Final Lowering and Boundary Pitch Movements in Spontaneous Japanese

Kikuo Maekawa [1]

[1] Department of Corpus Studies, National Institute for Japanese Language and Linguistics, Tokyo
kikuo@ninjal.ac.jp

Abstract

Standard theory of the prosodic structure in Tokyo Japanese treats both the final lowering and boundary pitch movements as the properties of utterance node. Validity of this treatment was examined by means of corpus-based analyses of spontaneous speech. The results showed that while final lowering could be treated as a property of utterance, boundary pitch movement could not. The latter should rather be treated as the property of accentual phrase. Based on these results, revised prosodic structure and annotation scheme were proposed.

Index Terms: final lowering, CSJ, X-JToBI, BPM

1. Introduction

Tokyo Japanese is one of the languages whose prosodic structure was the most thoroughly examined both theoretically and experimentally [1-3]. Most of the theories proposed for the language presuppose hierarchical prosodic structure as represented by the study of Pierrehumbert and Beckman [3] (P&B hereafter). The prosodic structure proposed in P&B involved, at least, three important nodes: utterance (Utt, the highest node), intermediate phrase (IP, the domain of catathesis or downstep), and, accentual phrase (AP, the domain of phrase-level accent manipulation and initial pitch rise). See the left panel of Fig. 4 below.

Among these, the most problematic is the Utt. The P&B theory stated that Utt was concerned with 3 distinct prosodic phenomena, i.e., F0 declination, final lowering, and boundary pitch movements (BPM). The aim of the present study is to examine the validities of the latter two possibilities based on the analysis of spontaneous speech. The examination of F0 declination will be the theme of an independent paper.

2. Data

2.1. CSJ-Core

Corpus of Spontaneous Japanese (CSJ) [4-6] was used in the present analysis. More specifically, it was the phonetically annotated part of the corpus called the CSJ-Core consisting of 44 hours of speech and involved a half million words that was analyzed in this study.

CSJ, and the CSJ-Core, involves four different types of speech, viz., APS, SPS, Dialogue, and Reproduction. APS, or academic presentation speech, is the live recording of various academic presentations. SPS, or simulated public speaking, is the extemporaneous speech done by laymen subjects on various every-day topics. The dialogues in the CSJ are mostly concerned with the interview on the contents of APS or SPS. Reproduction is the read-aloud of the transcription of APS or SPS done by the same speaker. In this paper, reproductions are excluded from the analysis, except in Table 1.

2.2. X-JToBI

The CSJ-Core was phonetically annotated both in terms of segments and prosody using the X-JToBI scheme [7]. Among the labels of X-JToBI, it was the tone labels for the boundary pitch movements and the BI (break indices) labels that were relevant in the present study.

BPM is the local characteristic of intonation that marks the end of an prosodic constituency. 4 BPMs, viz., ‘L%H%’ (rising), ‘L%HL%’ (rising-falling), ‘L%LH%’ (rising with extended low), ‘L%HLH%’ (rising-falling-rising) are recognized in the X-JToBI annotation, but the L%HLH% was not analyzed in this study because of too few occurrences. Note also that the falling intonation (L%) is not recognized as BPM.

The characteristic of the X-JToBI BI system consists in the use of intermediate values between the traditional values of 2 and 3 (as well as 1 and 2). The necessity of intermediate BI values will be shown in sections 4 and 5.

2.3. Clause boundary labels

Another important annotation is the clause boundary labels (CBL). As is well known, Japanese is a typical head-final language. Japanese clauses end, in principle, with a verb, adjective, or noun (including so called the stem of adjectival verbs) followed by a copula. Although these elements are modified by various particles and/or adverbs, it is possible to classify the degree of syntactic independence of clauses by means of morphological characteristics of the head of clauses. The CBL of CSJ was generated automatically by the CBAP-csj program [8].

As long as the CSJ-Core is concerned, detected clause boundaries and the attached CBLs were checked and, if necessary, modified by expert labelers. The CBAP-csj program recognizes 49 different clause types and classifies them into three classes, viz., ABS(olute), STR(ong), and WEAK.

The ABS class consists of typical sentence-ending expressions that introduce the deepest syntactic boundary. The STR class boundaries are not sentence-ending expressions, but are concerned with relatively deep boundaries introduced by various conjunctive (mostly juxtaposition) particles including /ga/, /keredo/, /keredomo/, /kedomo/, /kedo/, /si/, and /youni/. Boundaries of this class can sometimes end an utterance.

