



Prominence in Japanese Is Not Only Cued Acoustically

Shinobu Mizuguchi¹ and Koichi Tateishi²

¹Kobe University, Japan

²Kobe College, Japan

mizuguti@kobe-u.ac.jp, tateishi@mail.kobe-c.ac.jp

Abstract

This paper studies focal prominence in Japanese. Japanese is a mora-timed pitch-accented agglutinative language. It realizes its accent by a falling H*L bi-tonal contour, and lexical H*L pitch accents trigger downstep. In the literature there are two contradictory views on Japanese focus. One is that Japanese focus has acoustic focal cues: F0-rise on the focused word and reduction in the post-focal areas. The other is that Japanese accent is relatively weak and is not principally used to convey discourse information. In this paper, we would like to investigate which view is correct.

We conducted a production and a perception experiment on the focal prominence on the phrase-level. The previous works on Japanese focus were mainly on Accented (A-) words. Our study covered both A-words and Unaccented (U-) words so that we could analyze the acoustic properties of Japanese focal prominence overall. We also conducted an auditory rating task of prominence on spontaneous speech in Japanese to investigate focal prominence on the discourse level.

Our findings are (i) acoustic cues of F0-rise and post-focal reduction are not strong enough to make focus prominent in Japanese, and (ii) Japanese does not highlight prominence on the discourse level much, but, when it does, boundary rephrasing plays an important role.

We hope that our findings will lead to a cross-linguistic discussion on how pitch accents function on the discourse level.

Index Terms: Japanese, focal prominence, pitch accent, spontaneous speech

1. Introduction

Speakers produce, and listeners perceive certain parts of an utterance as more or less prominent based on prosodic and syntactic highlighting. The highlighted part is called ‘focus’ in semantics, the information in the sentence that is assumed by the speaker not to be shared by him and the hearer (cf. Jackendoff [1]). Baumann and Winter [2] classify prominence cues into three types: (i) prosodic parameters, such as frequency, intensity, and duration, (ii) contrastive prosodic categories, i.e. particular pitch accent types, and (iii) non-prosodic factors. However, it is still unclear which linguistic variables have the most substantial impact on prominence; prominence cues vary among individuals and languages. English employs duration and German relies on F0 and the rising tone as prominence cues (cf. [2], Cole et al. [3]). In French and Spanish, prominence plays a lesser role in signaling information structure than in English (cf. Cole et al. [4]).

This paper considers focal prominence in Japanese. Japanese is a mora-timed pitch-accented agglutinative language. Japanese realizes its accent by a falling H*L bi-tonal contour

and has Accented (A) and Unaccented (U) words, which are lexically determined. (1) shows examples of Tokyo Japanese accent patterns, where the accent mark indicates the accent on the immediately preceding mora.

- (1) Accented (A-) words: *a'ni* ‘big brother’, *a'me* ‘rain’
Unaccented (U-) words: *ane* ‘big sister’, *ame* ‘candy’

According to the Japanese literature, focus is realized by F0-rise on the focus word and reduction in the post-focal areas (cf. Pierrehumbert and Beckman [5], Kubozono [6], Ishihara [7], among many others), as observed in Figure 1.

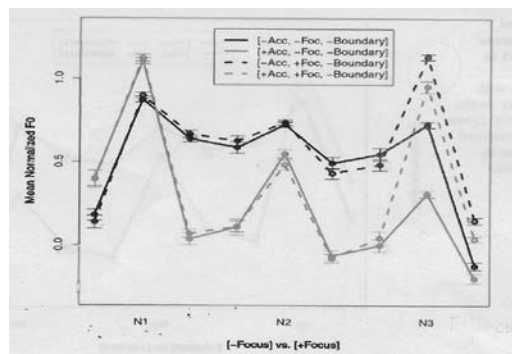


Figure 1: Mean normalized F0 in $[\pm\text{Accent}, \pm\text{Focus}]$ conditions (cited from Ishihara [8: p.1406])

This view contradicts the claim by Lee et al. [9]. They conducted production experiments, using a paradigm based on 10-digit strings, as in (2), in 7 languages (American English, Mandarin Chinese, Seoul Korean, Kyungsan Korean, Suzhou Wu, Tokyo Japanese, Standard French).

