PHONETIC INVESTIGATION OF BOUNDARY PITCH MOVEMENTS IN JAPANESE

Kazuaki Maeda [†⋆] *and Jennifer J. Venditti* [⋆††]

[⋆]Bell Labs – Lucent Technologies
[†]University of Pennsylvania
[††]Ohio State University

ABSTRACT

Pitch movements at the boundaries of sentence-medial and final phrases in Japanese can provide a cue to the speaker's intention. For example, the phrase /Nagano-de/ 'in Nagano' can be uttered with different rising and/or falling pitch movements on the the final mora /de/ to convey meanings such as clarification, incredulity, prominence in the discourse, insistence, etc. The identification of these movements is important not only for spoken language understanding systems, but also for natural-sounding speech synthesis. The current study examines F0 shape, height, and alignment characteristics of four distinct sentence-medial boundary rises. We compare these types with accented and focused unaccented words containing identical phonetic segments, and discuss a number of different possible phonological analyses of the pitch movements.

1. INTRODUCTION

There are a number of different boundary pitch movements in Tokyo Japanese. Just considering boundary rises, Kawakami identifies as many as five different rise types ([5], inter alia). His categories include *futsû no jôshôchô* 'normal rise', *ukiagarichô* 'floating rise', *hanmon no jôshôchô* 'return question', *tsuyome no jôshôchô* 'emphatic rise', and *tsuriagechô* 'hook rise'. Although the number of categories is large, it is difficult to understand which contours are intended, as no detailed phonetic description is provided. In this paper, we follow Maekawa [7] and Kori [6] in investigating the quantitative as well as qualitative aspects of boundary pitch movements. We limit our study to a set of carefully-elicited rises occurring in sentence-medial phrases.

Consider the following minimal pair:

hontô ni Nara no na no? 'Is this really the one from Nara?'
hontô ni Nara no na no! 'This is really the one from Nara!'

The first is a *question* (asking for information), and the second is *insisting* (asserting information). The only difference between the two is the final pitch movement. Both rise at the end from a low point to a higher F0, but the question rises higher. Japanese ToBI [9] describes these as phonologically the same: both are transcribed with a H% boundary tone at the right edge of the phrase. Another boundary movement transcribed with H% in J_ToBI is what we call the *prominence-lending rise*. These rises lend some kind of discourse prominence to a phrase, and are similar to those described by Muranaka and Hara [8] as 'prominent particles' in their analysis of Japanese narrative discourses:

(in the Momotarô story)
obâsan WA ... 'the old woman-TOP' ([8], p.397)

There are three potential hypotheses regarding the phonological nature of these boundary pitch movements:

- H-BT The rise is due to a H% boundary tone associated with the edge of the phrase [9].
- ACC The rise is an accent whose fall is not realized [2].
- PHR-H The rise is due to the H- phrase tone of an unaccented accentual phrase, with the following fall to low not realized [1].

After presenting the analysis results, we discuss the plausibility of each of these phonological accounts. Due to space constraints, we focus only on the phonetic and phonological characterizations of the boundary pitch movements. See [10] for a discussion of modeling these shapes for speech synthesis.

2. CORPUS 1

2.1. Design

In order to compare these rises on the same segmental material, we constructed mini-discourses in which the target phrases in the responses were kept prosodically identical, and the target morae at the end of the phrases were segmentally identical, with respect to phonemic class (/ni/ or /mi/, both abbreviated as 'NI' here). Examples of the target phrases are: *Na'oya ni?*, *Na'oya ni!*, etc. The different boundary rises on each target phrase were elicited by the previous discourse turn (speaker A), and were modeled when necessary using an audio prompt containing entirely different words. Utterances both with and without a pause following the target phrase were included in the experiment. In no-pause cases, the target mora NI was immediately followed by the mora /wa/ in the next word.

