Perception of Boundary and Prominence in Spontaneous Japanese: An RPT Study

Shinobu Mizuguchi¹, Koichi Tateishi²

¹Kobe University (Emeritus) ²Kobe College
mizuguti@kobe-u.ac.jp, tateishi@mail.kobe-c.ac.jp

Abstract

Traditional studies on prosody argue that prominence is highly tied to changes of F0 but recent perceptual research of utterance-level prosodic prominence using Rapid Prosody Transcription (RPT) shows that perception strategy is much more complex, as it involves not only phonetic cues but also phonological, semantic and information cues. This paper considers Japanese in the RPT framework. Since it is a mora-timed pitch language and uses pitch both for lexical accent and utterance-level prosody, it is expected that Japanese has a different perception strategy from some Indo-European languages that use pitch movement for utterance-level prosody only. It is also expected that our study will provide concrete data for the hot topic in Japanese literature, ‘Does focal prominence reset a phrase boundary?’ based on the utterance-level perception experiment.

We will show that (i) contra literature on Japanese focus, acoustic features of F0 and intensity are not strong prominence cues in Japanese, (ii) perceived prominence is strongly tied to pitch movement and its location in an utterance, and (iii) not only content words but also function morphemes get highlighted in Japanese. Perception strategies vary among languages, as predicted.

Index Terms: perception, spontaneous Japanese, boundary, prominence, Rapid Prosody Transcription (RPT)

1. Theoretical Background

This paper studies how boundaries and prominences are perceived in Japanese spontaneous speech, based on a perception experiment. We will provide theoretical background and the purpose of the study in Section 1. Section 2 shows our experiment, followed by a discussion in Section 3 and a conclusion in Section 4.

1.1. Accent and Pitch System in Japanese

Japanese is a pitch language, with its accent realized by a falling binalonal H*-L pitch contour. It has Accent (A-) and Unaccented (U-) words, which are lexically determined, as in (1).

(1) a. Initial-Accented: i’noi ‘life’
   b. Medial-Accented: koko’ro ‘heart’
   c. Final-Accented: atama ‘head’
   d. Unaccented: sakana ‘fish’

(N.B. ‘ following a vowel shows the existence of an accent on the syllable with the vowel).

Above the word level, the two phrasal levels, Minor Phrase (MiP, a.k.a. Accentual Phrase) and Major Phrase (MaP, a.k.a. Intermediate Phrase) are distinguished. The former is the domain of initial rise %LH and the latter is the domain of ‘downstep’, i.e., the phenomenon that an H*-L pitch accent lowers a falling F0 contour of the following material (cf. Figure 2). Downstep gauges the degree of relative pitch range expansion and influences focus perception. On the utterance level, the pitch gradually lowers towards the end of the utterance in Japanese, which is called ‘downtrend’ (a.k.a. catathesis) in the literature.

There has been a heated discussion in Japanese literature on how focus effects would interact with downstep, and we find two conflicting theories. One is that focus is strong enough to initiate its own MaP. Thus, the existence of focus resets the effects of downstep and initiates a new MaP (cf. [1]). The other theory is that focus simply raises the pitch and lowers the following F0 but that the effects are not phrasal and not strong enough to identify them with a phrasal break (cf.[2]-[4], among others). The former locates focus only at the phrase-initial position, and the latter allows the focus to appear in the middle or the final position of an Intonation Phrase (IP).

We often take it for granted that Japanese prosody has been much studied, but papers on the topic are mostly theoretical work on the word level or the phrase level at most. We, on the other hand, are going to conduct a perception experiment on the utterance level and investigate how IPs and prominences interact in Japanese. We will follow the method of Rapid Prosody Transcription (RPT) in our experiment.

1.2. Rapid Prosody Transcription (RPT)

In the perception study of prosody, the research subjects have long been restricted to the phoneme-, word- and phrase-level. RPT is a challenge to broaden a research target to cover spontaneous speech. Cole developed Rapid Prosody Transcription (RPT) with her colleagues (cf.[5]). It is a system for collecting prosodic annotations from untrained listeners; listeners mark prominences and boundaries on transcribed speech, as in (2), while listening to recorded speech.

(2) i really don’t know i think in today’s world what they call the nineties that uh it’s just like everything is changed like when i grew up …
Boundary (b-) score and prominence (p-) score ranging from 0 to 1 indicate the proportion of participants who underscore the respective word. RPT has no restriction on its materials or subjects, and the previous studies employ spontaneous speech or the read speech, L1 or L2 listeners (cf. [5] – [9]).

The previous RPT work has found that boundaries are much easier to perceive than prominences in English; [5] reports that the inter-speaker agreement is higher in boundaries than in prominences (κ0.63 and κ0.402 on Fleiss’ kappa, respectively). The median value of the inter-speaker agreement suggests that prominence is generally not easy to perceive in spontaneous speech in English.

