Speech Prosody 2004
Nara, Japan
March 23-26, 2004

Rhythmic Boost and Recursive Minor Phrase in Japanese

Takahito Shinya, Elisabeth Selkirk and Shigeto Kawahara

Department of Linguistics
University of Massachusetts, Amherst
{tshinya; Selkirk; kawahara}@linguist.umass.edu

Abstract

Kubozono [1] found that in sequences of four accented words in a syntactic-phrase-internal uniformly left-branching (LB) structure, the f0 peak of the third word is realized at the same height as or higher than the preceding word, showing no apparent catathesis (or downstep). He argues for an f0 raising effect, called *rhythmic boost,* that is the consequence of the organization of this 4-word syntactic structure into (recursive) two prosodic *minor phrases* (MiP, aka accentual phrase) that branch into two MiPs each. In this paper we report on two experiments. Experiment 1 demonstrates the rhythmic boost effect experimentally, giving solid evidence for Kubozono's claim about boost in these sequences. We find that the rise at the third word in LB sequences of four accented words is significantly higher than at the third accented word of 3-word phrase-internal LB sequences, and we find too that this rise is also significantly lower than the one found with a third word that initiates a syntactic maximal projection (XP) and would therefore be at the edge of prosodic *major phrase* (MaP, aka intermediate phrase). Experiment 2 investigates a possible influence of length of MiP on the reported boost in f0, to see if that boost could be derived from the *anticipatory length-based f0 raising* (ALR) effect found by Selkirk et al. [7]. We find that that ALR effect cannot explain the magnitude of the rise at the third noun in the 4-noun sequences, and conclude that there is indeed a place for a branching-sensitive rhythmic boost in f0.

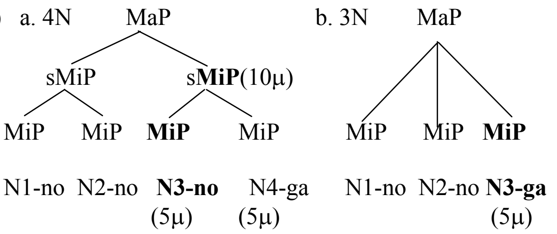
10.21437/SpeechProsody.2004-81

1. Introduction

In the past twenty years of research on Japanese intonation, a large body of work has accumulated on f0 downtrends during the course of an utterance [1, 3, 4, 8]. One interesting phenomenon first reported by Kubozono [1] is rhythmic boost, a disturbance in an expected downtrend. In sequences of four accented words in a syntactic phrase with uniformly LB structure, he shows that the f0 peak of the third word is realized as high or higher than the peak of the second word. This is an unexpected effect when considering that sequences of accented words in that same type of structure are held to undergo successive catathesis and yield a gradually descending staircase pattern [3, 4].

Accented nouns in sequence in Japanese each form a MiP, marked by the presence of the initial LH rise called "initial lowering." Kubozono proposes that the f0 boost on the third noun (N3) in a LB four accented sequence is the consequence of the prosodic "restructuring" of the accented N3 and N4 in such LB structures into a (superordinate) MiP, call it *sMiP*. This restructuring creates a recursive and branching MiP structure within the MaP that groups the LB four-word sequence, as in (1a). Kubozono claims that a phonetic implementation rule raises the pitch at the left edge of a right-branching structure, therefore at the edge of the second sMiP structure coinciding with the left edge of N3 in (1a). Rhythmic boost would not be present at N3 in the structure (1b), since there is no left edge of a right-branching constituent there.

(1)



The sMiP structure in (1a) could be understood as the consequence of a prosodic structure markedness constraint favoring binary-branching, such as Kubozono's *Principle of Rhythmic Alternation*. The constraint would force a 4-MiP sequence into a binary sMiP-sMiP sequence, giving a sMiP edge at N3. N3 in the 3-MiP structure would never fall at the edge of a right-branching sMiP. It would be organized as in (1b), or, if two nouns in the three-noun target sequence were paired in a sMiP, it would either be final in a sMiP or to the right of one — $\{$N1 (N2 N3 $)_{SMiP}\}_{MaP}$ or $\{($ N1 N2 $)_{SMiP}$ N3$\}_{MaP}$.

