Perceived prominence and downstep in Japanese

Hyun Kyung Hwang¹, Manami Hirayama², Takaomi Kato³

¹University of Tsukuba
²Seikei University
³Sophia University

hwang.kyung.gu@u.tsukuba.ac.jp, hirayama@fh.seikei.ac.jp, t-kato@sophia.ac.jp

Abstract

Perceived prominence as a function of fundamental frequency (f0) is examined with special attention given to downstep in Standard Japanese, a lowered f0 of the material following an accented word compared with that following an unaccented word. This lowering does not signal a reduction in prominence, whereas f0 compression triggered by focus is associated with prominence. The purpose of this study is to investigate how much f0 lowering is perceived as equally prominent to understand a perceptually acceptable pitch range of Japanese downstep.

A prominence perception test was conducted with varying peak f0s of two successive phrases. Results of the test reveal that there is a particular f0 level of the following phrase to be perceived as equally prominent as the preceding phrase, regardless of the peak f0 of the preceding phrase. This result indicates that f0 differences between the preceding and the following phrases are greater as the f0 of the preceding phrase increases, corroborating the effect of f0 on prominence perception found in the literature. Furthermore, an interesting asymmetry is suggested between production and perception of downstep in Japanese.

Index Terms: prominence, perception, downstep, focus, Japanese

1. Introduction

Prominence perception is closely influenced by fundamental frequency (f0). It is well known that f0 excursion size is correlated with perceived prominence ([1], [2], [3], [4], [5], [6], [7]), indicating that compressed f0 yields lower perceived prominence. In Standard Japanese (henceforth Japanese), there are two independent f0 compression patterns which are observed in a sequence of accented words—downstep and post-focus compression.

Information structure is often encoded by prosody. Particularly, prosodic patterns associated with the focus in Japanese have been extensively documented. In addition, there is a general consensus that the prosodic marking of focus in this language typically involves boosted f0 of a focused phrase with substantially compressed f0 of post-focus phrases ([8], [9], [10], [11], among others). It is widely acknowledged that this prosodic pattern of focus makes the focused material perceptually more prominent ([11]). Yet, it is still unclear how much f0 compression is required for listeners to perceive focus in Japanese.

Japanese downstep, on the other hand, is a prosodic phenomenon triggered by accentedness of a word: an accented word following another accented word is realized in a lower pitch register compared to one following an unaccented word ([8], [9], [12], [13]). Schematic pitch curves of Japanese downstep are illustrated in (1). Acute marks indicate pitch accent. The phrase nagái is downstepped in (1a) as its pitch is lower than (1b) due to the accentedness of the preceding word shirói.

(1) Schematic illustration of downstep (adopted from
[14])

a. Accented Accented

b. Unaccented Accented

Unlike post-focus compression, however, downstep does not signal a reduction in prominence. Two words following an accented and an unaccented word were perceived to have the same prominence ([15]). Thus, if further f0 lowering is made, listeners may interpret the pitch movement as post-focus compression.

However, perceptual aspects of downstep have rarely been experimentally tested. To our knowledge, [15] is the only experimental investigation where the relationship between f0 peaks and relative prominence was explicitly tested. Specifically, synthesized stimuli were utilized where f0 peaks of two successive phrases (P1 & P2) were manipulated. Nineteen participants judged which phrase was more emphasized. It should be noted that the 50% crossover point was assumed to represent that the P1 and P2 were equal in their perceived prominence. The results relevant to the current study confirmed the dependency of P2 peaks on P1 height. This pattern is in accordance with previous findings that for the higher P1 peak, the greater f0 difference between P1 and P2 is required for the two phrases to have equal perceived prominence ([2], [3], [4], [5], [16], [17]). Although this result appears consistent with prior findings, more strictly controlled experimentation is necessary because dialects of the participants were not controlled. Perhaps more importantly, the assumption that 50% crossover point as the equal prominence perception holds true only if the number of the stimuli for each speaker is large enough.

Thus, an experimental investigation is necessary to further explore prominence perception of prosodic patterns involving f0 compression. In the current study, we aim to investigate the relationship between the degree of f0 lowering and perceived prominence by providing explicit responses of equal perceived prominence through a perception test. Specifically, we address
the question as to how much f0 lowering is acceptable as downstep in Japanese.

2. Methods

A perception experiment was performed to test whether the degree of f0 lowering may affect listeners’ prominence perception.

2.1. Material

The tested phrase is given in (2). ACC and SFP represent an accusative marker and a sentence final particle, respectively.

(2) Trigger (P1) Target (P2)

\textit{Nagāi nēgi-o katta-yaō}

long leek-ACC bought-SFP

“(I) bought a long leek.”

