Revisiting focus prosody in Japanese*

Albert Lee, Yi Xu

Department of Speech, Hearing and Phonetic Sciences, University College London

kwing.lee.10@ucl.ac.uk, yi.xu@ucl.ac.uk

Abstract

This paper revisits Japanese focus prosody by comparing the realization of contrastive focus vs. neutral focus across 4 focus positions and 2 accenthoods. Post-focus F0 Range compression (PFC) is found only in accented stimuli, while focus in unaccented stimuli is marked by Min F0 lowering. Penultimate focus differs from other focus positions, in terms of its post-focus behavior. Intensity and Duration manifest patterns of focus marker distribution comparable to those of F0. These results combined confirm that focus realization interacts closely with accenthood and focus position, and have implications that may enhance our understanding of Japanese pitch accent in general.

Index Terms: focus, post-focus compression, Japanese

* This paper is adapted from the first author's unpublished MRes dissertation (University College London, 2011) under the supervision of the second author.

1. Introduction

The prosody of focus in Japanese has been extensively studied, especially within the Autosegmental-Metrical framework (e.g. [1], [2]). In general, it is agreed that Japanese marks focus with on-focus fundamental frequency (F0) raising and post-focus reduction. However, most studies have only looked at sentence-medial focus, leaving our understanding of focus in other positions rather limited. The present study therefore revisits this issue with a more comprehensive experiment design, by including focus position and also accenthood (i.e. initial accent vs. unaccented) as variables. This allows us to verify whether our current understanding of focus realization holds true in different contexts.

Previous studies on Japanese focus have mainly looked at how it is formally represented in phonology. One popular view is that focus causes (i) insertion of an Intermediate Phrase (or Major Phrase) boundary, which in turn leads to pitch resetting and blocks downstep, and (ii) post-focus dephrasing which reduces F0 movement. This view has since been modified. Specifically, it has been found that pitch is in fact not reset under focus [3], putting the Major Phrase boundary insertion claim in question. This leads Kubozono [4], for example, to look into the possibility of an additional layer of phrasing exclusively for focus.

Recently reignited attention and work on this topic have revealed more interesting details: Post-focus compression (PFC) is found in the first post-focus initial rise regardless of accenthood [5], contrastive focus appears to manifest pre-focus compression while WH-question does not [6], and, although only indirectly relevant to our present discussion, in Turkish PFC takes place only after initial focus [7].

2. Methodology

2.1. Stimuli

Our stimuli consist of 2 **non-sense** sentences (Table 1), one with 4 quadri-moraic words with initial accent (AA), and the other without accent (UA). The 4 words in each sentence take turn to bear contrastive focus, each elicited by a leading question that contains a piece of wrong information in the corresponding position. For example, to elicit initial focus in the AA sentence, the leading question would be '*Okada-ga* *kuru-made ninki-o kaiteru no*?', where the first word is UA. This yields 10 question-answer pairs (i.e. neutral focus in addition to the 4 focus positions), each of which is repeated 10 times, totalling 100 tokens from each speaker.

AA (HLLL)	緒方が来るまで任期を書いてる Ogata-ga / kuru-made / ninki-o / kaiteru Ogata-NOM / come-until / *ninki*-ACC / writing
UA (LHHH)	岡田が車で人気を欠いてる Okada-ga / kuruma-de / ninki-o / kaiteru Okada-NOM / car-by / popularity-ACC / lacking

Table 1. *Target sentences used in this study*

The 2 target sentences differ from each other (almost) only in terms of accenthood. Our intension is to avoid any potential confound related to intrinsic fundamental frequency. Nevertheless, by doing so, admittedly we have not been able to control for syntactic structure. For example, while Word 2 in UA is a noun, that in AA is a verb, which arguably has a different intonation pattern from other word classes. That said, because the goal of this study is to look at contrastive focus vs. neutral focus, differences in syntactic structure is unlikely to affect our results.

2.2. Procedures

Data from 7 native speakers were analyzed. These speakers come from the Greater Tokyo area (Tokyo, Saitama, Kanagawa, and Chiba) where Tokyo Japanese is spoken. They were 5 female and 2 male, and their ages ranged from 24 to 38 (mean age 32.14). Recording took place in a soundproofed room at UCL, where subjects were seated in front of a computer screen. A microphone was placed approximately 20 centimetres in front of them. The subjects were briefed beforehand and were given time to familiarise themselves with the target sentences.