In the first part of this paper, we experimentally (re)demonstrate this rhythmic boost effect by showing that the f0 values for N3 in the 4-MiP structure, putatively (1a), are significantly higher than in the 3-MiP structure. We also show that the f0 values for N3 in (1a), are nonetheless not as high as the values for N3 when it falls at the left edge of a syntactic XP (cf. (2a)). The interpretation we make of this is that the four-accented-noun syntactic-phrase-internal sequence is indeed all contained within a MaP, that within that MaP there is catathesis on successive accented words, but that the effects of catathesis on N3 are mitigated by the boost appearing at the edge of the complex sMiP. The magnitude of this boost is not enough to make the effects equivalent to the upward pitch resetting that is found at the right edge of MaP [3, 5, 6].

The second goal of this paper is to determine whether the boost in f0 seen at N3 in the four noun phrase could be attributed to another independently attested effect, rather than to a branching-sensitive rhythmic boost. Selkirk et al. [7]. show that there is a mora-counting anticipatory length-based f0 raising (ALR) effect at the MiP level whereby longer MiPs exhibit greater initial rise and higher peak values. Given the recursive MiP structure described above, it could be hypothesized that the f0 peak values of N3 in (1a) are higher than those of N3 in (1b) because the righthand upper sMiP in the former is longer in mora count (by 5 moras (5μ)) than the last MiP in the latter

In Experiment 2, however, we show that only simplex MiPs, and not the upper sMiP show the ALR effect.

2. Experiment 1

2.1 Experimental materials

The structure of the experimental sentences is given in (2). Each contained seven words, began with a sentence adverb and ended with a verb. The accented 4-word target sequences consisted of (a) a 2-noun subject noun phrase (2N) followed by a 2-noun accusative noun phrase, (b) a 3-word LB subject phrase (3N) followed by an accusative phrase with N4 as the first member, and (c) a 4-noun LB subject phrase (4N). Postposition is indicated by a "p".

(2)
a. 2N
[[Adv] [[**N1-no**] **N2-ga**] [[[N3-no] N4-o] [[N5-p] Verb]]]
{ }{ }{ }{ }=MaP

b. 3N
[[Adv] [[[**N1-no**] **N2-no**] **N3-ga**] [[[N4-no] N5-o] Verb]]
{ }{ }{ }=MaP

c. 4N
[[Adv] [[[[**N1-no**]**N2-no**]**N3-no**]**N4-ga**] [[N5-p] Verb]]
{ }{ }{ }=MaP

All nouns consisted of 5 CV syllables/moras with lexical accent on the second mora. Thus two accents were separated by 4μ so that we could identify the L following the accent H and the boundary L. The prosodic structure of the experimental sentences can partly be predicted from their syntactic structure: It is shown that a MaP boundary occurs at the left edge of a syntactic XP [5, 6]. A MaP boundary is expected between N2 and N3 in the 2N case, between N3 and N4 and between N4 and N5 in the 3N and 4N cases, respectively. Three different lexical sets were prepared for each set of noun phrases in order to avoid particular lexical properties influencing the f0 patterns.

2.2 Speakers

Three native speakers of Tokyo Japanese (KK, RH, SS) were recruited from the University of Massachusetts community. They were all female and in their mid twenties.

2.3 Procedures and analysis

In order to solicit natural utterances, the experimental sentences were each embedded in narratives, which were read as a whole by the speakers from index cards in Japanese orthography. Another set of narratives served as fillers. Recording sessions were carried out on two different days in a sound-attenuated booth. The total number of sentences was 54 (3 phrase types × 3 lexical sets × 3 repetitions × 2 sessions), with 18 tokens corresponding to each phrase type. The sentences were recorded to CD, digitized with a 11.025 KHz sampling rate and 16 bit quantization level, and submitted to F0 measurement using *PitchWorks* (Scicon R&D).

