible reasons for consonants behaving as tone gestures, as well as some issues with F0 trajectories as articulatory trajectories.

5.1 Diphthongs

The patterns of coordination in the diphthong target words point to two distinct TBUs; as predicted, *V1* and *V2* serve as gestures that tone and consonant gestures are coordinated with. Tone gestures pattern in two distinct ways, much like (non-moraic) consonantal gestures: in-phase coordinated with a vowel, and anti-phase coordinated with a vowel. In-phase coordination is demonstrated by *T1*, which acts like the second member of a complex onset in the co-selection set {*m-V1-T1*}. Anti-phase coordination is demonstrated by *T2*, which acts like a coda in the co-selection set {*V2-(coda)-T2*}. Unexpectedly, however, the lag between *V1* and *V2* varies depending on the presence of a coda, while the lag between *T1* and *T2* remains stable. This suggests that there is some direct relationship between *T1* and *T2*, instead of their relative timing being the result of other coordinative relationships.

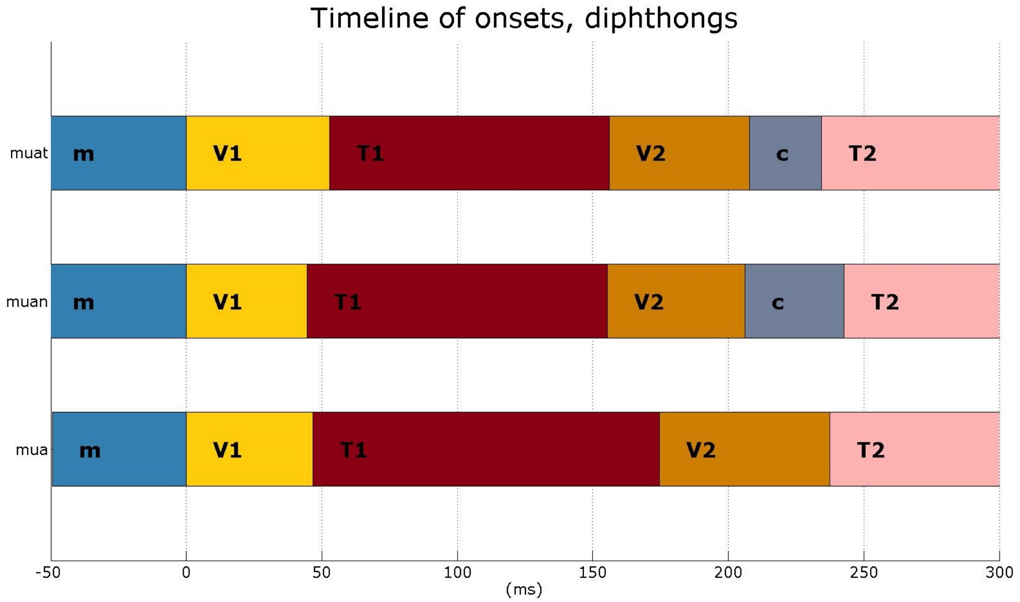


Figure 18: The timeline of onsets for diphthong target words. The left edge of each rectangle is the onset of that gesture. *V2* in *mûa* is significantly later than *V2* in *mûan* and *mûat*.

Since *V1* and *V2* are landmarked using the same trajectory (TBy, or the vertical component of the tongue body), it is possible that the difference in *V1*-*V2* lag is a reflection of a shortening of *V1*. This shortening could be due to the more complex structure in the following mora; a similar acoustic result was found by Munhall et al. (1992), where vowels with complex codas were shorter than vowels with simple codas. Both Munhall et al. (1992) and Marin and Pouplier (2010) only considered monosyllabic words, and so could not compare the duration of a preceding mora in the same word. However, in contrast to Munhall et al., Marin and Pouplier did not find that vowels were shorter with simple codas than with complex codas. In the current study, the lag between *V2* and the first coda-like gesture (*T2* for *mûa*; *coda* for *mûan* and *mûat*) is the same across target types [F(2,174) = 0.38, p = 0.6858], which indicates that *V2* itself is not shortened due to a more complex structure in the mora.

One other possibility is that the use of the tongue tip in the coda gestures perturbs the absolute position of the tongue body, giving the appearance of an earlier onset of *V2* when there is a coda. In this case, the rise of the tongue tip would have to cause the tongue body to lower, or at least the section of the tongue body where the TB sensor is placed. If the earlier onset of *V2* with codas is, in fact, due to a mechanical coupling of the tongue tip and the tongue body, there would be a shortening effect of complex codas on *V2*, as opposed to *V1*. This possibility cannot be ruled out without the further examination of tokens like *mûam*, where the coda articulator would be independent of the vowel articulator.