To investigate boundary cues, [5] annotates each word for a boundary code (b-) and prominence (p-) score ranging from 0 to 1, indicating the proportion of participants who underscore and give the token distribution of each part of speech. Our material contains 980 morphemes in total, and Table 1 gives the token distribution of each part of speech.

The cross-linguistic study is getting advanced in the RPT framework; not only segmental acoustic cues of F0, duration, and rms intensity but also non-segmental cues like pitch movement, especially LH*, are crucial in prominence perception in German (cf. [6]). [7] shows that English, French, and Spanish are statistically different in prominence ratings (χ2(2) =12.18, p=0.002). The three languages are similar in the use of acoustic/phonetic cues, but they vary in how non-phonetic cues are involved in prominence perception.

To the authors’ knowledge, no RPT experiment has been conducted on Japanese. Our research questions (RQ) are: (i) What are the cues of boundary marking and prominence marking in Japanese? (ii) Does prominence reset an IP?

We will present our methodology, materials, and results of our RPT experiment in the next section.

2. RTP Experiment on Japanese Spontaneous Speech

2.1. Method and Materials

Our method is RPT described in the previous section, and the materials are spontaneous speech. Our experiment was approved by the Ethical Committee of the National Institute of the Japanese Language and Linguistics (NINJAL). We used 12 materials of Corpus of Spontaneous Japanese (CSJ), released by NINJAL. CSJ is a corpus of pseudo-lectures and monologues collected from adult Tokyo Japanese speakers. Since Japanese is an agglutinative language, we segment our data-set on the morpheme level (e.g., (3)).

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

Our material contains 980 morphemes in total, and Table 1 gives the token distribution of each part of speech.

2.2. Results

The inter-listener agreements were κ0.638 on boundary (b-) score and κ0.359 on prominence (p-) score. The correlation between p-score and b-score is weak (r=0.12 on Person’s Correlation). We will report the result of boundary perception first, and then prominence perception below.

2.2.1. Boundary Perception

Table 2 shows the results of the regression analyses between b-scores and acoustic cues of Maximal (Max) F0, duration, rms Intensity, and range F0, measured on the morpheme level. We see that acoustic cues are weak predictors of boundaries.

| Table 1: Token distribution of part-of-speech categories (morpheme-based) |
|-----------------------------|---------------------|---------------------|
| Content | morphemes | Function morphemes |
| (N=533) | (N=447) |
| Noun | 264 | Particle | 177 |
| Verb | 137 | Auxiliary Verb | 89 |
| Conj. | 37 | Discourse Marker | 60 |
| Adv. | 27 | Suffix | 41 |
| Adj. | 22 | Complementizer | 31 |
| WH | 2 | Numeral | 27 |
| Pronoun | 1 | Topic marker | 25 |
| | | Demonstrative | 16 |
| | | Classifier | 6 |
| | | Negation | 3 |
| | | Quantifier | 1 |
| | | Disfluency | 10 |
| | | Interjection | 4 |

Out of the 980 morphemes, 232 are accented, and 339 are unaccented. The rest of the 309 morphemes are function morphemes with no lexically given accents. Our materials were 18 to 41-second long.

We conducted our experiment on Yahoo! Crowd-Sourcing Service. We recruited 35 Tokyo Japanese listeners (mean age 24.8, SD=0.5). They had no hearing difficulties. After the exercise session, they listened to each material twice via PC, while marking prominences and boundaries by clicking a mouse. Their responses were saved on the computer via LME, developed by Mahiri [10]. It took about 30 minutes for our participants to complete the task. They were paid by Yahoo! points.

To investigate how syntactic categories function to mark boundaries in Japanese, we calculated b-scores at left- and right-edges, following [5]. Figure 1 shows the result. We can see that syntactically higher categories of Sentence and
Conjunction are important boundary cues. Syntactically lower categories of Verb Phrase and Auxiliary are also important boundary cues, contra English (cf. [5]). Japanese is an agglutinative head-final language and auxiliaries often appear following a verb at the phrase final position. Japanese has topic marker wa and Topic Phrase also makes an independent IP (cf. Figures 2-3) and marks boundaries in Japanese.

Figures 2-3 show examples of the pitch movements perceived as prominent. The content words chocho 'the major key' and tancho 'the minor key' at the IP-initial and the IP-mid position respectively are perceived as prominent in Figure 2. They are assigned lexical accents H*L and the pitches are locally boosted. Figure 3 shows the highlighted topic marker wa at the IP-final position. Wa is a function morpheme and has no lexical accent. It is perceived as prominent when it is assigned prominence-lending tone HL.