We have already mentioned the first three types of pitch movements examined in this study:

- Question rise (*question*): *Na'oya ni?*
- Insisting rise (*insisting*): *Na'oya ni!*
- Prominence-lending rise (*prom*): *Na'oya ni/*

In addition, to help identify the similarities (or differences) of these rises to other types of rises in Japanese (see hypotheses ACC and PHR-H), we also included the following types:

- Unaccented word (*unaccented*): Monomoraic unaccented word /mi/ 'a nut'. The accentual phrase-initial rise (H-) is realized when the word is emphasized. *to'chi no MI*
- Rise-fall boundary movement (*rise-fall*): aka. J_ToBI HL%. This is commonly observed in conversation, especially among young speakers. *Na'oya ni∧*

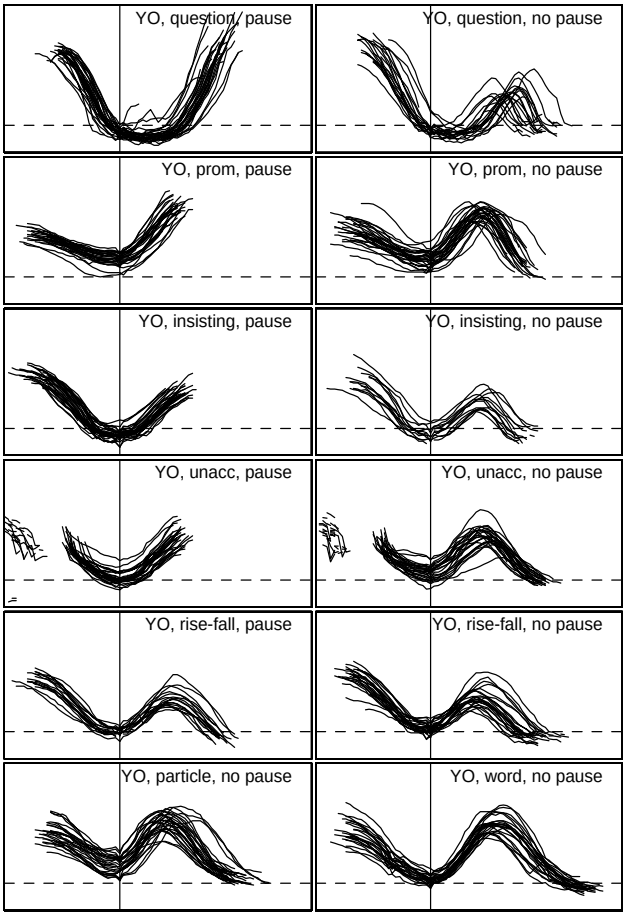

Figure 1: F0 shapes of the contours types in Corpus 1, pause and no-pause conditions. The solid line marks the onset of the target mora NI, and the dashed line marks a fixed arbitrary reference F0 height.

- Accented particle (*particle*): Compound particle /ni'-wa/ (DAT+TOP). *Na'oya ni'wa*
- Accented word (*word*): Bimoraic accented word /ni'wa/ 'two birds'. *na'ya no ni'wa*

Two native speakers of Tokyo Japanese participated in this production experiment: one female (YO), and one male (KM, the first author). Due to space constraints, we will report only data from YO here. The mini-discourses were presented in Japanese orthography on 3x5 file cards, and were recorded in a sound-attenuated room. Sound files were digitized at 16 KHz (16-bit resolution) on SUN and SGI Workstations, and analyzed using Entropic *Waves+* software. Labels were placed by a trained phonetician at key point in the F0 contours and at segmental landmarks.

2.2. Results

General shapes and F0 heights Figure 1 shows the F0 contours of all tokens of each tune type elicited in Corpus 1. The lines trace the F0 values of each frame between the peak of the preceding accented word and the peak of the target rise (pause cases) or the phonological low after the fall (no-pause cases).

Contours in the pause condition show 3 distinct contours shapes:

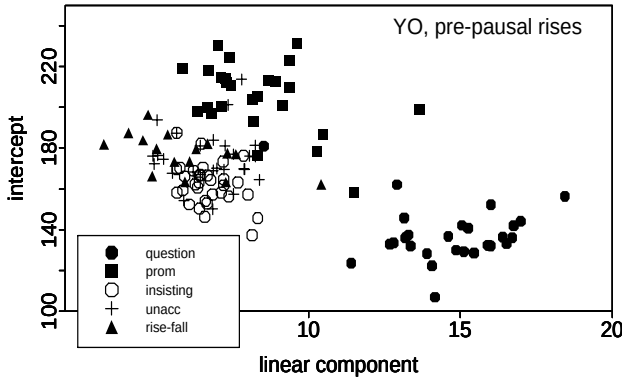

Figure 2: Linear components and intercepts of lines fit to rise portion (rise start ("elbow") to rise end).

1) the question contours fall to a low level and remain there for an extended duration before rising sharply to a high F0 target height at the phrase edge. 2) The prominence-lending, insisting, and unaccented word contours do not fall so low, and rise right at the onset of NI to a lower F0 height. 3) The rise-fall contours start off similar to class 2), but there is also a subsequent F0 fall. The no-pause contours are all similar to one another, with the exception of the question rises which show an extended low region before the rise.

