

Timing and Alignment: A Case Study of Lai*

Ian Maddieson

University of California, Berkeley

The timing of events in speech is a critical element of the phonetic and phonological organization of language and of individual languages. In this paper some aspects of segment duration, the syllable-internal timing relationships among segments, and the alignment of tones and segments in a controlled set of data from Hakha Lai are analyzed. Hakha Lai, although spoken in Northern Burma and India, has syllable structure and tone patterns that are reminiscent of the languages of the 'Sinosphere'. Among the effects observed in this language are considerable duration compensation between a preceding vowel and a coda consonant, and a smaller lag in tone alignment than has been observed in many others. In the discussion, an effort is made to partition the observed effects into those that are universal, including those that can be predicted from Lai's place in a quantitative typology of languages, and those that are language-specific.

Key words: speech timing, tone alignment, quantitative typology, Hakha Lai (Tibeto-Burman language)

1. Introduction

Several decades ago it was recognized that as a necessary preliminary to the development of adequate systems for synthesizing speech some kind of integrated timing model for the target language must be developed. Among the classic examples of this kind of work are the model for American English by Dennis Klatt (1976, 1979) and that for French by Doug O'Shaughnessy (1981). A partial model for Mandarin Chinese was developed by Ren Hongmo in 1985 following the prototype established by the

* This work was supported by grant BCS-9817345 for "Phonetic Studies of Endangered Languages" from the National Science Foundation to the University of California, Berkeley (Ian Maddieson, Principal Investigator). Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author and do not necessarily reflect the views of the National Science Foundation. The essential contribution of Kenneth VanBik to this study is gratefully acknowledged. Valued insights were also contributed by Larry Hyman, David Mortensen, and Jie Zhang. Sincere appreciation is expressed to Professor Ho Dah-an and the Institute of Linguistics (Preparatory Office) of Academia Sinica for their invitation to present it at the *8th International Congress on Chinese Language and Linguistics*.

foregoing. Every speech synthesis system must contain a component to generate appropriate timing of segments and appropriate alignment of acoustic events of different types with each other. Research inspired by this need often represents the most complete view of linguistic timing patterns in different languages. Unfortunately much recent work on these aspects is protected by commercial concerns and not made public, or is embedded in systems in which the information of linguistic interest is hidden or opaque.

It is also true that, because of their practical orientation, models of this kind generally do not distinguish between timing regularities that are associated with different levels of explanation. Linguists might be interested in distinguishing the universal from the language-specific in both phonological and phonetic domains, but this distinction is of little concern to a speech technology engineer. Some timing patterns may be typical of languages with certain phonological properties, such as segmental quantity distinctions or given canonical syllable structures. But other languages may deviate from what is typical in idiosyncratic ways. Some timing patterns may be so widespread and well-motivated that they should be attributed to universal phonetic effects, while others need to be stated as aspects of the individual phonetic grammar of a language.

Attempts to partition timing regularities between such different levels can draw on practically-oriented research of the kind cited above as well as on a considerable number of studies on timing for more basic descriptive purposes, such as Elert (1964) on Swedish or Farnetani (1981) on Somali. Particularly useful are studies that compare or contrast different languages. Some of these are undertaken with the object of assisting with language learning by drawing attention to differences between the learner's native language and the target language to be learned, such as the comparative studies of English and Finnish by Wiik (1965) and Suomi (1980). Other comparisons may be motivated by testing predictions of theoretical linguistic models and validating concepts such as rhythmic categories, moraicity, or syllable weight (e.g., Dauer 1983, Broselow et al. 1997, Ham 1998).

All of the work mentioned above most often deals with a careful and somewhat formal style of speech—more rapid or colloquial speech styles raise additional issues that require their own special handling. And they most often deal only with data in the acoustic domain. Articulatory timing, as well as the perceptual impact that particular durations might create, are much less often considered. (However, the acoustic durations give important indications of the articulatory and auditory organization.)

With this background in mind, this paper will take a tiny fragment of data from the Tibeto-Burman language Hakha Lai, spoken around the town of Hakha in Chin State, Burma (Myanmar), and will test whether it is possible to partition the acoustic timing of the events that can be recognized in this data between that which is explicable by

phonological patterns which create a universal type of timing, that which is due to language-specific phonology, that which is due to universal phonetic patterns, and that which is due to language-specific phonetic patterns. A variety of the timing patterns observed will be first described and then compared with what is known to be general across languages, or appears to be variable and therefore language-specific. The paper is thus as much an experiment in whether we can identify universals as it is an experiment on a single language. Lai has a number of the typological characteristics common among many of the languages of what Matisoff (e.g., 1990) refers to as the “Sinosphere”, such as a substantially monosyllabic lexicon, a limited set of permitted coda consonants, and the presence of underlying contour tones, even though it is geographically outside the area of direct Chinese influence. It is thus hoped that this study will provide useful insights to linguists looking into phonetic and phonological timing in any of the languages of this area.

2. Hakha Lai: data

The data is obtained from examining a small set of verbs in just one context. The verb forms are chosen to represent a variety of syllable and tone patterns. As has been described elsewhere, Hakha Lai has a quantity distinction in vowels (Melnit 1997, Hyman and VanBik 2002, Hyman 2002). Long and short vowels are, however, only lexically contrastive in closed syllables: open syllables in lexical forms have long vowels. Coda consonants fall into three regular classes and one special class. The three regular classes are voiceless unaspirated stops /p, t, k/, voiced sonorant true consonants in the form of nasals and liquids, and the vocalic glides /w, j/. These can be preceded by either a long vowel or a short vowel. The special class consists of the glottal stop and glottalized sonorants. Glottalized sonorants occur only in the coda position and are limited to the transitive/completive “Form 2” of verbs. They are always preceded by a short vowel, as is a glottal stop in the coda. This class of consonants is not further discussed in this paper, but will be studied in some detail in work to follow. (Some preliminary data are given in Roengpitya 1997 and Plauché et al. 1998).

Syllables with long vowels, excepting those with final stops, have a potential for tonal contrast between low level, rising, and falling tones. Those with final stops have low tone. Short closed syllables may also have low level, rising, and falling tones, unless they are stop-final. Short stop-final syllables do not have lexical tone contrast, but are pronounced in isolation and the context examined in this paper with a (truncated) falling contour. Short open syllables only occur in clitic-like elements and are phonologically toneless. Further details on syllable structure and tone are given in Hyman and VanBik (2002) and Hyman (2002), including discussion of alternations that are outside the

scope of the data addressed in the present paper.

Although there are redundancies that would allow a simpler transcription to be used, tone will be fully marked in this paper following IPA conventions. Lexical falling tone is shown by a circumflex accent above a short vowel (/â/) and by a sequence of acute and grave accents on two vowel symbols in the case of a long vowel (âà). Rising tone is shown by a sequence of grave and acute accents over two vowel symbols (âá). Low tone is shown by grave accents over one or two vowel symbols (à/âà). The different syllabic and tonal patterns provide the timing variation that will be the focus of this study.

Table 1: Hakha Lai verbs examined in this study.

		Word	Gloss
Short vowel	Final stop (CVK)	pâk	smoke
		kâp	stick to
	Final sonorant (CVR)	tâm	be plentiful
		pâl	step on
		kâl	go
Final glide (CVG)	kâw	call	
Long vowel	Final stop (CVVK)	kâàp	shoot
		pâàk	be single
	Final sonorant (CVVR)	pâàm	be wasteful
		tâàl	struggle
	Final glide (CVVG)	kâáw	be wide
		kâàw	open
	Open syllable (CVV)	pâá	be manly
tâà		hone	

The words selected for examination are given in Table 1, which is laid out to show the different syllable structures exemplified. The abbreviations that will be used below to refer to the classes of syllables are shown in parentheses in the second column: **-K** indicates a final stop, **-R** a final resonant (nasal or lateral), **-G** a final glide. All the words begin with unaspirated voiceless stops and contain the low central vowel /a/, so that different onsets or vowel qualities are not contributing to variation of timing except in so far as differences between bilabial, dental, and velar place of articulation create such differences. The different tonal possibilities are sampled, but not all possible combinations of syllable type and tone are included.

