Pitch range control of Japanese boundary pitch movements

Yosuke Igarashi\textsuperscript{1}, Hanae Koiso\textsuperscript{2}

\textsuperscript{1}Graduate School of Letters, Hiroshima University, Hiroshima, Japan
\textsuperscript{2}Dept. Linguistic Theory and Structure, National Institute of Japanese language and linguistics, Tokyo, Japan
igarashi@hiroshima-u.ac.jp, koiso@ninjal.ac.jp

Abstract
Boundary pitch movements (BPMs) are tones that occur at the end of prosodic phrases in Japanese and contribute to the pragmatic interpretation of certain utterances such as questioning, continuation, and emphasis. This paper investigates pitch range control of BPMs, which has received little attention in previous studies. Specifically, it examines whether downstep is observed in BPMs by analyzing large-scale Japanese spontaneous monologues. The results reveal that ordinal downstep was not found in BPMs, and that pitch range control of BPMs differs across their types. The results suggest that currently accepted frameworks of Japanese intonation cannot deal with pitch range control of BPMs.

Index Terms: downstep, pitch range, boundary pitch movement

1. Introduction
Boundary pitch movements (BPMs) are tones that occur at the end of accentual phrases (APs, see 2.1) in Japanese and contribute to the pragmatic interpretation of certain utterances such as questioning, continuation, and emphasis [7]. A number of researchers have inquired into how many BPMs exist in Japanese and what functions BPMs have [1,2,6], but there has been little investigation into how the pitch range of BPMs is controlled.

One of the main factors that determine the pitch range of utterances in Japanese is catathesis, or downstep, in which preceding lexical pitch accents compress the pitch range of APs [5]. Does downstep also occur in BPMs, or is it limited to the main parts of the APs? If downstep is not observed in BPMs, then pitch range control of BPMs is to some degree independent from that of the main parts of the APs. This putative independence of BPM pitch range has not been reported in previous literature. If this independence is indeed observed, then it will highlight that currently accepted models of Japanese intonation cannot correctly describe pitch range of BPMs.

The goal of this paper is to examine whether or not downstep is observed in BPMs through the analysis of monologues of The Corpus of Spontaneous Japanese (CSJ), a large-scale annotated database of spoken Japanese [3]. The following section will introduce several notions necessary to describe Japanese intonation and will clarify the research issues.

2. The prosodic system of Japanese

2.1. Prosodic phrasing in Japanese

In X-JToBI [4], the intonation labeling scheme that CSJ adopted, prosodic phrasing in Japanese is assumed to occur at two levels.

The lower level is the AP. AP is defined as having a delimitative rise in fundamental frequency (F0) at its left edge and as having at most one lexical pitch accent, realized as a sharp fall in F0. The right edge of the AP is always characterized by a low F0, regardless of whether the AP has an accent or not, though the right edge is always higher in unaccented APs than in accented APs. An AP consists of one or more morphological words. A typical AP contains one content word followed by function words (if any).

In X-JToBI, the AP initial rise is transcribed as a combination of two labels, i.e. [%L] and [H-]. An accentual fall is represented by [A], which is given at the beginning of the fall. The AP-final low is transcribed as [L%].

2.2. Boundary pitch movements (BPMs)

There has been debate over how many BPMs there are in Japanese, but as yet no consensus has emerged. X-JToBI identifies four main types of BPM as shown in Figure 1. H% is a BPM where F0 simply rises. HL% is a BPM with an F0 rise followed by a fall. LH% is similar to H% in that F0 rises, but the rise in LH% is preceded by a sustained low. HLH% has a rise-fall-rise contour.