Slope and shape of rise The differences in contour shapes can be shown more precisely by a quantitative analysis of the slopes and shapes of the rises. Figure 2 shows a scatter plot of the linear components and intercept values of lines fit to the rise portions of the contours (pause condition). The value of the linear component gives an idea of the slope (steeper slopes have higher values), and the intercept value reflects the timing of the rise onset (assuming similar slopes). This figure quantifies what can be visualized in Figure 1: that question rises have steeper slopes than the other types, and the rises are more delayed. Prominence-lending rises also pattern separately from the others, in that their higher intercepts indicate that the rise starts earlier. In the no-pause condition (plot not shown here), the only observable pattern is the delayed rise in questions (the slopes did not differ in this case).

Figure 3 shows a scatter plot of the linear and quadratic components of quadratic functions fit to the rises. The value of the quadratic component gives an idea of the curvature of the rise (positive=concave, negative=convex, and zero=linear). This method of quantifying shape also shows that question rises pattern apart from the other types: their shape is more concave (due to the long low stretch preceding the rise). In the no-pause case, it is also only questions that pattern separately.

Duration and timing Figures 1-3 above show that F0 height, slope, and shape of the rise are important for distinguishing question rises from the other types. Figure 4 shows the durations of the phones in the target mora NI, and the time-course of the rise with respect to these durations. Both pause and no-pause cases show similar effects: question and rise-fall types show a marked lengthening of the vowel /i/, while the durations of their nasal onsets are not lengthened and are comparable with the other types. As for the alignment with the F0 contour, non-question types start to rise within the onset consonant. In questions, the rise begins

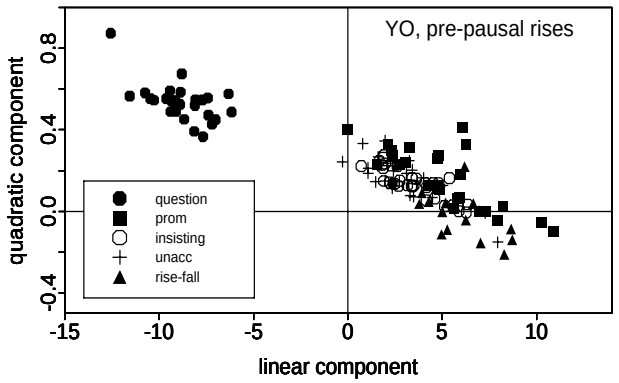

Figure 3: Linear and quadratic components of curves fit to rise portion (onset of NI to rise end).

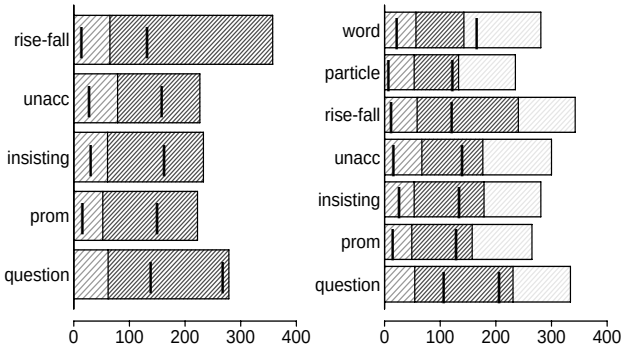

Figure 4: Mean durations (ms) of /n/ and /i/ (pause cases, left) and also /wa/ (no-pause cases, right). Vertical bars mark mean locations of rise start ("elbow") and rise peak.

well into the vowel portion. The duration of the rise (elbow to peak) is remarkably consistent across types: there is no significant difference in pause cases [$F(4,125)=2.851, p>.01$], and with the exception of the accented word type, there is no difference in no-pause cases either. While the peak location varies across types, the fact that the rise duration is invariant suggests that the rise is "anchored" by its onset at the beginning of the target mora in non-questions, and by its offset at the end of the mora in questions. This invariance also explains the differences in slope observed in Figure 2: the question contours are realized in an expanded range (the low onset is lower and the high offset is higher than in other contours), and this range combined with the invariant duration produces a steeper slope.

3. CORPUS 2

3.1. Design

Data from Corpus 1 show that question rises pattern separately from the other boundary pitch movements and accented/unaccented phrases. In the course of our analysis, we have noticed that the context used in the question mini-dialogs elicited a response that can be considered *incredulous*, and this incredulity could be the reason behind the marked shape of this rise, rather than any intrinsic feature of questions.