2.2.2. Prominence Perception

Table 3 shows the results of multiple regression analyses among p-scores, acoustic cues, IP positions, and pitch movement types. Max F0, duration, and range F0 are acoustic cues, IP-initial, IP-mid, and IP-final are positional cues, and H, HL, and LH are pitch movement cues. We have chosen these three pitch movements since they are classed as ‘prominence-lending’ tones in [11].

Table 3: Adjusted \( r^2 \) of multiple regression analyses among p-scores X IP-positions X acoustic cues/pitch movement (\( p<0.001 \), except in shaded columns)

<table>
<thead>
<tr>
<th>IP- initial</th>
<th>( \hat{\beta}_0 )</th>
<th>( \hat{\beta}_1 ) (acoustics/ tone)</th>
<th>( \hat{\beta}_2 ) (IP- position)</th>
<th>Adjusted ( r^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>MaxF0</td>
<td>-0.09</td>
<td>0.0008</td>
<td>0.33</td>
<td>0.375</td>
</tr>
<tr>
<td>duration</td>
<td>0.055</td>
<td>0.341</td>
<td>0.322</td>
<td>0.367</td>
</tr>
<tr>
<td>rangeF0</td>
<td>0.077</td>
<td>0.002</td>
<td>0.299</td>
<td>0.408</td>
</tr>
<tr>
<td>H</td>
<td>0.13</td>
<td>0.141</td>
<td>0.377</td>
<td>0.315</td>
</tr>
<tr>
<td>HL</td>
<td>0.126</td>
<td>0.245</td>
<td>0.336</td>
<td>0.375</td>
</tr>
<tr>
<td>LH</td>
<td>0.131</td>
<td>0.097</td>
<td>0.34</td>
<td>0.308</td>
</tr>
<tr>
<td>IP-mid</td>
<td>( \hat{\beta}_0 )</td>
<td>0.0009</td>
<td>0.216</td>
<td>0.272</td>
</tr>
<tr>
<td>MaxF0</td>
<td>-0.029</td>
<td>0.421</td>
<td>0.228</td>
<td>0.274</td>
</tr>
<tr>
<td>duration</td>
<td>0.039</td>
<td>0.002</td>
<td>0.21</td>
<td>0.346</td>
</tr>
<tr>
<td>rangeF0</td>
<td>0.066</td>
<td>0.021</td>
<td>0.267</td>
<td>0.164</td>
</tr>
<tr>
<td>H</td>
<td>0.13</td>
<td>0.036</td>
<td>0.246</td>
<td>0.25</td>
</tr>
<tr>
<td>HL</td>
<td>0.13</td>
<td>0.219</td>
<td>0.218</td>
<td>0.23</td>
</tr>
<tr>
<td>IP-final</td>
<td>( \hat{\beta}_0 )</td>
<td>0.001</td>
<td>0.224</td>
<td>0.201</td>
</tr>
<tr>
<td>MaxF0</td>
<td>-0.043</td>
<td>0.451</td>
<td>0.197</td>
<td>0.179</td>
</tr>
<tr>
<td>duration</td>
<td>0.045</td>
<td>0.002</td>
<td>0.152</td>
<td>0.256</td>
</tr>
<tr>
<td>rangeF0</td>
<td>0.147</td>
<td>0.185</td>
<td>0.241</td>
<td>0.08</td>
</tr>
<tr>
<td>H</td>
<td>0.142</td>
<td>0.3</td>
<td>0.133</td>
<td>0.161</td>
</tr>
<tr>
<td>LH</td>
<td>0.13</td>
<td>0.1</td>
<td>0.34</td>
<td>0.308</td>
</tr>
</tbody>
</table>

3. Discussion

Our research questions are: (i) What are the cues of boundary marking and prominence marking in Japanese? (ii) Does prominence reset an IP? Our answer to RQ (i) is: boundaries are cued not by acoustic features but by syntactic categories (cf. Table 2, Figure 1). As for prominences, we can claim that MaxF0 and rangeF0 are hardly prominence predictors in Japanese since the standardized partial regression coefficients are very low in Table 4. Traditional studies on prosody (cf. [12][13], among many others) argue that prominence is highly tied to the changes of F0, and prominence perception is affected by the excursion size of the F0 peak; the contour in Figure 4Aa involves more prominence than the one in Figure 4Ab. It is also known that the location of the F0 peak affects perceived prominence: Pierrehumbert [14] shows that when two peaks sound equivalent, the second F0 must be lower. Listeners know that an F0 peak that appears later in an utterance is lower than an earlier peak due to declination, and compensate it in perception.
Japanese realizes a downturn and both the topline and the baseline lower as the utterance goes on, as in Figure 4B. Listeners compensate and normalize F0 in Japanese (cf. Figure 2) and we can claim that Japanese does not cue prominences by MaxF0 and range F0. Table 3 leads us to claim that the prominences in Japanese are cued by pitch movements and positional cues more than acoustic cues. It is not, however, clear how these cues interact; the standardised partial regression coefficient is higher with an IP-position than a pitch movement in some combinations while in other combinations, it is higher with a pitch movement. We need further research before we conclude.