Figure 1: BPMs. Are ne ‘That was...’ with no preceding accent (upper panel), and I'ma ne. ‘Right now...’ with a preceding accent (lower panel). Produced by the first author. Pitch ranges of BPMs are indicated by squares.
We need multiple labels to describe time and F0 values of the multiple turning points that characterize each BPM. Table 1 shows the tone labels defined in X-JToBI. As can be seen, the first labels are necessarily [L%], which stands for the AP final low. This F0 event usually occurs at the beginning of the AP-final mora when it is accompanied by a BPM. Tone labeling of BPM is exemplified in Figure 1.

<table>
<thead>
<tr>
<th>BPM types</th>
<th>Tone labels</th>
</tr>
</thead>
<tbody>
<tr>
<td>H%</td>
<td>[l%], [lH%]</td>
</tr>
<tr>
<td>HL%</td>
<td>[l%], [pH], [HL%]</td>
</tr>
<tr>
<td>LH%</td>
<td>[l%], [pL], [LH%]</td>
</tr>
<tr>
<td>HLH%</td>
<td>[l%], [pH], [LH%]</td>
</tr>
</tbody>
</table>

2.3. Downstep

Downstep is a phonological process in which a lexical pitch accent in an AP iteratively lowers the F0 peak of the following APs [5]. As discussed below, downstep refers not only to such iterative lowering of the F0 peaks, but also to iterative compression of pitch ranges.

Figure 2 shows waveforms and F0 contours of the utterances uma‘i mame‘-ga arima‘si-ta, ‘There were sweet beans.’, and uma‘i ame-ga arima‘si-ta ‘There were sweet candies’. Since the first AP uma‘i has an accent, downstep occurs in the second AP and its F0 peak is lowered in both utterances. The second AP, on the other hand, has an accent (mame‘-ga) in the utterance in the left panel, but has no accent (ame-ga) in the utterance in the right panel. Downstep in the third AP (arima‘si-ta) is, therefore, present only in the utterance on the left. In what follows, the accent-induced lowering of F0 peaks is referred to as “peak downstep”.

Since the lowering effect of downstep at the bottom of APs is not as strong as at the peak, pitch range (i.e. the difference between the F0 peak and bottom) is iteratively compressed as the number of preceding accents increases. In what follows, the accent-induced compression of pitch ranges is referred to as “range downstep”.

Downstep is blocked by various factors including focus and specific syntactic boundary [2,5,7]. The block of downstep is called pitch reset. In Pierrehumbert and Beckman’s model [5], a boundary of a higher-level prosodic phrase is postulated at the location where pitch reset occurs. In X-JToBI, the prosodic phrase at issue is called IP. The IP is thus defined as a prosodic phrase that functions as a domain of downstep.

2.4. Downstep of BPM

Does downstep occur in BPMs? Pierrehumbert and Beckman [5] compare two utterances with H%, mootyotto migigawa-ga agerareru? ‘Couldn’t the right side be raised a little bit?’, and

mootyotto migigawa-ga sagerare’ru? ‘Couldn’t the right side be lowered a little bit?’. They found that the peak of H% is lower in the latter utterance (with an additional accent in the final AP sagerare’ru) than in the former (without an accent in the final AP agerareru). This means that, at least in these specific utterances, peak downstep is observed in H%.

At the same time, they also found accent-induced lowering at the bottom of H% (i.e. the beginning of the rise). It appears from their data, therefore, that the pitch range of H% (defined as the F0 difference between the peak and bottom) remains constant regardless of the number of preceding accents. Since they do not further discuss its pitch range, it is unclear whether range downstep occurs in H%. Moreover, it is not clear whether their results can be generalized to every type of BPM.

It can be seen from Figure 1, showing four main types of BPMs, that the peaks of the BPMs are far lower in the utterances with a preceding accent (lower panel) than in the utterances without the accent (upper panel). Peak downstep is therefore observed in BPMs, at least in these specific utterances. However, except in the case of HL%, the lowering of the F0 peaks are accompanied by the lowering of the F0 bottoms, and as a result, pitch ranges of BPMs are not compressed, even when preceded by an accent. Range downstep, therefore, appears to be absent in all types of BPM except HL%. The results of a visual inspection of Figure 1 suggest firstly that BPMs behave differently from the main parts of AP with respect to downstep, and secondly that BPM types differ in terms of whether downstep is observed or not.

