The Alignment of Gesture and Intonation in Pwo Karen

by

Jessica Hsieh

Drs. Claire Bowern and Jelena Krivokapić, Advisors

Yale University

Department of Linguistics

New Haven, CT

May 10, 2012

Table of Contents

Abstract

Although research on the relationship between gesture and prosody began two decades ago, the studies to date have been conducted exclusively on non-tonal languages (e.g. de Ruiter 2000, Loehr 2004, Yasinnik et al. 2004, Jannedy and Mendoza-Denton 2005, Leonard and Collins 2011). The studies mentioned above offer evidence that manual gesture strokes, which are characterized by an abrupt stop in movement called a "hit", and words bearing phrase-level pitch accents coincide with one another in non-tonal languages.

I argue that discrete manual gestures are similarly aligned with prosodically prominent words in Pwo Karen, a tonal Tibeto-Burman language. This paper offers a brief account of the phonology and intonation of Pwo Karen, and investigates the nature of interaction between gesture and intonation in that language. I annotated three videos of narratives by native Pwo Karen speakers for discrete gesture hits and prosodic prominence. I found that the majority of discrete gestures in each video were associated with prominent words. Although the pattern of alignment is far less consistent than that in previously studied languages, the results suggest the existence of coordination between the vowel onset and the apex of the gesture hit .

Acknowledgments

My deepest gratitude goes to both of my advisors for their guidance throughout the process.

Conducting research on Pwo Karen would have been impossible without the dedication of Kala Noo, who devoted many weekends over the past year to helping me record and translate videos. I am very grateful to the speakers in this study and in other projects for their time and for their interest in the documentation of Pwo Karen.

My family and my suitemates were sources of constant support and enthusiasm. Special thanks to Melissa Lavoie, Kate Mayans, and Pim Wacharaprapapong for proofreading help, and to Nick Huang for spending four hours helping me make Figure 15 and advising me on all the statistics.

This study was generously supported by a Mellon Grant from Jonathan Edwards College at Yale University.

1. Introduction

This study provides a description of the phonology and prosody of Pwo Karen, a Tibeto-Burman language, in addition to investigating the nature of the alignment between gesture and intonation in that language. It will test the hypothesis that abrupt gesture hits and prosodically prominent words will align, as has been previously demonstrated for English, and that the cognitive planning mechanisms implicated in gesture-prosody alignment in English will also be supported by evidence from Pwo Karen.

1.1 Background on Pwo Karen

Pwo Karen is a member of the Karen branch of the Tibeto-Burman language family. It is spoken by approximately 1 million people, most of whom live in southeastern Burma and northern Thailand (Ethnologue). The language may be separated into at least two dialect groups, Eastern and Western, which are not mutually intelligible and differ substantially in their phonology (Kato 1995). The Karen branch also includes Sgaw Karen, a closely related but more widely spoken language (Ethnologue). This essay will examine the Eastern dialect spoken near Hpa-an, the capital of Kayin State in Burma.

Previous research on the language, especially on Eastern Pwo Karen, is sparse even in comparison to the body of work on Sgaw Karen. Atsuhiko Kato has produced a grammar of Pwo Karen in Japanese, as well as a sketch grammar of the Eastern variety in English and a number of articles in both languages on the syntax, morphology, and comparative phonology of Pwo Karen. Audra Phillips has published several papers on the varieties of Pwo Karen spoken in Thailand. In addition, the language was the subject of a seminar on linguistic field methods at Yale University, taught by Claire Bowern in the spring of 2011. Data on the phonology of Eastern Pwo Karen gathered in the Yale field methods class differed substantially from the phonology described by Kato, and the little research that has been conducted on the phonetics of the language occurred during the same class, in addition to subsequent research by the author.

1.1.1. Pwo Karen phonology

Table 1: Consonant inventory of Pwo Karen; consonants in parentheses are used only in loan words.

	Bilabial	Labiodental	Dental	Alveolar	Palatal	Velar	Glottal
Plosive	p b p^{h}			t d t^{h}	c c^{h}	k k^{h}	ʔ
Affricate							
Nasal	m			n			
Fricative		(f)	θ	s		x ɣ	
Approximant	w			(ɹ)	j		
Lateral approximant				l			

Table 2: Consonant minimal pairs

θ	/θa^{5}/	fruit	x	/xa^{1}/	insect
c	/ca^{5}/	life	b	/ba^{1}/	to worship
k	/ka^{5}/	difficult	d	/da^{1}/	see
ʔ	/ʔa^{5}/	many, much	k^{h}	/k^{h}a^{1}/	to break
m	/ma^{5}/	mistake	ɣ	/ɣa^{1}/	person (num. clf.)
n	/na^{5}/	to drive	w	/wa^{1}/	bamboo
l	/la^{5}/	leaf	j	/ja^{1}/	hundred
s	/sa^{5}/	dark			
p^{h}	/p^{h}a^{53}/	male	p	/po^{1}/	story
t^{h}	/t^{h}a^{53}/	drum	s	/so^{1}/	to think
c^{h}	/c^{h}a^{53}/	to hurt			
t	/taĩ^{1}/	create	f	/fo^{3}/	phone
t^{h}	/t^{h}aĩ^{1}/	branch	ɹ	/əmeɹika/	America [tones uncertain]

The phoneme chart and minimal pair sets given in Tables 1 and 2 are derived from work by the author of this paper as well as the other authors of Bowern et al. (2011) with native Pwo Karen speakers living in Hartford, CT. Oral stops include voiceless unaspirated, voiceless aspirated, and voiced unaspirated series. The voiceless unaspirated stops occur in five places of articulation, namely bilabial, alveolar, palatal, velar, and glottal. /c/ and /cʰ/ are listed among the oral stops, but are realized phonetically as the alveolo-palatal affricates [tɕ] and [tɕʰ] in regular speech; they are pronounced /s/ and /sʰ/ in formal contexts. /θ/ may be voiced intervocalically. The spirantized allophone of the palatal glide /j/, [ʝ], occurs when the phoneme is emphasized. /f/ and /ɹ/ appear only in loan words from English, e.g. those derived from the words "phone" and "America." As our consultants all live in the United States and have some familiarity with English, it is unclear whether these phonemes would appear in the speech of non-English speakers in Burma or Thailand (Bowern et al. 2011).

Table 3: Vowels in Pwo Karen; modal voice (left) and creaky/nasal voice (right).

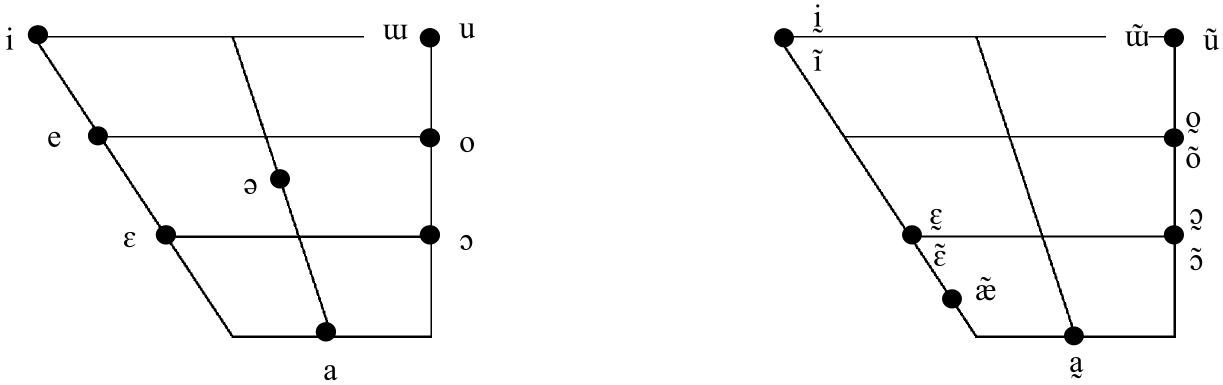


Table 4: Vowel minimal pairs

i	/pʰi^{53}/	grandmother	o	/do^{1}/	town, city
u	/pʰu^{53}/	grandfather	ɔ	/dɔ^{1}/	to fight
ɯ	/phɯ^{53}/	to jump			
a	/pʰa^{53}/	male			
e	/ɣe^{5}/	house	ə	/pə/	1PL
ɛ	/ɣɛ^{5}/	spicy			
Nasalization			Creaky voice		
	tʰɔ^{5}	to finish		di^{5}	egg
	tʰɔ̃^{5}	upward		dḭ^{5}	frog

To date, nine phonemic vowels have been definitively identified. Of these, seven may be nasalized, and five may carry creaky phonation. The precise phonotactics for nasalized and creaky vowels have yet to be determined. Since creaky vowels occur only in a small subset of the lexicon, it is possible that more creaky vowels may be identified as more vocabulary is compiled. Diphthongs occur widely in the lexicon, and some nasalized and creaky vowels, particularly /ã/ and /a̰/, appear much more frequently in nasalized diphthongs (in this case, /ãĩ/ and /a̰ḭ/). Although Kato and some of the authors of Bowern et al. (2011) posit the existence of an additional high unrounded modal vowel /ɨ/, no conclusive minimal pairs have been identified; /ɨ/ may simply be an allophone of /ɯ/.

All syllables in Pwo Karen are based on an open syllable structure with a mandatory consonant onset, or C(C)V(V). Onsets may consist of up to two consonants, the second of which must be an approximant. The nucleus contains either a single vowel or a diphthong. Word-level stress is not contrastive. Monosyllabic and disyllabic words are common, words of three or more syllables less so.

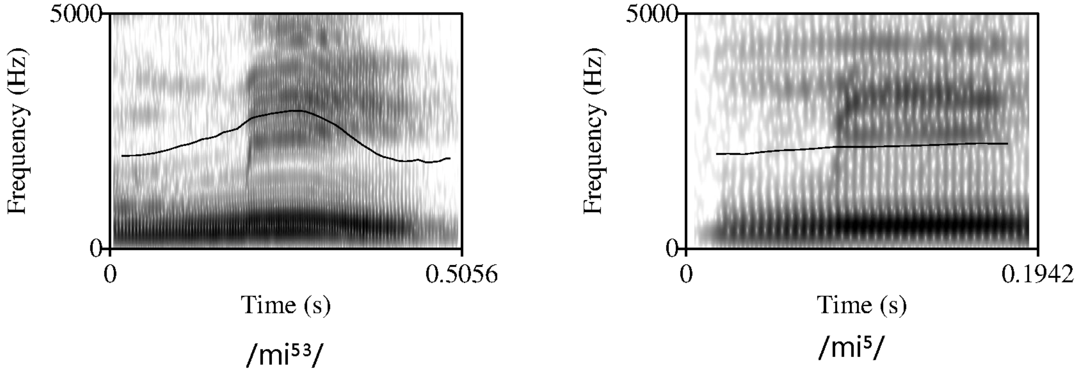
Pwo Karen distinguishes between four phonemic tones. One tone applies to each syllable, and every syllable possesses a tone, with the exception of syllables bearing neutral tone. Tone markings are based on the IPA tone system commonly used to describe Asian languages, in which a set of numbers ranging from 1 (lowest) to 5 (highest) represent five pitch levels. In comparison, the standard IPA system for indicating tones is less intuitive and less compatible with a wide variety of software. A single number indicates a level tone, while two numbers indicate a contour tone whose relative starting and ending pitches are indicated by the first and second numbers, respectively. The high falling tone (^{53}) may shorten to (^{5}) in fast speech or before another high (^{5}) or high falling (^{53}) tones. In Pwo Karen, the mid tone (^{3}) may also surface as a rising tone (^{24}) in citation form or at the ends of intonational boundaries (see Figures 1 and 2); the interactions between this contour variation and boundary tones will be discussed below.

Like Burmese and Mandarin Chinese, Pwo Karen also possesses a neutral tone, which occurs only on the vowel /ə/ and is negatively defined as the absence of any lexical tone. The majority of syllables containing the vowel /ə/ bear neutral tone; however, a small number of exceptions, such as /chənə^{3}/ "cow," have been recorded. Pwo Karen neutral tone patterns almost identically to the neutral tone in Burmese (Green 1995): it occurs only on /ə/, syllables bearing it may not occur word-finally in a multisyllabic word, and it may not co-occur with nasalization or creaky phonation. Green (1995) distinguishes between two types of Burmese syllables, major (with a non-/ə/ nucleus) and minor (with a nucleus of /ə/). Green posits that major syllables are bimoraic, while minor syllables are monomoraic, an analysis which likely also holds for Pwo Karen.

