INTERSPEECH 2016
September 8–12, 2016, San Francisco, USA

The Acoustic Manifestation of Prominence in Stressless Languages

Angeliki Athanasopoulou, Irene Vogel

University of Delaware
angeliki@udel.edu,ivogel@udel.edu

Abstract

Languages frequently express focus by enhancing various acoustic attributes of an utterance, but it is widely accepted that the main enhancement appears on stressed syllables. In languages without lexical stress, the question arises as to how focus is acoustically manifested. We thus examine the acoustic properties associated with prominence in three stressless languages, Indonesian, Korean and Vietnamese, comparing real three-syllable words in non-focused and focused contexts. Despite other prosodic differences, our findings confirm that none of the languages exhibits stress in the absence of focus, and under focus, no syllable shows consistent enhancement that could be indirectly interpreted as a manifestation of focus. Instead, a combination of boundary phenomena consistent with the right edge of a major prosodic constituent (Intonational Phrase) appears in each language: increased duration on the final syllable and in Indonesian and Korean, a decrease in F0. Since these properties are also found in languages with stress, we suggest that boundary phenomena signaling a major prosodic constituent break are used universally to indicate focus, regardless of a language's word-prosody; stress languages may use the same boundary properties, but these are most likely to be combined with enhancement of the stressed syllable of a word.

Index Terms: focus, enhancement, prominence, stressless language, Indonesian, Korean, Vietnamese

1. Introduction

Word level prosody is often used to classify languages in relation to stress and tone [1, 2], creating somewhat of a conundrum for languages that have neither property. In this paper, we investigate two such languages, Indonesian and Korean, as well as a third that is claimed to be stressless, but with tone, Vietnamese. Specifically, we address the questions of whether and how these languages can acoustically express (narrow) focus since it is widely observed that focus has the effect of enhancing the stressed syllable of a word, even though other elements may also show some changes. Put simply, the question is how a language that does not have lexical stress, and thus does not provide a particular syllable to enhance as a manifestation of focus, actually expresses focus.

2. Methodology: Measuring Prominence in Indonesian, Korean and Vietnamese

In order to examine the manifestation of focus in the three languages, we first examined the acoustic properties of vowels in different syllable positions within a word, in the absence of focus. Specifically, we examined the properties typically associated with the acoustic manifestation of prominence: F0 patterns, duration, intensity and vowel centralization. In this way, we were able to verify that there is, indeed, no evidence of lexical stress, or prominence, on a given syllable position. We then compared the vowel properties in the non-focus condition with those in a focus condition to determine if there is preferential enhancement of any of the syllables in a word.

2.1. Hypotheses

For each of the three languages under investigation, we first tested the prediction that if there is lexical stress, one of the three syllable positions would show consistent enhancement.

> Hypothesis 1: In the absence of focus, if lexical stress is present, one syllable position in a word will consistently show enhancement of any of the acoustic properties associated with prominence (i.e., F0, duration, intensity, vowel centralization).

As a more subtle, or indirect, way of determining whether a particular syllable is stressed, we subsequently tested the prediction that focus would enhance the acoustic properties of some syllable in a word if there is, in fact, a stressed syllable.

> Hypothesis 2: Placing focus on a word, if lexical stress is present, will result in the enhancement of the acoustic properties associated with prominence (i.e., F0, duration, intensity, vowel centralization) of the stressed syllable.

Since we measured a range of acoustic properties, it was also possible to further examine the findings in order to determine whether there were any other patterns that might serve as an indication of focus, even if they did not constitute enhancement a particular syllable.

2.2. Participants

For each language, the participants were 10 university educated speakers (approximately half female) aged 18-25. They spoke the standard dialect of Indonesian (Jakarta), Korean (Seoul) and Vietnamese (Hanoi). The data of one Vietnamese speaker were excluded due to technical issues with the recording; two more speakers were excluded due to the atypical absence of creaky phonation with the ngã tone (Tone 5).