- (2) Q: Is Mary’s number 215-418-5623?
A: No, the number is 215-417-5623.

The acoustic cues of duration, intensity and F0 are strong in languages like American English and Mandarin Chinese, but not in languages like Seoul Korean and Tokyo Japanese, as shown in Table 1.

Table 1: Median z-score values of focused digits (cited from Lee et al. [9])

	Ame.English,	Man.Chinese,	S. Korean	T.Japanese
Duration	0.95	1.19	0.13	0.10
Intensity	1.28	0.36	0.24	-.24
F0	2.96	3.13	0.62	0.6

They concluded that Japanese focus is acoustically weak and is not principally used to convey discourse information.

Our research question is, ‘Which view is correct?’ To answer this question, we conducted production and perception experiments on focal prominence on the phrase-level in Section 2 and a Rapid Prosody Transcription (RPT) experiment on spontaneous Japanese speech in Section 3.

2. Experiment 1: Focal Prominence on the Phrase Level

We conducted production and perception experiments on two-noun sequences. Based on the previous theoretical studies on narrow focus (cf. [5]-[8]), we predicted that i) focus on both A and U words would be boosted, and ii) post-focal reduction would be observed in focus on A words only, due to the absence of lexical fall in U words. We also predicted that iii) A words would be easier to perceive with more acoustic cues than U words.

2.1. Production experiment

The materials of our production experiment consist of two nouns W1W2 of similar segmental characteristics; the choice of words covers the four possible accentual sequences in Japanese, UU (e.g. *amai ume* ‘sweet plum’), UA (e.g. *amai u’ni* ‘sweet urchin’), AU (e.g. *ao’i ume* ‘blue plum’), and AA (e.g. *ao’i u’ni* ‘blue urchin’). With a [+Focus] condition, one stimuli set comes in 12 conditions: UU, U[+F]U, UU[+F], UA, U[+F]A, UA[+F], AU, A[+F]U, AU[+F], AA, A[+F]A, and AA[+F]. Two different stimuli sets were used, mixed with 24 fillers, in a pseudorandomized order. We recruited five Tokyo-Japanese speakers (F2, M3). We asked the subjects to produce an answer to wh-questions like (3).

(3) a. Narrow Focus (NF)

Q: Kore-wa aoi u’ni desu-ka?
this-TOP blue sea-urchin COPULA-Q
"Is this blue sea-urchin?"

A: Iie. Ao’i ume desu.
no blue plum COPULA
"No. (It’s) blue plum."

b. Broad Focus (BF)

Q: Kore-wa nani desu-ka?
this-TOP what COPULA-Q "What is this?"

A: Ao’i ume desu.
blue plum COPULA "(It’s) blue plum."

Our materials were 48 in total and were presented by PowerPoint to the participants. Their production was recorded at 44.1KHz, 16 bits directly to a computer.

For the analysis we follow Ishihara [8], and take the normalized F0-means of the six measurement points: the 1st F0-minimum, F0-maximum, the 2nd F0-minimum of W1 (L1-1, H1, L1-2), along with those of W2(L2-1, H2, L2-2).

Figure 2 shows the result. Contra our predictions, only a few focus effects were observed; F0-rise is statistically significant only in AU[+F] and UA[+F] (W=7, p=0.026 on Mann-Whitney Test for both conditions). Statistically significant Post-focal reduction is observed only in A[+F]U (W=41, p=0.038 on Mann-Whitney Test).