Table 5: Tone minimal pairs

/mi^{53}/	to sleep	/xwi^{5}/	hair
/mi^{5}/	fire	/xwi^{3}/	to boil
/mi^{3}/	tail	/xwi^{1}/	cockfight
/mḭ^{1}/	rice	/xwḭ^{1}/	to buy
/mə/	(future tense particle)		

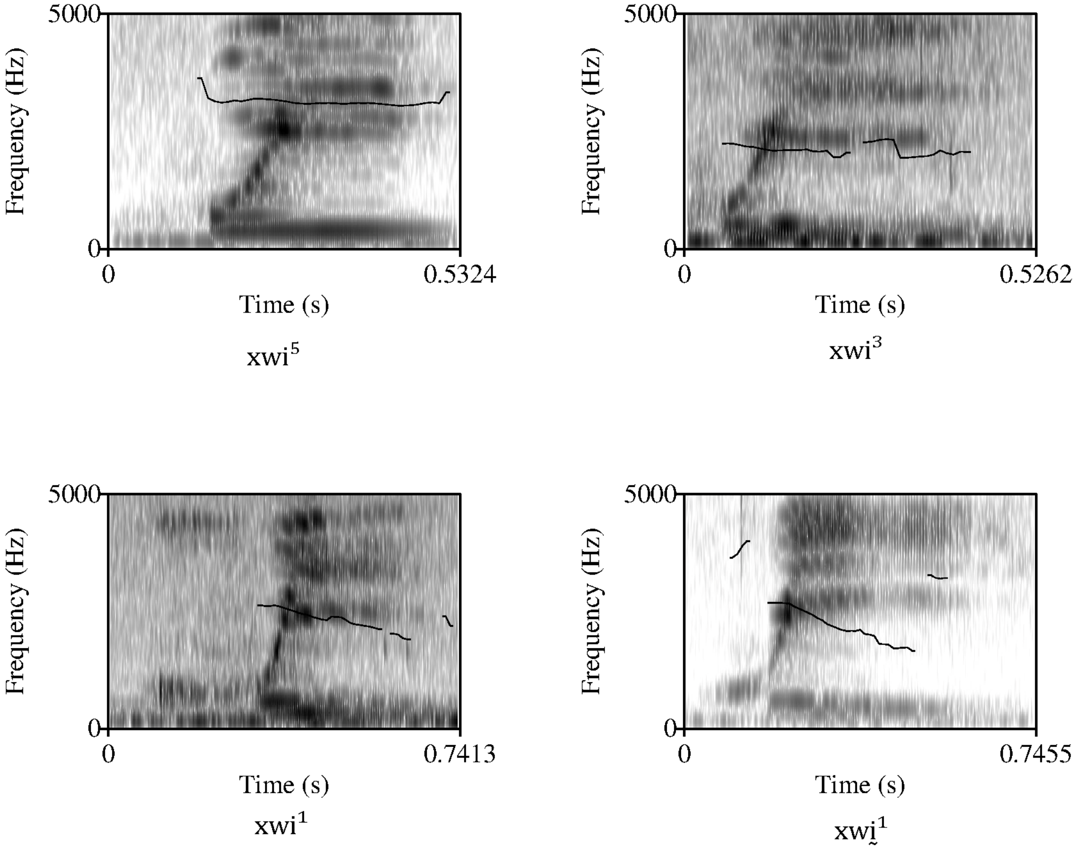
Figure 1: Tone minimal pairs /mi^{53}/ "to sleep" and /mi^{5}/ "fire (Speaker A)



/mi^{53}/

/mi^{5}/

Figure 2: Tone minimal pairs /xwi^5/ "hair," /xwi^3/ "to boil," /xwi^1/ "cockfight", and /xwḭ^1/ "to buy" (Speaker A)



1.1.2 Pwo Karen intonation

In Hsieh (2011), the author proposed a transcription system for some basic elements of Pwo Karen. That system is based on ToBI, a transcription system that has been adapted for a variety of languages. A number of dialogues (given in the Appendix) were created to elicit the intonational patterns associated with unmarked questions, declarative sentences, echo questions, contrastive focus, lists, and various boundary tones, and to examine the interactions between phonemic tone and those patterns. In sentences containing lists, for example, an effort was made to use words ending with the same phonemic tone in the lists. The dialogues were edited and translated with the assistance of Subject A, who read them aloud while being recorded with a lapel microphone. Since Subject A portrayed all the characters in the dialogues, the intonational patterns will likely sound artificial to a native speaker, and possible negative effects of this artifical method of elicitation will be discussed below.

Like Mainstream American English ToBI, Pwo Karen ToBI (PK_ToBI) makes use of both tones and break indices. It combines aspects of both MAE_ToBI and Pan-Mandarin ToBI. Only three tonal languages/language groups – Mandarin, Cantonese, and Taiwanese – have been described in publication thus far. After examining the results of the intonational elicitation described above, the author concluded that Pwo Karen intonation most closely resembles that of Mandarin, and that Pan-Mandarin ToBI would offer the best foundation for constructing PK_ToBI.

1.1.2.1 Tones

As in other tonal languages, Pwo Karen intonation consists of prosodic contours that overlay the pitch contours of lexical tones. A low boundary tone may co-occur with a final syllable bearing high lexical tone; boundary tones, phrase accents, and pitch range effects all raise or lower the pitch of a lexical tone relative to its usual pitch range, while maintaining its characteristic shape relative to other lexical tones.

The tones include the fundamental boundary tones and phrase accents used in MAE_ToBI: H%, L%, H-, and L-. L% marks the default intonational contour, in which there is a pitch downtrend across the entire phrase. Both positive and negative declarative statements and basic questions utilize this contour. The final word in both Figures 3 and 4, /nɔ^3/, ends at almost exactly the same pitch (~230 Hz), and the intonational contours of the two sentences differ only where the verb is replaced with a verb + in-situ Wh-question.

Although the pitch contour rises at the end of both Figures 3 and 4, apparently indicating the presence of a H% boundary tone, in fact the rise is the result of the /^{3}/ tone manifesting as its tonic allophone /^{24}/ at the end of an intonational boundary. Although the pitch of /nɔ^3/ does have a rising contour, the contour does not rise to the level of the tones earlier in the sentence; the tone of /nɔ^3/ would also be expected to have a higher pitch than the low tone on /ʔənai^{1}xu^1/, and that is the case in these examples. In addition, the pitch of /nɔ^3/ in Figures 3 and 4 should be compared with the pitch of the same word in Figure 5, where the high boundary tone H% causes the word to have a correspondingly higher pitch. These analyses may, however, be confounded by some characteristics of the stimuli. Many tones early in the sentences in Figures 3 and 4 are /^{5}/ or /^{53}/, which would have raised the overall pitch at the beginning of the sentence regardless of the intonational contour. The syllable directly before /nɔ^3/ in Figure 5 also carries /^{53}/ tone, which could have raised the pitch contour of /nɔ^{3}/ without the presence of H%, but the pitch track in Figure 3 seems to accord with the H% by remaining generally high even on low /^{1}/ tones.

Figure 3:

ʔəwe^{53}	mi^{53}	lə	mi^{5}po^{1}	ʔənai^{1}xu^{1}	nɔ^{3}
3SG	sleep	at	hearth	next-to	FOCUS

"He is sleeping next to the hearth." [1.6]

[The numbers next to each gloss correspond to the sentence's number in the dialogues in the Appendix.]

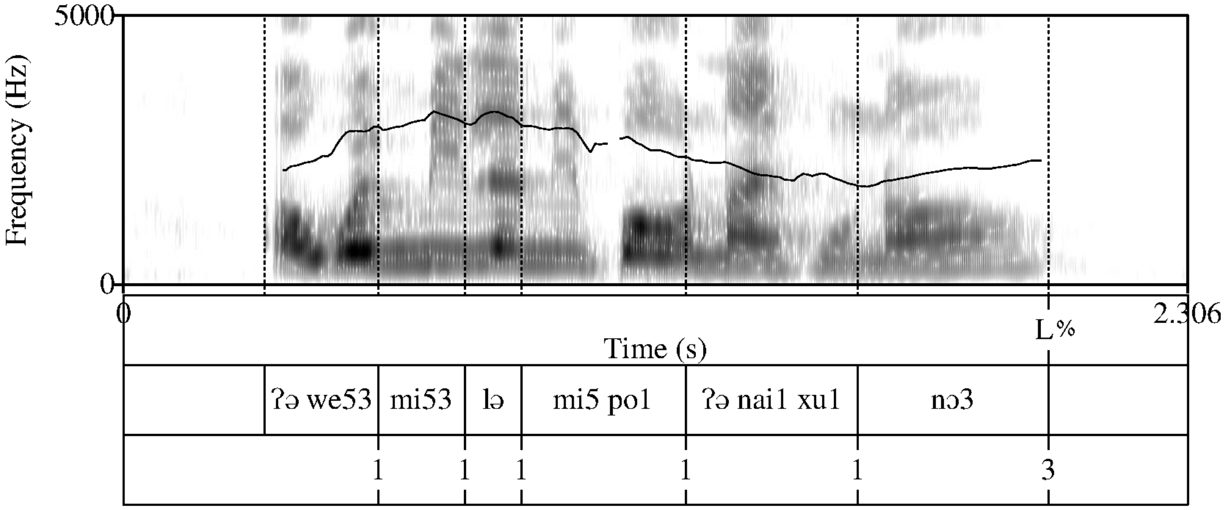
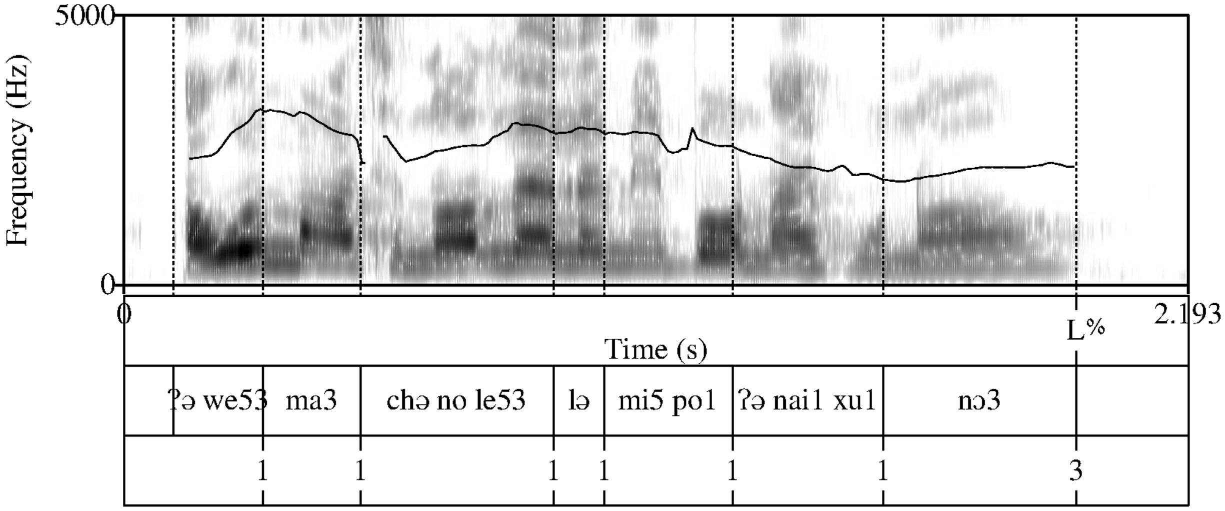


Figure 4:

ʔəwe^{53}	ma^{3}	chənole^{53}	lə	mi^{5}po^{1}	ʔənai^{1}xu^{1}	nɔ^{3}
3SG	do	what	at	hearth	next-to	FOCUS

"What is he doing next to the hearth?" [1.5]

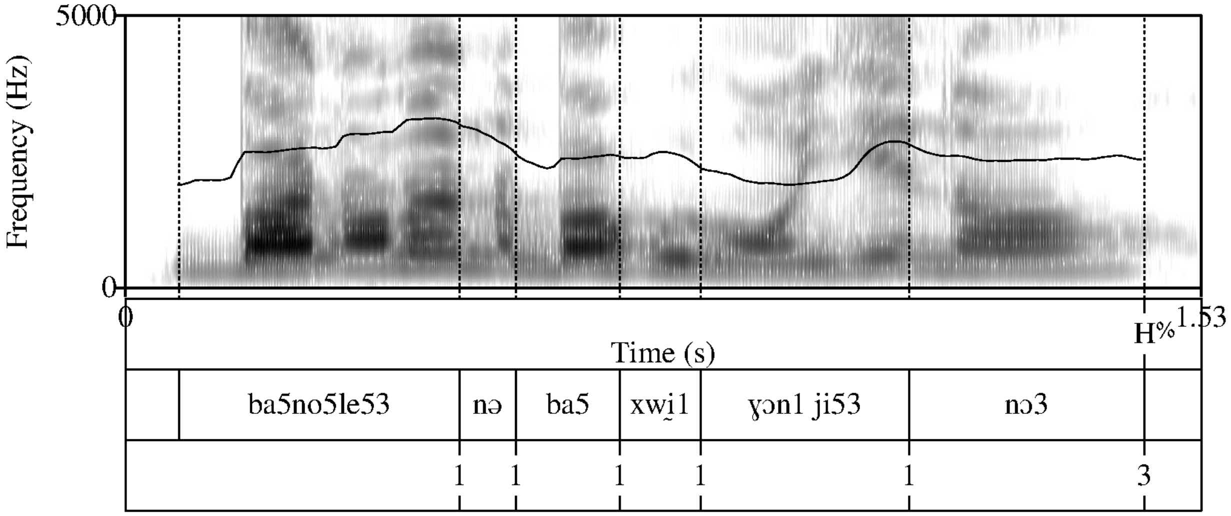


The high boundary tone H% is illustrated in Figure 3 in the context of a question indicating disbelief or disapproval.

Figure 5:

ba^{5}no^{5}le^{53}	nə	ba^{5}	xwi̯^{1}	ɣɔn^{1}ji^{53}	nɔ^{3}
why	2SG	must	buy	lemongrass	FOCUS

"Why did you buy lemongrass?" [2.4]



As in MAE_ToBI, phrase accents are marked at the ends of intermediate phrases. The sentence in Figure 6 was designed to elicit H- in the context of a list, with each list item ending with a low /^{1}/ tone. The long pauses in Figure 6, however, are one problematic consequence of the artificial elicitation process. The tone on the word /nai^{1}/ "type of basket" could be most accurately described as a high boundary tone followed a pause, for example, which would account for the dramatic rise in the pitch contour before the pause. The second word in the list, /mi^{5}dwai^{1}/ "matches", more accurately represents a H- contour; the phrase accent draws the pitch of the low tone upward. The final L% in Figure 6 is difficult to discern due to the presence of creaky phonation on /mi̯^{1}/ "cooked rice," but the pitch of /de^{3} mi̯^{3}/ "with cooked rice" is still lower overall than that of the previous two phrases.

L-, the low phrase accent, behaves much like the L% boundary tone (Figure 7). Again, a slight rise in the pitch of /nɔ^{3}/ can be attributed to the allophonic /nɔ^{24}/ form.