During the recording, target sentences were displayed on the computer screen with a Javascript-based sentence randomizer. The subjects were asked to read aloud each leading question and its answer in pairs, in order to elicit the correct prosodic focus. Where the experimenter deemed that an error (e.g. wrong focus position, wrong lexical accent, placing focus in a neutral focus sentence) had occurred, the subject would be alerted to the error, given time to practise, then reattempt. If the error persisted after 3 retrials, the trial closest to the experimenter's expectation would be accepted for further analysis. All cases of such persistent "failure" were related to subjects not producing the sentence final verb with the correct citation accenthood.

2.3. Data extraction

The recordings were analyzed using a custom-written Praat script [8]. The script allows manual rectification of the markings of individual vocal pulses before generating the final data for analysis. When the script was run, two windows, one with vocal pulse markings and the other with text grid and the spectrogram, were displayed. The vocal pulse markings generated by Praat were inspected and corrected manually when necessary. In the text grid window, boundaries can be added at any time point to mark syllables/segments/ utterances, depending on the purpose of the experiment. In the present study, boundaries were added to mark vowels and the moraic nasal but not initial consonants.

Note that data of the last 2 moras of some utterances were not obtainable. Of our 700 tokens, it was not possible to label the final mora *-ru* for 43, nor the final 2 moras *-teru* for a further 7. This is because many of the female informants produced these 2 moras with creaky voice, resulting in absence of visible vocal pulses. Anecdotally, Konno and colleagues [8] also point out that towards the end of utterances in Japanese, mora duration tends to shorten. This too contributed to the failure to obtain some of the utterance-final data.

3. Results

3.1. F0

Below Figure 1 and Figure 2 show the averaged F0 contours of all tokens. The dark dashed curve represents neutral focus, whereas the other curves represent various focus conditions, each identifiable by its peak location. Time (normalized) is plotted along X axis, while Y is F0 in Hz. The vertical grid lines show word boundaries.

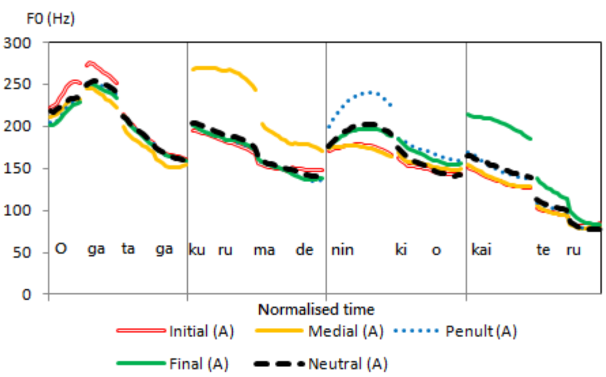

Figure 1. *Averaged F0 contours of accented stimuli*

Note that in Figure 2 there is a falling contour from the beginning of Word 4, where the green curve starts to diverge from neutral focus. This is because many speakers, despite our explicit instructions, produced an accented (HLLL) version of the final verb (47% of all tokens), instead of the unaccented (LHHH) citation form. Informants reported that in the given context the HLLL intonation was easier to say, even though they knew LHHH was the correct form.

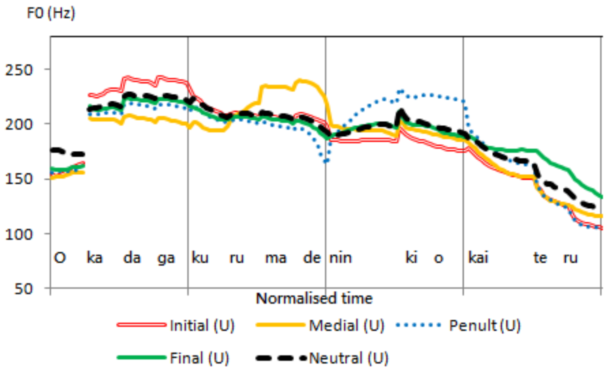

Figure 2. *Averaged F0 contours of unaccented stimuli*

Results from ANOVA reveal that, the main effect of focus condition is non-significant on pre-focus **F0 Range**, F(1,6)=0.597, p=0.469; **Max F0** F(1,6)=5.734, p=0.054; or **Min F0** F(1,6)=4.281, p=0.084. This non-significance of main effects means that F0 Range, Max F0 and Min F0 are not significantly different before focus and when they are under neutral focus.

