Acoustic cues to prosodic boundaries in Yami: A first look

Li-Fang Lai [1], *Shelome Gooden* [2]

[1] University of Pittsburgh, USA
[2] University of Pittsburgh, USA
lil91@pitt.edu, sgooden@pitt.edu

Abstract

It is well known that in many Indo-European languages speakers manipulate acoustic cues to encode different prosodic phrase boundaries. However, no such attempt has been made to investigate these effects in Austronesian languages. Therefore, this paper reports on preliminary research on the prosodic structure of Yami, an endangered Austronesian language spoken on Orchid Island, Taiwan. Two acoustic parameters were examined: pre-boundary syllable duration and phrase-final F_0 slope. The results provide evidence for a three-layered prosodic hierarchy in Yami: Word, AP, and IP. These levels differ significantly in syllable duration, particularly when the pre-boundary syllables bear lexical stress. The phrase-final F_0 slope also serves as a cue to prosodic boundaries since the IP boundary tones (H% and L%) are characterized by steeper F_0 curves, whereas the corresponding AP tones (Ha and La) are typified by steady pitch curves. The results we report not only offer a basic understanding of the prosodic structure of Yami, but contribute to the research on Austronesian language prosody more generally.

Index Terms: prosodic boundary, pre-boundary lengthening, normalization, F_0 slope, Yami

1. Introduction

A number of studies have indicated that acoustic cues such as pitch lowering, pausing, and segmental lengthening are manipulated by speakers to distinguish among distinct levels of prosodic boundaries [2, 8]. English, for instance, demonstrates a three-way distinction of pre-boundary lengthening to mark prosodic boundaries. [9] reported that the normalized duration was longest at the Intonational phrase (IP) boundary, followed by intermediate phrase (ip) boundary and then word. In addition to boundary, [9] shows that in English, pitch accent provides another source of segmental lengthening, as the duration is significantly longer for the accented nuclei than the unaccented ones for all levels of prosodic boundaries. In Persian, which has Japanese-style pitch accent rather than English-style stress contrast, longer syllable duration is also interpreted as side-effects of the pitch accent placement. [15] points out that despite the contrasting locations between cliticized and non-cliticized words, whose pitch accent is aligned within the antepenultimate and penultimate syllables respectively, accented syllables are overall longer than unaccented ones. Previous work also suggests an interaction between final lengthening and accentual lengthening. Specifically, there are increasing lengthening effects when the final syllable receives a pitch accent [2, 3, 5, 6, 9, 17]. A partial pre-boundary lengthening effect was also found when the pitch accent falls on the penultimate syllable because the domain of accentual lengthening begins from the onset of the stressed syllable and extends rightward until the end of the word [9].

Prosodic constituency may also differ from one language to another. For instance, in languages like French and Bengali, there is a smaller prosodic unit, the accentual phrase (AP), between word and ip [7, 10]. In these languages, an AP may optionally contain a content word plus its surrounding functional particles. In French, duration is used to encode prosodic boundary differences. [12] shows that the right edge of an ip-boundary is marked by significant durational cues that are stronger than those associated with an AP-boundary and weaker than those found at the vicinity of an IP-boundary.

Bengali [10] exhibits a similar lengthening pattern in which the IP-final syllables are far longer than those of the same word found in the AP-final position, while the final syllable of a word found ip-finally shows moderate lengthening. [10] further pointed out that the three prosodic boundaries also differ in tone height. The pitch of an IP boundary tone (H% or L%) is more extreme than that of the corresponding ip tone (H- or L-), which is more extreme than that of the corresponding AP tone (Ha or La).

However, the durational cue is not consistently found in all languages that have APs. For example, [11] uses ambiguous digit strings, each of which consists of five syllables, to examine what acoustic cues are exploited in prosodic phrasing in Korean. More specifically, the digit string /imanimano/ allows two interpretations based on how syllables are grouped: a 2 + 3 phrasing means the sequence of numbers 20000 and 20005, while a 3 + 2 phrasing represents the sequences 20002 and 10005. The results show that in Korean, speakers and listeners used F_0 rather than speech timing to distinguish the temporal structure between 2 + 3 and 3 + 2 phrasing. This suggests that duration may not be used as a major cue for AP segmentation for Korean speakers.