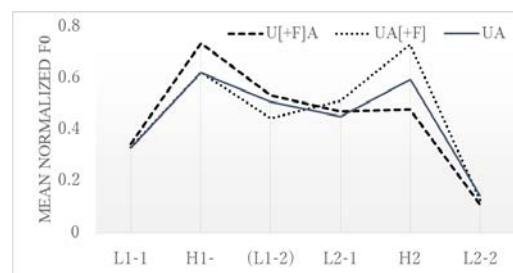
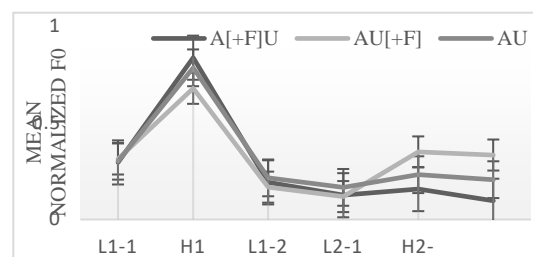
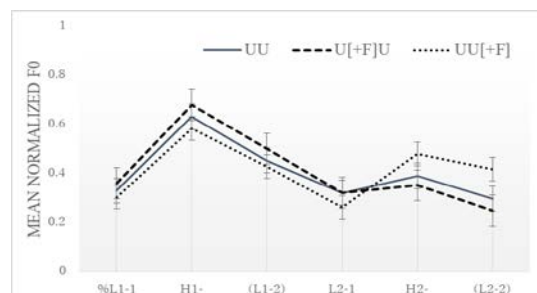
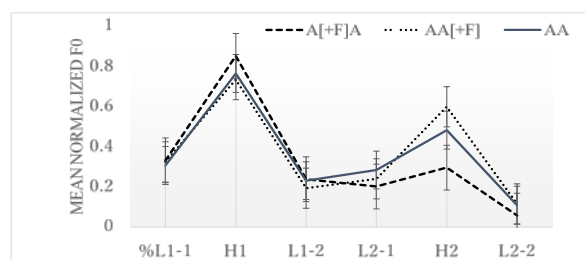


Figure 2: Mean normalized F0 of AA, UU, AU and UA (from top to bottom)

We can conclude that the focal effects in Japanese are not strong.

2.2. Perception experiment

We conducted a perception experiment of narrow focus (NF) in Japanese. We used the materials recorded in our production experiment and recruited 23 L1 Japanese (F12, M12, mean age 19.65). The task was to mark the words which they thought were focused. For broad focus (BF), the participants were instructed to mark ‘‘The whole sentence’’ as focused.

Tables 2 and 3 show the results; the focus identification ratio varies, depending on the context (cf. Table 2). Also narrow focus on A-words is identified better than that on U-words in general (cf. Table 3). However, the difference is statistically significant only in the context preceding and following U-words (cf. Figure 3).

Table 2: Confusion matrix of narrow focus and broad focus identification ratio (in percentage values)

Perceived	Target			Perceived	Target		
	A[+F]A	AA	AA[+F]		A[+F]U	AU	AU[+F]
A[+F]A	68.5	38	9.8	A[+F]U	77.2	51.1	16.3
AA	21.7	15.3	12	AU	22.8	31.5	18.5
AA[+F]	9.8	46.7	78.2	AU[+F]	0	17.4	65.2
Perceived	Target			Perceived	Target		
	U[+F]A	UA	UA[+F]		U[+F]U	UU	UU[+F]
U[+F]A	57.6	23.9	4.3	U[+F]U	64.1	37	10.9
UA	23.9	42.4	4.3	UU	26.1	21.7	13
UA[+F]	18.5	33.7	91.4	UU[+F]	9.8	41.3	76.1

Table 3: The focus identification ratio in WIW2 conditions

	Max(%)	Min(%)	Mean(%)	SD
NF on A words	91.4	68.5	76.8	16.1
NF on U words	76.1	57.6	65.8	4.6
BF	42.1	15.3	36.8	5.4

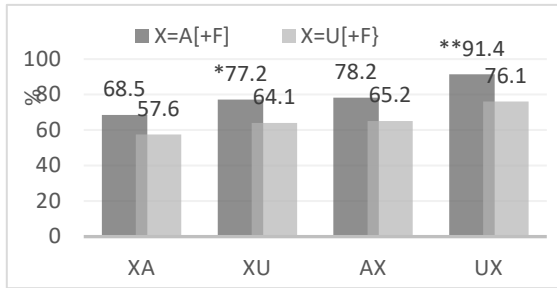


Figure 3: The narrow focus identification ratio in % (* and ** indicate significant difference $p < 0.05$ on T-test)

2.3. Discussion

Our predictions i) to iii) were not borne out; only a few cases of F0-rise and post-focal reduction were statistically significant. Also the difference was not statistically significant either between the identification ratio of the focus on A-words and that on U-words.