These words were recorded in a simple frame sentence in which they were

preceded by the third person singular clitic /ʔa-/ and followed by the perfective aspect marker /tsâŋ/ (Peterson n.d.). This context preserves the tone heard on the verb in isolation, but, as we shall see later, /tsâŋ/ alternates between falling and low tone according to the preceding tone. These sentences were spoken three times in sets of three by Kenneth VanBik after being cued by the English gloss. For most words nine tokens are available for analysis. However, an error in cueing the words lead to only six tokens of /kâp/ “stick to” and twelve tokens of /kâáw/ “be wide”. The session was recorded on audiotape and digitized at 22 kHz for acoustic analysis. The major acoustic time points that can be identified in each token were then marked, based on inspection of a simultaneous display of the waveform and a spectrogram of the token. These time points were entered into a statistical package and the durations between them calculated.

Each of the following time points was identified in the target verbs and the following /ts/ whenever possible:

1. The release of the initial stop
2. The onset of voicing for the following /a/ vowel
3. The offset of the vowel (= the onset of closure for any following true consonant)
4. The end of a coda consonant (when present)
5. The release of the stop component of the affricate /ts/ of the perfective.

These five points are indicated in figure 1 by bold vertical lines numbered correspondingly.

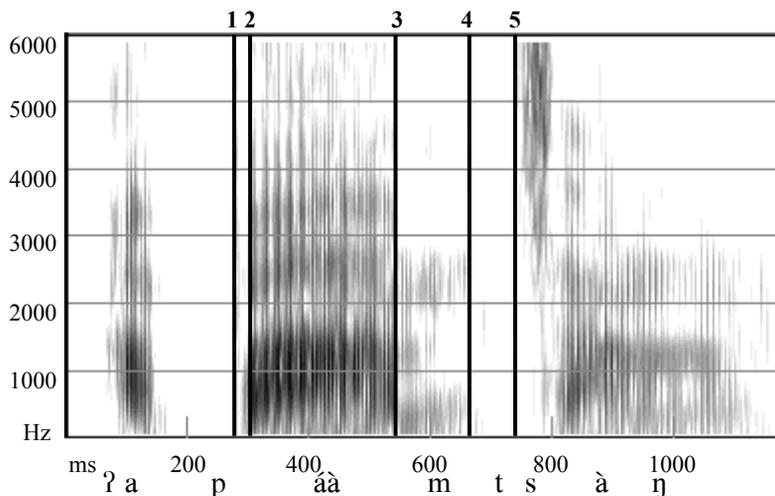


Figure 1: Example spectrogram illustrating acoustic segmentation.

Figure 1 illustrates one of the sentences including the verb /páàm/ ‘be wasteful’. In tokens of this type with a final sonorant (CVR, CVVR), all five time points are quite readily identifiable. Open-syllable words (CVV) obviously do not contain a time-point 4, marking the end of the consonant. In words with a coda stop (CVK and CVVK), there is usually no acoustic signal marking the end of the stop consonant (although a few final /-k/ tokens do have an identifiable consonant release transient), so that time-point 4 cannot normally be identified in these words either. In words with final glides (CVG, CVVG), a division cannot be reliably established between the vowel and the glide, so no time-point 3 is measured in these tokens. Because of these differences in measurement possibilities across the different categories, the durations that can be calculated vary by syllable type. This naturally will shape the presentation of results.

In addition to the time-points mentioned above, the fundamental frequency contour was obtained for the verb and the following /tsâŋ/. Average contour values as well as means at certain time points were calculated from this data. Greater detail on the procedure followed will be provided later. At this point it will only be mentioned that a general rule of Lai phonology modifies a falling tone after another falling tone to be low level. Hence the tone on the aspect marker in Figure 1 is shown as low. The subject clitic /ʔa-/ has a non-contrastive mid-range pitch in these sentences.

3. Voice onset interval

The first interval to examine is that between the initial consonant release (time-point 1) and the onset of the vowel (time-point 2), normally referred to as the voice onset time (VOT). There are significant differences in this duration according to the place of articulation of the consonant, as shown in Figure 2. If the initial consonant is /k/ the voice onset time is longest (mean 24.0 ms), and if it is /p/ it is shortest (mean 12.3 ms), while if it is /t/ it falls between these two, although closer to /k/ (mean 20.5 ms).

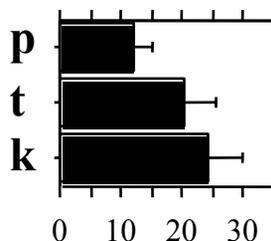


Figure 2: Mean voice onset time according to place of articulation (in ms).
 Bars show one standard deviation.

This pattern is very familiar from other languages and some comparative data is provided in Table 2. In every language for which VOT has been measured in voiceless unaspirated stops, a bilabial stop always has a significantly shorter VOT than a velar one. Stops made in the dental/alveolar area show somewhat more variable patterns, probably reflecting the fact that the closure for stops in this area can be formed in a number of different locations and the shape of the tongue can be varied. These differences of VOT according to place are so widespread that they are taken to be due to automatic and universal phonetic processes (Maddieson 1997, Cho and Ladefoged 1999), even though their precise causation is still open to some debate.

Table 2: Some comparisons of VOT of initial voiceless unaspirated stops (in ms).

	Mandarin (Shimizu 1996) 3 speakers	Burmese (Shimizu 1996) 2 speakers	Finnish (Suomi 1980) 10 speakers	Bunun (Maddieson et al., ms) 14 speakers
p	/_ i, a, u 7	7	9	11
t	/_ i, a, u 12	16	11	11
k	/_ a, u 19	31	20	25

Voice onset time also differs slightly in the Lai data depending on whether a long or a short vowel follows the onset, being slightly longer in the long vowel case. In an analysis of variance with consonant place and vowel length as main effects, vowel length has a significant effect. This is largely due to the short VOT in tokens of /tâm/ where the mean is only 16.1 ms. This might be because a short vowel before a nasal anticipates the nasality of the final consonant and thus the pressure build-up behind the stop closure is partly dissipated through the nose. Understanding of the speech production mechanism as well as descriptive studies of other languages suggest that further automatic variation in VOT is to be expected due to the height of the following vowel, but this factor was not varied in the data used in this paper.

4. Vowel duration

The mean duration of the vowel from the consonant release (time-point 1) to the end of the vowel (time-point 3) is shown in Figure 3 individually for each word. Words are ordered by decreasing durations, with long and short vowels distinguished by shading. Only words in which the end of the vowel can be determined are included, that is, those not ending in a glide. In this figure the voice onset interval is included in the vowel duration. Opinions differ as to whether this is the best measurement strategy for assessing acoustic vowel length or whether the duration of a vowel should exclude the

voice onset interval (see, for example, Port and Rotunno 1979, Beckman 1982). However, since the quality of a following vowel can be fully recognized from hearing the noise interval following the release, this choice has good motivation. Vowel durations calculated including the voice onset interval were compared with durations calculated without including the voice onset interval, that is, from time-point 2 to time-point 3. This reduces the difference between the pair *pâk/kâp* from 11 to 2 ms, but increases the difference between the pair *pààk/kààp* from 17 to 32 ms. Durations calculated omitting voice onset time do not therefore appear to be more consistent than those including voice onset time. In all subsequent discussion in this paper, voice onset time will be counted as part of the vowel duration.