The WEAK boundaries are much weaker than the STR boundaries in that they do not usually end an utterance, although there are some rare cases where an utterance ends with a WEAK boundary. The WEAK boundary includes the forms like /tara/, /to/, /nara/, /reba/, /kara/, /node/, /tari/, /te/, /toka/, /noni/, /toiu/, /de/, etc. Conjunctions and interjections are classified into this class.

Note that all prosodic boundaries other than the three mentioned above are classified as the NONE boundary in terms of CBL in the present analysis. The NONE boundaries do not occur at the end of an utterance.

2.4. F0 measurement points

Fig. 1 shows the F0 measurement points used in this study. The upper panel of Fig.1 shows the schematized F0 contour of an AP ending in a rising BPM, with (real line) and without (broken line) accent. Upper and lower boxes below the contour show respectively the X-JToBI labels and the names of F0 measurement points used in this study. ILT stands for initial low tone, IHT is for initial high tone, ACC is for accent, FLT is for final low tone, and, FHT is for final high tone. No ACC value is specified in the case of unaccented AP.

The lower panel of Fig.1 shows the case of initially accented AP ending with an rising-falling (L%HL%) BPM. Note that in the case of initially accented AP, it is impossible to make distinction between the H- tone and accent. In this study the initial peak in the F0 contour of initially accented AP is treated simply as an IHT. Note also that, in the case of rising-falling BPM, FLT and HLT stand respectively for the beginning and end (peak) of the rising part of the BPM. The last tone (FLT') will not be analyzed in the present study.

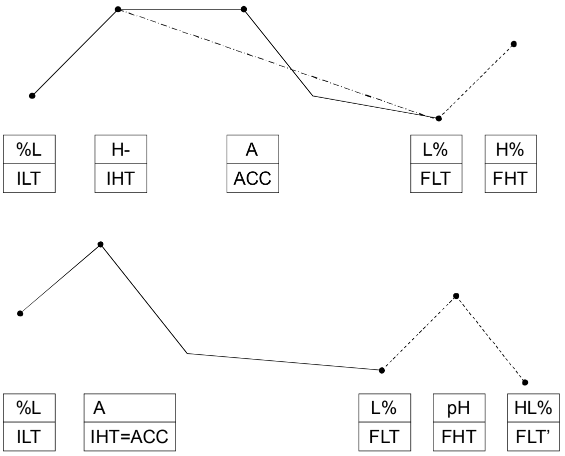

Figure 1: *F0 measurement points in this study.*

3. Analysis of Final Lowering

3.1. Existence of final lowering

Existence of final lowering in spontaneous speech is examined by comparing the mean FLT value across the 4 CBL classes. Table 1 shows the mean FLT values (in Hz) as a function of the sex of speaker, accentedness of AP (Accented and Unaccented), and 4 CBL classes (including NONE). The column named 'FLT' shows the overall means of spontaneous speech (APS, SPS, and dialogue), while the 'FLT-R' column showed the mean of reproduced (read) speech, with N standing for the number of observations. Consistent tendency is observed in mean FLT to increase in the order of ABS < STR < WEAK < NONE. The only exception to this is the relation of ABS and STR in the male unaccented AP of reproduced speech.

It is interesting to note that, as far as unaccented APs are concerned, the mean FLT value can be classified into two groups; ABS and STR on the one hand, and, WEAK and NONE on the other. This suggests strongly the interpretation that the difference between the two groups is due to the presence and absence of final lowering. The reason why binary grouping was not clearly observed in accented APs is an open question.

Table 1. Mean FLT values (Hz) across CBL and BPM

SEX	ACC	CBL	FLT	N	FLT-R	N
F	A	ABS	149	3026	127	291
		STR	163	2166	155	110
		WEAK	171	6837	167	293
		NONE	173	26069	168	1905
	U	ABS	159	395	135	13
		STR	171	574	167	7
		WEAK	205	2853	194	107
		NONE	208	12421	196	604
M	A	ABS	95	4556	79	255
		STR	100	2865	79	85
		WEAK	108	8456	89	316
		NONE	110	33546	109	1598
	U	ABS	101	520	123	14
		STR	104	637	75	16
		WEAK	132	3550	113	161
		NONE	131	17023	120	608