Last but not least, other pitch cues such as F_0 slope and pitch reset were also proved to be significant predictors of boundary categories (no boundary vs. boundary) in list intonation. For instance, [18] examined how pitch cues are employed by native English and Chinese speakers to distinguish the lists of food items with or without boundaries (e.g., bacon-salad and wine (no boundary) vs. bacon, salad, and wine (boundary)). The results show that in English, F_0 slope was an effective predictor for boundary categories, as a rising tone is usually used when speakers produce a list of items with boundaries. In Chinese, F_0 slope has the same dimension as lexical tones, so speakers cannot readily change pitch slope since the change could possibly result in a different word. Thus, in Chinese, pitch information was primarily conveyed via a reset of the pitch declination when the list of food items are produced with boundaries.

doi: 10.21437/SpeechProsody.2016-128

Despite these findings, few attempts have been made to investigate cues to prosodic phrase boundary in Austronesian (AN) languages. As far as we know, there are only two studies briefly discussing prosodic phrasing in AN languages such as Western Cham [14] and Tongan [16], in which two prosodic units above the word level were defined: the AP and IP. As defined earlier, an AP in Tongan and Western Cham contains one content word and often incorporates surrounding functional particles, and an IP contains one or more APs.

To address the dearth of research on this issue in AN languages, this study reports on preliminary research on prosodic phrase boundaries in Yami, an endangered AN language spoken on Orchid Island, Taiwan. Following previous work on the prosodic structure of AN languages [14, 16] and from more recent work on Yami [20], we postulate three levels of prosodic constituents in Yami: Word, AP, and IP. In addition, based on the findings that duration [9, 10, 12] and pitch cues [10, 18] are important cues to boundary strength, we examine whether and how these acoustic parameters are used by Yami speakers to signal layered prosodic domains. Two hypotheses are proposed. First, we expect that syllable duration will increase with the strength of the prosodic boundary (IP > AP > Word). Second, since Yami is typologically similar to English (stress language), we expect that the F_0 slope would also facilitate cueing boundary strength [18].

2. Methods

2.1. Corpus

Ten pairs of native Yami speakers (20 total), ages 34 to 58 at the time of recording, were recruited for this project. In order to ensure the naturalness of the utterances, (semi-) spontaneous speech data were collected using a modified map task [1]. In this task, one participant serves as an *asker* to produce *wh-questions*; the other participant serves as an *instruction-giver* to respond his/her partner what (s)he sees in a certain village, thereby producing *declarative sentences*. The *asker* may also produce yes-no questions in case the instructions are not clearly delivered.

Table 1. *Conversation for data elicitation.*

Asker	***Ka mangay do anjin?*** 'Where are you going?'
Instruction-giver	***Ko mangay do __.*** 'I am going to __.'
Asker	***Ikong mo nimacita?*** 'What did you see?'
Instruction-giver	***Ko nimacita so __ kano __.*** 'I saw __ and __.'

Each pairs 'travelled' across the six villages on Orchid Island and worked together to find ten differences between their maps. The participants contributed approximately 50 minutes of recording for a total of 287 utterances, the bulk of which were produced in a L% boundary tone condition.

2.2. Duration normalization

In order to prevent confounding influences induced by differences in speech rate, we performed duration normalization. It is worth reiterating that due to the game design, we collected the same *wh-questions* and *declaratives* across the *askers* and *instruction-givers* respectively. These were relatively short and had identical number of syllables and syllable composition. This allowed us to diminish segmental effects on syllable duration. Therefore, we used the normalization method employed in [13], in which syllable durations are normalized as a proportion of the sentence duration for each speaker using the formula:

$$d' = d/s \qquad (1)$$

where d = syllable duration , s = sentence duration, and d' = normalized syllable duration.