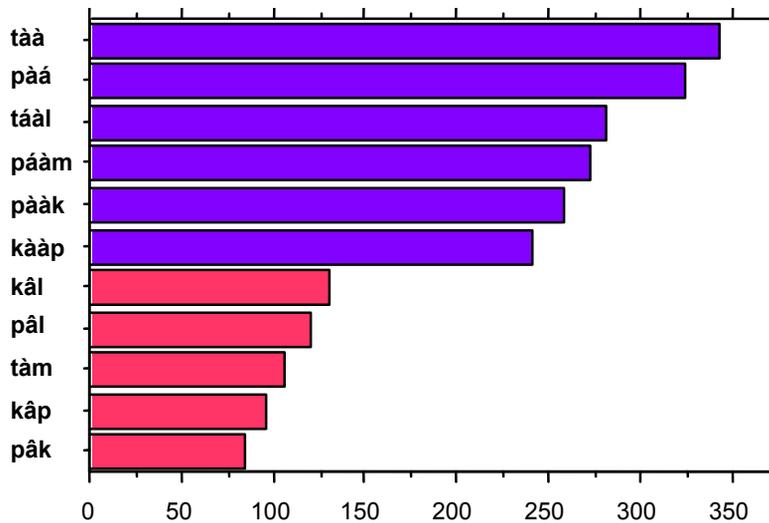


Figure 3: Vowel duration (including VOT) by word.

Vowel Quantity. As expected, Figure 3 shows that long vowels are substantially longer than short vowels. The grand mean of the durations in this data set for all the short vowels is 108 ms ($n = 42$) and is 287 ms ($n = 54$) for all the long vowels, giving a ratio of about 1:2.7 between them. In closed syllables with long vowels, the mean vowel duration is 264 ms, so the ratio between short and long vowels in the environment in which they contrast is closer to 1:2.4 and this ratio is better taken as representative of the relationship between long and short vowel durations.

Comparison data from several languages are given in Table 3. The languages cited are ones in which there is a clear quantity distinction between long and short vowels, with only minor and expected qualitative differences involved. Despite some differences in materials and methodology between the studies, the ratio of the durations is

consistently near or above 1:2. A ratio of this order of magnitude appears to be typical of a stable vowel quantity distinction, and can be taken as a standard way of implementing such a contrast.

Table 3: Some comparisons of distinctive short and long (stressed) vowel durations (in ms).

	Tamil (Balasubramanian 1981) 4 speakers	Somali (Farnetani 1981) 4 speakers	Finnish (Wiik 1965) 5 speakers	Aleut (Rozelle 1997) 4 speakers
short	112	90	99	78
long	221	227	232	151
ratio	<i>1 : 2.0</i>	<i>1 : 2.3</i>	<i>1 : 2.3</i>	<i>1 : 1.9</i>

Within the data presented in Figure 3 there are quite salient differences within each quantity category between vowels in different syllable structures. Long vowels in open syllables are longer than those in sonorant-final syllables, which are in turn longer than those in stop-final syllables (means 334 ms, 277 ms and 250 ms respectively). Short vowels in sonorant-final syllables are longer than those in stop-final syllables (119 ms vs. 89 ms). Each of these means is significantly different from all others at better than a .0001 level. The mean durations of vowels grouped by different syllable types are plotted graphically in Figure 4, with long and short vowel syllables distinguished by shading.

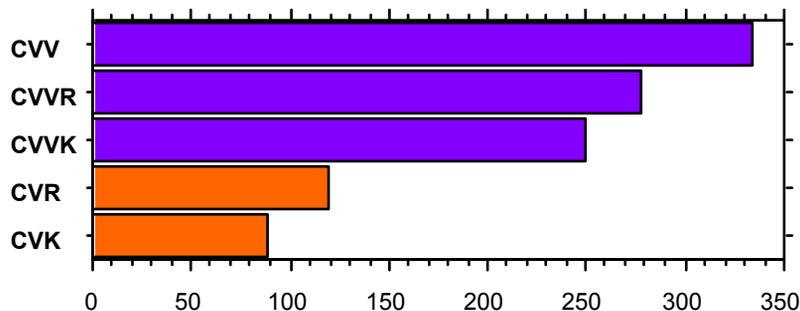


Figure 4: Vowel duration by syllable type.

Figure 4 makes clear that the class of final consonant has an essentially linear rather than a proportional effect on these durations. The mean duration difference between the vowels in stop-final and sonorant-final syllables is 27 ms for the long vowels and 30 ms for the short vowels. A long vowel before a stop thus has 90% of the duration of a long vowel before a sonorant, whereas a short vowel before a stop has only 75% of the duration of a short vowel before a sonorant.

Individual final consonants within a class do not have consistent effects, as the data in Figure 3 illustrate. In the short vowel pair *pâk/kâp* the second word of the pair has a slightly longer mean vowel duration, but in the long vowel pair *pààk/kààp* it is the first member of the pair that has the longer duration. There is a weak indication in these results that vowels before a lateral coda may regularly be longer than those before a nasal coda, but the difference between vowel length before /-m/ and before /-l/ in /*pâàm/* and /*tââl/* is less than that between the two lateral-final words /*kâl/* and /*pâl/*. It therefore seems more likely that the variation seen between the sonorant-final words is unsystematic.

It appears that the influence of syllable type on vowel duration shown in Figure 4 should be partitioned into universal and language-specific effects. Closed-syllable vowel shortening is widespread enough in the world's languages that it is appropriately regarded as a universal. The shorter duration of a vowel before a tautosyllabic consonant seems likely to be one of the perceptually significant indications of syllable structure (Maddieson 1985). However, the magnitude of the shortening effect appears to be quite variable across languages, indicating a language-specific component. Cross-language comparisons based on existing literature are difficult because the data is often unmatched (and sometimes unmatchable because of dissimilar phonological possibilities), but some comparative data is given in Table 4. In Mandarin Chinese, for example, the mean duration of vowels before final nasals, the only possible coda consonants, is substantially shorter than the duration of the same vowels when no final consonant is present. (Cf. Ren 1988.) A smaller but still substantial difference was found by Engstrand between a sample CV vs. CVC word pair in a Saami ("Lappish") dialect. However, a much smaller difference occurs in the Arabic of the Eastern Mediterranean. Averaged results for three speakers calculated from data in Broselow et al. (1997) show only a modest duration difference between long /a:/ in open and in closed syllables and none for short /a/. A similarly small difference was found in Hausa vowels before medial singleton and geminate consonants (details in Maddieson 1985). Shortening does not occur in Japanese (e.g., Smith 1992), nor in data from two Hindi speakers reported by Broselow et al. (1997) and data from two Hungarian speakers reported by Ham (1998).

Table 4: Some comparisons of closed syllable vowel shortening effects (in ms).

	Mandarin (Ren, p.c.) 4 speakers	Saami (Engstrand 1987) 1 speaker	Arabic (Broselow et al. 1997) 3 spkrs	Hausa (Lindau, p.c.) 10 speakers
open	CV 363	/la/ 346	/Xa:.bX/ 133	CV.CVV 68
closed	CV _n , CV _ŋ 219	/lat/ 267	/Xa:b.X/ 114	CVC.CVV 53
difference	<i>144</i>	<i>79</i>	<i>19</i>	<i>15</i>

Table 4 suggests considerable cross-language variability in the degree of closed vowel shortening; but, because there are many influences on duration, this data cannot be taken uncritically as a demonstration of the scale of the differences between these languages. The materials measured in the first two languages are monosyllabic words, whereas the Arabic and Hausa data are from trisyllabic and disyllabic words respectively. Longer words frequently have shorter durations of their component segments, and may show reduced influence of the factors controlling vowel duration variation. Moreover, different consonants in coda position may have quite varied effects, especially, it appears, within languages such as English (Wiik 1965) and French (O'Shaughnessy 1981) that allow for a very large number of distinct codas.