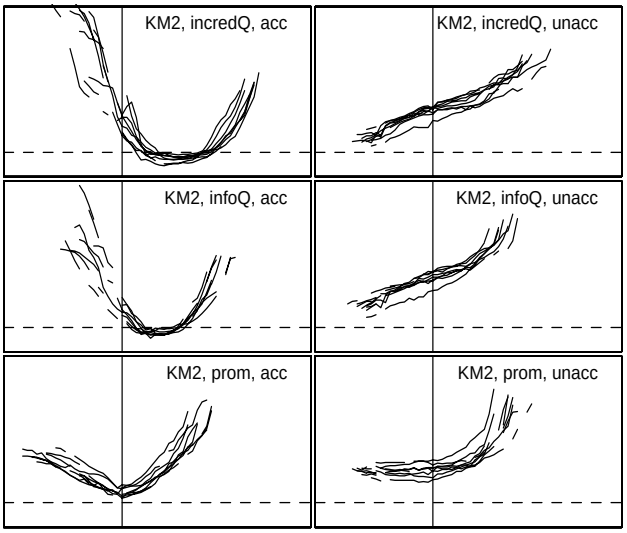

Figure 5: F0 shapes of the contour types in Corpus 2, preceding word accented (left) and unaccented (right).

In order to investigate the contribution of incredulity to the rise shape, we designed a small follow-up study which included three boundary pitch movements types in the pause condition: incredulity question (*incredQ*), confirmation question (*infoQ*), and the prominence-lending rise (*prom*). Also, we examined accented *Na'oya* vs. unaccented *Manami* preceding nouns. Speaker KM recorded 7-8 repetitions of the different types, following a procedure similar to that described in Section 2.1.

3.2. Results

F0 shapes Figure 5 shows the F0 contours of all tokens of each tune type elicited in Corpus 2. In unaccented cases (right), the plotted contour starts at the middle of the 2nd mora, instead of at the preceding "peak", as there was no observable peak in these cases. This graph shows that both question types are distinct from the prominence-lending rise, in that they include an extended portion of low F0 before the rise at the end of the phrase. In contrast, the prominence-lending type begins to rise (from a higher F0 value) at the onset of the target mora NI. Both results are consistent with Corpus 1.

Duration and timing Figure 6 shows segment durations in the target mora NI, and the locations of the start and end of the rise (the start of the rise is not marked for unaccented questions, since no elbow was observed). While only the incredulity type has a marked lengthening, the timing of intonational events (start and end of rise) is similar for both questions. As in Corpus 1, the start of the prominence-lending rise is within the nasal onset. In the question cases, the rise starts in the vowel portion, and appears to be anchored to the end of the phrase. In this corpus, the rise durations of the two question types are not different, though the rise in the prominence-lending case is longer. These results suggest that the distinct features of questions observed in Corpus 1 are representative of questions in general, be they incredulous or purely confirmation seeking.

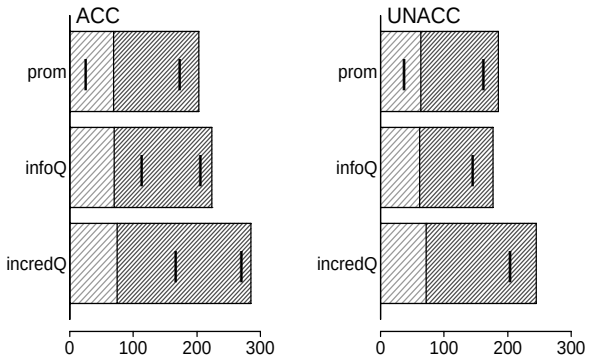

Figure 6: Mean durations (ms) of /n/ and /i/ in preceding accented (left) and unaccented (right) cases. Vertical bars mark mean locations of rise start ("elbow") and rise peak.

4. DISCUSSION

In this paper, we report on F0 shapes, durational patterns, and alignment characteristics of four Japanese boundary pitch movements, and comparable accented and focused unaccented words. Based on these phonetic descriptions, we can attempt to evaluate the potential phonological accounts introduced in Section 1.

In addition to these three hypotheses, one other possibility arises. The long, very low F0 stretch in the question contours suggests that this rise type is in fact a LH% boundary tone, in contrast to a simple H% in the prominence-lending and insisting types. We call this hypothesis LH-BT.