F0 peaks and valleys that appeared in each MiP were measured, with the assumption that the peak values represent the accent-related H in a MiP and the valley values the boundary L at the left edge of a MiP. For the measures of comparison we used the amount of initial rise at the left edge of N3 (H3-L3) and that of the descent from N2 to N3 (H2-H3).

2.4 Results and Discussion

For all three speakers, the values of H3-L3 in the 4N cases were distinctively higher than those in the 3N cases and were also remarkably lower than those of H3-L3 in the 2N cases (Fig 1 and Fig 2a, 2c & 2e). The values of H2-H3 for the 4N cases show the same pattern: N3 in 3N shows a greater descent than in 4N, and N3 in 2N much smaller descent than in 4N (Fig 1 and Fig 2b, 2d & 2f).

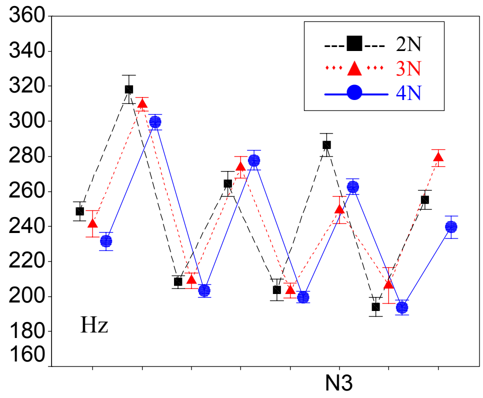


Figure 1 *Means of peaks and valleys of the first 4 nouns for speaker KK. Error bars represent 95% confidence intervals.*

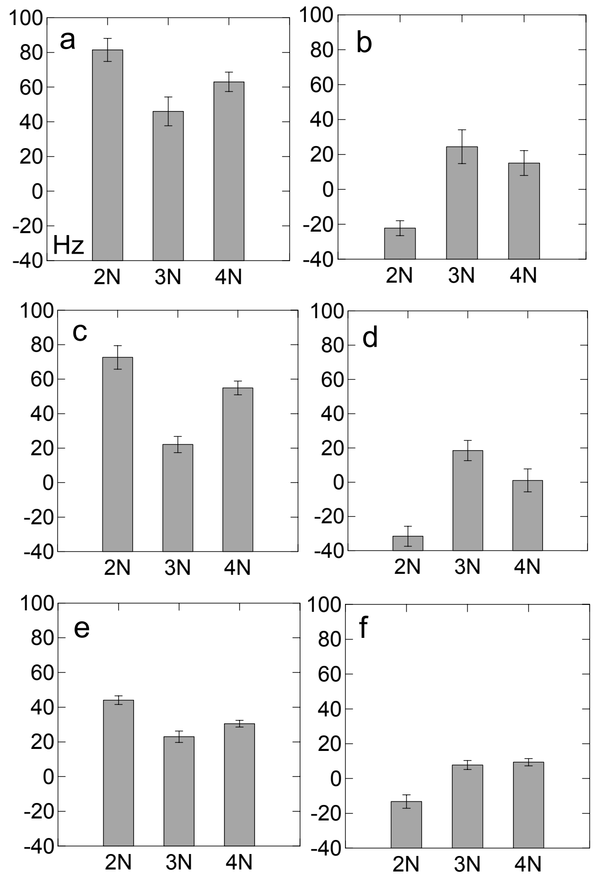


Figure 2 *Means of H3-L3 (a, c, e) and of H2-H3 (b, d, f) for N3 in the 2N, 3N and 4N cases. Graphs (a) & (b) for KK, (c) & (d) for RH, and (e) &(f) for SS.*

In order to assess the statistical significance of these differences, the data obtained from each speaker were submitted to a one-way ANOVA with phrase types as the independent variable. The results were highly significant for both measures for all speakers ([H3-L3]: $F(2,51) = 93.84$, $p<.0001$ for KK; $F(2,51) = 30.79$, $p<.0001$ for RH; $F(2,46) = 95.21$, $p<.0001$ for SS. [H2-H3]: $F(2,51) = 91.12$, $p<.0001$ for KK; $F(2,51) = 50.54$, $p<.0001$ for RH; $F(2,46) = 88.59$, $p<.0001$ for SS). A post-hoc multiple comparison on H3-L3 shows that for all speakers 4N cases are significantly different from the 2N and 3N cases, respectively (p values from .002 to <.0001, with the Bonferroni procedure applied, so α=.017 and α=.003 for the .05 and .01 levels, respectively.). Significant results were obtained on the measure of H2-H3 for all speakers, with insignificant results for 4N vs 3N of RH and SS ($p = .189$ for RH and $p = .951$ for SS).