However, one may argue that, judging from Figure 2 and results from a paired-sample t-test, pre-focus Max F0 lowering should not be discarded as a reliable focus marker in Japanese. Indeed, with data of our accented stimuli removed, the main effect of focus condition immediately becomes significant, F(1,6)=7.696, p=0.032. That said, if individual variability is taken into account, out of 21 cases of pre-focus vs. neutral focus contrast (i.e. 7 speakers * 3 focus positions * 1 accenthood), 5 (23.81%) do not show pre-focus lowering. Thus, although it may be associated with focus in this language, it is at best an optional strategy that is consistently used. See also Xu and colleagues' work [9], [10] for inconsistent pre-focus patterns in Mandarin and English.

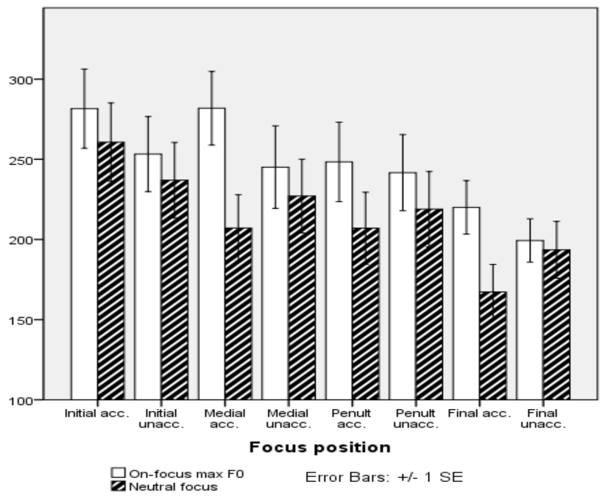

Figure 3. *On-focus Max F0 vs. Neutral focus (unit: Hz)*

Meanwhile, on the focused item, there is consistent expansion of **F0 range** (F(1,6)=44.992, p=0.001), as well as raising of

Max F0 (F(1,6)=68.143, p<0.001) and **Mean F0** (F(1,6)=37.011, p=0.001). However, focus does not seem to affect **Min F0** in any observable way. These observations are found across all 4 focus conditions and in both accenthoods, with the exception of unaccented final focus, where no significant on-focus F0 range expansion is found.

After focus, F0 behavior becomes more complex. Both accenthood and focus condition appear to influence how focus is marked by pitch. **Post-focus Mean F0** is significantly lower than its neutral focus counterpart, F(1,6)=10.190, p=0.019, except for post-penultimate focus, of which contrast with neutral does not reach statistical significance. **Post-focus Max F0** is lowered in non-penultimate accented focus, F(1,6)=7.034, p=0.038, with individual variations. **Post-focus Min F0** is consistently lower than neutral in the unaccented stimuli, F(1,6)=6.363, p=0.045. Finally, post-focus **F0 Range** compression is found only in the accented stimuli, but again, except post-penultimate focus, F(1,6)=6.667, p=0.042. In sum, accenthood plays a big role in determining how pitch behaves after focus, and in the sentence final position, post-focus F0 behavior is often not manifest.

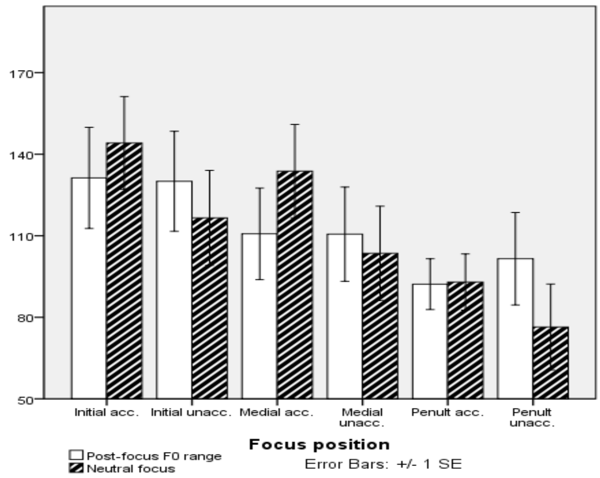

Figure 4. *Post-focus F0 Range vs. Neutral focus (unit: Hz)*

3.2. Intensity

The distribution of intensity-related focus markers is similar to that of F0 above. Specifically, pre-focus markers, if at all statistically significant, have high individual variations; and post-focus intensity behavior is not distinguishable from neutral at the sentence-final position.