Figure 6:

jə	ɣe^{53}	chu^{1}	nai^{1},	mi^{5}dwai^{1},	de^{3}	mḭ^{1}
1SG	come	bring	type-of-basket	matches	with	cooked-rice

"I am bringing a *nai* basket, matches, and cooked rice."

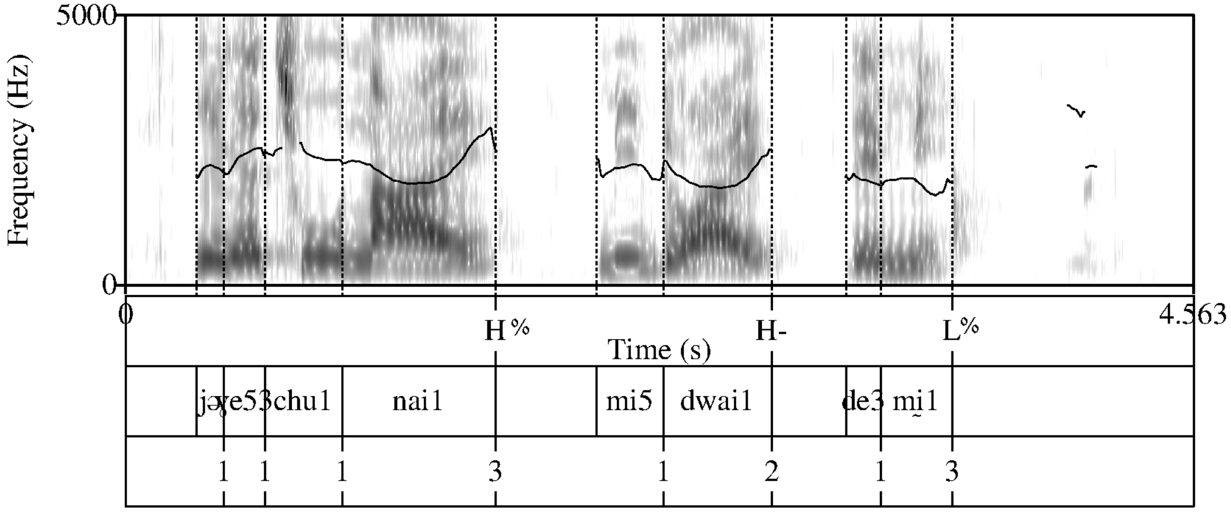
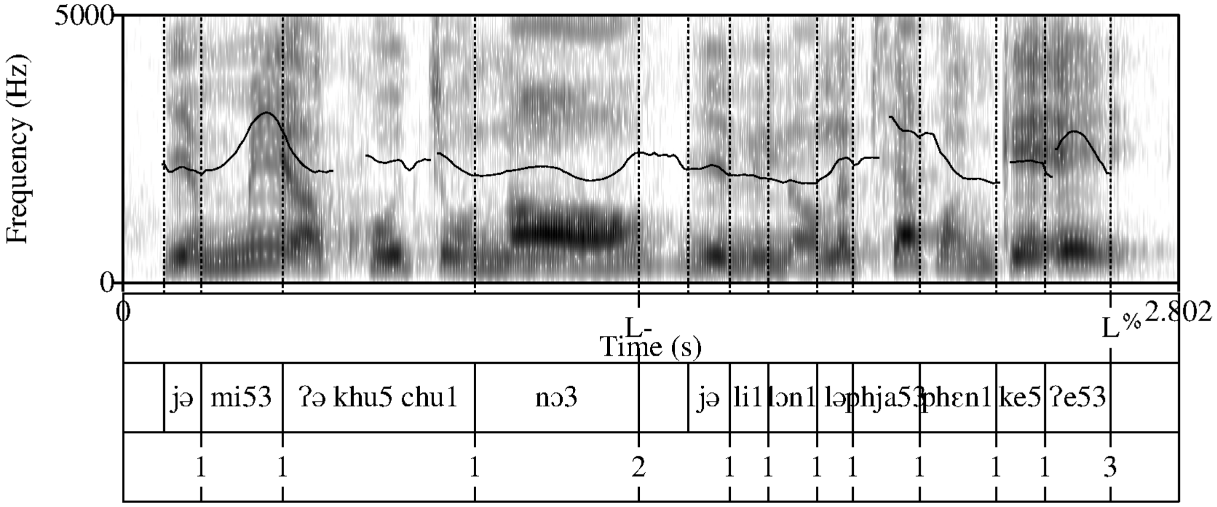


Figure 7:

jə	mi^{53}	ʔəkhu^{5}chu^{1}	nɔ^{3},	jə	li^{1}	lɔn^{1}	lə	phja^{53}	phɛn^{1}
1SG	sleep	because	FOCUS	1SG	go	down	at	market	in
	ke^{5}	ʔe^{53}							
	want	NEG							

"Because I am sleeping, I do not want to go to the market." [3.3]



The pitch range effects used in PK_ToBI are adapted from the Pan-Mandarin ToBI, and represent "backdrop contours" that apply to entire phrases or parts of phrases (Jun 2005). All pitch range effect tags are notated at the beginning of the domain of their effect, and the marked effect continues until it encounters another pitch range effect or tone.

%q-raise describes the general raising effect that is exemplified by echo questions like 1.7 (Figure 8). Emphatic prominence on a given syllable is indicated by greater word duration as well as an expanded tonal pitch range. Comparing Figures 3, 7, and 9, the word /mi^{53}/ "to sleep" is more than twice as long in duration when emphasized. The pitch range effect tag %e-prom marks the beginning of local prominence, while %compressed marks the beginning of the compensatory pitch range compression that follows a prominent section. Figure 9 is not an ideal example, however, since the speaker paused after %e-prom; the pause might affect not only the the duration of the prominent syllable, but also the compression afterward.

Figure 8:

ma^3	chənole^{53}	lə	mi^{5}po^1	ʔənai^{1}xu^1	nɔ^3
do	what	at	hearth	next-to	FOCUS

"***What*** is he doing next to the hearth?" (echo question) [1.7]

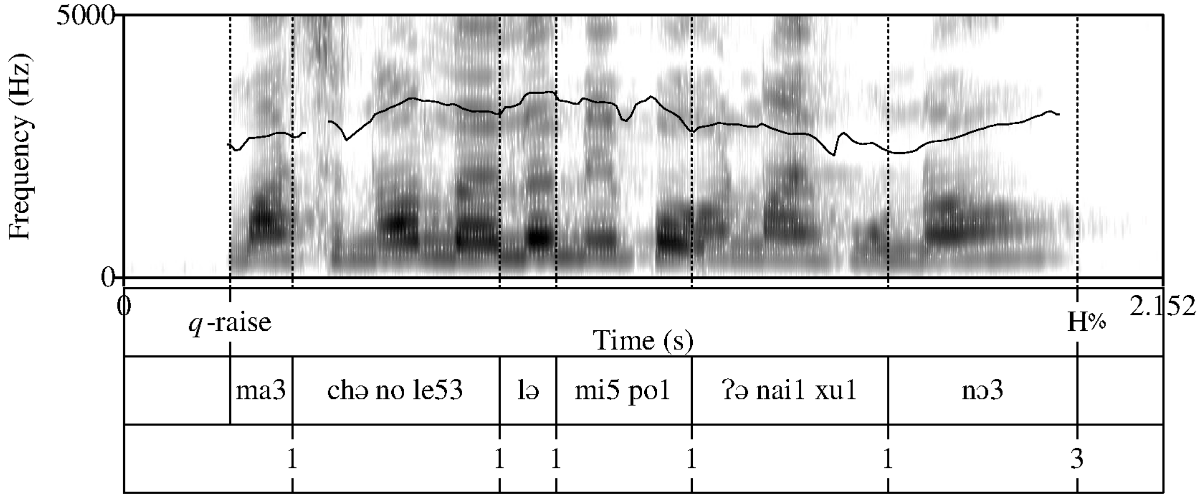
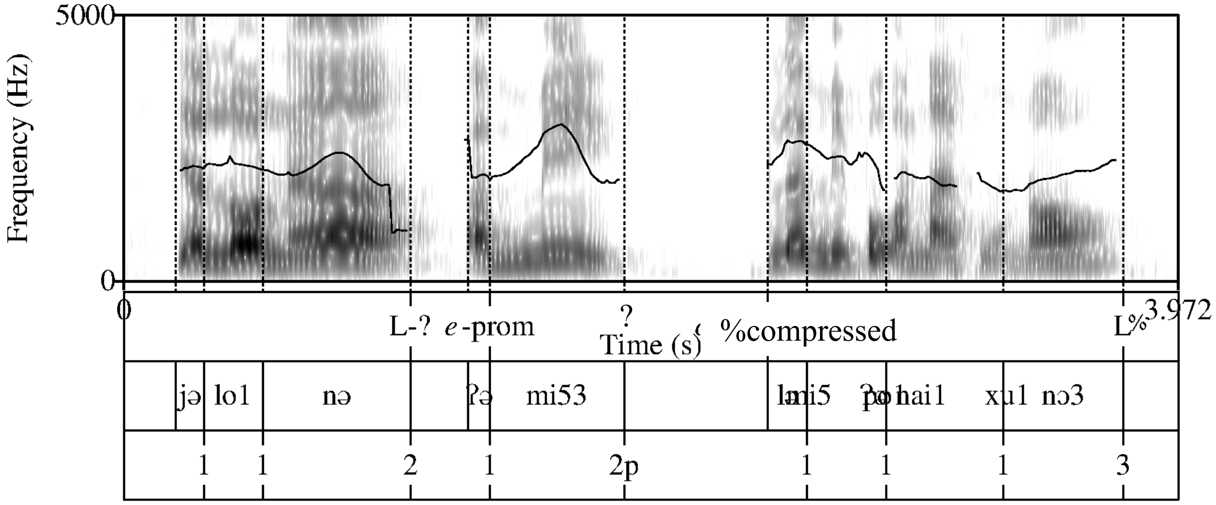


Figure 9:

jə	lo^{1}	nə,	ʔəwe^{53}	mi^{53}	lə	mi^{5}po^{1}	ʔənai^{1}xu^{1}	nɔ^{3}
1SG	tell	2SG	3SG	sleep	at	hearth	next-to	FOCUS

"I told you, he is ***sleeping*** next to the hearth." [1.8]

(emphatic prominence; pause after /mi^{53}/ "to sleep")



Another type of local prominence was identified in the course of analyzing the data for this paper, and was not discussed in Hsieh (2011). The types of local prominence elicited in Hsieh (2011), such as contrastive focus and emphatic prominence, carry semantic content; they supplement the contrastive meaning or pinpoint a specific word for the listener's attention. This paper, however, examined the correlates of all pitch accent in Pwo Karen, including pitch accents whose semantic content may not be so obvious. These include what the author perceived to be the Pwo Karen correlates of American English ToBI's H* and L* (Beckman and Elam 1997), which might occur in the context of simple declarative or interrogative statements with no special emphasis placed on any particular words. The tonal language correlates of these unmarked pitch accent-like prominences were not discussed in the Pan-Mandarin ToBI in Jun (2005) or in Chen and Gussenhoven (2008), and it is unclear whether they exist in Mandarin Chinese. Although the author could perceive these prominences, the author's Pwo Karen consultant could neither confirm nor deny their existence. This type of prominent pitch range effect has been tentatively termed %prom, though much work remains to be done on the distinguishing factors between %e-prom and %prom. The applicability of %compressed, a reduced pitch range effect that follows %e-prom, to %prominence is also uncertain.

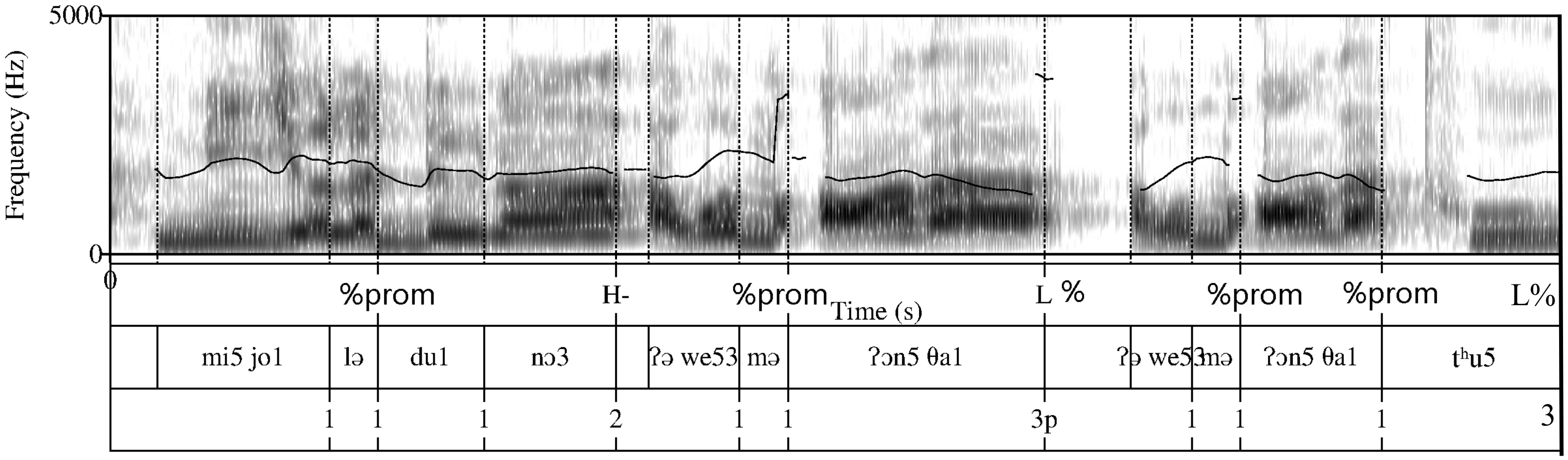
The acoustic characteristics of %prominence are more subtle than those of %e-prom. In Figure 10, the pitches of the prominent low /^{1}/ tone syllables (/du^{1}/ and /θa^{1}/) are lower than the non-prominent low tone in /mi^{5}jo^{1}/ in all three instances. The high /^{5}/ tones syllables (ʔɔn^{5} and tʰu^{5}), however, are not higher than the non-prominent high tone in / mi^{5}jo^{1}/, which could be the result of tone downstepping leading to the low boundary tones. Duration is also relevant to prominence in this section, but relative duration is not a reliable characteristic in all cases. A native speaker of Thai, which contains five level and contour tones, independently confirmed the author's perception of the placement of %prom in Figure 10, though she is unfamiliar with Pwo Karen.