The results described above provide solid experimental evidence for an f0 boost effect at N3 in the 4N structure. N3 in utterances containing the 4N target structure is significantly higher than N3 in the 3N structure. This gives support for the claim that the MiPs appearing in the third position of four LB words do not have the same prosodic structure as the MiPs that appear in the last position of utterances consisting of three LB nouns. It can be taken as supporting the presence of the upper sMiP nodes in (1a). Our results also support Kubozono's claim that catathesis still appears between the second and third words the 4N structure, despite the appearance of that extra f0 boost on N3. This may be concluded from the fact that the f0 peak values are substantially lower for N3 in the 4N structure than in the 2N structure. N3 in the 2N structure appears at the left edge of a syntactic XP, hence at the edge of a MaP [5]. The rise in that MaP-edge context exhibits upward pitch resetting; any catathesis effect is nullified. The significantly lower values for N3 in the 4N structure therefore suggest that it is still within a MaP, and as a consequence, still within the domain of catathesis. This indicates that the f0 values on N3 here are the result of the interaction of catathesis and the boost effect.

3. Experiment 2

Kubozono [1] accounts for rhythmic boost by positing a recursive prosodic structure where the first two and last two simplex MiPs in the 4N structure each form a higher sMiP-level branching constituent. The third simplex MiP in that structure is found at the left edge of a right-branching structure, and Kubozono's proposal is that this branching triggers extra f0 boost on the third MiP. In Experiment 2, we examine the possibility that the upper sMiP is the domain of the ALR effect found by Selkirk et al. [7] for simplex MiPs. If the upper sMiP serves as the domain of the ALR effect, then we might be able to do away with the rhythmic boost effect itself. In so doing we would remove support for the notion that it is the branchingness that has an effect on f0 realization.

3.1 Experimental materials

The experimental materials were similar to those in Experiment 1. They contained LB subject noun phrases consisting of sequences of three and four accented nouns. The first noun phrase, seen in (3a), consisted of three nouns—all 5μ-long nouns—and is referred to as "3N5". The second three-noun phrase, in (3b), differed from the 3N5 set in that N3 was 7μ long instead of 5μ (= "3N7"). The remaining target noun phrase consisted of three 5μ nouns and a final 7μ noun ("4N7", as in (3c)).

(3)

a. 3N5

[[[[**N1-no**] **N2-no**] **N3-ga**] [[N4-ni] [[N5-o] Verb]]]

{ 5μ }{ }{ }=MaP

b. 3N7

[[[[**N1-no**] **N2-no**] **N3-ga**] [[N4-ni] N5-o] Verb]]]

{ 7μ }{ }{ }=MaP

c. 4N7

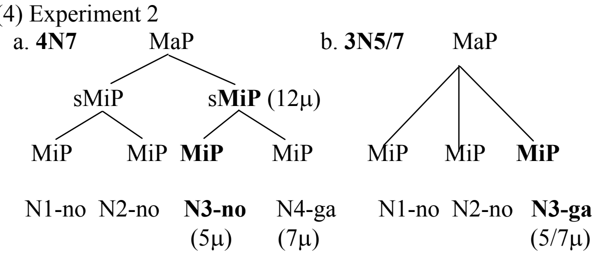
[[[[[**N1-no**]**N2-no**]**N3-no**]**N4-ga**] [[N5-o] Verb]]]

{ 5μ 7μ }{ }=MaP

The overall number of words in a sentence was controlled so that each sentence had six elements (five nouns followed by a verb). All of the 5μ nouns had pitch accents on the second mora while with the 7μ noun it was on the third mora. The reason for this difference is that there were not many 7μ words in Japanese with accent appearing on the second mora. Just as in Experiment 1, three different lexical sets were prepared for each phrase type.