3. Methods

3.1. Corpus

The data were a subset of CSJ Core (a part of CSJ which contains a wide variety of annotations) [3]. Specifically, we analyzed about ninety hours of “academic presentation speech” (70 speeches) and about twenty hours of “simulated public speech” (107 speeches). They were all monologues.

3.2. APs and BPMs

The APs we analyzed were those which belong to an IP having at least one BPM. In what follows, the part of the F0 contour other than the BPM will be referred to as “AP body”.

F0 characteristics of AP bodies were measured as a function of the number of preceding accents (henceforth, “#PrecAcc”, see 3.3). In addition, they were analyzed separately for unaccented and accented APs. The APs preceded by five or more accents were excluded from the analysis due to lack of a sufficient number. In effect, we analyzed 32,363 APs in total.

The F0 characteristics of BPMs were measured as a function of #PrecAcc separately for each BPM. The BPMs preceded by six or more accents were excluded from the analysis for lack of a sufficient number. For the same reason, we excluded the LH% BPMs with four or more preceding accents and all the HLH% BPMs. In effect, we analyzed 32,353 BPMs in total.

3.3. Method of counting the number of preceding accents

The method of counting #PrecAcc was as follows. In the case of AP bodies, it was the number of accents in the preceding APs that belong to the same IP. In the case of BPMs, #PrecAcc was...
first counted in the same way as in the case of AP bodies, and one accent was then added to #PrecAcc if the AP that the BPM belongs to is accented.

3.4. F0 measurements

We exploited F0 values stored in XML files that CSJ contains. Missing values were excluded from the analysis.

In order to minimize gender and individual differences, F0 values were transformed into Z scores for each speech.

F0 values of AP bodies were extracted in the following way. We regarded the value of [H-] or [A] as the peak. If the AP had both [H-] and [A], the higher value of the two was selected. If the AP had none of the two labels, then the higher value of the two labels, [%L] and [L%], was considered as the peak. As for the F0 bottom, we compared values of [%L] and [L%], and regarded the lower one as the bottom.

The F0 extraction method for BPMs differed according to their types. As the peak, we adopted the value of [H%] for H%, that of [pH] for HL%, and that of [LH%] for LH%. We regarded the value of [L%] as the F0 bottom for both H% and LH%. The pitch range of these two BPMs was defined simply as the difference between the peak and bottom. Since HL% is characterized by the rise and fall, its has two bottoms, i.e., the beginning of the rise ([L%]) and the end of the fall ([HL%]). We measured both, and will refer to the former as “bottom1”, and to the latter as “bottom2”. The pitch range of this BPM was measured in two ways: the difference between [L%] and [pH] (“rise range”, henceforth), and that between [pH] and [HL%] (“fall range”, henceforth).

3.5. Statistical analyses

All the data were averaged within each speech, and then submitted to one-way analyses of variances accompanied by post-hoc Bonferroni tests. The results are shown in Figures 3-5 (*p < .05, **p < .01, ***p < .001).

4. Results

4.1. Downstep in AP bodies

First, the data were examined for the presence of downstep in AP bodies. For lack of space, we will provide a summary of results. The peaks and bottoms consistently decreased as #PrecAcc increased, but the degree of lowering was larger for the peaks than bottoms. Pitch ranges were therefore compressed as #PrecAcc increased. The results confirm range downstep in AP bodies, consistent with previous findings [5,7].

4.2. Downstep of BPMs

Figure 3 depicts the results for H%. We can see that the peaks became lower as #PrecAcc increased, and that the bottoms also became lower as #PrecAcc increased. Pitch ranges remained constant regardless of #PrecAcc.