2.3. Measurement domain

Duration measures were taken from the final syllable of a word at three prosodic levels: Word, AP, and IP, as illustrated in Figure 1.

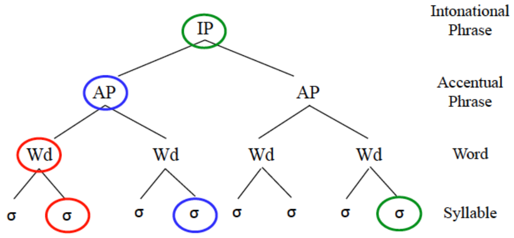

Figure 1. *Measurement domain.*

2.4. F_0 slope

Before calculating F_0 slope, pitch normalization was done using the ProsodyPro Praat script [19] to produce time-normalized F_0. Next, the pitch contour changes for phrase-final syllables were computed using the following formula [18]:

$$F_0 \text{ slope} = f_{max}\text{–}f_{min}/t_{max} - t_{min} \text{ (s.t./s)} \qquad (2)$$

where f_{max} and f_{min} are the maximum and minimum F_0 of the target syllables, and t_{max} and t_{min} are the times at which the minimum and maximum F_0 occurred. The numerator f_{max}–f_{min} are converted from Hz into semitones (s.t.) by (3):

$$Hz = 12*[\log(f_{max}/f_{min})/\log(2)] \qquad (3)$$

2.5. Coding

A ToBI-style Break Index labelling system [4] was used to notate the strength of prosodic boundaries. The Word level was assigned a break index value 1; AP, 2, and IP, 3. In some cases, a value 0 was labeled wherever there is clear coarticulation across a word boundary, as illustrated in Figure 2.

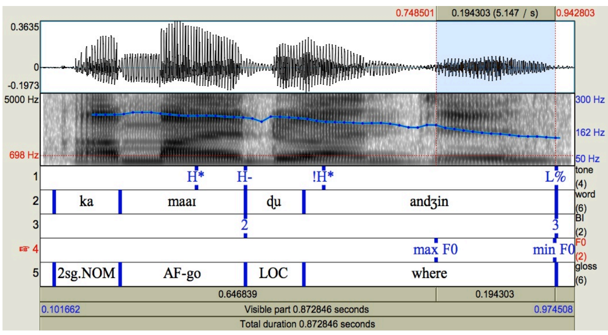

Figure 2. *Illustration of Break Index labeling. 'Where are you going?'*

2.6. Target word-final syllables

All duration measurements were taken from the word-final syllables (underlined in Figure 3). Two word stress conditions – final stress and penultimate stress are considered. When the final stressed syllable also receives a pitch accent, it is classified as a *final accented* syllable, represented as 3(a). Whereas when the stress and pitch accent fall on the penultimate syllable, the pre-boundary syllable is interpreted as a *final post-accented* one, illustrated in 3(b).

Measuring both contexts is important because it allows us to test for final lengthening effects and also to see whether accent-induced lengthening extends beyond the stressed syllable to the end of the word; namely, the final post-accented syllables.

(a) […σ σ <u>σ́</u>] Boundary (b) […σ σ́ <u>σ</u>] Boundary

Figure 3. *Schematic diagram of the two target pre-boundary syllables.*

3. Results

The number of tokens for analysis under the two conditions of stress and pitch accent and three boundary levels are presented in Table 2.

Table 2. *Frequency table.*

Pre-boundary syllable	Boundary	Final stress 3(a)	Penultimate stress 3(b)
No pitch accent	Word		14
	AP		49
	IP		51
Pitch accent	Word	88	
	AP	112	
	IP	202	
	N	402	114

3.1. Duration

The statistical analysis suggests a significant boundary effect on the final accented syllables [$F(2, 399) = 96.05$, $p < .001$]. Post hoc comparisons reveal a three-way durational distinction: syllables in IP-final position are significantly longer than those in AP- and Word-final positions ($p < .001$ respectively). AP-final syllables were also significantly longer than the Word-final ones ($p = .031$) (Figure 4).