A comparison between the f0 values at N3 in the 3N subject cases 3N5 and 3N7 will permit us to see if the ALR effect is attested in these materials as well, cf. (4b). A comparison between the f0 values at N3 in the 4N case 4N7 and the f0 values at N3 in 3N5 and 3N7 will allow us to see if the boost expected at N3 in 4N7 is of the order of magnitude expected on the basis of the 12μ total length of the putative sMiP posited for that structure, as in (4a).

(4) Experiment 2



3.2 Speakers

The speakers of Experiment 2 were KK and RH from Experiment 1.

3.3 Procedure

The basic procedures were identical to Experiment 1. In Experiment 2 only one recording session took place for each speaker. Unlike Experiment 1, the experimental sentences were not embedded in narratives. Also, fillers were not used. There were 6 repetitions for each sentence type/lexical choice pair. The total number of experimental sentences was 54 (3 phrase types × 3 lexical sets × 6 repetitions).

3.4 Results and Discussion

In the schematic f0 contours in Figure 3 we can see that for both speakers the mean f0 peak values for N3 are lower in the 3N5 case than for phrase types 3N7 and 4N7. Moreover the peak values for 3N7 and 4N7 show similar f0 values. As for the f0 values of the rise at N3, Figure 4 also shows that the same trends are observed: the initial rises in 3N7 and 4N7 show a comparable pattern but those in 3N5 are distinctively lower. These observations suggest that the 3N7 and 4N7 cases are intonationally indistinguishable with respect to the rise appearing at the left edge of the third MiP.

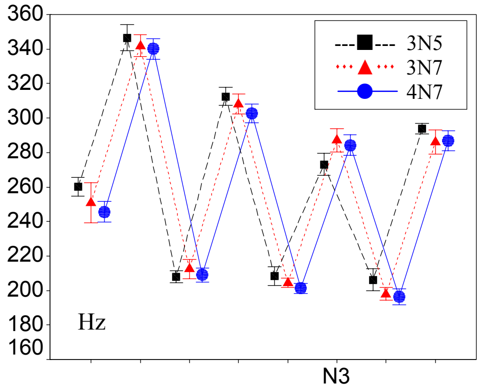


Figure 3 *Means of peaks and valleys of the first 4 nouns for speakers KK..*

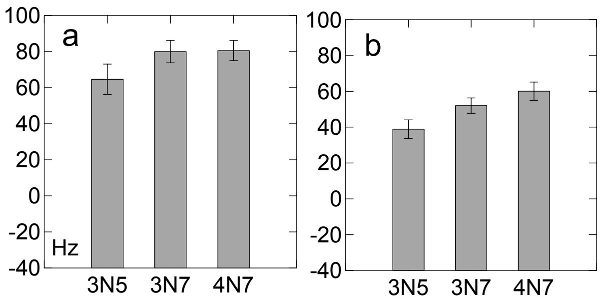


Figure 4 *Means of rise at N3 (H3-L3) in the 3N5, 3N7 and 4N7 cases for KK (left panel) and for RH (right panel).*

Just as with Experiment 1 a one-way ANOVA was performed with phrase type as the independent variable. The results show that for all speakers there was significant effect for phrase type ([H3-L3]: $F(2,51) = 9.62$, $p=.0003$ for KK; $F(2,51) = 11.53$, $p<.0001$ for RH; $F(2,51) = 4.01$, $p<.024$ for SS. [H2-H3]: $F(2,50) = 16.69$, $p<.0001$ for KK; $F(2,50) = 20.10$, $p<.0001$ for RH). The data were further analyzed using a post-hoc comparison test, (with the Bonferroni correction, $\alpha=.017$ and $\alpha=.003$ for the .05 and .01 significance levels, respectively). It shows that for both speakers there was a significant difference between 3N5 and 3N7 (p<.001 for KK, p<.003 for RH) and 3N5 and 4N7 (p<.001 for KK, p<.001 for RH), but no significant difference is found between 3N7 and 4N7 (p<1.0 for KK, p<.044 for RH).