Our answer to RQ2 is ‘No’. We have provided concrete data to show that prominences are perceived in the IP-mid and the IP-final position (cf. Figures 2-3). We have found an answer for the theoretical hot issue of Japanese literature and proved that the reset theory is not correct, based on the perception experiment on spontaneous speech.

Besides the RQs, we have a new finding; not only content words but also function morphemes are subject to highlight in spontaneous Japanese (cf. Figure 3). The RPT literature (cf. [5]-[8]) claims that only content words, not function morphemes, get highlighted. In Japanese literature, highlighted particles have long been observed on the sentence level (cf.[15]-[18]). Our study is the first report of the highlighted function morphemes in spontaneous speech.

What is interesting is that highlighted function morphemes appear only at the IP-final position. We need to clarify two things to account for this phenomenon: one is the function of prominence in Japanese, and the other is what appears at the IP-final position.

Let us start our discussion with the first issue: the function of Japanese prominence. Though there is not much work on the function of focal prosody in Japanese, Kuno [17] claims that when particle wa is highlighted, it is contrastive. He classifies two types of topic marker wa: thematic wa and contrastive wa. It is the latter that gets prominent. Tomioka [18] also observes that contrastive wa gets prominent.

(4) Q: Dare-ga ukat-ta-no?  
who-NOM pass-PAST-Q  ‘Who passed?’
A: Ken-wa/Ken-WA ukat-ta.  
‘Ken passed/(At least) Ken passed.’ (cited from [18])

In this Q/A pair, the answer without prominence on wa gives new information as an answer to the question, but with prominence, the speaker implies that among the examinees including Ken, he knows that Ken passed but does not know about other people’s results. Contrastive wa induces an alternative set {Ken, Naomi, Lisa, …} in the sense of Alternative Semantics (cf. [19]), and provokes the pragmatic implicature. Thematic wa, on other hand, does not provoke this implication and gives only new information in (4A).

Note here that having prominence on a particle does not indicate that the particle is in contrast with other particles. The scope of the focal prominence is larger than the particle itself and covers the whole syntactic phrase it is contained. This phenomenon is accounted for when we assume that Japanese is an agglutinative head-final language. Fukui [20] proposes that case (Kase in his term) is the head of Kase Phrase (KP), a Japanese equivalent to DP, and occupies the phrase-final position. Contrastive wa is the head of KP (cf. (5a)) and determines the domain of focal prominence FoC (cf. (5b)). In the framework of Alternative Semantics, FoC assignment induces an alternative set (cf. (5c)) and prominence is realized in the head (cf. (5d)). Refer to Figure 3 for the pitch movement.

(5) Phrasal prominence realized on the head
a. syntactic labeling: [wA [S [daimei]-wA] [wa] [FoC]]
   title TOP
b. FoC assignment: [wF [daimei]-wF [wa] [FoC]]
   a. alternative set: [wA [daimei-wA] [FoC] = {wA. Topic y in
   w | y ∈ D1}] = {title, contents, language, …}
   b. prosodic prominence: daimei-wa
   HL

In our experiment, 183 boundaries and 236 prominences are marked. Only 35 out of 236 are prominences marked on function morphemes and the rest of the prominences are found on content words. We suspect that the way to realize prominence is different between content words and function morphemes. Japanese is a pitch language and its accent is lexically determined (cf. (1)). Since lexical accents are kept on the utterance level, the only way to highlight accented content words is a local boost (cf. Figure 2). But one-mora function morphemes are devoid of accents by definition and have no lexical accents. They are given prominence-lending pitches of HL, LH, and H at the IP-final position (cf. Figure 3). The latter is much easier to perceive than the former for native listeners of Japanese since the local boost in Japanese is not so strong (cf. [21]).

4. Conclusion

We have discussed the perception of boundaries and prominence in spontaneous Japanese, based on an RPT perception experiment. Our findings are (i) boundaries are cued not acoustically but syntactically, (ii) prominences are predicted by the correlation of pitch movement and the locus of prominence in an IP, (iii) both content words and function morphemes are highlighted. We have shown that Japanese, a pitch language, has its own perception strategy. It is a first step toward a cross-linguistic view on perception biases among languages.

5. Acknowledgements

This work is partially supported by the collaborative research project fund of NINJAL, given to the first author. We sincerely thank Timothy Mahrt for his help and suggestions in the course of our experiment.
6. References