5.2 The Target Word *mân*

The target word *mân* is unique in this study in that it is the only token that has a moraic coda. Interestingly, the *μn-T2* lag is not significantly different from the *V2-T2* (for *mûa*), *n-T2* (for *mûan*), and *n-T2* (for *mûat*) lags [F(3,232) = 0.34, p = 0.7994]. This is a direct comparison of the lags between TBU/mora gestures and their first coda-like gestures. However, the *μn-T2* lag is significantly different from the *n-T2* (*mûan* only), [F(1,115) = 59.9, p $<$ 0.0001]. This suggests that the moraic status of /n/ in *mân* changes the relationship with the non-moraic gestures that are coordinated to it.

One possibility is that the *moraic n* has lower gestural stiffness, and is more similar to vowels. This would push the onset of *T2* back in time, as it would take longer for the gesture to reach 180°. This should result in a longer interval from the onset of *moraic n* to the target of *moraic n*, in comparison with the interval between *non-moraic n* onset and *non-moraic n* target. However, this is not the case: when considering all /n/ codas, the interval means are significantly different [F(2,175) = 6.65, p = 0.0016]; however, the *moraic n* interval is shorter than the *non-moraic n* of *mâan*, but there is no difference between *mân* and *mûan* or *mûan* and *mâan*.

5.3 Long Monophthongs

The patterning of *mân* again suggests that it is *T1* and *T2* that are coordinated with each other, rather than first being coordinated with some TBU. However, the long diphthongs do not support this hypothesis. When only considering long monophthong target words, the *T2-T1* lag is significantly different between target words [F(2,156) = 7.21, p = 0.001]. In these target words, *mâan* is significantly longer than both *mâa* and *mâat*; *mâa* and *mâat* are not significantly different. Additionally, the *T2-T1* lag is significantly different between nucleus types (aa, a, and ua) [F(2,360) = 25.52, p $<$ 0.0001]; /ua/ is significantly longer than both /aa/ and /a/, but there is no difference between /aa/ and /a/. This is more in line with the hypothesis that tone gestures coordinate with a TBU/mora gesture. In this case, /aa/ would be shorter than /ua/ due to a re-selection of the same gesture for *V2*, as opposed to the selection of a gesture with a different target.

One further point of interest is that *mâan* is not significantly longer in acoustic duration than *mâat*, but the *T2-T1* lag is greater for *mâan* than for *mâat*. This is due to a shorter *T2-coda* lag being for *mâat*, though *T2* comes after *coda* in both target words. This difference could again be due to changes in the F0 contour from the voicelessness of the /t/ coda, which would cause differences in landmarking. If this is the case, the *T2-T1* lag could potentially not be significantly different between *mâan* and *mâat*; however, the lag for *mâa* is still significantly shorter. This is in contrast with the diphthong target words, where the *T2-T1* lag is consistent across all three target words.

The inconsistency with landmarking F0 trajectories for *mâat* and *mûat* points to a potential problem with the use of F0 trajectories as the articulatory trajectory. Instead of measuring changes in 3-D space, as the articulatory sensors do, F0 is measuring an acoustic effect of some group of muscles. This is akin to using acoustic segmentation to find gestural onsets and targets, as opposed to articulatory trajectories. As noted in the results section, acoustic landmarking does not provide the same insight on coordinative structures; for example, the C-center effect is obscured when using acoustic measures, while it is consistent and clear using articulatory measures. While it may be the case that the F0 trajectory is an acoustic measure, as opposed to an articulatory one, the coordinative patterns are largely consistent with those found with just consonant and vowel gestures, such as the C-center effect and the anti-phase coordination of *T2*.

5.4 Possible Origins of Tone as a Consonant-like Gesture

The results above confirm previous findings that tone gestures behave as consonantal gestures. In addition to the C-center effect, as found in complex onsets, tone also seems to coordinate as a coda consonant. This is somewhat unintuitive, as tones demonstrate behavior that is more similar to vowels than to consonants, such as spreading and harmony (Yip 2002). However, the mechanisms of tonogenesis provide a possible explanation for a consonantal origin, and thus the consonantal behavior, of tones. Ohala (2001) has proposed that voicing contrasts in consonants cause phonetic perturbations on F0 of following vowels, which are then reinterpreted as meaningful changes in F0. Gradually, these perturbations are phonologized into tone, while the voicing contrast on the consonants is leveled and disappears. Such a mechanism is possible for Thai, which has collapsed voiced and voiceless sonorants into one voiced category, while maintaining the lexical contrast through tone (Pittayaporn 2009). Thus, the changes in F0 had a consonantal origin, and consequently reorganize as another consonant, resulting in a C-center structure.