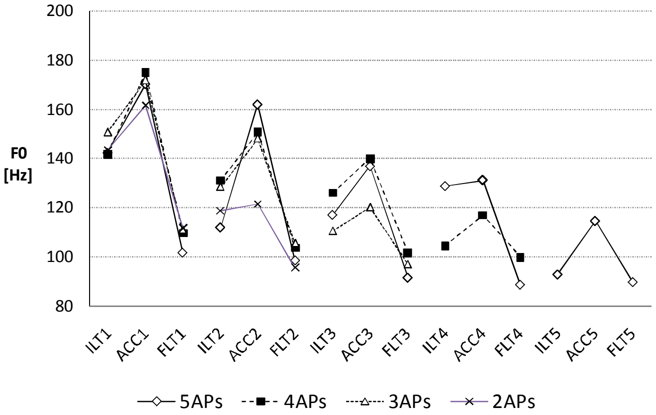

Figure 2: *Analysis of utterances consisting only of accented APs.*

3.2. Domain of final lowering

There is disagreement between the study of Poser [2] and P&B as to the interpretation of the domain of final lowering. In the case of utterances without BPM, Poser supposed that it was only the FLT (i.e., the L%) that was affected by the final lowering, while P&B supposed that the effect of final lowering could cover much wider area of utterance.

This issue was examined in Fig. 2. This figure shows the mean values of F0 measurement points included in the utterances consisting of only accented APs uttered by male speakers. Here an utterance is defined as the string of APs truncated by the ABS/STR CBL labels at the beginning and end, and having only WEAK/NONE labels inside.

Utterances consisting of 2 to 5 accented APs were analyzed. The symbol in the abscissa stands for the measurement points. ACC2, for example, stands for the accentual peak in the second AP. It is interesting to see that all measurement points in the final AP are always much lower compared to the measurement points in the non-final

APs. For example, at the third AP, the measurement points in the utterances consisting of 3 APs (which are in the final AP of the utterances) are clearly lower than the corresponding measurement points in the utterances consisting of 4 or 5 APs (which are not in the final AP). It is also interesting to note that the pitch range in which the tones in the final APs were realized stays roughly speaking the same, i.e., 100-120Hz.

These facts suggest strongly the conclusion that it is the tones in the final AP that are affected by the final lowering. Exactly the same tendency as in Fig.2 was observed in the corresponding female data. All these findings support the view of P&B stated at the beginning of this section.

4. Analysis of BPM

The prosodic hierarchy proposed by P&B supposed that last boundary tones of BPM (i.e., H%, HL%, etc.) are introduced by the Utt node. They also supposed that Utt is the highest node of the prosodic tree. It follows from these that the presence of any BPM automatically triggers resetting of pitch range, because presence of a BPM implies presence of an Utt node that presupposes the presence of an IP node that defines the domain of catathesis and triggers the resetting of pitch range (See the left panel of Fig.4).

If this hypothesis is correct, two accentual peaks in the string of two accented APs punctuated by a BPM will be nearly identical, or the second peak is higher than the first one, because the effect of catathesis is blocked by the IP node introduced by the BPM.

Fig. 3 plots the mean ACC2-ACC1 value (i.e., the difference between the two accentual peaks in Hz) as a function of BPM types. Contrary to the prediction done by the P&B's hypothesis, all values are negative except for the case of LH in the female data (where the number of samples was as low as 5, and 2 of them occurred at the ABS boundary where pitch range resetting seems to be almost obligatory).

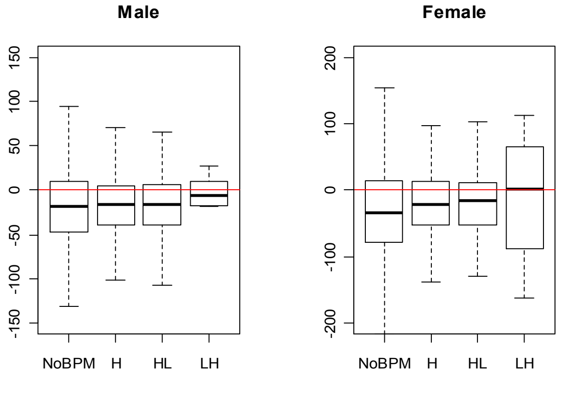

Figure 3: *ACC2-ACC1 as a function of BPM types.*

Fig. 3 showed that presence of BPM does not imply the occurrence of pitch range resetting. It is incorrect, however, if we interpret that pitch range resetting does not occur when there is BPM. Table 2 shows the distribution of utterances as a function of the speakers' sex, relationship between ACC1 and ACC2, and the presence of BPM. Roughly speaking, about one-third of samples show clear pitch resetting regardless of sex and the presence of BPM.

The conclusion that can be drawn from Fig.3 and Table 2 is a straightforward one: the presence of BPM has nothing to do with the prediction of pitch range resetting. From a point of view of the study of prosodic structure, this implies that the node in a prosodic tree to which the tones of BPM are attached should be lower than the IP node. The only phrase-level node that fulfills this condition is the AP node. Venditti, Maekawa and Beckman [9] stated virtually the same view, but without showing experimental evidence.