\textit{Nagāi}, constituting P1, is a potential trigger, whereas \textit{nēgi}, P2, is a target of downstep. Note that both words bear an accent.

To vary the information structure of the phrase, it was uttered as an answer to a different prompt question as in (3). LOC and Q indicate a locative and interrogative marker, respectively.

(3) Prompt questions for different information status

a. Broad focus

\textit{yaoya-de nāni sita-no!}

vege store-LOC what did-Q

“What did (you) do in the vegetable store?”

b. P1 focus

\textit{yaoya-de dōna nēgi-o katta-no?}

vege store-LOC what.kind.of leek-ACC bought-Q

“What kind of leek did (you) buy in the vegetable store?”

c. P2 focus

\textit{yaoya-de nagāi nāni-o katta-no?}

vege store-LOC long what-ACC bought-Q

“What is the long one (you) bought in the vegetable store?”

Downstep is expected to occur in P2 in (2) as an answer to the broad focus prompt in (3a). P1 and P2 focus conditions were included since the f0 movements in these conditions can be a reference when manipulating the f0 curves.

In the recording, two speakers of Japanese, a male and a female, exchanged conversions. The female speaker, who was from Tokyo, produced the target phrase. Each question-answer pair was repeated twice, and the rendition that exhibited smoother pitch curves was chosen for f0 manipulation.

2.2. Stimuli

Manipulation was made based on an utterance of the broad focus condition where downstep is expected to be present without further f0 compression by prosodic focus. The natural utterance exhibited f0 peaks originally at 264 Hz and 244 Hz for P1 and P2, respectively. P1 and P2 peaks were manipulated using the “Pitch Synchronous OverLap Add (PSOLA)” method in Praat; P1 was varied in 4 steps (264, 284, 304 & 314 Hz), and P2 in maximally 12 steps (204-314 Hz) with a step size of 10 Hz, as shown in Figure 1. The maximum or minimum f0 was determined based on the pitch movement of P1 or P2 focus conditions in (3). The step size of 20 Hz for P1 and 10 Hz for P2 were determined based on the results of pilot tests. The peak f0 values of P1 and P2 tested are summarized in Table 1. It is worth noting that P2 was not higher than P1 in the stimuli as such a pitch pattern was not observed in the recording even when P2 received focus. Combinations that were not included are shaded in Table 1. As a result, a total of 37 test phrases were created.

Table 1: Tested peak f0s of P1 and P2.

<table>
<thead>
<tr>
<th>P1</th>
<th>204</th>
<th>214</th>
<th>224</th>
<th>234</th>
<th>244</th>
<th>254</th>
<th>264</th>
<th>274</th>
<th>284</th>
<th>294</th>
<th>304</th>
</tr>
</thead>
<tbody>
<tr>
<td>P2</td>
<td>204</td>
<td>284</td>
<td>304</td>
<td>314</td>
<td>204</td>
<td>284</td>
<td>304</td>
<td>314</td>
<td>204</td>
<td>284</td>
<td>304</td>
</tr>
</tbody>
</table>

![Figure 1: Example f0 curves tested (P1 = 314 Hz).](image)

The slightly diverging boundary tones were the artifact of f0 manipulation in Praat. It was confirmed that this did not influence the perceived prominence of P1 or P2.

Eight fillers of various structures were also prepared. Three speakers, including the female speaker who uttered the target phrases, took part in the recording of fillers. In creating fillers, an effort was made to assign the prosodic patterns of broad focus.

Before the test, the intensity level of the stimuli was scaled on Praat to maintain the mean intensity of each stimulus at 70 dB. Stimuli were divided into 5 blocks, and each block involved 7-8 test phrases and 8 fillers. Within a block, stimuli were placed to alternate test and filler phrases. To circumvent the order effect, two versions were created: Version 1 presented from Block 1 to Block 5, while Version 2 was in the opposite order. Then we randomly assigned half of the participants to Version 1 and the other half to Version 2.

2.3. Participants and task

Twenty-two speakers (14 females) of Japanese participated in the test. Their ages ranged from 20 to 48 years old at the time of the test.

The test was conducted online on Cognition.run, a platform to run online experiments, using a JavaScript coding framework jsPsych (library version 6.3.0). Participants were asked to wear headphones/earphones and informed that they would hear Japanese sentences. The task was to choose which word was more prominent/emphasized after hearing a phrase twice. Three choices were provided on a computer screen: P1, P2, and neither. The choice neither is crucial as it explicitly indicates that speakers perceive P1 and P2 as equally prominent.
3. Results and Discussion

The results of a prominence perception test reveal that f0 peaks are highly correlated with perceived prominence. Figure 2 shows the number of P2 responses (dashed lines) and equal prominence (solid lines) as a function of P2 height in four different P1 height conditions—264 Hz, 284 Hz, 304 Hz, and 314 Hz. For example, when P1 was 284 Hz and P2 ranged from 204-244 Hz, more than half of the participants heard no prominence on either P1 or P2, but when P2 was higher than 255 Hz, the majority of the participants heard prominence on P2.