Figure 10:

mi^{5}jo^{1}	lə	du^{1}	nɔ^{3},	ʔəwe^{53}	mə	ʔɔn^{5}θa^{1}…
cat	one	CLASS	FOCUS	3SG	FUT	want-to-eat?[1]

ʔəwe^{53}	mə	ʔɔn^{5}θa^{1}	tʰu^{5}.
3SG	FUT	want-to-eat?	bird

"The cat, he will want to eat…he will want to eat the bird."



[1] The precise translation of this word, and by extension this sentence, is unclear. The morpheme /θa^{1}/, which means approximately "to feel" or "feeling," may have been added to /ʔɔn^{5}/ "to eat"; alternatively, this word may in fact be a serial verb construction comprising two separate words.

Table 6: Tones and pitch range effects used in the tone tier

Basic tones tags:	
H%	high boundary tone
L%	low boundary tones
H-	high phrase accent
L-	low phrase accent
Pitch range effect tags:	
%q-raise	beginning of raised pitch range
%prom	beginning of expanded pitch range caused by pitch accent-like local prominence
%e-prom	beginning of expanded pitch range caused by focus prominence
%compressed	beginning of reduced pitch range following the expansion under %e-prom

1.1.2.2. Break indices

The break indices system is based on that of the original ToBI system for Mainstream American English. Four basic break values are distinguished, in addition to three supplementary diacritics. Break indices and diacritics in brackets have been casually observed but not formally recorded. A break index of 0 indicates a closer-than-normal word juncture within a phrase. Regular inter-word junctures within phrases are indicated by a break index of 1, the "default" boundary that is used in the absence of more marked criteria. Break indices 2 and 3 represent intermediate phrase-level and intonational phrase-level boundaries, respectively. A break index of 2 must co-occur with a phrase accent, and vice versa; the same is true of a break index of 3 and boundary tones. In addition, three diacritics are used to indicate uncertainties or disfluent junctures.

Table 7: Break index levels and other diacritics used in the break indices tier

Basic break index values:	
[0	Reduced inter-word juncture]
1	Ordinary phrase-internal word end
2	Intermediate phrase end
3	Intonational phrase end
Diacritics (marked after the break index value):	
?	Break index uncertainty
p	Disfluent juncture
[m	Mismatch between the strength of the disjuncture and the tonal event]

1.2. Background on gesture and intonation

In a 1983 interview, Noam Chomsky claimed that "there are certain obvious interconnections between the verbal and gestural systems…They're in tandem, and some common source is obviously controlling them both" (Rieber 1983). Two years later, David McNeill published a widely-cited paper, titled "So You Think Gestures Are Nonverbal?", in which he drew on evidence from studies on the semantic content of gestures, child development, and aphasia, among other sources, to argue that "gestures and speech are parts of the same psychological structure and share a computational stage" (McNeill 1985). McNeill in turn credited Adam Kendon with the insight that gesture and speech are coordinated (Kendon 1972); McNeill's paper states that many psychologists of the time were convinced that the connection between gesture and speech existed, although some in the linguistics community remained skepical (McNeill 1985). Gesture studies prior to the 1970's mostly focused on the role of gesture in rhetoric and culture; without the ability to easily record synchronized audio and video, those studies were inevitably subjective (McNeill 1992).

Today, there is little doubt that the production of gesture is linked with the production of speech, even if the nature of that link is hotly debated (Leonard and Cummins 2011). Gesture co-occurs with language 90% of the time; gestures pattern with speech in aphasics; the development of both occurs together in children; and gesture and speech are connected in terms of semantics, pragmatics, and timing (Esteve-Gibert and Prieto 2011). Even children who have been blind from birth use gesture, which suggests that gesture does not exist solely for the benefit of the listener (Iverson and Goldin-Meadow 1997). David McNeill's lab at the University of Chicago has formulated a theory of gesturing that focuses on the "Growth Point," a hypothetical unit that combines imagery with categorical linguistic information (McNeill and Duncan 2000). McNeill posits a "thinking-for-speaking" hypothesis, in which speakers of a given language pattern their thinking to match the demands of speaking that language. Even authors who propose that gesture and speech exist separately in the brain, as in Rochet-Capellan et al. (2008), agree that the two systems are somehow coordinated with each other.

A number of studies have demonstrated a connection between prosody and gesture. Two studies, Cave, Guaitella, Bertrand, Santi, Harlay, and Espesser (1996) and Keating, Baroni, Mattys, Scarborough, Alwan, Auer, and Bernstein (2000), have identified a correlation between increased height in eyebrow movements and increased F0 in a pitch track. Keating et al. (2000) suggested that exaggerated face and head movements are used by listeners to assist the perception of phrasal stress in English.

In particular, several recent studies have found that pitch accents and the “stroke” of a manual gesture, the segment of a gesture involving the greatest output of effort, often coincide in English. In English, pitch accents are defined as syllables bearing relatively greater prominence than other syllables in a given phrase (Beckman and Elam 1997).

Yasinnik, Renewick, and Shattuck-Hufnagel (2004) concluded that pitch-accented syllables and manual gesture strokes align in English by analyzing videos of academic lectures, a result that has been replicated in a number of other studies. In his 2004 dissertation, Daniel Loehr used videos of spontaneous conversation to investigate the relationship between intonation and gesture in American English. Loehr focused on a section of the gesture stroke that he terms the "apex," or "the kinetic 'goal' of the stroke"; the gesture stroke is an interval of time, while an apex is a single target point within that interval, making it ideal for comparing against pitch accents. He filmed groups of speakers engaging in spontaneous conversation for one hour each, and found that alignment was most common between pitch accents and manual gestural apices. In an analysis of real-life political speeches in American English, Jannedy and Mendoza-Denton (2005) showed that 95.7% of all gesture apices were accompanied by a pitch accent, demonstrating that the extremely high co-occurrence rate of gesture strokes and pitch accents in laboratory studies also applied in natural speech. Leonard and Cummins (2011) showed specifically that the apex of a beat gesture aligns with the peak of the pitch accent

To date, the majority of studies on gesture, including all studies on the relationship between gesture and prosody, have been conducted on non-tonal languages such as English, Dutch, and Italian (Loehr 2004, De Ruiter, Jan P. and D. Wilkins 1998, Sansavini 2010); even the few non-Indo-European languages on which gesture studies have been conducted, such as the indigenous Australian language Arrernte (De Ruiter, Jan P. and D. Wilkins 1998), have all been non-tonal. Some tonal languages, such as Mandarin Chinese and Cantonese, have been shown to indicate prosodic prominence differently from non-tonal languages, and in general to employ distinct strategies for producing intonational patterns (Jun 2005). Chen and Gussenhoven (2008) claim that emphasis in Mandarin Chinese is indicated by increased word duration and exaggerated tonal contours.

Systematic descriptions of prosodic emphasis in tonal languages reject the presence of pitch accents, which occur on only one syllable, in favor of broad emphatic pitch effects that may cover more than one syllable. Analyses of gesture and intonation in non-tonal languages may thus not be generalizable to tonal languages, especially if there may be multiple adjacent prominent words in one phrase. Based on the existing data, however, there is no basis for assuming that gesture in tonal languages will behave differently from gesture in non-tonal languages. In addition, if gesture is indeed the product of a linguistic faculty in the brain, it should hold that gesture aligns with speech in all languages.

2. Methods

2.1. Subjects

Video recordings of three native speakers of Pwo Karen, all of whom immigrated to the northeastern United States within the past four years, are used in this study. Each video is between 3.5 and 6 minutes long. Speaker A is a 23-year-old woman who grew up in the Hpa-an region of Burma and subsequently a Karen refugee camp in Thailand, and is proficient in English, Sgaw Karen, Burmese, and Thai in addition to Pwo Karen. Speaker B is a 26-year-old man who grew up in a refugee camp in Thailand, but whose parents were native to the Hpa-an region. Speaker B had only limited English proficiency. Speaker C is a 57-year-old woman who lived in

Burma (the author is unsure of the precise region) for most of her life before moving to a refugee camp in Thailand. Speaker C has very little familiarity with English. The proficiency of Speakers B and C in other languages is unknown, although most Pwo Karen speakers of similar background have some ability in Sgaw Karen due to its larger speaker base. One other speaker was also recorded for this study, but that speaker's video was omitted from the analysis due to an insufficient amount of manual gesturing.

2.2. Stimuli

The stimulus for the study was an extract from a 1949 episode of *Looney Tunes*, the popular American animated cartoon series. The plot of "Canary Row," which runs for seven minutes, focuses on the rivalry between Tweety Bird and Sylvester the Cat. First used for gesture elicitation in McNeill and Levy (1982), the cartoon is useful for cross-linguistic gesture studies because it contains little dialogue and abundant movement. The semantic content of the subjects' narratives was not essential to the analysis; the video was intended as a prompt to elicit narratives of similar length and gesture frequency for each subject, in addition to providing a connection to the many other studies that have used this video. "Canary Row" has been used in dozens of experiments in the fields of gesture studies, sign-language studies, and second language acquisition; a sampling of those studies is given here: McNeill (1992), Cassell, McNeill, and McCullough (1998), Stam (1998), Ozyurek (2002), Casey and Emmorey (2008), Sandler (2009), de Kok and Heylen (2010), and Brown and Gullberg (2011).

2.3. Procedure

Each subject was told that they would be shown a video of a children's cartoon, and that they would be asked to recount the plot of the cartoon after viewing the video. After "Canary Row" was shown on a laptop computer, the subjects were asked to describe the main characters of the cartoon, followed by the general plot, to a listener seated next to the video camera. Subjects were told that they would hear some English dialogue in the cartoon, but that they were not expected to understand or remember the dialogue. The author gave instructions in English to and acted as listener for Speaker A, while Speaker A gave instructions to Speakers B and C in Pwo Karen due to their low level of English proficiency, and acted as listener for both. Speakers A and B were filmed in the home of Speaker A; Speaker C was filmed in her own home.

2.4. Equipment

Video recordings were made using a Sony HDR camcorder connected to an Audio Technica PRO-88W-MT830 wireless microphone system. The wireless transmitter was connected to Audio Technica ATR831 series lapel microphones, which were clipped to the shirt of the subject. The video camera was placed on a tripod directly across from the subject at a distance of approximately 10 feet.

2.5. Measurements

Praat, a free phonetic analysis program (http://www.fon.hum.uva.nl/praat/), was used to make prosodic annotations and to transcribe the utterances. Gestures were analyzed using the frame-by-frame viewing function in Windows Movie Maker, which moved at a rate of approximately 30-40 milliseconds per frame. In order to avoid perception interference, prosodic annotations in Praat were made using only the audio track extracted from the original video, and gesture transcriptions were made with the volume muted on Windows Movie Maker.

Data were recorded in the form of time stamps, converted into seconds, from video frame numbers or audio tracks. Two decimal places were recorded for each time stamp because Windows Movie Maker, the less precise of the two programs, was precise only to 10 milliseconds.

2.5.1. Manual gesture transcription

David McNeill categorizes gestures in terms of their semantic function as well as their form. The fundamental distinction is between imagistic and non-imagistic gestures; beats, the gestures relevant to this paper, belong to the second category. In addition, gestures are broken down into three movement phases: (1) preparation, in which the hands move away from rest position, (2) stroke, the moment of peak effort and the only obligatory movement phase, and (3) retraction, in which the hands return to rest position.

McNeill's definition of a beat gesture is given in the form of a "beat filter," a test used to differentiate iconic and metaphoric gestures from discrete gestures when the gesture occurs between two rest positions. For each "yes" answer to a question, 1 point is added to the total score; a score of 0 means that the gesture is likely a beat. The beat filter is as follows: "(1) Does the gesture have other than two movement phases (i.e., either one phase or three phases, or more)? (2) How many times does wrist or finger movement or tensed stasis occur in any movement phase not ending in a rest position? (add this number to the score). (3) If the first movement is in a non-center part of space, is any other movement performed in center space? (4) If there are exactly two movement phases, is the space of the first phase different from the space of the second?"