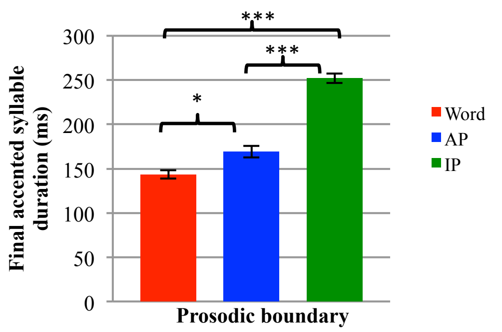

Figure 4. *Effect of prosodic boundary on normalized syllable duration under the final accented condition.*

Under the final post-accented condition, syllable duration also differs significantly [$F(2,111) = 71.21$, $p < .001$]. Post hoc comparisons indicate that syllables in IP-final position are significantly longer than those in AP- and Word-final positions ($p < .001$ respectively). However, the durational effect was not found between AP and Word boundaries ($p = .880$) (Figure 5).

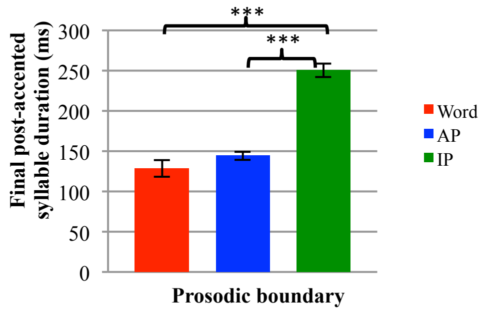

Figure 5. *Effect of prosodic boundary on normalized syllable duration under the final post-accented condition.*

3.2. F_0 slope

The statistical analyses on phrase-final F_0 slopes reveal that degree of pitch contour change varies with boundary strength as well. As demonstrated in Figure 6, both high edge tones are characterized by positive slopes, yet their steepness differs. The mean F_0 slope for high IP boundary tones (M = 12.93, SD = 6.309) is steeper than that of the corresponding AP tones (M = 10.12, SD = 7.578) [$t(145) = 17.34$, $p < .001$]. With regard to low tones, both low edge tones have negative slopes. Again, Yami speakers produced much sharper F_0 slopes in low IP boundary tones (M = -16.49, SD = 10.776) than in low AP tones (M = -5.96, SD = 2.829) [$t(179) = 18.11$, $p < .001$].

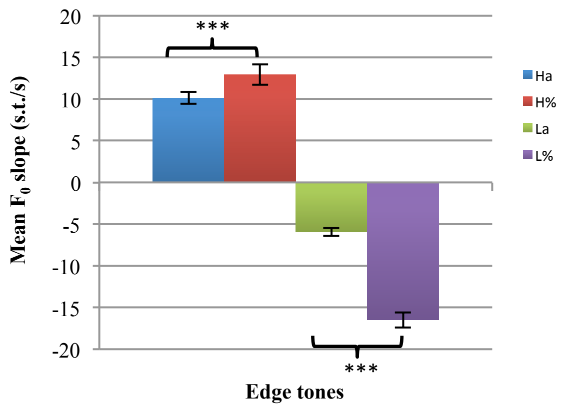

Figure 6. *Phrase-final F_0 slopes of edge tones.*

4. Discussion

As discussed earlier, boundary as well as accent are the major sources of lengthening [2, 3, 5, 6, 9, 15, 17] crosslinguistically. These results on pre-boundary syllable duration offer strong initial evidence supporting our proposal for a three-layered prosodic hierarchy in Yami: Word, AP, and IP. We showed that under the final accented condition (3a), in which the final syllable bears prosodic prominence, syllable duration serves as a reliable cue to boundary strength. Syllables in IP-final position are significant longer than those in AP-final position, which are in turn significantly longer than those occurring word-finally. On the other hand, under the final post-accented condition (3b), where the final boundary syllable does not carry prosodic prominence, only a two-way durational contrast was found. Syllables appearing in the IP-final position are longer than those in both AP- and Word-final positions, but no observable difference was found in syllable duration between the AP and Word levels. Other than syllable duration, pitch contour changes also facilitate cuing the prosodic boundaries. As predicted, larger prosodic constituents show steeper, be it positive or negative, phrase-final F_0 slopes. The high and low AP tones, on the contrary, are realized with gentle slopes (Figure 6). These results are very important because they provide a foundational basis on which we can continue the research to tease apart more precisely the cues to prosodic phrasing in Yami.