The results for HL% are shown in Figure 4. It can be seen that in general, the peaks and the bottom1s became lower as #PrecAcc increased. However, the bottom2s were at almost the same level, except when no accents preceded. The rise range remained more or less constant regardless of #PrecAcc, whereas fall range generally became smaller as #PrecAcc increased.

Figure 5, showing the result for LH%, reveals that in general, the peaks and the bottoms became lower as #PrecAcc increased. Pitch ranges were larger when there were no preceding accents than when there were one or more preceding accents.
should be interpreted with caution.

was far smaller in number than other BPM types, the results
AP body with LH%. Nevertheless, because the LH% analyzed
characteristics could largely be derived from the F0 values of the
peaks downstep was observed in LH% and F0
The peaks, bottoms, and pitch ranges of LH% behaved similarly
to that of H%. No range downstep was observed in LH% and F0
characteristics could largely be derived from the F0 values of the
AP body, specifically on L%. Given that pitch ranges of H%
should in fact result from the downstep occurring in the AP body.

5.2. HL%

Similar results were obtained for the rising part of HL%. Both
the bottoms and peaks became lower as #PrecAcc increased, indicating that
peak downstep occurred in HL%. However, the same sort of
iterative lowering was also observed in the bottoms, resulting in
pitch ranges remaining constant. Therefore, range downstep was
absent in HL%.

The bottom of H%, by definition, coincides with the F0 value of L%, i.e. the value around the onset of the final mora of the AP
that accompanies the BPM. The iterative lowering in the bottom of H% can, therefore, be regarded not as the effect that the
preceding accents had on H%, but as the effect they had on the AP body, specifically on L%. Given that pitch ranges of H%
remain constant, the iterative F0 lowering of the peaks of H% should in fact result from the downstep occurring in the AP body.

5.3. LH%

The peaks, bottoms, and pitch ranges of LH% behaved similarly
to that of H%. No range downstep was observed in LH% and F0
characteristics could largely be derived from the F0 values of the
AP body with LH%. Nevertheless, because the LH% analyzed
was far smaller in number than other BPM types, the results
should be interpreted with caution.

5.4. General discussion

Pitch range control of H%, LH%, and the rising part of HL% differed from that of the falling part of HL%. The differences can be generalized as those between rising and falling
movements. Pitch ranges of rising movements are more or less constant irrespective of #PrecAcc. The bottoms (and peaks) of
the rises are predictable from the AP body with the BPM; as the bottom of the AP body is lowered due to downstep, both the bottom and peak of the BPM are also lowered. It can be concluded, therefore, that while peak downstep is observed in rising movements, they exhibit no range downstep.

Pitch ranges of falling movements, on the other hand, largely undergo the accent-induced iterative compression. The peaks of
the falls are lowered as #PrecAcc increases, but the bottoms remain more or less constant. This suggests that the pitch range control in the falling movements of BPMs is comparable to that in the AP bodies which exhibit range downstep.

6. Conclusion

This study examined if downstep is present in Japanese boundary pitch movements (BPMs) by analyzing large-scale spontaneous monologues. The results suggest that, while the iterative lowering of peaks (peak downstep) was found in all types of BPMs, the F0 range compression (range downstep) is limited to a falling part of a specific type of BPM (HL%).

The results showing that pitch range of BPMs is controlled differently from that of the main parts of accentual phrases (AP bodies) suggest that currently accepted frameworks of Japanese intonation cannot deal with pitch range control of BPMs, since they are proposed through of the analysis of the AP bodies. A new theory of pitch range control of Japanese utterances should incorporate the findings in this study, though further research is clearly necessary to fully understand the behaviors of BPMs.

7. Acknowledgements

This study was supported in part by Grant-in-Aid for Young Scientist to YI, and by NINJAL research projects “Foundation of Corpus Japanese Linguistics” and “Empirical Study on the Role of Prosodic Features in Conversations” to YI and HK.

8. References