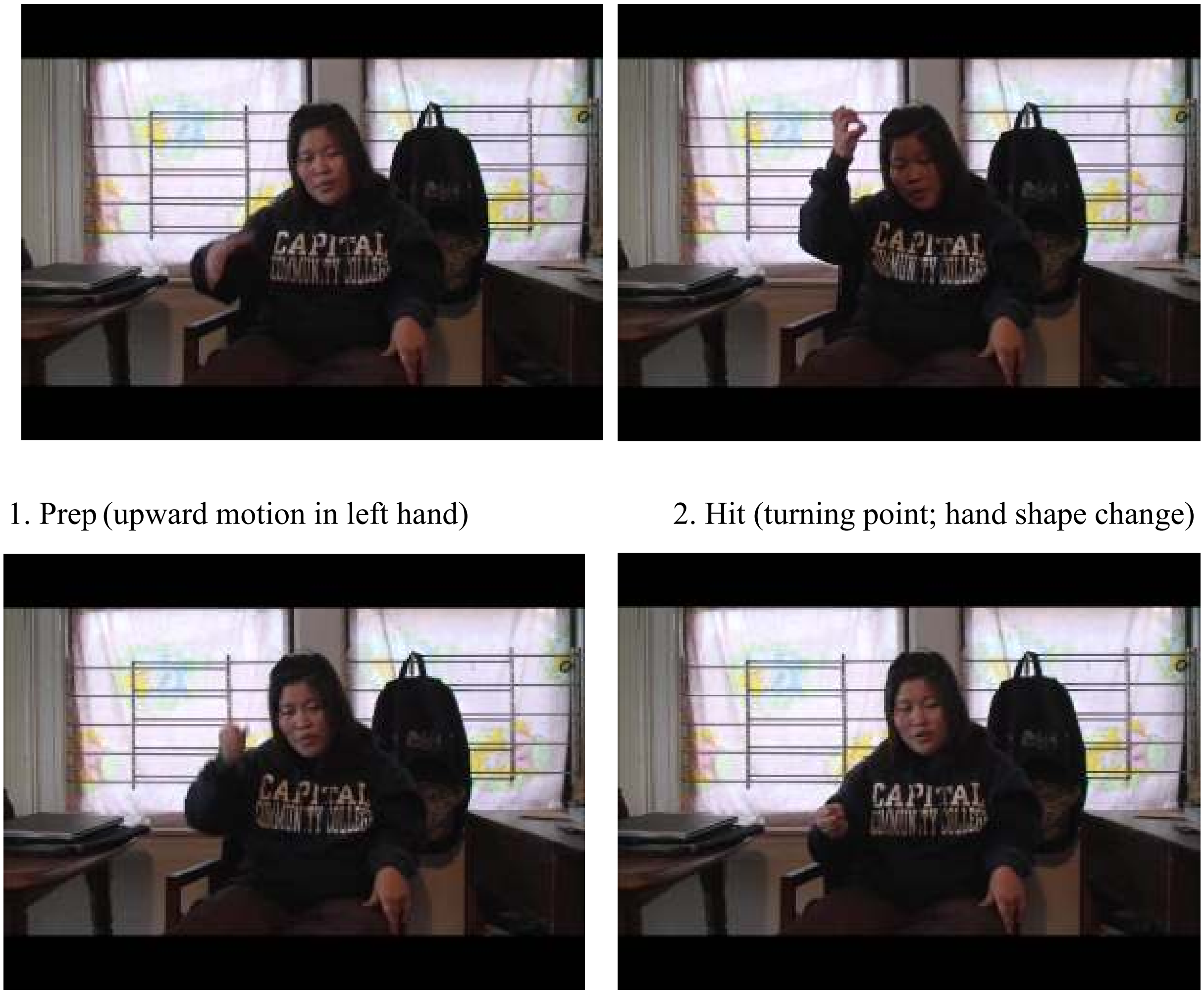
The system used to transcribe manual gesture in this study is based on that used in Yasinnik, Renwick, and Shattuck-Hufnagel (2004), a very similar study whose methodology is informed by McNeill's.

The gesture categories used in Yasinnik et al., in contrast, are much more broadly defined: they comprise only "discrete" and "continuous" categories, where discrete gestures are characterized by "an abrupt stop or pause in movement, which breaks the flow of the gesture during which it occurs." This abrupt stop is called a "hit," and is equivalent to the stroke phase described in McNeill 1992. The discrete gesture category contains the gesture type that McNeill calls beats, but does not exclude gestures that do not meet the strict formation criteria of the beat filter.

Since the discrete gesture category was designed for ease of pinpointing a single point in time, and therefore ease of making quantifiable comparisons with pitch accents, the criteria for discrete gestures were used to identify relevant gestures for this study. Each gesture hit was analyzed as comprising two sections: the abrupt motion of the hit, and the apex or target point at which motion was aimed and where it ceased. Functional movements of the hands, such as adjusting hair or scratching an itch, were not measured.

The hit of each discrete gesture was measured at the apex. When the hands were held in place following a gesture hit, rather than being retracted, the hit was measured when the hands ceased moving. The movement for a gesture hit was often apparent in the video frames as a blurring of the hands; the cessation of the blurring was used as a guide to the location of the target point (Yasinnik et al. 2004). Figure 8 illustrates the phases of a discrete gesture; the hit begins in frame #2 and proceeds rapidly downward until abruptly halting at frame #4, whose time stamp was recorded. The beginnings of the preparation phase and hit for each gesture were not recorded due to the greater subjectivity and difficulty of pinpointing those points.

Figure 11: Gesture phases (Speaker A)



1. Prep (upward motion in left hand)

2. Hit (turning point; hand shape change)

3. Hit (blurred downward motion)

4. Apex (no blurring)

2.5.2. Prosodic transcription

In English, each intonational phrase obligatorily contains at least one pitch accent, a single syllable that receives sentence-level stress and is more prominent than other syllables in the phrase. The pitch range effects indicating prominence in Pwo Karen and pitch accents in English are distinguished by the fact that pitch accents may only occur on one syllable, while pitch range effects may extend over a number of syllables. Although most prominent words were monosyllabic, a small number contained two or more syllables; these were analyzed as having a prominence pitch range effect extending across the entire word (see section 5). Words that carried the pitch range effect %prom were measured.

For each prominent word, the beginning of the word, beginning of the vowel, and end of the word were measured. The initial consonant or consonant cluster of each prominent word was also recorded. The beginning of a syllable was marked at the beginning of frication or vibration for an initial fricative or sonorant. When the syllable on onset was a plosive, the beginning of silence following the previous syllable was marked for a voiceless stop, while voiced stops were marked at the cessation of formants from the previous syllable. If the prominent word contained more than one syllable, the word onset and vowel onset were measured in the first syllable, and the word end was measured at the end of the last syllable. The beginning of a vowel was measured either at the cessation of a fricative or sonorant or at the stop release of a plosive. Since all Pwo Karen syllables take the structure C(C)V(V), the end of a syllable was marked at the end of vibration for the vowel if the syllable occurred phrase-finally, or at the beginning of the following syllable if it occurred phrase-internally.

2.5.3. Alignment

To determine whether gesture hits and prominent words were aligned, the time stamps for the gesture hits were compared against the time stamps for prominent word starts and ends in an Excel spreadsheet. Although previous studies only counted overlapping gesture hits and syllables as being aligned, over a third of the gesture hits in this study occurred after the codas of syllables with which the author perceived them to be associated (see section 4). Therefore, in the absence of any other established standard, any prominent word that began 500 ms before or after a gesture hit was noted as a possible candidate for alignment. This initial approximation was confirmed by reviewing the video with the audio turned on.

Reviewing the video also brought to light more concrete evidence to support the idea that a gesture hit apex and a syllable could be aligned even if they did not overlap in time. For example, Speaker B describes a scene near the end of the cartoon, when Sylvester is trying to escape from Tweety by running across trolley cables strung above the street. Instead, Tweety pursues Sylvester in a trolley car; whenever the trolley car's pole connects with the cable that Tweety is standing on, Tweety levitates in the air for a moment as he is electrocuted. To illustrate the continuous pursuit, Subject B repeated the same sentence describing the scene, using identical sets of three gestures for each repetition: a discrete gesture outward to indicate the moving trolley car, a discrete gesture upward to indicate Sylvester's levitation, and a shaking gesture with both hands to indicate electrocution. The regularity of the rhythm in Subject B's production of the three gestures was mirrored in the rhythm of three prominent syllables per sentence, with which the gestures appeared to align; however, the two discrete gesture apices occurred outside the prominent syllables in both repetitions. Without a translation of the sentence, the semantic correspondances between the syllables and gestures cannot be examined. The simple correspondance between the the number of gestures and number of prominent syllables, repeated identically, and the subjective perception of alignment, however, suggest that overlap might not be the only criterion for alignment.

3. Results

In total, the three videos contained 115 discrete gestures, of which 69, or 60%, were aligned with prominent words. In comparison, studies such as Loehr (2004), Yasinnik et al. (2004), and Jannedy (2005) all found an alignment rate of over 95%. Although gesture hits have clear physical correlates that translate well on video, perceived prosodic prominence did not always co-occur with physical evidence of prominence (e.g. greater vowel duration). Since prominence was marked conservatively to avoid mislabeling, employing a native speaker of Pwo Karen to mark prominence would likely increase the proportion of aligned gestures. In addition, discrete gestures occasionally occurred during pauses or speech disfluencies.

One-way ANOVAs were used to test for differences between the mean values for the timing of gesture hits for each subject. No significant difference was found for each of the following measures: time between word onset and gesture hit, F(2,66) = 0.447, p = 0.641; time between vowel onset and gesture hit, F(2,66) = 0.164, p = 0.849; and time between word end and gesture hit, F(2,66) = 0.913, p = 0.406.

Table 8: Data summary

Subject	Video duration (sec)	# gestures	# word + gesture alignments	% gestures aligned
A	352.79	46	27	59%
B	186.62	43	22	51%
C	146.00	26	20	77%
Total	685.41	115	69	60%

As shown in Figures 12 and 13, a general pattern of alignment emerged, although the precise timing of the gesture hit varied widely. The gesture hit always fell after the word onset, at an average distance of 282.393 ms for all subjects; the gesture hit also tended to occur after the onset of the vowel (Figure 12), at an average distance of 190.523 ms, although four gestures occurred during the consonant onset.

Figure 12

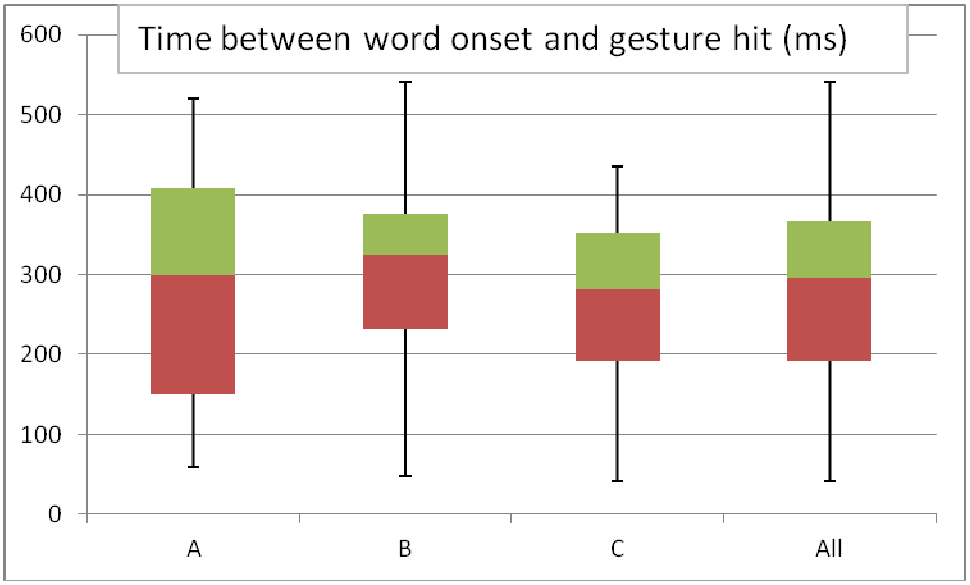
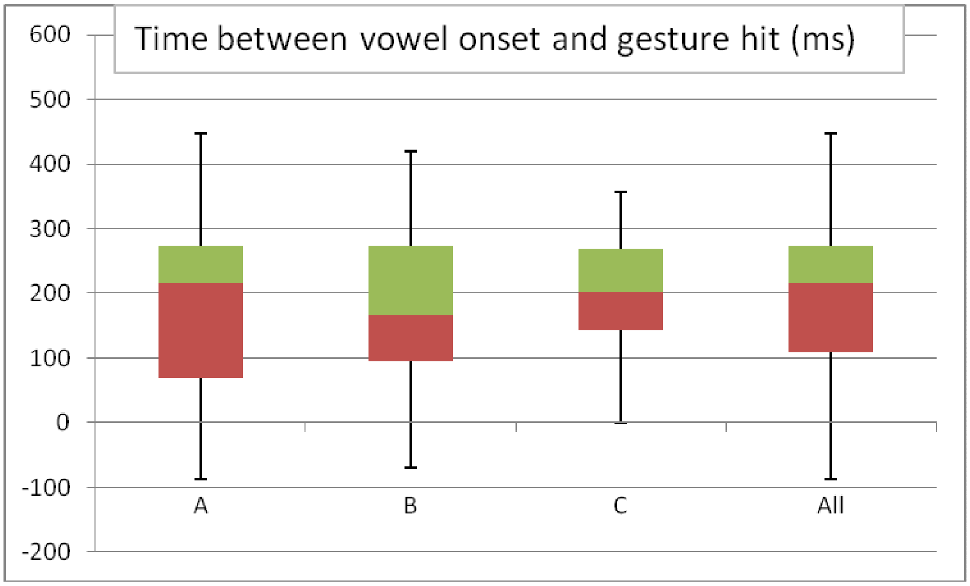
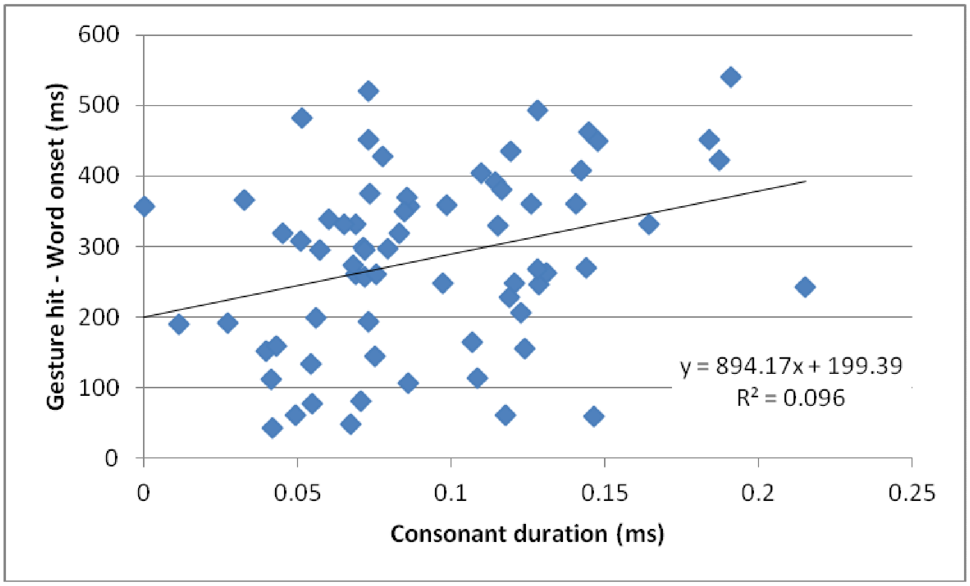


Figure 13



A moderate positive correlation was found between the distance between the word onset and the gesture hit and the duration of the consonant onset, r(67) = 0.310, p < .05 (Figure 14). That is, the gesture hit was more likely to fall further away from the word onset if the onset consonant was longer. Word duration was weakly positively correlated with the onset-to-hit distance, r(67) = 0.271, p < .05. Vowel duration showed no significant correlation with the same distance.