2.3. Stimuli

To permit comparison, the same types of stimuli and structures were tested in each language. The stimuli were real three-syllable words, ideally, CVCVCV. While some non-target syllables had codas, the target vowels always appeared in an open syllable. One word could have more than one target vowel.

In Indonesian, each target vowel /i, u, a/ appeared in 10 items in each of the three syllables, as illustrated in Table 1.

http://dx.doi.org/10.21437/Interspeech.2016-1424

Table 1: *Examples of Indonesian Stimuli for vowel /a/.*

Indonesian		
Syllable 1	b**a**tako	'concrete brick'
Syllable 2	nab**a**ti	'concerning plants'
Syllable 3	bahas**a**	'language'

In Korean, each target vowel /i, o, a/ appeared in 5 items following tense, lax and aspirated onset consonants in each syllable position. We used a normalization procedure to pool the results, so the different onsets are not analyzed here. Examples of the Korean stimuli (in IPA) are given in Table 2.

Table 2: *Examples of Korean Stimuli for vowel /a/ with aspirated onset.*

Korean		
Syllable 1	[t^{h}**a**ʤagi]	'type writer'
Syllable 2	[tɛt^{h}**a**ʤa]	'pinch hitter'
Syllable 3	[wɛʤok^{h}**a**]	'maternal niece/nephew'

In Vietnamese, 8 items for each target vowel /i, u, a/ appeared with each of the two tones with rising trajectories (modal Tone 3 (*sắc*) and creaky Tone 5 (*ngã*)) in the first and second syllables; however, there were no cases of T5 with /u/ in the second syllable. In the third syllable, only /ɯ/ appeared with T5. Given the morphological system of Vietnamese, all the target words were compounds, as illustrated in Table 3.

Table 3: *Examples of Vietnamese Stimuli for vowels /a/ and /ɯ/ with tone 5.*

Vietnamese		
Syllable 1	x**ã** trưởng nữ	'female mayor'
Syllable 2	hà m**ã** cái	'female hippopotamus'
Syllable 3	vũ công n**ữ**	'female dancer'

Since the present investigation involves a comparison of the vowels in all three syllables, we only discuss the findings pertaining to T5, since this tone is the one that appears in all of the syllable positions in our data.

In order to assess the acoustic properties of focus, we compared the productions of the stimuli in both non-focus and focus contexts. Two short dialogues were used to elicit these contexts, as illustrated in English in Table 4; only the target vowels in the answers were measured.

Table 4: *Elicitation Dialogues. The focused word of each sentence is bolded.*

	Focus	Non-Focus
Q	What did Maria say in the morning?	Did Maria say "[target]" in the **afternoon**?
A	Maria said "**[target]**" in the morning.	No. Maria said "[target]" in the **morning**, not the afternoon.

The wording of the dialogues varied somewhat according to the language; however, in all cases the target appeared in the middle of the response sentence. The usual type of structure involving a target in a carrier sentence places the target in focus, as seen in Table 4. Thus, in order to have the speakers produce the target without focus, we used a carrier sentence in which a word following the target was focused. The target was not placed after the focus to avoid possible complications from post-focal compression. Given the two focus conditions, the number of targets per speaker were thus: Indonesian (N=180), Korean (N=270), Vietnamese (N=136).

2.4. Procedure

The participants were tested individually in a quiet room by a native speaker in their home country. After training and practice items, the participants read the test dialogues, alternating with distractors (common objects that had to be named), in a PowerPoint presentation. The speech was recorded to the computer on which the experiment was presented using a head-mounted microphone.

3. Analyses

3.1. Acoustic Analysis

The data were segmented and analyzed using Praat [3] for Indonesian and Korean and VoiceSauce [4] for Vietnamese. For each target vowel, measurements were made for: Duration (dur), mean F0, F0 change from beginning to end (ΔF0), Intensity (int) and Vowel Centralization (cent). For Vietnamese we added phonation measurements (H1-H2, H1-A1, H1-A2, H1-A3, harmonic-to-noise ratio (HNR), subharmonic-to-harmonic ratio (SHR), and cepstral peak prominence (CPP)) to further assess the patterns with T5, which is typically produced with creaky voice. The measurements were normalized using z-scores to permit the results to be pooled across the speakers and vowels in each language, and to be used as the basis of the statistical analysis.