As expected, we have found an ALR effect in comparing the values for the simplex MiPs in 3N5 and in 3N7. The initial rise is higher for the latter than the former, replicating the results obtained by Selkirk et al. [7]. The question now is whether the grammar has both an ALR effect and a rhythmic boost effect at the edge of a prosodic right-branching structure like the putative sMiP, or whether the rhythmic boost effect may be dispensed with in favor of an ALR effect, defined both on simplex MiP and on the upper sMiP as well. Our evidence suggests that the ALR effect is *not* defined over the domain of the upper sMiP, and so suggests that there must be a branching-sensitive rhythmic effect as well as an ALR effect at play. There was no significant difference between the f0 values of N3 in the 3N7 and 4N7 cases. If an ALR effect were defined over the 12μ sMiP that N3 initiates in (3c) (cf. (4a)), the f0 values in the 4N7 case should be significantly greater than for the 3N7 case. Since the values were not different, there can be no evidence for computing ALR on the longer and higher sMiP. Instead, we are left with a theory that includes an ALR effect only on simplex MiPs and a theory of rhythmic boost at the left edge of the branching higher sMiP.

Evidence for the presence of both these effects in the grammar in fact comes from the indistinguishability of the f0 values on N3 in the 3N7 and 4N7 cases, which are claimed to have the contrasting prosodic structures in (3). This indistinguishability might seem surprising in view of the results from Experiment 1, in which the f0 values for N3 in the 4N cases were significantly larger than in the 3N cases (cf. 1a and 1b), due to the rhythmic boost effect. Why is rhythmic boost not creating a contrast between the 3N7 and 4N7 cases in Experiment 2? The answer lies in the length of N3 in these cases. In 3N7, N3 is 7μ long, while in 4N7, N3 is only 5μ long. The ALR effect, defined over the length of the simplex MiPs containing N3, would predict that the longer MiP in 3N7 would be realized with higher f0 values. The rhythmic boost effect, defined over the right-branching structure of the sMiP seen in 4N7, would, on the contrary, predict that N3 in the 4N7 case should have higher f0 values. The fact that there is no significant difference at all in these cases suggests that the ALR effect and the rhythmic boost in effect simply cancel each other out. This allows us to conclude that the magnitude of the rhythmic boost effect is analogous to the magnitude of the ALR effect of a 2μ difference in MiP length.

*Acknowledgements: The work reported on this paper was supported by NSF grant BCS-0004038 to Elisabeth Selkirk.

4. References

[1] Kubozono, H., 1989. Syntactic and rhythmic effects of downstep in Japanese, *Phonology* 6, 39-67.

[2] Ladd, D. R., 1986. Intonational phrasing: the case for recursive prosodic structure, *Phonology Yearbook* 3, 311-340.

[3] Pierrehumbert, J. B; Beckman, M. E., 1988. *Japanese Tone Structure*. Cambridge, Mass: MIT Press.

[4] Poser, W. J., 1984. *The phonetics and phonology of tone and intonation in Japanese.* Doctoral Dissertation, MIT.

[5] Selkirk, E.; Tateishi, K., 1991. Syntax and downstep in Japanese, In Georgopoulos, C. and R. Ishihara (eds.) *Interdisciplinary Approaches to Language: Essays in Honor of S.-Y. Kuroda.* Dordrecht: Kluwer Academic Publishers, 519-43.

[6] Selkirk, E.; Shinya, T.; Sugahara, M., 2003 Degree of Initial Lowering in Japanese as a Reflex of Prosodic Structure Organization. *Proc. 15th ICPhS,* Barcelona, Spain, 491-494.

[7] Selkirk, E., Shinya, T.; Kawahara, S., 2003. *Phonological and phonetic effects of minor phrase length on F0 in Japanese*, Proc. *Speech Prosody 2004*, Nara, Japan.

[8] Sugahara, M., 2003. *Downtrends and Post-FOCUS Intonation in Tokyo Japanese.* Doctoral Dissertation, University of Massachusetts, Amherst.