Figure 14: Correlation between consonant duration and distance from word end to gesture hit



The distance between the word coda and gesture hit appears much more variable (Figure 15). Only 61% of all gesture hits occurred before the codas of their associated words, while twenty-seven gesture hits (39%) occurred during the following syllable or pause. However, the time between the coda and the hit was strongly negatively correlated with word duration, r(67) = -0.50, p < .01 (Figure 16). That is, a gesture hit was more likely to fall after the end of its associated word if that syllable was shorter in duration. The same distance between coda and hit was also strongly negatively correlated with vowel duration (Figure 17), r(67) = -0.535, p < .01, but not with consonant duration, r(67) = 0.079, p > .05.

These results suggest that the vowel, and not the consonant, is the most salient portion of the syllable for gesture alignment; this conclusion will be discussed below.

Figure 15

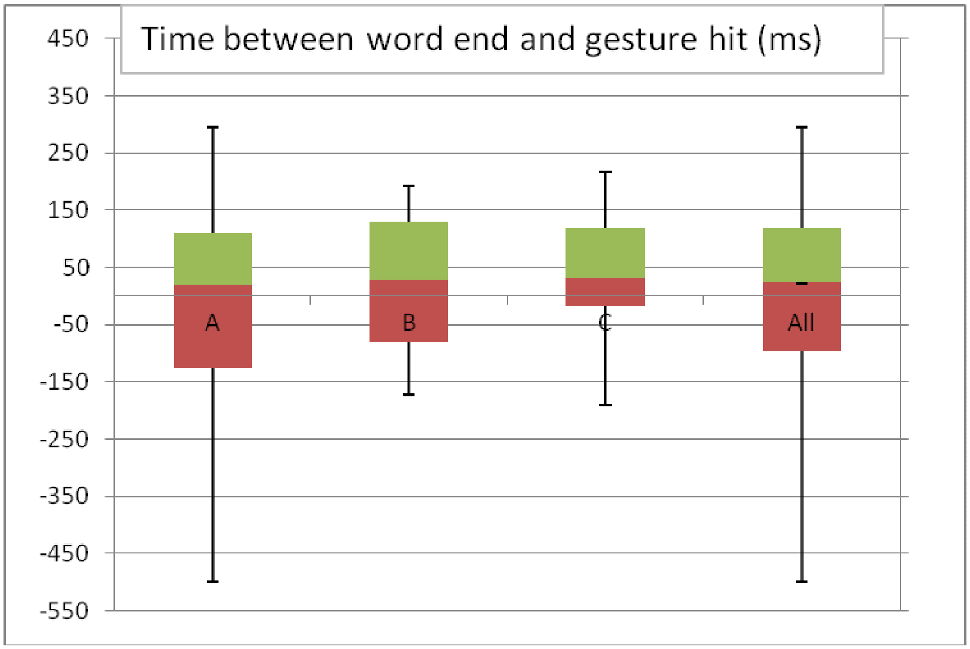


Figure 16: Correlation between word duration and distance from word end to gesture hit

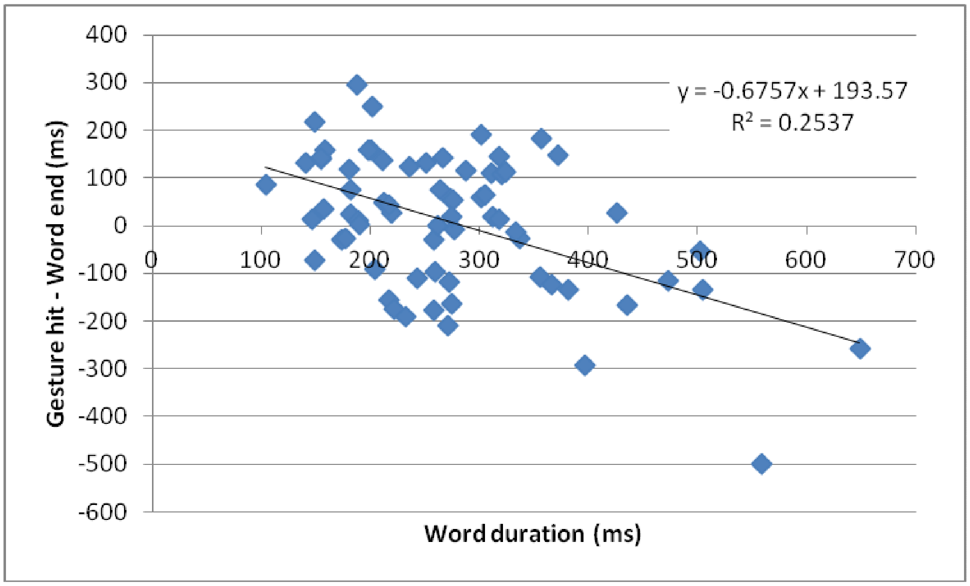
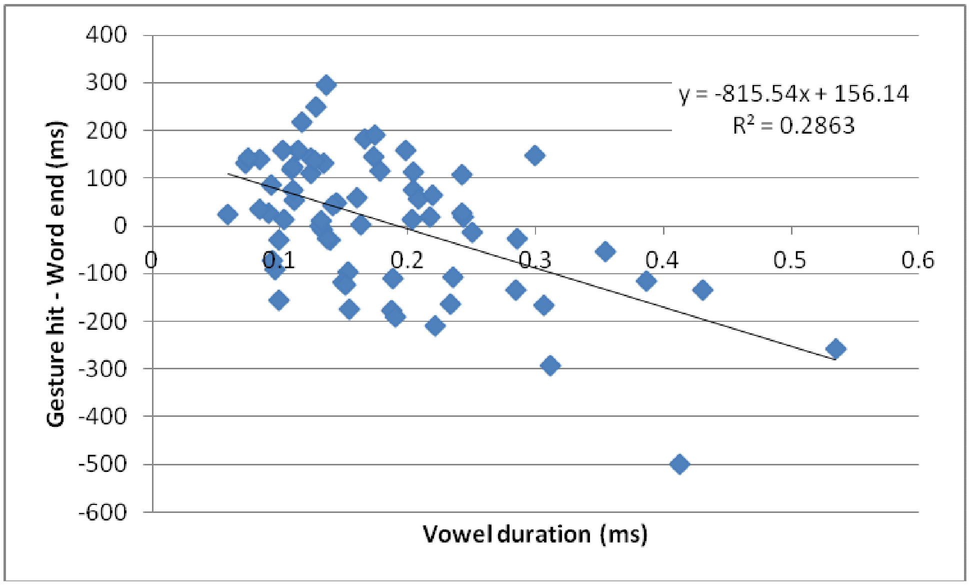


Figure 17: Correlation between vowel duration and distance from word end to gesture hit



4. Discussion

The results given above offer tentative support for the hypothesis that discrete gestures and prominent words are aligned in Pwo Karen. The majority of the discrete gestures in the videos were associated with a syllable, and the percentage of aligned gestures is likely to be higher in actuality.

The results also suggest that the vowel onset is coordinated with the gesture apex. The strongest statistical correlations were between word/vowel duration and and the distance from the word end to the gesture hit; the shorter the word/vowel duration, the more likely it was that the gesture occurred outside the syllable. Consonant duration was not implicated. The wide variation in distance from the apex to the word end also indicates that the word end is not a reliable landmark for coordination. The idea that the gesture is synchronized with some aspect of the vowel was noted in Leonard and Cummins (2011), where the authors discovered that the maximum velocity of the stroke reliably occurred within 100 ms of the vowel onset.

The precise mechanism for coordinating the vowel onset and the gesture apex, as well as the precise nature of the coordination, remain unclear. The correlations suggest that the gesture apex should occur at some constant distance from the vowel onset, but Figure 13 shows that that there was considerable variation in that distance even within the same subject.

Although gesture hit onsets were not measured for the study, the approximate onset of one gesture has been indicated as a green line in Figure 18 to demonstrate one type of hit-onset pattern that occurred in the data. The hit itself begins not within its associated word, but shortly before it, and the apex occurs in the following syllable. De Ruiter (2000) noted that gesture onsets tend to precede the onsets of their associated syllables by less than a second. Leonard and Cummins (2011) claimed that movement onsets for beat gestures began approximately 300 ms before the onset of the stressed vowel, which implies that the movement onsets may have preceded the whole word as well. It has been shown that the articulators in the vocal tract begin articulating a vowel during the production of the preceding segment (Browman and Goldstein 1992). In the absence of any data regarding those vowel and gesture preparations, it may be possible that the beginning of the gesture hit or the preparation for the gesture is coordinated with the beginning of vowel articulation in the vocal tract, and that the apex-onset coordination is incorrect.

The ability of a pitch range effect to extend over multiple syllables is crucial to this analysis, as it marks one of the distinctions between pitch range effects and pitch accents. Although the majority of prominent words in the data were monosyllabic, a small number of polysyllabic words were prominent and associated with gesture hits. Although there were too few instances for conclusive generalizations to be made, Figures 18-20 demonstrate that the gesture might coordinate with the first syllable in the word regardless of the number of syllables in the word. That pattern of coordination would imply that the whole word, not just a particular syllable in it, is associated with the gesture. Since Pwo Karen does not have word-level stress on the initial syllable or on any other syllable, and since initial weak syllables follow the same coordination pattern (Figure 21), the existence of pitch range effects in Pwo Karen is supported.

The vowel onset to gesture apex time was 263.209 ms in Figure 18 and 302.209 ms in Figure 19; in addition, the gesture apex occurred after the word coda in Figure 18, but before it in Figure 19. Although these figures are from different speakers, both of these discrepancies support the analyses above. If the entire time span between the vowel onset of the first syllable and the word coda, including the subsequent syllables, is the "vowel duration," then the vowel duration in Figure 18 is 243.136 ms, and the vowel duration in Figure 19 is 431.121 ms. The location of the gesture apex therefore corresponds with the vowel duration and perhaps the vowel onset.

Figure 18: Prosodic prominence on two-syllable word /$t^hu^5p^hu^5$/ "bird," with gesture apex marked with red line. Total word duration = 312.207 ms. (Subject A)

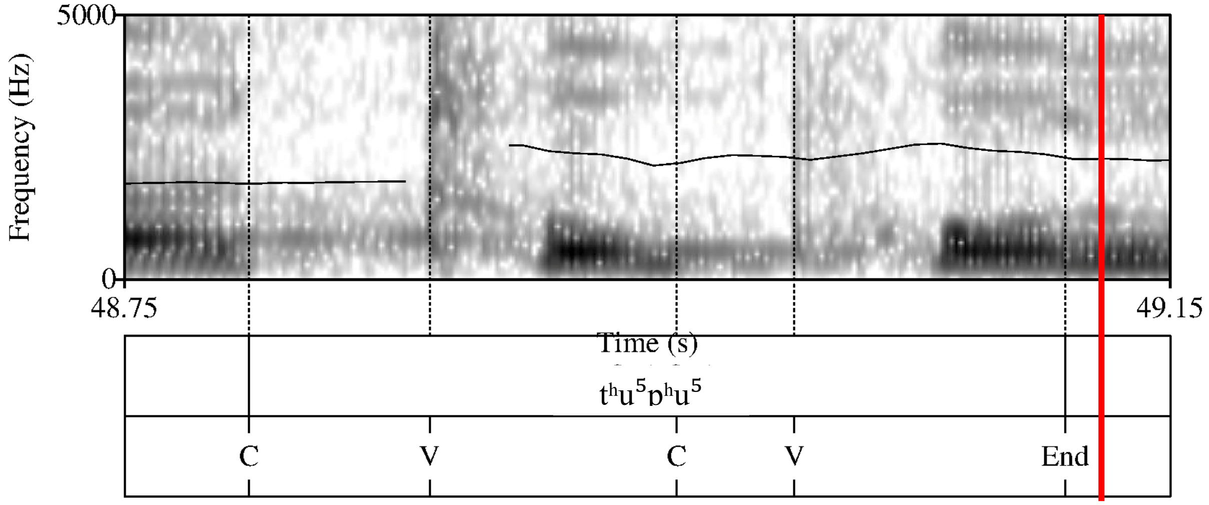
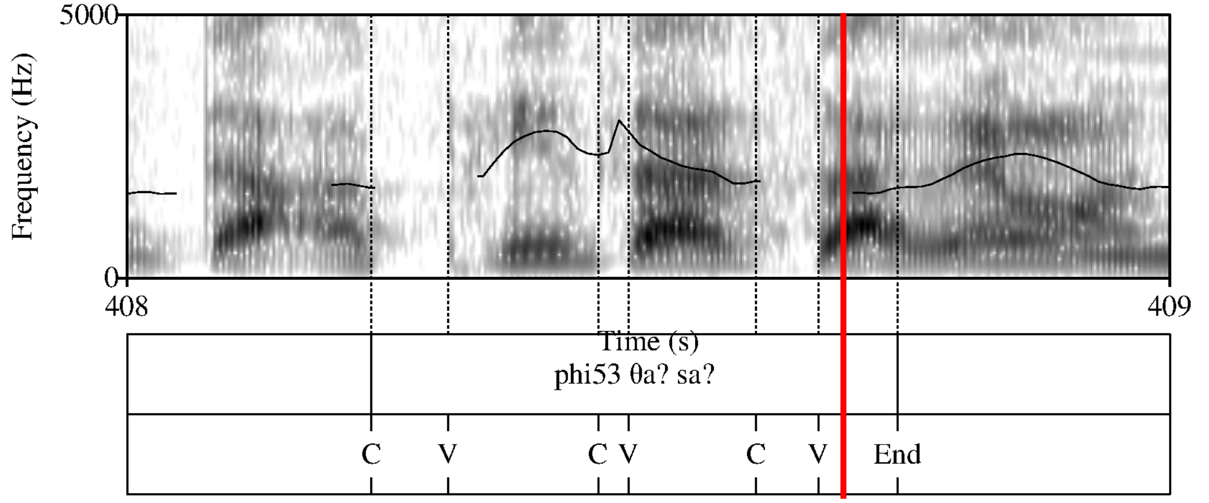