However, due to the nature of the data we collected in which wh-questions and declaratives are characterized by falling tunes [20], the low AP (La) and the high IP (H%) tones were relatively infrequent in our data (La comprises 21.1% of the low edge tones and H%, 18.5% of the high edge tones). Thus, eliciting speech containing more low AP tones and high IP tones would enable us to better assess the reliability of F_0 slope as a cue to prosodic phrase boundaries. With respect to lengthening due to accent, the prosodic conditions here are limited to final accented and final post-accented contexts (see Table 2), so pair-wise comparisons between final accented vs. final unaccented or penultimate accented vs. penultimate unaccented as was done in English [9] are not possible.

Nevertheless, the observed trends provide some support for the effects of accentual lengthening and final lengthening in Yami. For instance, overall, accented syllables at Word- and AP-final positions are longer than post-accented syllables in the same positions. This suggests that accent-induced lengthening is in effect but does not extend beyond the stressed syllable. Following this line of reasoning, we expected the post-accented IP-final syllables to be shorter than the accented IP-final syllables as well. Surprisingly, the results show that there is no significant difference between IP-final syllables in these two prosodic contexts, as both are long. This immediately raises questions about the potential source for the lengthening found in the final post-accented context in the IP. The durational effect seen here may be interpreted either as accent-induced lengthening that extends beyond the stressed syllable to the end of the word or simply a boundary effect. Further investigations are needed to test for these effects.

5. Conclusions

The prosodic structure of Yami has never been previously discussed. This research was aimed at gaining insights into this issue to see how layered prosodic domains are encoded through speakers' manipulation of duration and pitch. To our knowledge, this is the first attempt made to explore how acoustic cues are used to mark prosodic boundaries in AN languages. The results show that duration is used in conjunction with F_0 slope by Yami speakers to distinguish among distinct levels of prosodic boundaries under the final accented condition. The final post-accented condition is a bit more complex because the lengthening effect was limited to IP-final syllables. No significant durational contrast was found between AP and Word levels. While the results are promising, more data is needed to support the proposal for an AP boundary as distinct from a word boundary. The results on F_0 slope also suggest its significance in signaling prosodic boundary strength because IP-boundaries have steeper slopes than AP-boundaries.

Given that (Taiwanese) AN language prosody is severely under researched, the results we report not only allow us to have a basic understanding of the prosodic structure of Yami, but serve as a starting point for expanding research on AN language prosody more broadly.

6. Acknowledgements

We are grateful to our participants for providing us the invaluable data and to the INPAC, GPSG, and the Department of Linguistics at the University of Pittsburgh for providing us the summer research grants and fellowship to conduct the fieldwork. Also thanks to those who made constructive comments from which this study has greatly benefited.

7. References

[1] A. Anderson, M. Bader, E. Bard, E. Boyle, G. M. Doherty, S. Garrod, S. Isard, J. Kowtko, J. McAllister, J. Miller, C. Sotillo, H. S. Thompson, and R. Weinert, "The HCRC Map Task Corpus," *Language and Speech*, vol. 34, pp. 351-366, 1991.

[2] C. W. Wightman, S. Shattuck-Hufnagel, M. Ostendorf, and P. J. Price, "Segmental durations in the vicinity of prosodic phrase boundaries," *Journal of the Acoustical Society of America*, vol. 91, no. 3, pp. 1707-1717, 1992.