Our results lead us to claim that focus effects are not always realized as claimed in previous studies (cf. [5]-[8], among many others), and actual utterances may lack them. The remaining question is; if prosodic cues are not substantial in Japanese, how is prominence perceived? We are going to conduct a prominence perception experiment on spontaneous speech to answer this question.

3. Experiment 2: Focal Prominence on the Utterance Level

Cole and her colleagues [3] developed a Rapid Prosody Transcription (RPT) method to collect prosodic annotations from untrained listeners; listeners mark boundaries and prominences while listening to a recorded speech. Boundary (b-) score and prominence (p-) score ranging from 0 to 1 indicate the proportion of participants who underscore the particular word. It is a system to study inter-speaker and intra-speaker variations among a large number of participants.

We conducted an RPT experiment on spontaneous speech in Japanese. We predicted that p-scores would not be high, since acoustic focal cues are found not to be strong on the phrase level in our experiment in Section 2.

3.1. Method

We used 12 materials of the *Hanashikata Corpus*, released by the National Institute for Japanese Language and Linguistics (NINJAL). Since Japanese is an agglutative language, we segment our data-set at the morpheme level (e.g. (4)).

- (4) mazu watashi-no sukina mono inu-desu ...
 first I-GEN like thing dog-COPULA
 ‘First, what I like is dog. ...’
 (N.B. In the experiment, the text was written in Japanese *Hiragana*, i.e. Japanese cursive character)

Our material contains 980 morphemes in total, and Table 4 gives the number of each part-of-speech.

Table 4: Token distribution of part-of-speech categories (morpheme-based)

Content morphemes (532)	Function morphemes (448)
N 264, V 137, Adj 22, Adv.27, Conj 37, Numeral 27, Dem 16, WH 2	Part 177, Suf. 41, Aux 89, Top 25, CL 6, Comp 31, Neg 3, Pron 1, Quant 1, DM 60, Disfl 10, Int 4

(N.B. Part is Particle, Suf is Suffix, Top is Topic marker, CL is Classifier, Pron is Pronoun, Comp is Complementizer, Quant is Quantifier, DM is Discourse Marker, Disfl is Disfluency, and Int is Intensifier)

Out of the 980 morphemes, 232 are accented and 339 are unaccented, and the rest of the 339 morphemes are function morphemes which have no lexically-given accents.

We recruited 21 L1 Japanese speakers (F14, M7, mean age 20.9). After the exercise session, they listened to each material twice while marking prominences by a circle and boundaries by a vertical line, from a speaker in a quiet room, as in (4’).

- (4’) mazu | watashi-no sukina mono | inu-desu ...

Our materials were 18 to 41 seconds long each, and the experiment took about 30 minutes to complete.

3.2. Results

Table 5 shows the correlations between acoustic/non-acoustic factors and b-/p-scores. Acoustic factors of length, pitch and intensity are important prominence markers, and duration is a good boundary cue.

Table 5: Correlations between b-/p-scores and acoustic/non-acoustic factors (Unit: Peason’s r , ** $p < .001$)

variables	b-score	p-score
mora duration	0.562**	0.165**
mean F0	0.04	0.310**
max F0	-0.004	0.382**
range F0	0.112**	0.458**
intensity	-0.009	0.292**
log.word count	0.04	-0.183**

Cole et al. [4] regard Word Count, Part-of-Speech and Boundary as non-acoustic contextual factors of prominence. Table 5 shows that Word Count is neither a boundary nor prominence cue in Japanese. As for Part-of-Speech, Figure 4 illustrates that categories of Particle and Suffix are highlighted and given prominence in Japanese.