Nonetheless, it is fair to conclude that the Lai data show a relatively strong role for closed syllable vowel shortening with a magnitude of the effect in the mid-range of that reported in the sample of languages examined, and a small role for influence of coda consonant type. In the Lai monosyllabic words with long vowels, the mean vowel duration across the two closed syllable types is 264 ms, 70 ms less than in open syllables. Separating the two types, between CVV and CVVR structures the mean difference is 57, and between CVV and CVVK structures the mean difference is 84 ms. These values are of the same general magnitude as noted in Table 4 for Saami, and similar to that reported by Broselow et al. (1997) for long /aa/ vowels in Malayalam, where the difference is 64 ms between CVV.CX and CVVC.X structures (273 ms vs. 209 ms; 3 speakers). Broselow et al. suggest that this size difference in Malayalam is associated with the fact that this language occupies a particular place in a typology of syllable weight (namely, coda consonants are non-moraic and weight is only determined by vowel quantity). They predict from this that coda consonant duration will not vary according to the quantity of the preceding vowel since in both cases the consonant is regarded as sharing a mora with the preceding vowel, while vowels before a coda are shortened precisely because the vowel's mora is shared. If the proposed typology is sound and if Lai is like Malayalam, then Lai codas would not be expected to vary in duration as a function of preceding vowel quantity—yet they do.

5. Rhyme duration

As noted earlier, the onset and offset times of coda segments in target words in which these points can reliably be determined was recorded, that is, in the words containing final sonorants, and the interval between time-points 3 and 4 calculated. The relevant data is shown in Figure 5. In this figure the first shaded duration for each word represents the mean acoustic duration of the vowel, the second shaded duration the mean acoustic duration of the coda sonorant. Across the three words in the CVR category the mean coda consonant duration is 209 ms but in the two words of the CVVR category mean coda duration is 134 ms. A sonorant following a long vowel is substantially shorter than one following a short vowel. Consequently, the difference in total rhyme duration between long vowel and short vowel words is much less than the difference in the duration of the vowels themselves (although the VC duration in words with long vowels remains somewhat longer than in words with short vowels).

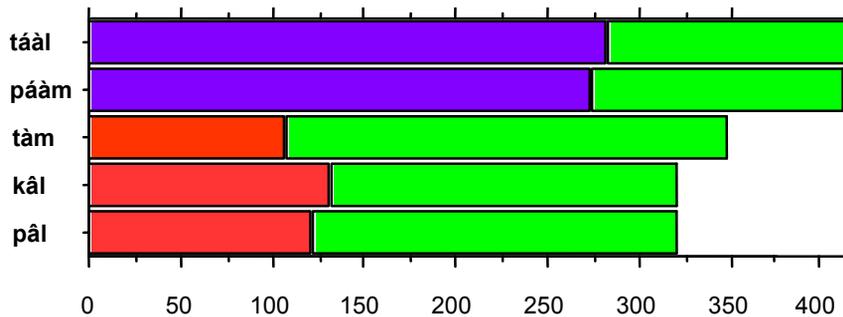


Figure 5: Vowel duration (including VOT) and final sonorant duration, by word.

A second view on this issue can be obtained by calculating the duration from the onset of the vowel in the target vowel to the release of the closure in the onset of /tsâŋ/ (time-point 5). This duration corresponds to the rhyme duration of the target words plus the closure duration of the stop component of the onset of /tsâŋ/ and can be calculated for all syllable types, including in stop-final words where point 4 is not identifiable and in glide-final words where time-point 3 cannot be identified. Comparison with CVV syllables can also be made. It is assumed that variations in this duration largely reflect variation in the rhyme duration, with the [t]-closure remaining essentially constant. Results are shown in Figure 6 by syllable type.

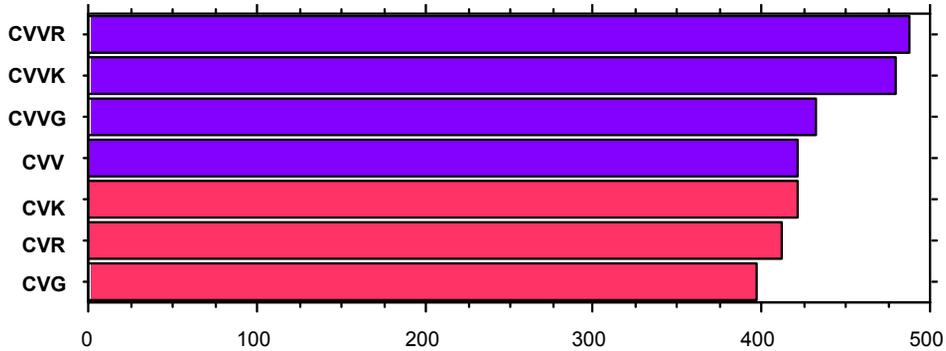


Figure 6: Syllable onset to syllable onset duration, by syllable type.

Figure 6 confirms that the durations of different rhyme types show smaller differences than that between long and short vowels, although a one-way analysis of variance shows that syllable type does have a significant effect on this syllable-onset to syllable-onset duration. Post-hoc comparison of means shows that this is due to the greater duration of stop and sonorant-final cases (CVVK and CVVR), which are significantly longer than all others. None of the short vowel types and open or glide-final types are significantly different from any of the others (significance threshold $p < .01$). Having a final true consonant adds some 40 to 50 ms to the duration of a long vowel rhyme. However, most of the difference in duration between long and short vowels is compensated for by duration adjustment of the coda consonant. Mean vowel duration in CVVK is 250 ms and in CVK it is 89 ms, a difference of 161 ms. In CVVR mean vowel duration is 277 and in CVR it is 119 ms, a difference of 158. But the difference between syllable-onset to syllable-onset duration in CVVK vs. CVK cases is only 67 ms, and between CVVR and CVR cases is only 75 ms, so the rhyme durations differ by less than half the difference between the vowel durations.

Glide-final rhymes with long and short vowels are even closer in duration. In the glide-final cases the rhyme duration can be directly measured; the mean is 358 ms in CVVG, and 320 in CVG, a difference of only 35 ms. (Recall that open-syllable rhymes fall in between these, with a mean duration of 334 ms). The fact that the glide is lengthened after a short vowel can be shown qualitatively by the spectrograms in Figure 7.

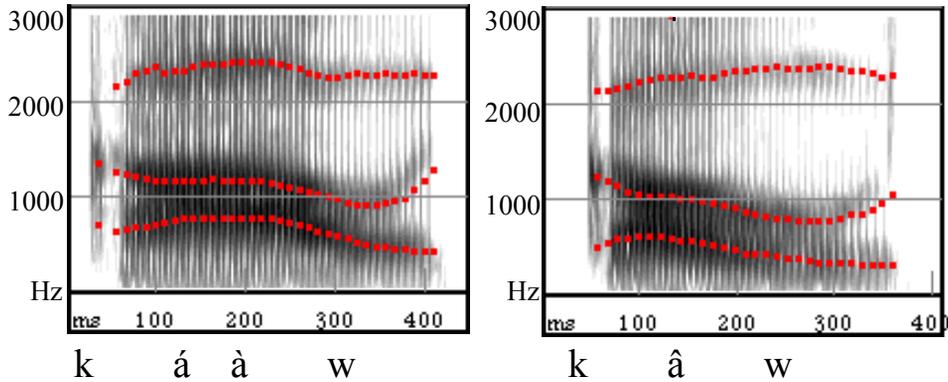


Figure 7: Spectrograms of /káàw/ and /kâw/ showing lowest three formants.

In these spectrograms the first three formants, estimated by an LPC procedure, are marked by a string of dots. The transition from the vowel /a/ to the glide /w/ is reflected in the lowering of two lowest formants, and by the reduction of amplitude most easily visible in these displays as the light shading between the second and third formants, and above the third formant. In the token of /káàw/ shown on the left of the figure there is a relatively sharp change in amplitude at around 280 ms and the first and second formant frequencies start falling appreciably before this time. In the token of /kâw/ shown on the right, the amplitude drop is not so well-defined but clearly occurs much earlier, as does the fall in formant frequencies. Strikingly, the frequencies of the first two formants are considerably lower throughout the short vowel in /kâw/ in comparison to the long vowel in /káàw/, indicating that the rounding of the glide is anticipated in the short vowel. The vowel, which might be transcribed as [ɔ] in this context, is thus identifying the /aw/ rhyme right from its beginning.