Figure 19: Prosodic prominence on three-syllable word /phi^{53}θasa/ (tones uncertain) "grandmother," with gesture apex marked with red line. Total word duration = 504.627 ms. (Subject C)



The words in Figures 18 and 19 both contained only strong syllables. Figure 20 illustrates a two-syllable combination with one weak (schwa) syllable and one strong syllable, a common occurrence in Pwo Karen. The first syllable's vowel (minus consonant aspiration) is much shorter than the second syllable's, and the author could only perceive prominence on the second syllable. The gesture apex occurs within the first syllable, and Figures 12 and 13 show that the gesture apex always occurs after the onset of the associated word, and almost always occurs after the vowel onset. Therefore, the evidence suggests that the first, weak syllable is prominent as well, despite the author's inability to perceive prominence there. This argument is somewhat circular, but since the gesture apices occurred after the word onset in every other instance, it would be anomalous for the gesture to be associated with only /ʔɔ¹/; it is more consistent to assume that the gesture is associated with the entire word, beginning with /cʰə/. The absence of a pitch track on /cʰə/, as well as its neutral tone, make it difficult to conclude independently that it bears %prom. A comparison with a non-prominent iteration of the same word in Figure 22 shows that the prominent word is approximately 20 ms longer than the non-prominent one; the fact that both iterations in Figures 21 and 22 occur phrase-medially and in fast speech may explain the small size of the difference between the two.

Figure 21: Prominence on two-syllable word /cʰəʔɔ¹/ "monkey" with weak and strong syllables, with gesture apex marked with red line. Total word duration = 275.053 ms. (Subject A)

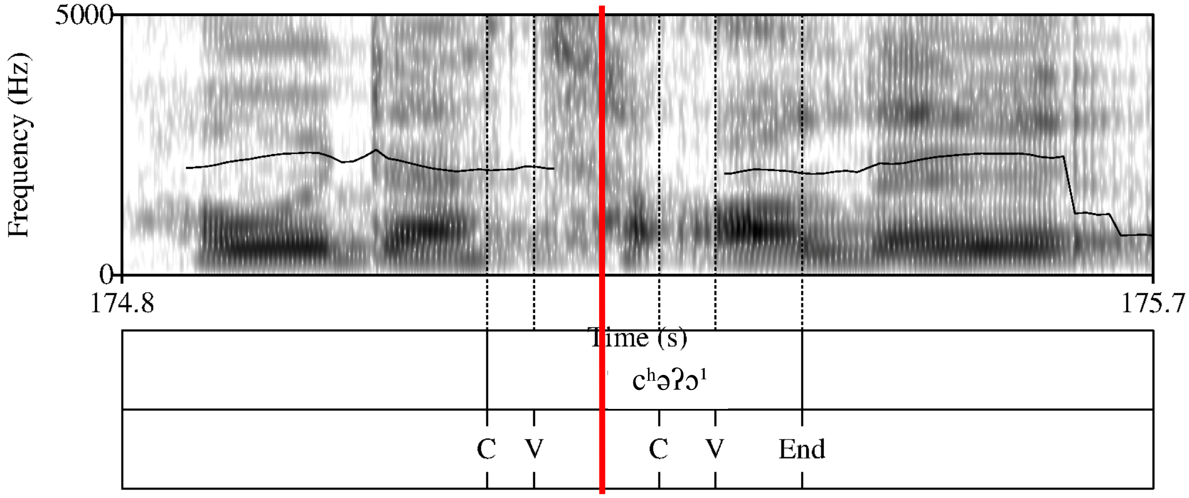
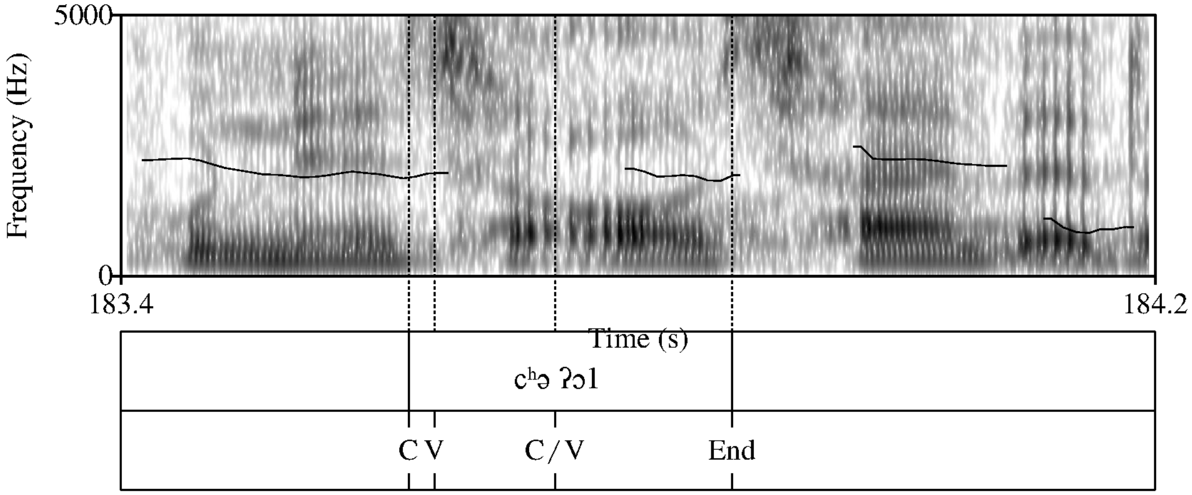


Figure 22: /c^{h}əʔɔ^{1}/ "monkey" without prominence. Total word duration = 254.906 ms. (Subject A)



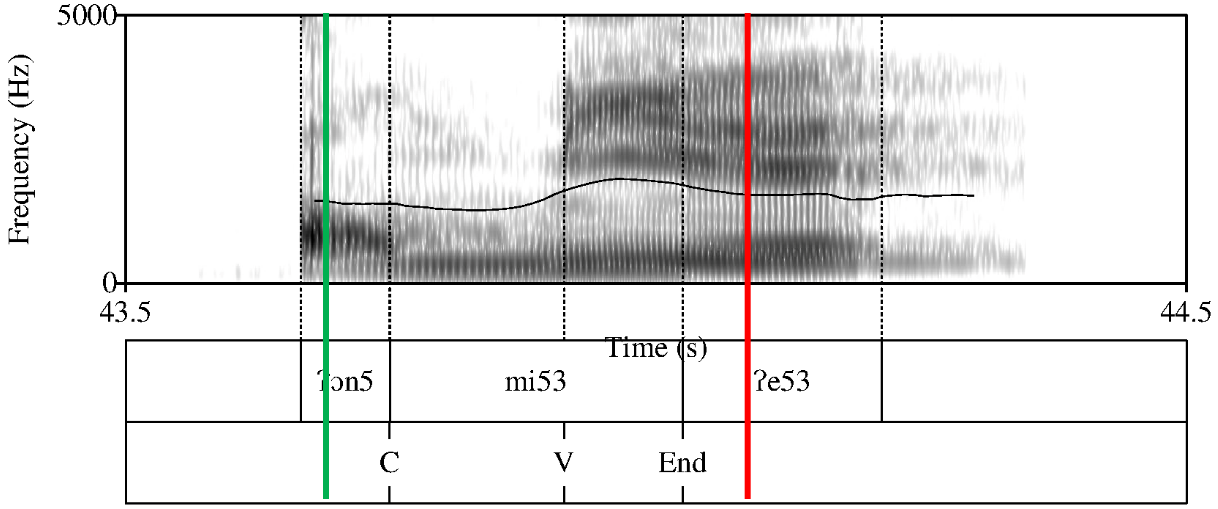
The question of why the gesture apex does not pattern as consistently in Pwo Karen as in other languages remains unresolved. If other languages in this region, or other tonal languages, also do not demonstrate consistent gesture coordination with intonational prominence, current theories of the connection between gesture and speech in the mind will need to be reevaluated. In the absence of other evidence, however, several theories might account for the discrepancy between the results of this study and those of other studies. The presence of phonemic tone in Pwo Karen per se would not have any obvious effect on gesture distribution, but several other aspects of Pwo Karen phonology and prosody might be culpable.

One major difference between the languages used in previous studies on gesture and intonation and Pwo Karen is that the former contain many words of three or more syllables that occur frequently. Pwo Karen contains mostly mono- and disyllabic words, and most of the aligned words in this study were monosyllabic. The short word duration in Pwo Karen may make it more difficult to ensure that the gesture apex occurs within the word, especially if the connection between production of the gesture and production of the vowel is universal. On the other hand, the gesture apex was "regularly reached within the stressed syllable" in Leonard and Cummins (2011), the only study that provided this detail. An illustration of a gesture hit that occurs outside its associated word is given in Figure 23; the relatively short duration of the vowel (112 ms) contributed to the placement of the gesture.

Figure 23: Prosodic prominence on one-syllable word /mi^{53}/ (meaning uncertain[2]), with gesture hit onset in green and apex in red. Total word duration = 275.761 ms. (Speaker B)

[The time scales in Figures 18-22 are given in terms of the time stamps from the videos. C, V, and End mark the word onset, vowel onset, and word coda, respectively.]

ʔɔn^{5} mi^{53} ʔe^{53}
eat [?] NEG
"…[does] not eat _?_"



More salient to this study is the fact that the languages used in previous studies all employ word-level stress, even if stress is not phonemic. It could be that stressed syllables, and therefore pitch accents, offer a stronger target with which the gesture can coordinate. If so, any language that employs pitch range effects as opposed to pitch accents would be affected.

Yasinnik et al. (2004) offers the best opportunity for direct cross-linguistic comparison. That study used samples from three videos of academic lectures by male native English speakers, of which two were examined for coincidence of pitch accents and gestures. The locations of the hits of discrete gestures were determined at a frame rate of 33 ms/frame. Pitch accents were

[2] The transcriptions and translations for some words were not confirmed due to time constraints.

marked according to the ToBI transcription conventions found in Beckman and Elam (1997), which specifies that "pitch accents are placed somewhere within the accented syllable, preferably within the interval that can be identified with the syllable's vowel." In Yasinnik et al., alignment was determined by overlap: pitch-accented words, rather than syllables, were examined for overlapping gesture hit video frames, and the words associated with gesture hits were examined for the presence of pitch accents.

For the one video in which Yasinnik et al. labeled pitch accents independently from gesture hits, the results bear some resemblance to the results of this study. Among polysyllabic hit-aligned words, 90% contained a pitch accent. Among monosyllabic hit-aligned words, however, only 65% contained a pitch accent, though the authors note of the 35% non-aligned words that "most of these were within 100 milliseconds of a pitch accented syllable and often considerably closer." That is, the percentage of non-aligned hits was comparable to the results for Pwo Karen. Based on the location of weak and strong syllables in relation to the non-aligned hits, Yasinnik et al. suggest that a "foot-like rhythmic grouping" might determine the location of a gesture hit. Although Pwo Karen does feature strong and weak syllables, the absence of word-level stress is an argument against the applicability of this model to Pwo Karen.

4.1. Potential issues

Two impediments to a true account of gesture and intonation in Pwo Karen are the author's lack of fluency in the language and the type of subjects employed for this study. Without a native speaker to confirm the locations of prosodic prominence, ambiguously prominent words must be excluded from the data set. In addition, both subjects are young and have spent the majority of their lives outside of Burma, offering many opportunities for interference from other languages, although both subjects have lived exclusively in Karen-speaking communities.

Measuring gesture onsets and offsets might also have given some insight into the overall alignment of the gesture with intonational prominence; the relation between those landmarks and associated words was successfully investigated in Leonard and Cummins (2011) for English. Most other studies besides Yasinnik et al. (2004) also restricted their analysis to beat gestures as defined by McNeill; although limiting the relevant gestures to beats would have lowered the number of gestures available for analysis in this study, it is possible that the kind of alignment explored here differs between beats and other gesture categories.

5. Conclusion

The preliminary results of this study suggest that prosodic prominent and gesture hits do align in Pwo Karen, though in a different manner than in non-tonal languages such as English. The nature of the alignment could not be fully determined within the bounds of this study, which could be due either to the restraints of the measurements or to a distinct alignment phenomenon in Pwo Karen. The vowel onset, however, is the most likely target for gesture coordination.