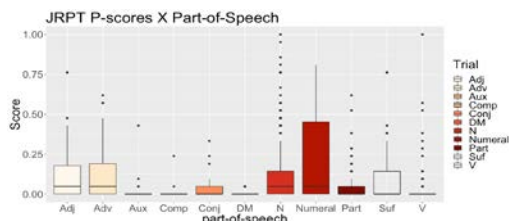


Figure 4: *P-scores X part-of-speech*

To measure agreement between listeners, we follow Cole et al. [2] and use Fleiss' kappa; Fleiss' kappa was 0.359 on p-score, and 0.638 on b-score in our RPT experiment. We observe moderate agreement in prominence marking. Note that in Japanese, prominence is not always highlighted in each Intonation Phrase (IP). Our RPT materials contain 161 IPs in total but only 65 prominence marks. Table 6 gives the number of prominence marks of Accented and Unaccented morphemes at the IP-initial, IP-mid and IP-final positions.

Table 6: *Number of prominence marks in the three IP-positions*

	IP-initial	IP-mid	IP-final	sum
Accented	28	8	2	38
Unaccented	19	5	3	27
Sum	47	13	5	65

4. Discussion

It is no surprise that Japanese uses acoustic cues of length, F0, and intensity to mark prominences. What is interesting with Japanese is that boundaries are marked more than prominences. Pinter et al. [10] conducted an RPT experiment on 12 materials from Buckeye Corpus on English native speakers and Japanese learners of English, and we have found that prominences were marked more frequently than boundaries by native speaker of English (cf. Table 7).

Table 7: *Boundary and prominence marks in English and Japanese*

	word-/morpheme-count	boundary	prominence
English (cf. [10])	644 words	94	< 111
Japanese (this study)	980 morphemes	161	> 65

Another difference between English and Japanese is that prominence is likely to be perceived at the IP-final word in English (cf. Cole et al. [4]) but in Japanese the IP-initial position is more likely to be highlighted (cf. Table 6).

We consider that boundary, a non-acoustic contextual cue, plays an important role for prominence marking in Japanese. Utterances in Japanese show downstepping, as observed in Figure 5.

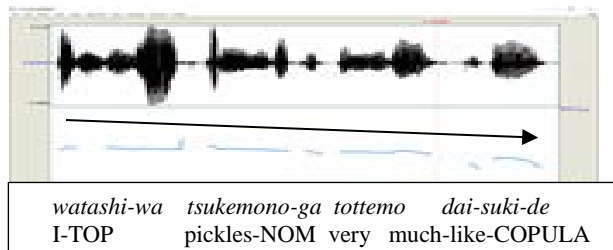


Figure 5: *Example of downstepping contour in Japanese*

The IP-initial position has the MaxF0 in each pitch contour, even without any acoustic prominence cues. To put an element at the IP-initial position is one of the tactics that Japanese realizes prominence. Japanese even allows a syntactic constituent to be divided into different IPs, as in (5).

- (5) [IP₁ ... tairyō-no] [IP₂ takuan-o ...] 'lots of pickled radish'
lots of-GEN pickled radish-ACC

In (5), 'lots of pickled radish' is a syntactic constituent of NP, but the speaker inserts pause between *tairyō-no* 'lots of' and *takuan* 'pickled radish' to make a new IP so that *takuan* is highlighted at the IP-initial position. The IP-initial position is endowed with the MaxF0 within the IP by position, and is given prominence without acoustic cues. This kind of 'rephrasing' is sufficient to start a new IP. This is probably why boundary marks exceed prominence marks in Japanese (cf. Table 7).

In the IP-mid position and the IP-final position, on the other hand, Japanese rises F0 to highlight prominence. Figure 6 is an example with the F0-rise at the IP-final position. The utterance is not a question but a statement and the speaker highlights a bound morpheme *ha* 'party' here.