Overall, the P2 responses increase, and the neither responses, i.e., equal prominence, decrease as the peak f0 of P2 increases across the different P1 conditions. This result suggests that larger f0 excursion size is correlated with greater perceived prominence, which was extensively addressed in prior research in several languages ([1], [2], [3], [4], [5], [7], [16]).

Strikingly, the responses of P2 prominence and that of equal prominence cross (P2 prominence = equal prominence) when the peak f0 of P2 is approximately 244-254 Hz regardless of the P1 heights. It seems that not only the relative f0 height but also a particular f0 level may play a role in perceiving prominence. In other words, at least in this speaker’s pitch range, 244-254 Hz may be the threshold of an expected f0 height without pitch modification triggered by focus, and further f0 drop is associated with the reduction in prominence.

This result also indicates that for a higher P1, the greater f0 difference between P1 and P2 is required for the two phrases to be perceived equal in prominence. This tendency is illustrated in Figure 3.

This result corroborates the dependency of P2 height on P1 height concerning perceived prominence, which has been extensively reported in previous studies ([2], [3], [4], [5], [15], [16], [17]). For two phrases, P1 and P2, the peak f0 differences between P1 and P2 should be larger as the peak f0 of P2 increases for the two phrases to be perceived as having equal prominence.

Further, in answering the question of how much f0 lowering is acceptable as downstep in Japanese, an obvious asymmetry is observed between production and perception of f0 lowering patterns. Note that P2 of the speaker in this study was approximately 20 Hz lower than P1. Compared to the degree of f0 lowering in downstep production, however, a much larger f0 difference was acceptable for two successive phrases to be equally prominent. For instance, when P1 is 314 Hz, f0 lowering by 60 Hz (P2 = 254 Hz) was still not large enough to be interpreted as post-focus compression, resulting in about half of the participants perceiving the two phrases as equally prominent.

Possibly, this asymmetry can be attributed at least partially to the lack of other prosodic cues. Greater f0 compression is required for focus perception when duration or intensity cues remain unchanged. However, for the perception in terms of pitch alone, it appears that listeners allow for quite a large pitch lowering for downstep, at least much larger than the...
pitch drop observed in the production of natural speech in Japanese.

Alternatively, this greater range of f0 lowering for equal prominence in perception might be related to the lack of discourse context. It is widely acknowledged that information structure, including the focus status, is constructed and signaled by various devices such as discourse context, syntactic constructions, word order, lexical items, or prosodic marking. Yet, in the controlled experiment in this study, no other contextual cues existed to facilitate the processing of focus, which might, in turn, be the source that the listeners required more salient prosodic cues for focus perception.

4. Conclusions

Most studies on Japanese downstep or prosodic implementation of focus have only paid much attention to the acoustic characteristics of the prosodic patterns. In the current study, we have focused on the perceptual aspects of downstep and investigated the relationship between the degree of f0 lowering and perceived prominence in a sequence of two accented words. Results of a prominence test reveal that there is a particular f0 level of the following phrase regardless of the peak f0 of the preceding phrase, suggesting that a certain f0 height may function as a threshold of equal prominence perception. Further, the empirical finding reported in this study suggests an interesting asymmetry between production and perception of downstep in Japanese. Specifically, the greater f0 lowering than the production of downstep is still acceptable as downstep in the perception.

Yet, to better understand the processing of prominence perception, it is necessary to expand the scope of the investigation to include other acoustic correlates of prominence, such as duration cues and their interactions with f0 cues. Another direction for further investigation is to examine the various syntactic structure and lexical items of different parts of speech, as those factors may influence prosodic patterns involving downstep ([14], [18], [19], [20]). Additionally, the interaction between focus and downstep should be explored as it has been argued that focus and downstep can occur simultaneously ([21], [22]). The results obtained in the current study will be fed into a study involving constructions with greater syntactic complexity or testing the interactions between focus and downstep in Japanese.

5. Acknowledgements

We are grateful to Prof. Haruo Kubozono for his valuable comments on this work, and for all the subjects for their participation. This research was funded by JSPS (Japan Society for the Promotion of Science) KAKENHI grant 19K00613 and 21K00498.

6. References