Table 2. Frequency of clear pitch range resetting.

SEX	RELATION	BPM	
		WITH	WITHOUT
F	ACC1=<ACC2	187	609
	ACC1 > ACC2	360	1383
M	ACC1=<ACC2	215	848
	ACC1 > ACC2	512	1738

This conclusion requires substantial revision of the prosodic structure of P&B. A revised prosodic structure should be something like the right panel of Fig. 4. There are 2 differences between the P&B (left panel) and new trees. First, the utterance final boundary H tone that consisting a part of L%H% BPM belongs to the AP node rather than the Utt node. Second, the utterance-initial boundary L tone does not belong to the Utt node anymore; it also belongs to the AP node.

The second revision can be problematic, because the corpus-based evidences that we obtained so far does not necessarily require this revision. It is theoretically possible to think of a prosodic structure in which the utterance-initial L belongs to the Utt, while the utterance-final H belongs to the AP. Here, the revised structure shown in Fig.4 was preferred over the asymmetric structure due to its affinity to the treatment of 'tonal phrase' in Japanese linguistics [10].

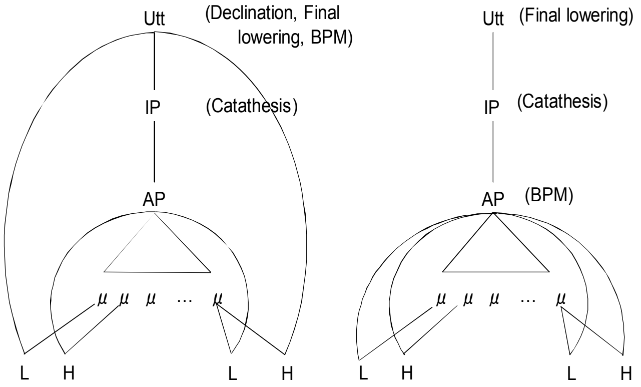

Figure 4: *Prosodic structure of P&B (left) and the revised prosodic structure (right).*

5. X-JToBI Intermediate Labels

As a matter of fact, the facts reported in section 4 were predicted at the time X-JToBI was designed [11], and the system allowing intermediate BI values was introduced as the solution to the problems [7].

Fig. 5 shows the typical case where the presence of a BPM does not block catathesis. The utterance shown here consists of 4 APs, viz., /hyoHka zji'QkeN wa/ 'evaluation experiment –TOPIC', /kono yo'H na/ 'like this', /tatemono ku'HkaN de/ 'in the building-space', /okonai ma'si ta/ 'conduct-POLITE-PAST' ('The evaluation experiment was conducted in a building-space like this'). In Fig.5, apostrophe in the word tier shows the locations of pitch accent. The corresponding accentual peaks in the F0 contour are indicated by the open triangles above the contour. Also, the first 3 AP boundaries are all marked with L%H% BPMs. The filled triangles on the F0 contour indicate the locations of H% tone.

The pitch range was reset at the end of the first AP. The second accentual peaks of this utterance have higher F0 values than the first accentual peak, suggesting that the boundary is a candidate of either an IP or Utt boundary.

At the end of the second AP, however, the pitch range was not reset regardless of the presence of BPM. The third accentual peak was clearly lower (downstepped) with respect to the second. In the same vein, the pitch range was not reset between the third and fourth APs.

In the X-JToBI annotation of BI, the first AP was labeled with a '3' label, and all following APs were labeled with '2+b' intermediate labels. This symbol means that the boundary in question is stronger than ordinary '2' boundary, because the boundary is marked with a BPM. Similarly, the '2+p' label is used for an AP boundary followed by a clearly perceptible pause, and the '2+bp' label is used for an AP boundary marked with a BPM and followed by a pause.

These intermediate labels are by no means exceptional in spontaneous Tokyo Japanese. CSJ-Core contains 7098 '2+b', 3456 '2+bp', and 7155 '2+p' labels as opposed to 42568 ordinary '2' labels. As reported elsewhere, the frequency of intermediate labels, especially the '2+b', played important role in the formation of 'tone-of-voice' of APS [12].

6. Conclusion

Corpus-based analysis of spontaneous speech brought about important findings about the phonological nature of the Utt node of the prosodic tree in Tokyo Japanese. As for final lowering, the predictions of the standard theory were tolerable. Final lowering was a property of Utt node and the domain of its application covered the last AP of utterance. As for BPM, however, the prediction turned out to be inaccurate. Contrary to the prediction, a BPM did not have any effect on pitch range resetting. This negative finding requires considerable revision of the traditional prosodic tree. For that matter, pilot corpus-based analysis of F0 declination, --another property of the Utt node left untouched in this study--, also revealed a need for considerable revision of the standard prosodic structure attenuating the presently powerful role of the Utt node.