[3] A. E. Turk and J. R. Sawusch, "The domain of accentual lengthening in American English," *Journal of Phonetics*, vol. 25, pp. 25–41, 1997.

[4] M. E. Beckman and G. Ayers Elam, *Guidelines for ToBI Transcription*, ver. 3.0, Available online at: http://www.ling.ohio-state.edu/~tobi/ametobi/, 1997.

[5] A. E. Turk and L. White, "Structural influences on accentual lengthening in English," *Journal of Phonetics*, vol. 27, pp. 171–206, 1999.

[6] T. Cambier-Langeveld and A. E. Turk, "A cross-linguistic study of accentual lengthening: Dutch vs. English," *Journal of Phonetics*, vol. 27, pp. 255–280, 1999.

[7] S.-A. Jun and C. Fougeron,"A Phonological model of French Intonation," in A. Botinis (ed.), *Intonation: Analysis, Modeling and Technology*, Kluwer Academic Publishers: Dordrecht, pp. 209-242, 2000.

[8] J. J. Venditti, "The J ToBI model of Japanese intonation," in S.-A. Jun (ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing*, Oxford: Oxford University Press, pp. 172-200, 2005.

[9] T.-J. Yoon, J. Cole, and M. Hasegawa-Johnson, "One the edge: Acoustic cues to layered prosodic domains," *Proceedings of the 16th International Congress of Phonetic Sciences, pp. 1017-1020, August 6-10, Saarbrücken, Germany,* 2007.

[10] S. D. Khan, *Intonational Phonology and Focus Prosody of Bengali*, Dissertation, University of California at Los Angeles, 2008.

[11] H.-S. Jeon and F. Nolan, "Segmentation of the accentual phrase in Seoul Korean," *Proceedings of Speech Prosody 2010, paper number 100023:1-4, May 10-14, Chicago, USA,* 2010.

[12] A. Michelas and M. D'Imperio, "Durational cues and prosodic phrasing in French: Evidence for the intermediate phrase," *Proceedings of Speech Prosody 2010, paper umber 100881:1-4, May 10-14, Chicago, USA,* 2010.

[13] C. Gabriel, I. Feldhausen, and A. Pešková, "Prosodic phrasing in Porteño Spanish," in C. Gabriel and C. Lleó (eds.), *Intonational phrasing in Romance and Germanic: Cross-linguistic and bilingual studies*, Amsterdam: Benjamins, pp. 153-182, 2011.

[14] K. *Ueki, Prosody and Intonation of Western Cham*, Dissertation, University of Hawai'i at Mānoa, 2011.

[15] V. Abolhasani Zadeh, C. Gussenhoven, and M. Bijankhan, "A pitch accent position contrast in Persian," *Proceedings of the 17th International Congress of Phonetic Sciences*, *pp. 189-191, August 17-21, Hong Kong,* 2011.

[16] G. Kuo and C. Vicenik, "The intonation of Tongan," *UCLA Working Papers in Phonetics*, no. 111, pp. 63-91, 2012.

[17] I. Yuen, F. Cox, and K. Demuth, "Interaction of vowel length, utterance-position and focus accent in 3-year olds' Australian English," *Proceedings of the 14th Australasian International Conference on Speech Science and Technology, December 3-6, Sydney, Australia*, pp. 17-20, 2012.

[18] X. Zhang, *A comparison of cue-weighting in the perception of prosodic phrase boundaries in English and Chinese*, Dissertation, University of Michigan, 2012.

[19] Y. Xu, "ProsodyPro — A Tool for Large-scale Systematic Prosody Analysis," *Proceedings of Tools and Resources for the Analysis of Speech Prosody*, 7-10, August 30, Aix-en-Provence, France, 2013.

[20] L.-F. Lai, and S. Gooden, "What does the question sound like? Exploring wh- and yes-no interrogative prosody in Yami," *Proceedings of the 18th International Congress of Phonetic Sciences, paper number 0790:1-5*, *August 10-14, Glasgow, UK,* 2015.