One possible origin of the consonant-like phasing is that the tone gestures are a reflection of the phasing of the glottal gestures. For example, the release of glottal abduction gestures on an aspirated /p/ is after the release of the lip closure; this release pattern is similar to the release pattern of consonant clusters. In this case, the glottal gesture is changed to a tone gesture. Browman and Goldstein (1992) depict the onset of glottal gestures as after the oral constriction gestures for aspirated stops; however, it is not entirely clear if the glottal gesture is anti-phase coordinated with the oral constriction gesture. A similar mechanism has also been proposed for coda consonants (Tang 2008, Yip 1995), where variation in glottalization of codas would provide a similar pitch perturbation, and reinterpretation of the F0 change as a consonantal gesture. However, it is somewhat unclear what the proposed phasing relationships of coda glottal gestures are.

5.5 Targets and Coordination of Tone Gestures

One further question is what an F0 target is—an absolute Hz value, or some relative change in Hz. Earlier treatments of tone have concluded that absolute, invariant F0 values are not the targets of tone (Ladd 2008, Gussenhoven 2004); rather, the precise Hz value of any tone can vary greatly and is influenced by speech rate, surrounding tones, and syntactic position, among other factors (Morén and Zsiga 2006, Nitisaroj 2006, Yip 2002). In Thai, overall tone shape is preserved in the face of pressures like stress and speed. However, both the excursion size and timing of F0 extrema are affected by stress type, but not by speed (Nitisaroj 2006). When normalizing for rhyme duration, the F0 peak in unstressed falling tones was later than the F0 peak of a stressed falling tone. Since Thai unstressed syllables are shorter than stressed syllables, the question remains if there is a significant difference in absolute timing of the F0 peak.

In this study, the *T1-T2* lag decreases with increased speed [$F(2,360) = 24.97$, $p < 0.0001$, fast < med < slow]. These lag patterns indicate that tone gestures are coordinated with segmental gestures, though there may be an additional level of coordination between tone gestures, as discussed earlier. One possibility is that *T1* must hit some threshold before *T2* can be activated, in some kind of competitively selected arrangement. The maximum rate of change of F0 is limited by physiological factors (Xu and Sun 2002); that is, in utterances with high time pressure, the target of *T1* would be achieved later with respect to the segmental gestures since the same change in Hz requires a minimum amount of time. Thus, the onset of *T2* would then be delayed relative to its carrier gesture. In this case, it is also possible that only the first of two tone gestures in a word must pass through an F0 threshold; the *T1-T2* unit would then be coordinated with the segmental gestures, and the entire word unit of tone and segments coordinated with other words. This would prevent a sort of "domino effect" of tones being realized progressively later through an utterance.

5.6 Implications for Phonological Representation

While the model of moraic co-selection sets accounts for motor implementation and acoustics, there still remains the issue of phonological patterning. As noted before, not all bimoraic words can carry contour tones; /mat/ can only have high or low tone, while /man/ can carry all five tones (mid, low, falling, high, rising). I propose that there are two constraints that govern the moraic structure in individual languages: (1) concerning how many co-selection sets at the moraic level are allowed to combine into one syllable, and (2) concerning what classes of co-selection sets are allowed to have tone gestures coordinated with them.

In Thai, rule 1 caps Thai syllables at two moraic co-selection sets, reflecting the bimoraic limit (see Fig. 17[3]). Instead of coordinating all gestures between the moraic co-selection sets, I argue that

[3] M(id), L(ow), F(alling), H(igh), R(ising)

Syllable Type	Example	Moras	Permissible tones
CV	ma	1	H, L
CVO	mat	2	H, L
CVV	maa	2	M, L, F, H, R
CVN	man	2	M, L, F, H, R
CVVN	maan	2	M, L, F, H, R

Table 17: Moraic weights in Thai

non-moraic gestures are coupled to moraic gestures, and the moraic gestures, acting as the "hub" of the co-selection set, are coordinated to each other. In this arrangement (illustrated in Fig. 19), the two moraic hubs are anti-phase coordinated with each other.