6. Compensatory duration adjustment in cross-linguistic perspective

We now return to the question raised by the typology proposed by Broselow et al. (1997) and proposals similar in spirit: To what extent are the intra-syllabic compensatory duration adjustments shown in the preceding section predictable from phonological quantity and rhythmic factors? In a number of languages, the durations of vowels and their following coda consonants show a pattern of compensation tied to phonological vowel quantity broadly similar to that seen in Lai. In Standard Thai, for example, Mixdorff et al. (2002) show that nasals are about 35 ms longer on average after a short vowel than after a long vowel in words with the syllable shapes shown in Table 5 (long

vowels are 79 ms longer than short vowels in these words, so CVVN rhymes remain about 45 ms longer than CVN rhymes).

Table 5: Thai VN durations

(Mixdorff et al. 2002; 2 speakers, 2 tone categories per segment string, 5 repetitions).

Short V	V dur	C dur	rhyme	Long V	V dur	C dur	rhyme
lon	131	117	248	loon	205	77	282
saŋ	139	123	262	saan	205	93	298
wan	116	113	229	waan	212	79	291
mean	128.7	117.7	246.4		207.3	83.0	290.3

These results seem very comparable to data on /an/ and /aan/ rhymes from two Cantonese speakers in a study by Zee (see Zee 1995) reported by Lee (1999:144-6), although specific numbers are not included. Limited data based on three pairs of nasal-final Cantonese words in the IPA Illustration of this language yield short and long vowel durations in nasal-final rhymes of 111 and 218 ms respectively (IPA 1999; recordings available on IPA website). The nasals in these words have durations of 130 ms and 89 ms respectively. Mean duration of CVN is thus 241 ms, and of CVVN 307 ms. The difference in rhyme durations is 66 ms.

Although the three languages compared here are similar in showing rhyme-internal duration compensation, the magnitude of the effect has a language-specific component. In our Lai data, coda sonorants after short vowels are actually longer than the vowels they follow. In neither the Thai nor the Cantonese material is this the case.

Broselow et al. (1997) suggest that coda duration compensation will be absent in languages in which a coda consonant does not contribute to phonological weight (as in Malayalam) or invariably contributes its own mora (as they argue is the case in Hindi). It is expected to be present in cases where VC and VVC rhymes are equally heavy, since in the first structure V and C have a full mora each, whereas in the second the final VC string shares a single mora. (Cf. Maddieson 1993.) Broselow et al. suggest that this type is illustrated by the duration pattern in non-word-final syllables of Levantine Arabic. Further they predict that in this case closed syllable vowel shortening will only be found in syllables with long vowels, since in both V and VC rhymes V has a full mora to itself. For Lai at this point it is not certain how codas should be evaluated with respect to contributing to syllable weight, but one consideration argues in favor of this conclusion. Lexical forms of CV shape are disallowed, whereas CVV and CVC are legitimate. This can be taken as evidence for a bimoraic minimum word template with codas contributing a mora. Recall that rhyme durations in CVV and CVC are equal in Lai. However, it is not clear if the also-permitted CVVC syllables are “heavier” in any

sense than CVV or CVC syllables or should be considered equal to these. As noted in section 6, Lai coda consonants are indeed shorter after long vowels than after short vowels. Using the typology of Broselow et al. as a guide, final C in CVVC syllables might therefore be taken as sharing the final mora, and not contributing to syllable weight. It thus looks as if Lai is of similar type to Levantine Arabic. However, a final piece of evidence in support of this conclusion must remain lacking. According to this typology, a weight contributing coda in a VC rhyme should not shorten the preceding short vowel. Since Lai does not allow lexical CV, this prediction cannot be tested.

Ham (1998) suggests that durational compensation of coda consonants is likely to be absent in languages in which there is a quantitative distinction among consonants (i.e., where singleton and geminate consonants contrast). This implies that it is normally to be expected in languages with vowel quantity contrasts, but lacking consonant quantity contrasts. Lai, Thai, and Cantonese all meet this condition. Durational compensation is also to be expected in languages that fall into the (admittedly still problematical) rhythmic category of being syllable-timed, since such compensation helps to make rhyme durations more equal. (Cf. Smith 1992.) These languages are also judged to be syllable-timed.

7. Duration and tone category

The preceding sections show that Lai fits quite well into several of the phonological typologies with a bearing on segmental durational patterns that have been proposed. Some of these make appeal to the theoretical construct of the mora and the associated concept of relative syllabic weight. Gordon (1999, 2002) argues that the mismatches between what is likely to count as “heavy” for different roles, such as stress-attraction as opposed to eligibility for tone-bearing, argue against such a construct and in favor of a more concrete phonetic account, including specifically the suggestion that relative durations of sonorant rhyme portions account for tone bearing ability. Zhang (2000) goes further in suggesting that contour tones require more duration for their realization than level tones and more complex contours require more duration than simpler ones. Indeed in many languages it has long been noted that the duration of segments in the rhyme varies according to the tones they bear, and in Mandarin sonorant rhyme duration matches quite well with tone contour complexity (Woo 1969); for example, tone 3 rhymes are longer than those of the other tones, and rising tone is longer than falling tone, a fall being considered simpler than a rise because of how the pitch control mechanism works. Lai provides a challenge to such a phonetically transparent account. Stop-final syllables with short vowels have the shortest sonorant rhymes, yet bear contour tones; and the relative durations of the all-sonorant rhymes with long vowels (VV,

VVR, VVG) pooled across the tokens in this study do not correlate with tone contour complexity, as is shown in Figure 8.

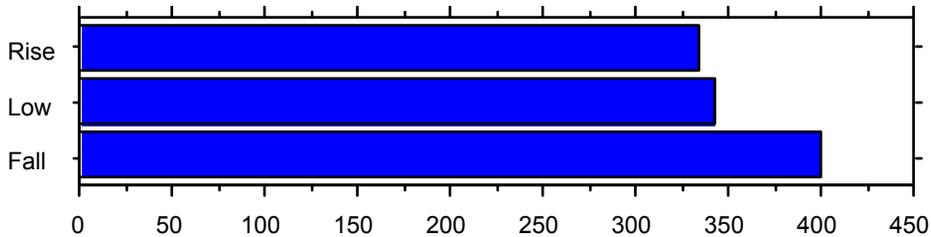


Figure 8: Duration of long all-sonorant rhyme durations by tone category.

8. Tone shape and alignment

We now turn attention to the alignment of Lai tonal contours with the segmental material. For all the utterances in this study fundamental frequency was estimated using an autocorrelation procedure in the Macquiner phonetic analysis package (www.SCICONRD.com). F0 estimates were obtained over 35 ms time windows at 10 ms intervals. Successive estimates were constrained to be within 20 Hz. Because of the length of the window—long enough to cover three to four periods of vocal fold vibration—this procedure means that there is a certain amount of smoothing of the actual period-to-period variation in vocal fold vibration. Mean amplitude values for each window were also examined. F0 values for those with amplitude below 9 dB were eliminated, as this threshold appeared to indicate that the window was largely voiceless.

In order to obtain a general idea of the characteristic F0 patterns, the F0 tracks were aligned and averaged. To preserve as much of the timing information as possible, the following procedure was employed. Repetitions of the individual target words were aligned with each other at their onsets (i.e., where calculable values for F0 were first reported). Across the (usually 9) repetitions of any given word the duration over which F0 was calculated naturally varied somewhat. However, in almost all cases there was a clear modal duration (five repetitions with the same number of F0 windows). Tokens with F0 values beyond this duration were truncated. In tokens with fewer than the modal number of F0 windows, additional values were interpolated based on the value in the last window and the trend seen in the other tokens.