Like in F0, intensity before focus is not consistently different from neutral. **Pre-focus Max Intensity** is significantly lowered, F(1,6)=10.698, p=0.017), but there are also individual speakers (19.05%) who show no lowering. **Pre-focus Mean Intensity** is consistently lower than neutral across all focus positions excluding unaccented final focus, thus the main effect of focus condition is highly significant, F(1,6)= 10.374, p=0.018. However, even so there exist individual variations (21.43%).

On-focus Max Intensity (F(1,6)=58.973, p<0.001) and **Mean Intensity** (F(1,6)=176.931, p<0.001) are significantly higher than their neutral focus counterpart. On the other hand, although **on-focus Min Intensity** is raised (F(1,6)=31.479, p=0.001), individual cases of non-raising (26.79%) also abound. Since both Max and Min Intensity are raised, **On-focus Intensity Range** is not significantly different from neutral.

After focus, **Max Intensity** (F(1,6)=27.729, p=0.002) and **Mean Intensity** (F(1,6)=37.455, p=0.001) are lowered, except after penultimate focus. For **Min Intensity**, focus condition has a significant main effect in the unaccented stimuli, F(1,6)=9.449, p=0.022.

3.3. Duration

Finally, Duration again shows patterns comparable to F0 and Intensity. On-focus Duration is significantly lengthened, F(1,6)=56.526, p<0.001, confirming results in previous studies (e.g. [11], [12]). Meanwhile, post-focus Duration is shortened, F(1,6)=14.077, p=0.009. Note, however, the final position behaves differently from other positions in both on-focus and post-focus Duration. Whether this is due to informants' creaky voice in the final verb, or that the final verb does behave differently will require further investigation. The present results conflict with previous findings [13] that pre-focus duration is shortened.

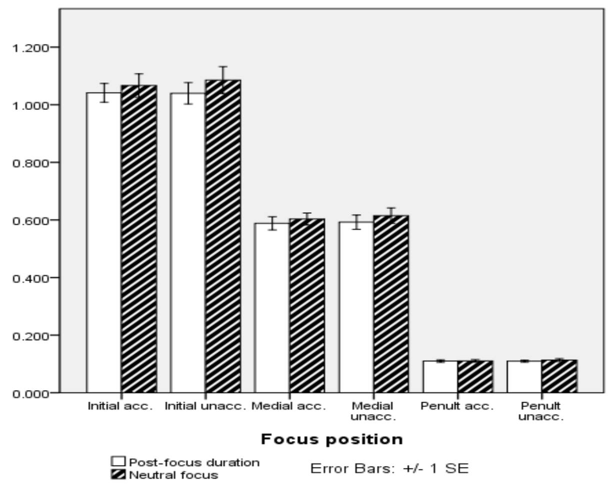

Figure 5. *Post-focus Duration vs. Neutral focus (normalized duration)*

4. Discussion

Our conclusion that pre-focus F0 behavior is disqualified as a reliable focus marker is at odds with Hwang's findings [6], where pre-focus F0 range is reported to be compressed in Tokyo Japanese. There are discrepancies between our results for accented stimuli and hers. The most obvious difference in experiment design is that in the present study, contrastive focus is being compared with neutral focus elicited with a leading Yes/No-question; whereas for Hwang, contrastive focus is compared with the "given" stimuli elicited by a Yes/No question. In addition, 7 speakers were analyzed in the present study, while Hwang tested four. The same is true for Duration, that our results do not see pre-focus shortening reported in Maekawa [13]. Although in our own data, there are instances of individual speakers producing pre-focus F0 range compression and shortening, the combined picture of all speakers does not show this. Until further tests with a larger number of speakers show the contrary, we posit that pre-focus F0 and duration are not a reliable cue to focus in Japanese.

On the other hand, that the sentence-final position manifests little post-focus F0 behavior is in line with results in a comparable study on Turkish [7]. Ipek reports significant post-focus lowering after initial focus, but not after medial focus. She suggests that in SOV languages, the sentence-final object-verb sequence naturally forms a falling F0 contour which resembles post-focus intonation, rendering focus indistinguishable from neutral near the end of a sentence. It would be interesting to see if other SOV languages also show a comparable behavior. Meanwhile, for Japanese it seems that as far as F0 encoding of focus is concerned, only the final position behaves differently. Our data shows that other positions generally behave in the same way, in line with what has been reported in the literature.