3.2. Statistical Analyses

Binary Logistic Regression Analyses (BLRA [5]) were used to assess the roles of the acoustic properties, together and individually, in distinguishing among the different syllable positions and between the two focus contexts in each language.

4. Results

We present the results of the BLRAs separately for each language. The descriptive acoustic patterns that are the most important in the BLRAs, duration and F0, are presented in a combined figure for all three languages.

4.1. Indonesian

To determine whether any of the three syllable positions in the words showed prominence, we conducted BLRAs comparing each syllable with the others in the non-focus context. As Table 5 shows, the strongest overall distinction, using all of the acoustic properties, was between the first and last syllables (83%). When used as the sole classifier, F0 is the strongest distinguishing cue for both BLRAs of Syll3 (81% and 76%).

Table 5. *Indonesian BLRAs in non-focus context: pairwise syllable comparisons; chance = 50%.*

Classification	Syll1 / Syll2	Syll1 / Syll3	Syll2 / Syll3
Overall	72%	**83%**	74%
Individual Properties	dur (68%), ΔF0 (61%), F0 (54%)	**F0 (81%),** ΔF0 (63%), dur (59%)	F0 (76%), cent (60%), dur (59%)

To determine whether Focus enhances a specific syllable position, we conducted BLRAs with the Focus vs. Non-Focus contexts. As Table 6 shows, Syll3 has the best classification (slightly better than Syll2), most strongly cued by duration.

Table 6. *Indonesian BLRAs: Focus vs. Non-Focus Contexts; chance = 50%.*

Focus vs. Non-Focus Classification			
Classification	**Syll 1**	**Syll 2**	**Syll 3**
Overall	n.s.	61%	**65%**
Individual Properties	n.s.	dur (58%), cent (56%)	**dur (64%),** ΔF0 (60%), int (57%)

Figures 1 and 2 provide normalized results for the duration and F0 properties of Indonesian, and the other languages, since these most consistently appear among the strongest classifiers.