The number of values between the last F0 value in each target word (before truncation or interpolation) and the first F0 value in the marker /tsaŋ/ was obtained and the mean number of 10 ms steps corresponding to the unvoiced portion (that is, /ts/ and any preceding voiceless stop) was calculated for all utterances containing a given target

word. The strings of F0 values for the /tsaŋ/ syllables were then aligned after this mean number of steps. Following alignment, the same truncation/interpolation procedure was applied to the end of the /tsaŋ/ syllables. Finally the F0 values at each time point were averaged for each utterance type with a given target word. The resulting F0 contours are plotted in Figures 8, 9, and 10 using different shading and point shapes to identify each type. Three figures are used to avoid too much superposition of the individual tracks. The time and frequency scales are the same for the three figures, so that different patterns may be readily compared.

Since all the target words and the /tsaŋ/ marker begin with voiceless obstruents, the local raising effect of this class of consonant on the first few periods of vocal fold vibration must be taken into account. As a rule of thumb, onset pitch can be expected to be some 10-15 Hz higher than it otherwise would be. The effect diminishes over the first 50 ms of the following vowel.

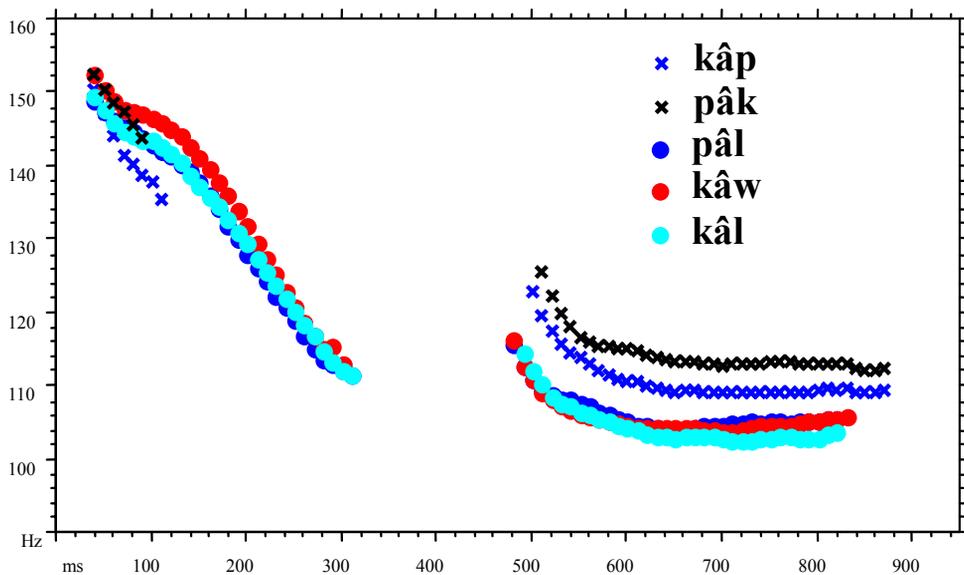


Figure 9: F0 in target words with short vowels and falling tones and following /tsaŋ/.

Figure 9 displays the target words with short vowels and falling tones. In those with coda sonorants the fall is mostly realized during the sonorant portion, which as noted earlier is longer than the vowel. (Cf. Figures 5 and 7.) The end-point of the fall in these cases is around 112 Hz, which is similar to the end-point of the falls in long-vowel rhymes in Figures 10 and 11. The short duration of the voiced portion in CVK syllables does not allow very much of a falling contour to be realized. The truncated

falls in /kâp/ and /pâk/ end closer to 140 Hz (± 5). Although the basic tone on /tsâŋ/ is falling, a regular phonological process in Lai realizes a falling tone as low when a falling or low tone precedes, as discussed by Hyman and VanBik (2004). Hence /tsâŋ/ has a low tone in all these utterances. However F0 in this syllable is notably higher after the stop-final words than after the sonorant-final words, so the truncation of the fall has a phonetic effect that persists throughout the syllable.

Figure 10 is designed primarily to enable rising, falling, and low tones in long-vowel target words to be compared. Rising and low tones start from very similar onset values. Rising tone begins to rise very early—in fact, if the effect of the local F0 raising by initial /p, k/ is allowed for, the rise can be viewed as beginning immediately at the syllable onset. Cross-language comparison suggests that this is unusual—rising tones generally begin to rise relatively late in the syllable. This pattern is reported, inter alia, for rising tones in Thai (e.g., Abramson 1962), Mandarin (tone 2, e.g., Shih 1988, Tseng 1990), Shanghai (tone 3, e.g., Rose 1993, Zhu 1999), Taiwanese (e.g., Peng 1997), and Cantonese (*yin shang* and *yang shang* tones, Lee 1999). Lee's data for four Hong Kong Cantonese speakers shows a particularly late onset of rising F0 in rising tone syllables, but in the other studies cited the rise often does not begin until at least 40% of the voiced portion of the syllable is completed. In Lai on the other hand the initial low of a rising tone onset appears to be aligned with the vowel onset.

Falling tones in the long-vowel syllables have a slightly lower overall rate of fall than those in short-vowel syllables, with a mean change of 1.2 Hz or less per 10 ms step, compared to a mean change of 1.4 Hz or more. The difference in rate of decline in long-vowel and short-vowel syllables can be modeled by viewing the low target for the end of the fall as being at the end of the rhyme coda consonant, rather than at a constant duration. Long-vowel falls are also marked by a more evident concave shape to the part that follows the initial portion raised by the onset consonant, but some concavity is also visible in the short-vowel syllables with final sonorants or glides in Figure 10. (Why the falling contour in /káàw/ is lower than that in /pâàm/ is unknown, but it leads to a lower pitch through much of the following /tsâŋ/ syllable.)

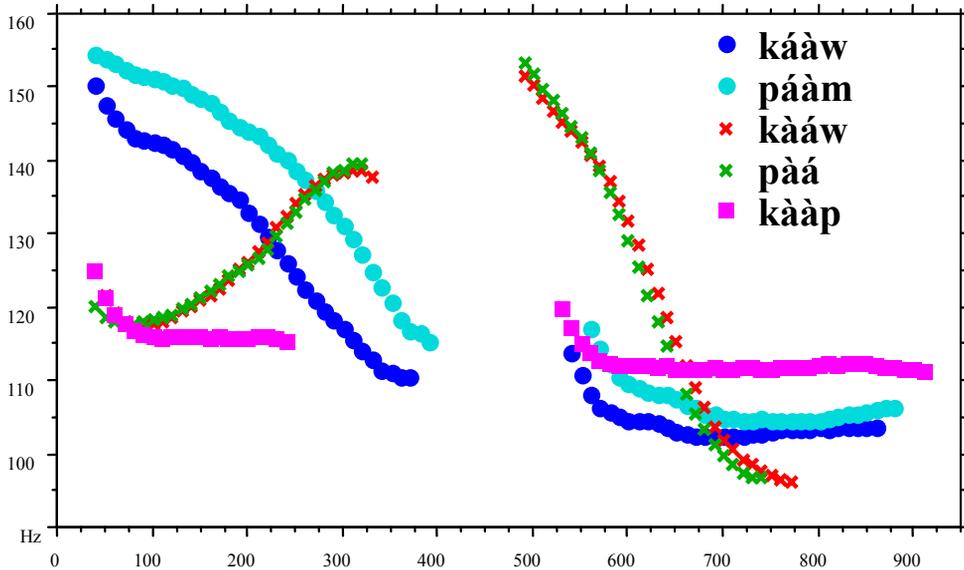


Figure 10: F0 in some target words with long vowels and following /tsàŋ/.

Figure 11 is primarily designed to compare the long- and short-vowel target words with low tone. To provide context, one of the falling tone long-vowel words is also shown on this figure. Low tones are very similar regardless of rhyme structure, except for the somewhat truncated duration of the voiced portion in stop-final rhymes, also illustrated by /kààp/ in Figure 10.