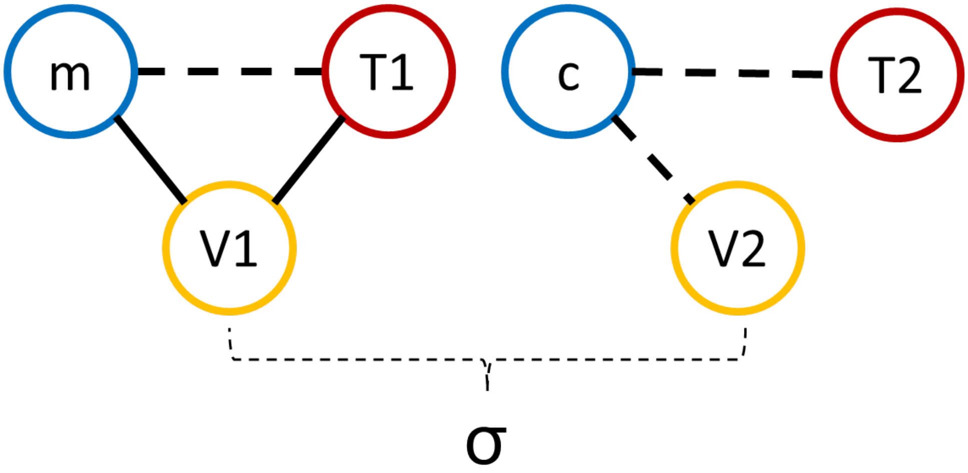


Figure 19: The proposed coupling relationships of moraic co-selection sets. At this level, the "hubs" of each moraic co-selection set, or the gestures that corresponds with the moraic segments, are anti-phase coordinated with each other.

The moraic and syllabic structure alone cannot account for the distribution of tones in Thai. In Thai, not all moraic segments can carry tone; as mentioned before, /mat/ cannot carry a contour tone (see Fig. 13). The variability in codas that can anchor a tone gesture is due to the phonetic implementation of the phasing relationships. The structure of *mat* is straightforward: {ma} and {t}. If the components of a contour tone were to be coordinated with each moraic co-selection set, *T2* would be anti-phase coordinated with *t*. However, as seen with *mân*, this would result in the onset of *T2* occurring after the onset of *t*, or sometimes even after the target of *t* (see Fig. 20). Thus, although historically a T2 gesture could have been present on syllables like /mat/, the consequences of the coupling relationships would have ultimately resulted in an imperceptible T2.

Finally, there is the question of how words are represented in the mental lexicon. One possibility is that words are stored with all coupling relationships made explicit; that is, *mûan* is stored as {m—u—H}-{a- -n- -L}. Another possibility is that all the segment-sized co-selection sets are labeled with their function—moraic hub, onset, coda, first tone, second tone, etc.—and then the phasing relationships are determined from the functions of each gesture.

5.7 Conclusions and Future Directions

The current study indicates that Thai syllables are organized into two moraic co-selection sets, and that T gestures are coordinated within each co-selection set. In this way, the "hub" of the moraic co-selection set acts as a gestural TBU. A moraic level of coordination could additionally account for

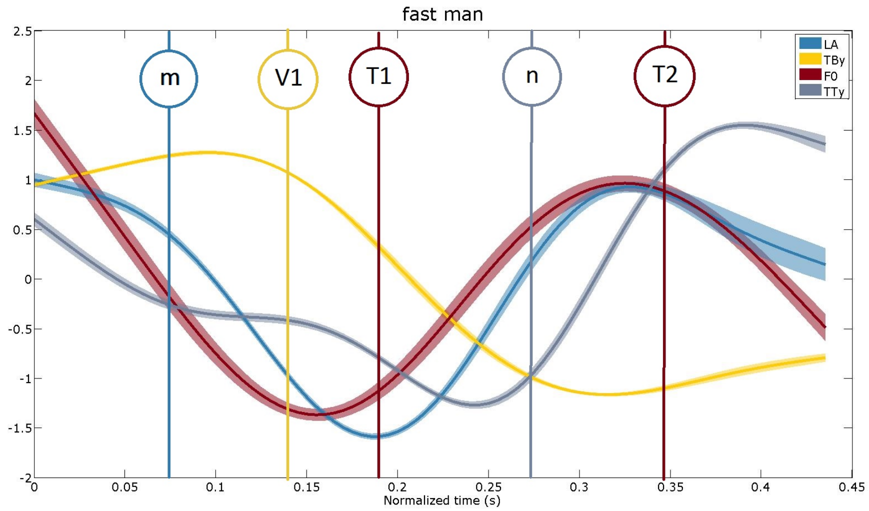


Figure 20: Trajectories for *mân*, at fast speed, with onsets marked

previously documented patterns of tone gestures and intonational gestures—i.e., that tone gestures behave like consonants, and intonational gestures like vowels. If tone gestures are coordinated at the same level as consonant gestures, they would interact with the actual consonant gestures to create a C-center effect; in contrast, intonational gestures would be coordinated at a level where the C-V phasing had already been determined, and thus a C-center effect would not occur. Regarding the proposed relationship of tone gestures to historical consonant gestures, further study is also needed on the cross-linguistic behavior of tone gestures. In particular, a study to compare and contrast different tone systems could shed light on tonal organization.