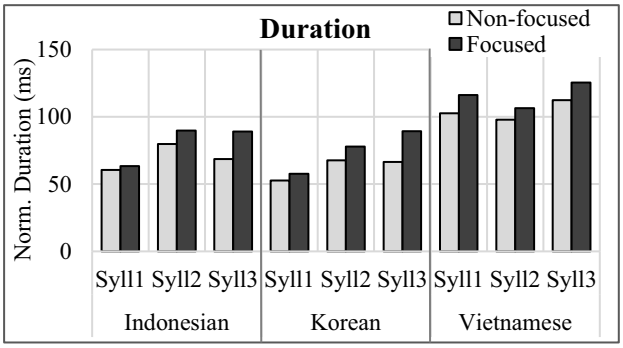

Figure 1. *Duration patterns for Indonesian, Korean, and Vietnamese.*

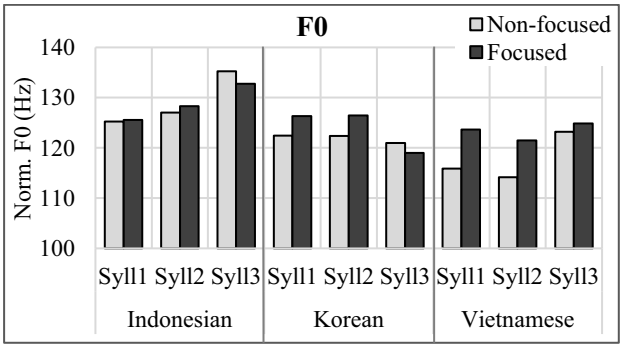

Figure 2. *F0 patterns for Indonesian, Korean, and Vietnamese*

As can be seen in Figure 1, in Indonesian, focus results in some lengthening in Syll3; and slightly less in Syll2. The fact that both Syll2 and Syll3 are lengthened, however, indicates that no single syllable is preferentially enhanced. Moreover, the fact that Syll3, without focus, shows raised F0 (Figure 2) but also shorter duration than the previous syllable, opposites with respect to enhancement, further supports the conclusion that there is no evidence for stress in Indonesian, consistent with other experimental studies [6, 7, 8, 9]. What the findings *do* suggest, however, is boundary marking at the right edge under focus, consisting of increased duration and lowered F0, compared to a higher F0 signaling continuation in the non-focus context. Together, these properties thus cue focus not by enhancement of a given syllable, but by signaling a strong prosodic boundary (e.g., Intonational Phrase) following a focused word.

4.2. Korean

BLRAs were also run with the Korean data. The normalized, pooled results for all onset types are shown in Table 7. Additional analyses found that the different onsets exhibited similar patterns overall, but unsurprisingly, the classification rates were higher for each onset type considered on its own.

Table 7. *Korean BLRAs in non-focus context: pairwise syllable comparisons; chance = 50%.*

Classification	**Syll1 / Syll2**	**Syll1 / Syll3**	**Syll2 / Syll3**
Overall	73%	**84%**	72%
Individual Properties	dur (73%), int (65%), cent (59%)	**int (73%),** ΔF0 (71%), dur (71%)	ΔF0 (69%), int (58%), cent (52%)

The overall classifications are remarkably similar to those of Indonesian, with the largest distinction, again, between the first and last syllables. Differently from Indonesian, however, where F0 stood out as the strongest single classifiers, Korean shows somewhat less consistency among the use of the cues.

Additional BLRAs determined whether Focus enhances any syllable position. As seen in Table 8, the overall results are once more quite similar to those of Indonesian. That is, Syll3 shows the best classification, followed by Syll2.

Table 8. *Korean BLRAs: Focus vs. Non-Focus Contexts; chance = 50%.*

Focus vs. Non-Focus Classification			
Classification	**Syll 1**	**Syll 2**	**Syll 3**
Overall	64%	69%	**74%**
Individual Properties	int (63%), cent (57%)	dur (63%), int (63%), cent (58%)	**ΔF0 (67%),** dur (64%), cent (58%)

Although there is more variation in the cues used to signal focus in Korean than in Indonesian, possibly due to somewhat different patterns with the different onset types, the overall classifications of both individual syllable pairs and focus vs. non-focus contexts are quite similar. In the absence of focus, the greatest distinction is between the first and last syllables, where (Δ)F0 is again a main cue, but in this case, with similar classifications also provided by intensity and duration. With regard to the manifestation of focus, as in Indonesian, Figure 1 shows that both Syll2 and Syll3 exhibit some lengthening, and Figure 2 shows a small drop in F0 on Syll3 (F0 contour is more important for Korean). Again, while the patterns do not reveal evidence for stress on any one of the syllable positions in the word, the combination of properties suggests boundary marking, with a lower final F0 and increased duration at the end of a word under focus.

4.3. Vietnamese

Table 9 provides the BLRA results for Vietnamese, showing only the comparisons involving T5, since, as mentioned, we did not have stimuli with T3 in the final syllable.

Table 9. *Vietnamese BLRAs in non-focus context: pairwise syllable comparisons; chance = 50%.*

Classification	Syll1 / Syll2	Syll1 / Syll3	Syll2 / Syll3
Overall	68%	77%	**81%**
Individual Properties	H1-A1 (65%), int (58%)	dur (69%), CPP (65%), F0 (61%)	**dur (77%)**, SHR (71%), F0=CPP (69%)

The range of overall classification rates is similar to Indonesian and Korean. As in the other two languages, too, in the absence of focus, the largest difference between syllables involves Syll3; however, in Vietnamese the difference is somewhat larger between Syll2 and Syll3, not Syll1 and Syll3.