Some of our data support hypothesis LH-BT, while others do not. First, the lengthened vocalic durations of questions is similar to rise-fall type (hypothesized to be HL%). In Tokyo Japanese (unlike Kansai Japanese), a contour-tone cannot be realized on a single mora. According to this restriction, the contour in the rise-fall case could be described as a bimoraic structure which realizes the H on the 1st mora (very similar to the rise in other types), and the L on the 2nd mora. Similarly, one could think of the question rises as shapes in which the target NI is lengthened to accommodate the two LH tones. Further support for this hypothesis is the fact that the low before the rise in questions is noticeably lower than in the other types, suggesting a separate low target in these cases. However, the fact that questions preceded by unaccented words (cf. Figure 5) show no low valley argues against this hypothesis. Unless the low of the LH% is severely undershot in these cases, the LH-BT hypothesis is untenable.

Hypotheses ACC and PHR-H propose that boundary rises can be thought of as the first half of an accented or focused unaccented accentual phrase, respectively. In our data, we found that non-Q boundary rises pattern together with *both* accentual phrase types (focused unaccented words were not distinct from accented words), in that the duration of the rise is invariant, and the F0 heights are comparable. To accomodate the question rises, one would need to account for the differences in rise alignment and expanded range (which is described to some extent in [2]).

Finally, hypothesis H-BT claims that each boundary rise can be represented by a H%. The potential problem with this analysis which is pointed out in [9], is that questions rise to a high level (aka. highH%), while insisting and prominence-lending types often rise to a mid level (aka. midH%). This difference could be accounted for in the J_ToBI approach by indicating that the overall pitch range in questions is expanded, while it is not in the other cases. The possibility that differences only in range (keeping the same phonological representation of tune) can cue different pragmatic meanings has a precedent in studies in other languages (e.g. [3, 4]. However, while the differences in F0 height can be accounted for by H% scaled in different pitch ranges, the durational differences cannot. A phonological transcription system such as Japanese ToBI is impoverished in that it concentrates heavily on the tonal aspect, and does not take into account the crucial contribution of categorical duration and alignment facts to the realization of the contours. As described above, the question rises are clearly distinct from the other two boundary rises in their durational/alignment characteristics.

We conclude that while the LH-BT account seems unlikely, the other proposals could explain the current data, assuming that categorical durational and alignment facts are incorporated, in addition to the description of F0 heights.

5. REFERENCES

1. Beckman, M. E. The parsing of prosody. *Language and Cognitive Processes 11* 1996, 17–67.
2. Fujisaki, H., Ohno, S., Osame, M., Sakata, M., and Hirose, K. Prosodic characteristics of a spoken dialogue for information query. In *International Conference on Spoken Language Processing (ICSLP)* (1994), pp. 1103–1106.
3. Hirschberg, J., and Ward, G. The influence of pitch range, duration, amplitude and spectral features on the interpretation of the rise-fall-rise intonation contour in English. *Journal of Phonetics 20*, 2 1992, 241–251.
4. Jun, S.-A., and Oh, M. A prosodic analysis of three types of Wh-phrases in korean. *Language and Speech 39* 1996, 37–61.
5. Kawakami, S. On phrase-final rising tones. In *A Collection of Papers on Japanese Accent.* Kyûko Shoin, Tokyo, 1995, pp. 274–298. [Originally published in 1963] (in Jpns).
6. Kori, S. Japanese intonation: Form and function. In *Japanese Phonetics 2: Accent, Intonation, Rhythm and Pause*, e. a. Tetsuya Kunihiro, Ed. Sanseido, 1997, pp. 169–202. (in Jpns).
7. Maekawa, K. Transmission of paralinguistic information by speech: From a linguistic point of view. In *Proceedings of the Acoustical Society of Japan (ASJ)* (1997). (in Jpns).
8. Muranaka, T., and Hara, N. Features of prominent particles in Japanese discourse: Frequency, functions, and acoustic features. In *International Conference on Spoken Language Processing (ICSLP)* (Yokohama, Japan, 1994), pp. 395–398.
9. Venditti, J. J. Japanese ToBI labelling guidelines, 1995. [http://ling.ohio-state.edu/Phonetics/J_ToBI/jtobi_homepage.html].
10. Venditti, J. J., Maeda, K., and van Santen, J. P. H. Modeling Japanese boundary pitch movements for speech synthesis. In *Proceedings of the 3rd ESCA Workshop on Speech Synthesis* (Jenolan Caves, Australia, 1998).