Additionally, it is unclear exactly how tone gestures are combined with each other and other gestures; some data indicate that T gestures are only coordinated with segmental gestures, while other data suggest that T gestures are additionally combined with each other. This question is related to the issue of what tone targets are. One possible way to get at this problem is to increase the time pressure, either by using unstressed words, which are naturally shorter in duration, or by manipulating the surrounding tones further to force rapid changes in F0.

References

Brookes, Mike, et al. 1997. Voicebox: Speech processing toolbox for Matlab. *Software, available [Mar. 2011] from www. ee. ic. ac. uk/hp/staff/dmb/voicebox/voicebox. html* .

Browman, Catherine P, and Louis Goldstein. 1989. Articulatory gestures as phonological units. *Phonology* 6:201–251.

Browman, Catherine P, and Louis Goldstein. 1992. Articulatory phonology: An overview. *Phonetica* 49:155–180.

Gao, Man. 2008. Mandarin tones: An articulatory phonology account. Doctoral dissertation, Yale University.

Gussenhoven, Carlos. 2004. *The phonology of tone and intonation*. Cambridge University Press.

Jeannerod, M. 1986. The formation of finger grip during prehension. a cortically mediated visuomotor pattern. *Behavioural Brain Research* 19:99–116.

Ladd, D Robert. 2008. *Intonational phonology*. Cambridge University Press.

Marin, Stefania, and Marianne Pouplier. 2010. Temporal organization of complex onsets and codas in American English: testing the predictions of a gestural coupling model. *Motor Control* 14.

Morén, Bruce, and Elizabeth Zsiga. 2006. The lexical and post-lexical phonology of Thai tones. *Natural Language & Linguistic Theory* 24:113–178.

Mücke, D, H Nam, A Hermes, and L Goldstein. 2011. Coupling of tone and constriction gestures in pitch accents. *Consonant Clusters and Structural Complexity* .

Munhall, Kevin, Carol H Fowler, Sarah Hawkins, and Elliot Saltzman. 1992. “compensatory shortening” in monosyllables of spoken English. *Journal of Phonetics* .

Nitisaroj, Rattima. 2006. Thai tonal contrast under changes in speech rate and stress. *Speech Prosody 2006, Dresden, Germany, May 2-5 2006* .

Ohala, John J. 2001. The phonetics of sound change. *Phonology: Critical Concepts in Linguistics* 4:44.

Pittayaporn, Pittayawat. 2009. The phonology of Proto-Tai. Doctoral dissertation, Cornell University.

Prieto, Pilar, Doris Mücke, Johannes Becker, and Martine Grice. 2007. Coordination patterns between pitch movements and oral gestures in Catalan. In *Proceedings of the XVIth International Congress of Phonetic Sciences, Pirrot GmbH: Dudweiler*, 989–992.

Prieto, Pilar, and Francisco Torreira. 2007. The segmental anchoring hypothesis revisited: Syllable structure and speech rate effects on peak timing in Spanish. *Journal of Phonetics* 35:473–500.

Tang, Katrina Elizabeth. 2008. *The phonology and phonetics of consonant-tone interaction*. ProQuest.

Tilsen, Sam. 2013. A dynamical model of hierarchical selection and coordination in speech planning. *PloS one* 8:e62800.

Tilsen, Sam. 2014. Selection and coordination of articulatory gestures in temporally constrained production. *Journal of Phonetics* 44:26–46.

Xu, Yi, and Xuejing Sun. 2002. Maximum speed of pitch change and how it may relate to speech. *The Journal of the Acoustical Society of America* 111:1399–1413.

Yip, Moira. 1995. Tone in East Asian languages. *The handbook of phonological theory* 476–494.

Yip, Moira. 2002. *Tone*. Cambridge University Press.

Robin Karlin
Department of Linguistics
Cornell University
Ithaca, NY 14853
rpk83@cornell.edu