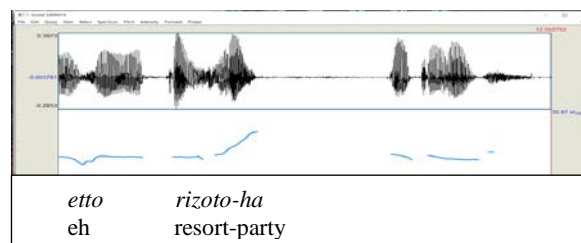


Figure 6: *Example of F0-rise at the IP-final position*

Our research question was to investigate which view is correct, the strong acoustic view or the weak acoustic view. We answer that neither is correct; in Japanese, acoustic cues are not strong enough to realize prominence by themselves. Japanese rephrases syntactic constituents to generate IPs and puts the rephrased element at the IP-initial position to be perceived as prominent. In conclusion, Japanese uses acoustic and non-acoustic cues to highlight prominence.

5. Conclusion

Based on production and perception experiments, we have found that (i) Japanese acoustic focal cues are not strong enough to highlight prominence by themselves, and (ii) Japanese rephrases constituents and puts the rephrased element at the IP-initial position, where the pitch is highest within the IP, due to Japanese downstepping pitch contour.

We hope that our observation will contribute to the much-discussed issue whether prosodic cues play a secondary role to morphological and syntactic marking in tone languages or not (cf. Fielder and Janney [11] for Ewe), Downing [12] for Chichewa, Wang et al. [13] for Chinese, among others).

6. Acknowledgements

We would like to express our gratitude to the experiment participants at Kobe University and Kobe College. We also thank Tim Mahrt for his analysis, comments, and suggestions on the earlier version of this paper.

7. References

- [1] R. Jackendoff, *The Semantic Interpretation in Generative Grammar*, The MIT Press, 1972.
- [2] S. Baumann and B. Winter, "What makes a word prominent? Predicting untrained German listeners' perceptual judgements," *Journal of Phonetics*, Vol.70, pp.20-38, 2018. <https://doi.org/10.1016/j.wocn.2018.05.004>
- [3] J. Cole, Y. Mo, and M. Hasegawa-Johnson, "Signal-based and expectation-based factors in the perception of prosodic prominence," *Language and Cognitive Processes*, Vol.1, pp. 425-452, 2010. <https://doi.org/10.1515/labphon.2010.022>.
- [4] J. Cole, J. I. Hualde, C. L. Smith, C. Eager, T. Mahrt and R. N. de Souza. Sound, structure and meaning "The bases of prominence ratings in English, French and Spanish." *Journal of Phonetics*, Vol.75, pp.113-147, 2019. <https://doi.org/10.1016/j.wocn.2019.05.002>.
- [5] J. Pierrehumbert and M. Beckmann, *Japanese Tone Structure*, The MIT Press, 1986.
- [6] H. Kubozono. *The Organization of Japanese Prosody*, PhD, University of Edinburgh, 1988.
- [7] S. Ishihara, *Intonation and Interface Conditions*, PhD, MIT, 2003.
- [8] S. Ishihara, "Japanese downstep revisited," *Natural Language and Linguistic Theory*, Vol.34, pp.1389-1443, 2016. <https://doi.org/10.1007/s11049-015-9322-8>
- [9] Y-C. Lee, B. Wang, S. Chen, M. Adda-Decker, A. Amelot, S. Nambu, and M. Liberman. "A crosslinguistic study of prosodic focus". In *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference*, pp.4754-4758, 2015. <https://doi.org/10.1109/ICASSP.2015.7178873>.
- [10] G. Pinter, S. Mizuguchi and K. Tateishi. "Perception of prosodic prominences and boundaries by L1 and L2 speakers of English," *Proceedings of InterSpeech 2014*, pp.544-547, 2014.
- [11] I. Fiedler and S. Iannedy, "Prosody of focus marking in Ewe," *Journal of African Languages and Linguistics*, 34(1), pp.1-46, 2013. <https://doi.org/10.1515/jall-2013-0001>.
- [12] L. J. Downing, "Focus and prominence in Chichewa, Chitumbuka and Durban Zulu," *ZAS Papers in Linguistics*, 49, pp.47-65, 2008.
- [13] T. Wang, J. Liu, Yong-hun Lee and Yong-cheol. Lee. "The interaction between tone and prosodic focus in Mandarin Chinese," *Language and Linguistics* 21:2, to appear.