With respect to the differences between syllables, duration is the strongest cue, followed by various phonation properties. That is, while the main difference between Syll2 and Syll3 is the longer duration of the latter, Syll3 is also distinguished by having less creaky phonation than Syll2. It is likely that the lesser role of F0 in Vietnamese compared to the other two languages is due to the fact that the rising contour is required for the contrastive value of T5 in Vietnamese, regardless of its position in a word. Moreover, since creaky phonation typically disrupts F0 [10], this may affect the role of F0 in the classifications. The stronger contribution of duration and the additional phonation properties may, however, compensate for the smaller role of F0.

The BLRAs shown in Table 10 compare the two focus conditions in each syllable to determine whether focus enhances any syllable position.

Table 10. *Vietnamese BLRAs: Focus vs. Non-Focus Contexts; chance = 50%.*

Focus vs. Non-Focus Classification			
Classification	**Syll 1**	**Syll 2**	**Syll 3**
Overall	**74%**	69%	72%
Individual Properties	**dur (66%)**, int (62%), F0 (60%)	dur (65%), cen (61%), int (58%)	dur (71%), int (62%)

Again, Vietnamese differs from the other two languages, with Syll1 showing the highest classification rate (74%); though, it is only minimally different from the rates of the other syllables. In all three positions, duration is the strongest focus cue, and as Figure 1 shows, the amount of lengthening is the same in Syll1 and Syll3, and only slightly less in Syll2. The observation that the F0 properties were not relevant for focus in Vietnamese is consistent with the fact that Tone 5 is a lexical rising tone. As such, it is not subject to manipulation for prominence as it is in the other two languages. In fact, this follows the more general observation that properties used for lexical contrasts in a given language will be avoided as cues for other phenomena in the language so as to preserve their contrastive value [11].

As in the other languages, when we take the combination of properties of Vietnamese into consideration, we find no indication of stress on any syllable position. Similarly, too, we find instead an indication of boundary marking. That is, even though focus increases the duration of all three syllables, its role in focus classification is somewhat greater in Syll3.

5. Discussion

Based on our BLRAs, Indonesian, Korean and Vietnamese all show a similar absence of acoustic properties that could be interpreted as evidence of stress on a specific syllable position. This is consistent with Hypothesis 1, which predicts that, in the absence of focus, enhancing acoustic properties will be found only if a particular syllable is stressed. The fact that we also did not find evidence of enhancement of one syllable under focus is consistent with Hypothesis 2, and a further indication of the absence of stress. That is, since focus is primarily manifested on a stressed syllable, since no syllable showed particular enhancement, none could be claimed to bear stress.

This is not to say that we found no acoustic effect of focus in the three languages. In fact, they all showed an increase in duration under focus, though it was not localized on a single syllable, as in a stress language (e.g., Greek [11]). Instead, in Indonesian and Korean, the last two syllables were lengthened; in Vietnamese, all three syllables were lengthened. The fact that the lengthening of the final syllable was coupled with a decrease in F0 in Indonesian and Korean, moreover, suggests that what is marking focus in these languages are the cues associated with a strong prosodic boundary (i.e., Intonational Phrase [12]) following the focused word. This is, furthermore, consistent with Nespor and Vogel's [12] proposal that placing focus or emphasis on a word may lead to an interruption of prosodic constituent structure, manifested by the presence of additional boundary phenomena. Vietnamese does not exhibit a similar word-final F0 decrease since T5 is a rising tone, and it is thus not free to lower to cue a boundary, as this could compromise its contrastive lexical function. It is thus possible that the greater role played by increased duration on Syll3 in Vietnamese, compared to the other languages, is a means of compensating for the lack of contribution of F0.