The shape and timing of the F0 contour during /tsaŋ/ in the different contexts can also be compared in Figures 9-11. As already noted, the F0 level in some of the target words has a perseverative effect on the phonetic level in following /tsaŋ/. But there are significant non-assimilatory effects. A falling contour is retained on /tsaŋ/ when a rising tone precedes. Naturally, this means that the onset F0 of /tsaŋ/ is considerably higher. (Onset F0 is also higher in /tsaŋ/ when the preceding target word is stop-final. The onset F0 values of all the individual tokens were analyzed according to the rhyme category of the preceding target words. This showed that three categories are significantly different at at least the .01 significance level—those following rising tone target words, stop-final target words, and all others. These group means are plotted in Figure 12. (The stop-final category is significantly different even if the short-vowel stop-final words are omitted, so this effect is not due to the truncation of falls noted in discussion of Figure 9.)

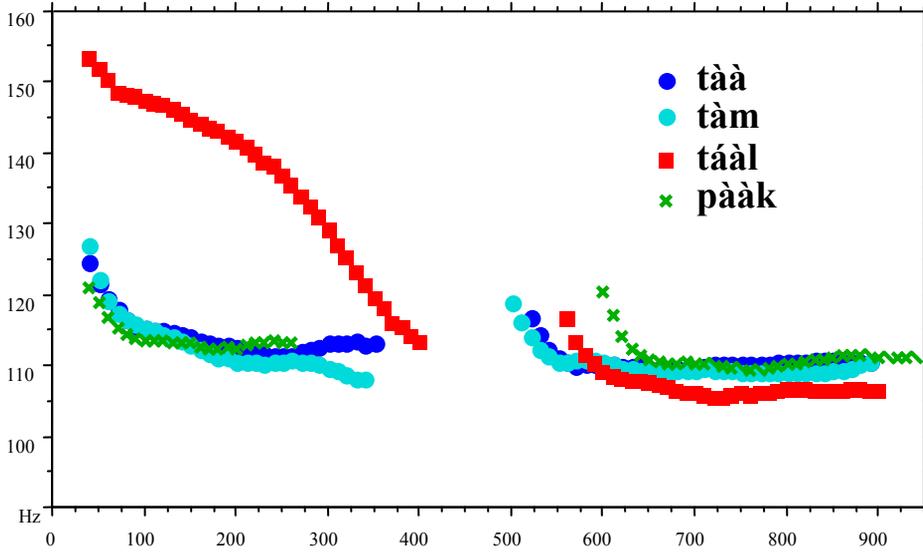


Figure 11: F0 in some target words with long vowels and following /tsàŋ/.

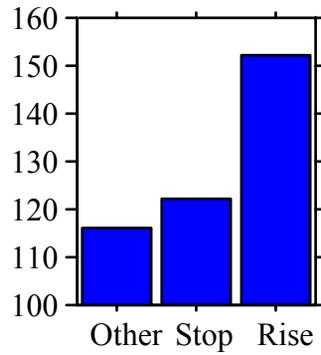


Figure 12: Mean onset F0 in /tsaŋ/ according to major significant preceding syllable categories.

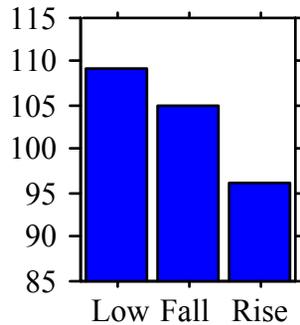


Figure 13: Mean lowest F0 in /tsaŋ/ according to preceding tone.

When /tsaŋ/ has falling tone, its onset F0 (mean 152.4 Hz) is at essentially the same level as the onset of falling tones in the target words (mean 151.7 Hz). But the rate of F0 fall is more rapid and the offset F0 value reached at the end of the fall is noticeably lower. In fact, the lowest F0 reached in /tsaŋ/ is significantly different for each of the preceding tonal categories. The lowest F0 in each individual token of /tsaŋ/ (before any truncation or interpolation) was obtained and analyzed according to preceding tone. The means are shown in Figure 13. The lowest F0 is significantly lower after a rising tone than after a low tone, as well as being lower still after a falling tone (threshold < .01). Thus a preceding High component on the target lowers the Low component of /tsaŋ/, and the closer the High is temporally the lower it goes. Sensitivity to the proximity of the nearest H element is also indicated by the fact that the lowest F0 is the final value when /tsaŋ/ is falling tone, but very often occurs well before the end of the rhyme when the tone is low.

Gandour et al. (1994) show a similar pattern in Thai for a lower minimum F0 in a low tone following a falling tone. By contrast, Xu (1997) finds that the F0 offset of a Mandarin falling tone (tone 4) is slightly higher after a rising tone (tone 2) than when a high tone (tone 1, assumed to have an earlier High target) precedes, and Laniran (1993) is among several authors to show that the end-point of the second Low in Yoruba Low-Low sequences is considerably higher than the end-point of High-Low sequences, which derive a falling F0 contour in the second syllable. A possible interpretation of the retention of the fall after a rise in Lai might be that it is purely a phonetic interpolation between the offset high of the rise and the low tone seen on /tsaŋ/ in other contexts. These comparisons with other languages suggest this is not the most satisfactory explanation. The differences in F0 minima of /tsaŋ/ argue both for a language-specific pattern in Lai and a clear falling target, with the Low end-point specified as being at the end of the rhyme.

9. Summary

In summary, the analysis of Lai timing patterns carried out thus far and the cross-linguistic comparisons made suggest the following partition of the influences on timing in this language. The alignment of laryngeal and oral movements at syllable onset produces a pattern of voice onset timing that is a clear phonetic universal. The implementation of the quantitative distinction of vowels is typical of such a contrast in the phonology and can be regarded as little influenced by language-specific phonetic implementation patterns. Lai shows both robust closed-syllable vowel shortening effects and rhyme-internal compensatory timing adjustments. This pattern would be unexpected in a language with both vowel and consonant quantity (with free combination of these), but the phonological typologies so far proposed do not yet seem to predict if other phonological properties of a language with quantitative distinctions only among vowels can account for the presence of these properties jointly. For the moment, closed-syllable vowel shortening is proposed as a language universal default requiring a language-specific override to be suppressed. Rhyme-internal duration compensation to the degree seen in Lai is best viewed as a language-specific pattern, though perhaps one having a widespread areal distribution in East and Southeast Asia. This conclusion is supported by the related and clearly language-specific anticipation of rounding in short-vowel /aw/ rhymes in Lai.

As for the timing of tone targets in Lai it appears that there are significant language-specific contributions to the way that tone targets are aligned with segments. In the majority of languages studied from this perspective it appears that contour tones have onset targets located well after the onset of the vowel of the carrying syllable. This is not the case for Lai. Contour tone offsets in Lai may well be constrained to remain at rhyme offset, but the data are not extensive enough to have much confidence in this conclusion. Nonetheless, the phonetic implementation of Lai tonal patterns seems to contain a good deal that is phonetically idiosyncratic.

The various timing patterns noted in the small amount of data examined here can thus be tentatively assigned to phonetic and phonological, and universal and language-specific layers of influence. However, a strong *caveat* needs to be issued about any study conducted with a single speaker and further research is needed on the timing patterns in Lai as well as on other languages so that more extensive and better-matched cross-language comparisons can be made. I hope this small excursion into acoustic timing relations will encourage others to help develop the larger database on many languages that we need for linguistically-informed work on timing to reach more general and interesting conclusions.