7. References

[1] McCawley, D., *The phonological component of a grammar of Japanese.* The Hague: Mouton,1968

[2] Poser, W., *The phonetics and phonology of tone and intonation in Japanese.* Ph.D. Diss. MIT, 1984.

[3] Pierrehumbert, J. and M. Beckman, *Japanese tone structure.* Cambridge: MIT Press, 1988.

[4] Maekawa, K., H. Koiso, S. Furui, and H. Isahara., "Spontaneous speech corpus of Japanese", *Proc. LREC 2000*, Athens, 947-952, 2000.

[5] Maekawa, K., "Corpus of Spontaneous Japanese: Its Design and Evaluation", *Proc. ISCA Workshop on Spontaneous Speech Processing and Recognition*, Tokyo, 7-12, 2003.

[6] Kikuchi, H. and K. Maekawa, "Construction of XML documents for the study of prosody using the Corpus of Spontaneous Japanese", *Proceedings of Oriental-COCOSDA*, 38-42, 2007.

[7] Maekawa, K., H. Kikuchi, Y. Igarashi, and J. Venditti. "X-JToBI: An extended J_ToBI for spontaneous speech", *Proc. ICSLP2002*, Denver, 1545-1548, 2002.

[8] Uchimoto, K., R. Hamabe, T. Maruyama, K. Takanashi, T. Kawahara, and H. Isahara. "Dependency-structure annotation to Corpus of Spontaneous Japanese", *Proc. LREC2006*, 635-438, 2006.

[9] Venditti, J., K. Maekawa, and M. Beckman, "Prominence marking in the Japanese intonation system." S. Miyagawa and M. Saito (eds.) *The Oxford Handbook of Japanese Linguistics*," Oxford University Press, 456-512, 2008.

[10] Kawakami, S. "On the relationship between word-tone and phrase-tone in Japanese." *Study of Sounds*, 9, 169-177, 1961.

[11] Maekawa, K. and H. Koiso, "Design of spontaneous speech corpus for Japanese", *Proc. International Symposium: Toward the Realization of Spontaneous Speech Engineering*, Tokyo, 70-77, 2000.

[12] Maekawa, K. "Study of paralinguistic information using the CSJ-Core," *Proc. 2009 Autumn Convention of the Acoustical Society of Japan*, 479-480, 2009.

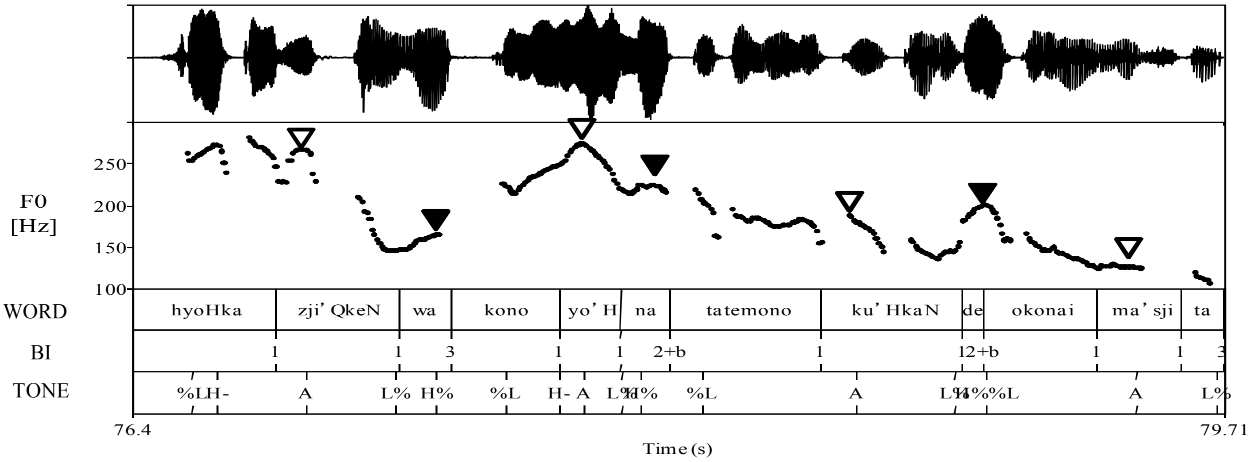

Figure 5: *Example of '2+b' boundary in an APS of CSJ-Core with X-JToBI annotation.*