Comparisons with other tonal languages would be valuable for determining whether the alignment pattern seen in Pwo Karen can be generalized to tonal languages as a whole. In particular, it would be useful to conduct studies both on a tonal language whose words are predominantly mono- and disyllabic, e.g. members of the Chinese language family, and on a tonal language with an abundance of polysyllabic words, to help rule out the potentially confounding factor of average word length. Tonal languages exhibiting some degree of word-level stress, such as Mandarin, would also be valuable sources of comparison (Kochanski, Shih, and Jing 2003).

Appendix : Stimuli for prosodic elictation experiment

Dialogue 1

1.1

A:	nə	thwi^{5}	nɔ^{3}	ʔo^{5}kəle^{53}
	2SG	dog	TOPIC	where

"Where is your dog?"

1.2

B:	jə	thwi^{5}	ʔo^{5}	lə	jə	ɣe^{5}	phɛn^{1}	lə	mi^{5}po^{1}	ʔənai^{1}xu^{1}	nɔ^{3}
	1SG	dog	have	at	1SG	house	in	at	hearth	next-to	FOCUS

"My dog is next to the hearth in the house."

1.3

A:	nə	thwi^{5}	ʔo^{5}	lə	nə	ɣe^{5}	phɛn^{1}	kole^{53}	kain	le^{53}?
	2SG	dog	have	at	2SG	house	in	where	[?]	QUES

"Where is your dog in the house?"

1.4

B:	jə	lo^{1}	nə,	jə	thwi^{5}	ʔo^{5}	lə	jə	ɣe^{5}	phɛn^{1}	lə	mi^{5}po^{1}
	1SG	tell	2SG	1SG	dog	have	at	1SG	house	in	at	hearth

ʔənai^{1}xu^{1}	nɔ^{3}
next-to	FOCUS

"I told you, my dog is next to the ***hearth*** in the house."

1.5

A:	ʔəwe^{53}	ma^{3}	chənole^{53}	lə	mi^{5}po^{1}	ʔənai^{1}xu^{1}	nɔ^{3}
	3SG	do	what	at	hearth	next-to	FOCUS

"What is he doing next to the hearth?"

1.6

B:	ʔəwe^{53}	mi^{53}	lə	mi^{5}po^{1}	ʔənai^{1}xu^{1}	nɔ^{3}
	3SG	sleep	at	hearth	next-to	FOCUS

"He is sleeping next to the hearth."

1.7

A:	ma^{3}	chənole^{53}	lə	mi^{5}po^{1}	ʔənai^{1}xu^{1}	nɔ^{3}
	do	what	at	hearth	next-to	FOCUS

"***What*** is he doing next to the hearth?"

1.8

B:	jə	lo^{1}	nə,	ʔəwe^{53}	mi^{53}	lə	mi^{5}po^{1}	ʔənai^{1}xu^{1}	nɔ^{3}
	1SG	tell	2SG	3SG	sleep	at	hearth	next-to	FOCUS

"I told you, he is ***sleeping*** next to the hearth."

1.9

A:	ʔə	khu^{5}xwi^{5}	phɔn^{5}ku^{1}	nɔ^{3}	chə	nɔ^{3}	ʔo	le^{53}
	3SG	head	on-top-of	FOCUS	thing	FOCUS	have	QUES

"What is on his head?"

1.10
B: ʔə mi^1 ʔo^5 lə ʔə khu^{5}xwi^5 phɔn^{5}ku^1 nɔ^3
3SG tail have at 3SG head on-top-of FOCUS
"His tail is on his head."

1.11
A: lə ʔə khu^{5}xwi^5 phɔn^{5}ku^1 nɔ^3 chə nɔ^3 ʔo le^{53}
at 3SG head on-top-of FOCUS thing FOCUS have QUES
"***What*** is on his head?"

1.12
B: ʔə mi^1 ʔo^5 lə ʔə khu^{5}xwi^5 phɔn^{5}ku^1 nɔ^3, jə lo^1 nə
3SG tail have at 3SG head on-top-of FOCUS 1SG tell 2SG
"His ***tail*** is on his head, I told you."

Dialogue 2

2.1
A: nə xwḭ^1 ʔɔn^5 chənole^{53}
2SG buy eat what
"What did you buy?"

2.2
B: jə xwḭ^1 ni^2 ɣɔn^{1}ji^{53}, θəwa^{53}, lɛ θə^{1}dɔn^3
1SG buy get lemongrass lotus and shrimp
I bought lemongrass, lotus, and shrimp.

2.3
A: nə xwḭ^1 ɣɔn^{1}ji^{53}
2SG buy lemongrass
You bought ***lemongrass***?!

2.4 ba^{5}no^{5}le^{53} nə ba^5 xwḭ^1 ɣɔn^{1}ji^{53} nɔ^3
why 2SG must buy lemongrass FOCUS
"Why did you buy lemongrass?"

2.5
B: ʔə xwḭ^1 ba^5 ʔəkh^{5}chu^1 nɔ^3 jə xwḭ^1
3SG buy NEG? because FOCUS 1SG buy
Because it was cheap, I bought it.

3. Miscellaneous Sentences

3.1 ʔəmi^{5}ja^5 lə nwe^{5}xa^5 nɔ^3 jə lo^1 da^{5}we^1 ʔə jə

first one week FOCUS 1SG tell to(?) 3SG
1SG

mə li^{1} lɔn^{1}
FUT go down

"Last week I told her I would go [there]."

3.2 jə lo^{3} lɔn^{1}ma^{1} lə phja^{53} phɛn^{1}
1SG rock lose at market in
"I lost my rock in the market."

3.3 jə mi^{53} ʔəkhu^{5}chu^{1} nɔ^{3}, jə li^{1} lɔn^{1} lə phja^{53} phɛn^{1}
1SG sleep because FOCUS 1SG go down at market in

ke^{5} ʔe^{53}
want NEG

"Because I am sleeping, I do not want to go to the market."

3.4 jə ɣe^{53} chu^{1} nai^{1}, mi^{5}dwai^{1}, de^{3} mḭ^{1}
1SG come bring type-of-basket matches with cooked-rice
"I am bringing a *nai* basket, matches, and cooked rice."

Bibliography

Beckman, M.E. and Gayle Ayers Elam. 1997. Guidelines for ToBI labeling. Available from <ling.ohio-state.edu/~tobi>

Bowern, Claire, Emily Gasser, Sophia Gilman, Dan Hansen, Jessica Hsieh, Ilkyu Kim, Sabina Matyiku, and Jason Zentz. 2011. Pwo Karen Sketch Grammar. Unpublished manuscript, Yale University, New Haven, CT.

Browman, Cathy and Louis Goldstein. 1992. Articulatory phonology: an overview. *Phonetica*, 49(3-4), 155-80.

Brown, Amanda and Marianne Gullberg. 2011. Bidirectional cross-linguistic influence in event conceptualization? Expressions of Path among Japanese learners of English. *Bilingualism: Language and Cognition*, 14(1), 79-94.

Capirci, O., Iverson, J.M., Pizzuto, E., Volterra, V. 1996. Gestures and words during the transition to two-word speech. *Journal of Child Language*, 23(3), 645-673.

Casey, Shannon and Karen Emmorey. 2008. Co-speech gesture in bimodal bilinguals. *Language and Cognitive Process,* 24, 290–312.

Cassell, Justine, David McNeill, and Karl-Erik McCullough. 1998. Speech-gesture mismatches: Evidence for one underlying representation of linguistic and nonlinguistic information. *Pragmatics & Cognition,* 6(2), 1-34.

Cave, C., I. Guaitella, R. Bertrand, S. Santi, F. Harlay, and R. Espesser. 1996. About the relationship between eyebrow movements and f0 variations. *Proceedings of the ICSLP*, 2175-2179. Philadelphia, PA, USA.

Chen, Yiya and Carlos Gussenhoven. 2008. Emphasis and tonal implementation in Standard Chinese. *Journal of Phonetics*, 36 (4), pp. 724-746.

de Kok, Iwan and Dirk Heylen. 2010. Differences in listener responses between procedural and narrative task. In *Proceedings of the 2nd international workshop on Social signal processing* (SSPW '10). ACM, New York, NY, USA, 5-10.

de Ruiter, Jan P. 1998. *Gesture and speech production.* Doctoral dissertation, Catholic University of Nijmegen, The Netherlands.

de Ruiter, Jan P. and D. Wilkins. 1998. The synchronization of Gesture and Speech in Dutch and Arrernte (an Australian Aboriginal language): A Cross-cultural comparison: In Santi, S. et al. (Eds.), *Oralité et Gestualité*, 603-607. Paris: L'Harmattan.

de Ruiter, Jan P. 2000. The production of speech and gesture. In D. Mc Neill (Ed.), *Language and gesture: Window into thought and action*, 284-311. Cambridge, UK: Cambridge University Press.

Duncan, Susan, Fey Parrill, and Dan Loehr. 2005. Discourse factors in gesture and speech prosody. Presented at the 2nd Conference of the International Society for Gesture Studies (ISGS), Lyon, France, June 2005.

Esteve-Gibert, N. and Prieto, P. 2011. The temporal alignment between prosody and gesture in Catalan babbling infants. Oral presentation at *Gesture and Speech in Interaction (Gespin) 2011*. Bielefeld University: Bielefeld, Germany, 5-7 September 2011.

Iverson, J. and Goldin-Meadow, S. 1997. What's communication got to do with it? Gesture in children blind from birth. *Developmental Psychology*, 33, 453-467.

Green, A.D. 1995. The prosodic structure of Burmese: a constraint-based approach. *Working Papers of the Cornell Phonetics Laboratory*, 10, 67-96.

Hsieh, Jessica. 2011. Towards a prosodic transcription system for Pwo Karen. Unpublished manuscript, Yale University, New Haven, CT.

Jannedy, Stefanie and Norma Mendoza-Denton. 2005. Structuring information through gesture and intonation. *Interdisciplinary Studies on Information Structure*, 3, 199-244.

Jun, Sun-Ah (ed.). 2005. *Prosodic Typology: The Phonology of Intonation and Phrasing.* New York: Oxford University Press.

Keating, P., M. Baroni, S. Mattys, R. Scarborough, A. Alwan, E. Auer, & L. Bernstein. 2003. Optical Phonetics and Visual Perception of Lexical and Phrasal Stress in English. *Proc.9 15th International Congress of Phonetic Sciences*, 3-9 August 2003, Barcelona, Spain, 2071-2074.

Kendon, Adam. 1972. Some relationships between body motion and speech. In A. Siegman and B. Pope (Eds.), *Studies in dyadic communication*, 177-210. New York: Pergamon Press.

Kochanski, G., C. Shih, H. Jing. 2003. Quantitative Measurement of Prosodic Strength in Mandarin. *Speech Communication*, 41(4).

Leonard, Thomas and Fred Cummins. 2011. The temporal relationship between beat gestures and speech. *Language and Cognitive Processes*, 26 (10), 1457-1471.

Loehr, Daniel P. 2004. Gesture and intonation**.** Doctoral Dissertation, Georgetown University, Washington, DC.

McClave, Evelyn. 1998. Pitch and manual gestures. Journal of Psycholinguistic Research, 27(1), 69–89.

McNeill, David. 1992. *Hand and Mind: What Gestures Reveal about Thought*. Chicago: University of Chicago Press.

McNeill, David & Duncan, Susan. 2000. Growth points in thinking-for-speaking. In D. McNeill (Ed.), *Language and Gesture*, 141-161. Cambridge: Cambridge University Press.

McNeill, David, & Levy, Elena. 1982. Conceptual representations in language activity and gesture. In R. Jarvella & W. Klein (Eds.), *Speech, place, and action,* 271-295. Chichester, England: Wiley.

Nobe, Shuichi. 1996. *Cognitive Rhythms, Gestures, and Acoustic Aspects of Speech*. Ph.D. thesis, Department of Psychology (Cognition and Communication), University of Chicago.

Ozyurek, Asli. 2002. Speech-gesture relationship across languages and in second language learners: Implications for spatial thinking and speaking. In B. Skarabela, S. Fish, & A. H. Do (Eds.), *Proceedings of the 26th annual Boston University Conference on Language Development*, 500-509. Somerville, MA: Cascadilla Press.

Rieber, Robert W. 1983. Dialogues on the psychology of language and thought. New York: Plenum Press.

Rochet-Capellan, A., Laboissière, R., Galván, A., and Schwartz, J. L. 2008. The speech focus position effect on jaw-finger coordination in a pointing task. Journal of Speech and Hearing Research, 51(6), 1507-1521.

Rochet-Capellan A., Vilain C., Dohen M., Laboissière R. and Schwartz J.L. 2008. Does the number of syllables affect the finger pointing movement in a pointing-naming task? In Proceedings of the 8th International Seminar on Speech Production, Strasbourg-France, 257-260.

Sandler, Wendy. 2009. Symbiotic symbolization by hand and mouth in sign language. *Semiotica* 174 (1/4), 241-275.

Sansavini, B., Guarini, S. & Stefanini, C. 2010. Early development of gestures, object-related actions, word comprehension and word production, and their relationships in Italian infants. *Gesture*, 10(1), 52-85.

Stam, Gale. 1998. Changes in Patterns of Thinking about Motion with L2 Acquisition. In Cavé, C., Guaïtella, I. & Santi, S. (Eds.). *Oralité at Gestualité: Communication Multimodale, Interaction*, 615-619. Paris: L'Harmattan.

Tuite, Kevin. 1993. The production of gesture. Semiotica, 93(1/2), 83–105.

Yasinnik, Yelena, Margaret Renwick and Stefanie Shattuck-Hufnagel. 2004. The timing of speech-accompanying gestures with respect to prosody. *From Sound to Sense: 50+ Years of Discoveries in Speech Communication*, 11-13 June 2004, Cambridge, MA.