6. Conclusions

In sum, despite differences in other prosodic properties of the languages under investigation, Indonesian, Korean and Vietnamese, we found confirmation that all three must indeed be considered stressless. That is, in the absence of focus, they all failed to show consistent enhancement of one of the three syllable positions in our stimuli that could be interpreted as a manifestation of stress. Moreover, they also all failed to show enhancement of a particular syllable under focus which could have been taken an indirect indication of stress in that position. Instead, all of the languages exhibited a similar combination of properties on Syllable 3 consistent with the right edge of a major prosodic constituent, assumed here to be the Intonational Phrase. That is, they showed somewhat increased duration on the final syllable, and a decrease in F0, except in Vietnamese where the word-final syllable had a contrastive rising tone. Since these same boundary properties are also found in languages with stress [13, 14]), we suggest that boundary phenomena associated with a major prosodic constituent break may be used universally to signal focus, regardless of the word-prosodic properties of a language. Interestingly, a similar conclusion about the universality of the relationship between prosodic prominence and boundary phenomena is reached by Prieto et al. [15], on the basis of differences between two types of languages with stress, those considered to be syllable- vs. stress-timed. Thus, while the specific stress and rhythmic properties of languages may vary, the manifestation of major prosodic structures appears to be fundamentally constant, and thus a robust cue for prominence.

7. References

[1] L. Hyman, "How (not) to do phonological typology: the case of pitch-accent," *Language Sciences,* vol. 31, no. 2-3, pp. 213-238, 2009.

[2] L. Hyman, "Word-prosodic typology," *Phonology,* vol. 23, pp. 225-257, 2006.

[3] P. Boersma and D. Weenink, "Praat: doing phonetics by computer," 2016. [Online]. Available: http://www.praat.org/.

[4] Y.-L. Shue, P. Keating, C. Vicenik and K. Yu, "VoiceSauce: A program for voice analysis," in *Proceedings of the ICPhS XVII*, 2011.

[5] I. Vogel, A. Athanasopoulou and N. Pincus, "Acoustic Properties of Prominence and Foot Structure in Jordanian Arabic," in *Arabic Perspectives*, John Benjamins, forthcoming.

[6] E. F. Ebing, Form and Function of Pitch Movements in Indonesian, vol. 55, Leiden: CNWS Publications, 1997.

[7] E. van Zanten and V. J. van Heuven, "Word stress in Indonesian: fixed or free?," *NUSA Linguistic Studies of Indonesian and other Languages of Indonesia,* vol. 53, pp. 1-20, 2004.

[8] L. Roosman, "Melodic structure in Toba Batak and Betawi Malay word prosody," in *Prosody in Indonesian Languages*, Leiden, Leiden University Center for Linguistics, 2007, pp. 89-115.

[9] R. Goedemans and E. van Zanten, "Stress and accent in Indonesian," in *Prosody in Indonesian Languages*, Leiden, Leiden University Center for Linguistics, 2007, pp. 35-62.

[10] P. Keating, M. Garellek and J. Kreiman, "Acoustic properties of different kinds of creaky voice," in *Proceedings of the 18th International Congress of Phonetic Sciences*, 2015.

[11] I. Vogel, A. Athanasopoulou and N. Pincus, "Prominence, Contrast and the Functional Load Hypothesis: an acoustic investigation," in *Accent and Stress*, Cambridge, University Press, forthcoming.

[12] M. Nespor and I. Vogel, Prosodic Phonology, Dordrecht: Foris, 1986.

[13] S.-A. Jun, "Prosodic Typology," in *Prosodic Typology: The Phonology of Intonation and Phrasing*, Oxford University Press, 2005, pp. 430-458.

[14] S.-A. Jun, "Prosodic Typology: By prominence type, word prosody and macro-rhythm," in *Prosodic Typology II: The phonology of intonation and phrasing*, Oxford University Press, 2014, pp. 520-540.

[15] P. Prieto, M. d. M. Vanrell, L. Asturc, E. Payne and B. Post, "Phonotactic and phrasal properties of speech rhythm. Evidence from Catalan, English, and Spanish," *Speech Communication,* vol. 54, p. 681–702, 2012.