References

- Abramson, Arthur S. 1962. *The Vowels and Tones of Standard Thai*. Bloomington: Indiana University Research Center in Anthropology, Folklore and Linguistics.
- Balasubramanian, T. 1981. Duration of vowels in Tamil. *Journal of Phonetics* 9:151-161.
- Beckman, Mary. 1982. Segment duration and the 'mora' in Japanese. *Phonetica* 39: 113-135.
- Broselow, Ellen, S-I. Chen, and Marie Huffman. 1997. Syllable weight: Convergence of phonology and phonetics. *Phonology* 14:47-83.
- Cho, Taehong, and P. Ladefoged. 1999. Variations and universals in VOT: Evidence from 18 languages. *Journal of Phonetics* 27:207-229.
- Dauer, Rebecca. 1983. Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics* 11:51-62.
- Elert, Claes-Christian. 1964. *Phonologic Studies of Quantity in Swedish*. Uppsala: Almqvist and Wiksell.
- Engstrand, Olle. 1987. Salient features of Lule Sami pronunciation. *The Sounds of Lappish*, ed. by C.-C. Elert. Umeå: Department of Phonetics, Umeå University.
- Farnetani, Edda. 1981. *Dai Trattati ai Parametri: Introduzione all'Analisi Strumentale della Lingua Somala*. Rome: Ministero del Affari Esteri.
- Gandour, J., S. Potisuk, and S. Dechongkit. 1994. Tonal coarticulation in Thai. *Journal of Phonetics* 22:477-492.
- Gordon, Matthew. 1999. *Syllable Weight: Phonetics, Phonology and Typology*. Los Angeles: University of California dissertation.
- Gordon, Matthew. 2002. A phonetically-driven account of syllable weight. *Language* 78:51-80.
- Ham, William H. 1998. *Phonetic and Phonological Aspects of Geminate Timing*. Ithaca: CLC Publications, Cornell University.
- Hyman, Larry M., and Kenneth VanBik. 2002. Tone and syllable structure in Hakha-Lai. To appear in Proceedings, 28th Annual Meeting, Berkeley Linguistics Society, Special Session on Tibeto-Burman and Southeast Asian Linguistics.
- Hyman, Larry M., and Kenneth VanBik. 2004. Directional rule application and output problems in Hakha Lai tone. *Language and Linguistics* 5.4:821-861.
- IPA (International Phonetic Association). 1999. *Handbook of the International Phonetic Association*. Cambridge: Cambridge University Press.
- Klatt, D. H. 1976. Structure of a phonological rule component for a speech synthesis by rule program. *IEEE Transactions in Acoustics, Speech and Signal Processing, ASSP-24*:391-398.

- Klatt, D. H. 1979. Synthesis by rule of segmental durations in English sentences. *Proceedings of the 9th International Congress of Phonetic Sciences*, Vol.2, 290-297. Copenhagen: Institute of Phonetics, University of Copenhagen.
- Lee, Wai-Sum. 1999. *A Phonetic Study of the Speech of Cantonese-Speaking Children in Hong Kong*. Hong Kong: City University of Hong Kong dissertation.
- Lehiste, I. 1970. *Suprasegmentals*. Cambridge: MIT Press.
- Maddieson, Ian. 1985. Phonetic cues to syllabification. *Phonetic Linguistics*, ed. by V. A. Fromkin, 203-221. Orlando: Academic Press.
- Maddieson, Ian. 1993. Splitting the mora. *UCLA Working Papers in Phonetics* 83:9-18.
- Maddieson, Ian. 1997. Phonetic universals. *Handbook of Phonetic Sciences*, ed. by W. Hardcastle & J. Laver, 619-639. Oxford: Blackwell.
- Maddieson, Ian, Jie Zhang, and Richard Wright. (ms). The phonetics of Bunun.
- Matisoff, James A. 1990. On megalocomparison. *Language* 66:106-120.
- Melnit, N. 1997. The sound system of Lai. *Linguistics of the Tibeto-Burman Area* 20.2: 9-19, 125-127.
- Mixdorff, H., S. Luksaneeyanawin, H. Fujisaki, and P. Charnvivit. 2002. Perception of tone and vowel quantity in Thai. *Proceedings of the 7th International Conference on Spoken Language Processing*, 753-756. Denver.
- Mortensen, David, and K. VanBik. 2002. Vowel length and quality in Hakha Lai: A perceptual study. Paper presented at the 2003 Annual Meeting of the Linguistic Society of America, Boston.
- O'Shaughnessy, Douglas. 1981. A study of French vowel and consonant durations. *Journal of Phonetics* 9:385-406.
- Peng, Shu-hui. 1997. Production and perception of Taiwanese tones in different tonal and prosodic contexts. *Journal of Phonetics* 25:371-400.
- Peterson, David. A. n.d., 2000? Hakha Lai. Manuscript. Max Plank Institute for Evolutionary Anthropology, Leipzig.
- Plauché, M., R. Beam de Azcona, R. Roengpitya, and W. F. Weigel. 1998. Glottalized sonorants: A linguistic universal? *Proceedings of the Special Session on Phonetic and Phonological Universals, Berkeley Linguistic Society 24th Annual Meeting*, 381-390.
- Port, Robert, and R. Rotunno. 1979. Relation between voice-onset time and vowel duration. *Journal of the American Acoustical Society* 66:654-662.
- Ren, Hongmo. 1985. Linguistically conditioned duration rules in a timing model for Chinese. *UCLA Working Papers in Phonetics* 62:34-49.
- Ren, Hongmo. 1988. Temporal structure of the syllable with nasal ending. *Report of Phonetic Research, Institute of Linguistics* 18:53-79.
- Roengpitya, Rungpat. 1997. Glottal stop and glottalization in Lai (connected speech).

- Linguistics of the Tibeto-Burman Area* 20.2:21-56 (text), 115-123 (captions), 128-162 (spectrograms).
- Rose, Phil. 1993. A linguistic-phonetic acoustic analysis of Shanghai tones. *Australian Journal of Linguistics* 13:185-220.
- Rozelle, Lorna. 1997. The effect of stress on vowel length in Aleut. *UCLA Working Papers in Phonetics* 95:91-101.
- Shih, Chilin. 1988. Tone and intonation in Mandarin. *Working Papers of the Cornell Phonetics Laboratory* 8:83-109.
- Shimizu, Katsumasa. 1996. *A Cross-Language Study of Voicing Contrasts of Stop Consonants in Asian Languages*. Tokyo: Seibido.
- Smith, Caroline. 1992. *The Timing of Vowel and Consonant Gestures*. New Haven: Yale University dissertation.
- Suomi, Karl. 1980. *Voicing in English and Finnish Stops*. Turku: Department of Finnish and General Linguistics, University of Turku.
- Tseng, Chiu-yu. 1990. *An Acoustic Phonetic Study on Tones in Mandarin Chinese*. Taipei: Institute of History and Philology, Academic Sinica.
- Wiik, Kalevi. 1965. *Finnish and English Vowels*. Annales Universitatis Turkuensis, Series B, Vol.94. Turku: University of Turku.
- Woo, Nancy. 1969. *Prosody and Phonology*. Cambridge: MIT dissertation.
- Xu, Yi. 1997. Contextual tonal variations in Mandarin. *Journal of Phonetics* 25:61-83.
- Zee, Eric. 1995. Temporal organization of syllable production in Cantonese. *Proceedings of the 13th International Congress of the Phonetic Sciences, Stockholm*, Vol.3, 250-253.
- Zhang, Jie. 2000. *The Effects of Duration and Sonority on Contour Tone Distribution—Typological Survey and Formal Analysis*. Los Angeles: University of California dissertation.
- Zhu, Xiaonong S. 1999. *Shanghai Tonetics*. München and Newcastle: Lincom Europa.

[Received 11 April 2003; revised 15 June 2004; accepted 19 June 2004]

Department of Linguistics
1203 Dwinelle Hall, #2650
University of California, Berkeley
Berkeley, CA 94720-2650
USA
ianm@socrates.berkeley.edu

時量與連綴：以俛語爲例

Ian Maddieson

加州大學柏克萊校區

時間計量是人類語言中語音與音韻組成不可或缺的要素。本文從哈卡俛語的時計效應探討音段時長、音節內音段間的時計關係、以及聲調與音段的連綴。哈卡俛語雖然是印度和緬甸北部的一種藏緬語，但是音節結構與聲調型態卻頗類似“漢語圈”的語言。哈卡俛語有值得注意的前行元音與韻尾輔音之間的抵補音長，而其聲調連綴時的時阻亦較其他語言爲短。本文在討論中，將著重分析俛語的時量效應何者來自語言通則、何者屬於個別語言特徵、以及這些效應在計量類型學上的地位等問題。

關鍵詞：言語時量，聲調連綴，計量類型學