Our observation that PFC is absent after an unaccented focus confirms Ishihara's view [14]. In other words, PFC is present after an unaccented focus, but only very subtly, i.e. in the first post-focus initial rise [5]. When the whole post-focus domain is considered, as is the case in this study, no PFC is found in our UA stimuli. This of course poses problems for representational frameworks, but for phoneticians this is no less puzzling. To claim that PFC applies to different scopes under different accentual environments would be hard to sustain, because so far this has not been attested in enough languages. An alternative account would be to argue that PFC always applies to the first initial rise in Japanese, but that would be equally arbitrary. Perhaps the next step is to revisit the underlying F0 target of the pitch accents in the language, to find out what is so special about unaccented words that lead to such different focus intonation.

5. Conclusion

This paper revisits the F0/Intensity/Duration realization of Japanese focus by comparing 2 accenthoods and 4 focus positions. Our data confirms results in earlier studies on this topic, but also found that PFC is absent in our unaccented stimuli, and that the final verb behaves differently from other focus positions. Intensity and Duration show similar patterns to F0 in general. Also, pre-focus markers do not seem to be reliable cues to focus due to their high individual variability. The asymmetric focus behavior of different accenthoods calls for more work on the nature of Japanese pitch accents.

6. References

[1] Pierrehumbert, J. B., "The phonology and phonetics of English intonation", PhD Thesis, MIT, 1980.

[2] Pierrehumbert, J. B., & Beckman, M. E., "Japanese Tone Structure", MIT Press, 1988.

[3] Poser, W. J., "The phonetics and phonology of tone and intonation in Japanese", PhD Thesis, MIT, 1984.

[4] Kubozono, H., "Focus and intonation in Japanese: Does focus trigger pitch reset?", Interdisciplinary Studies on Information Structure (ISIS), 9:1-27, 2007.

[5] Sugahara, M., "Downtrends and post-focus intonation in Tokyo Japanese", PhD Thesis, University of Massachusetts, Amherst, 2003.

[6] Hwang, H. K., "Distinct types of focus and wh-question intonation", Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS 2011), Hong Kong, 2011.

[7] Ipek, C., "Phonetic realization of focus with no on-focus pitch range expansion in Turkish", Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS 2011), Hong Kong, 2011.

[8] Xu, Y., "ProsodyPro", University College London, Retrieved from www.phon.ucl.ac.uk/home/yi/ProsodyPro/, 2005-2011.

[9] Konno, K., Sugishita, M., & Hirose, H., "Prosody in Broca's aphasia: A pilot study", Annual Bulletin of the Research Institute of Logopedics and Phoniatrics, 18:169-177, 1984.

[10] Xu, Y., "Effects of tone and focus on the formation and alignment of f0 contours", Journal of Phonetics, 27(1):55-105, 1999.

[11] Xu, Y., & Xu, C. X., "Phonetic realization of focus in English declarative intonation", Journal of Phonetics, 33(2):159-197, 2005.

[12] Wu, W.-L., & Xu, Y., "Prosodic focus in Hong Kong Cantonese without post-focus compression", Proceedings of the 5th International Conference on Speech Prosody (SP2010), Chicago, IL, 2010.

[13] Lee, Y.-C., & Xu, Y., "Phonetic realization of contrastive focus in Korean", Proceedings of the 5th International Conference on Speech Prosody (SP2010), Chicago, IL, 2010.

[14] Maekawa, K., "Effects of focus on duration and vowel formant frequency in Japanese", In Y. Sagisaka, N. Campbell, & N. Higuchi [Ed], Computing prosody: Computational models for processing spontaneous speech, 129-153, Springer, 1997.

[15] Ishihara, S., "Focus prosody in Tokyo Japanese wh-questions with lexically unaccented wh-phrases", Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS 2011) , Hong Kong, 2011.

7. Acknowledgements

We thank Dr. Shinichiro Ishihara (Uni-Frankfurt) for valuable comments on the design of this experiment, and Dorothea Hackman (UCL) for comments on an earlier draft. All inadequacies and errors remain our own.