



Conditional realisation of post-focus compression in Japanese

Albert Lee¹, Yi Xu²

¹University of Tokyo, Japan

²University College London, United Kingdom

alee@aless.c.u-tokyo.ac.jp, yi.xu@ucl.ac.uk

Abstract

The typological feature ‘post-focus f_0 range compression’ (PFC) is often considered an all-or-nothing phenomenon, being completely absent in -PFC languages but applied across-the-board in +PFC languages. This paper presents production data from Japanese and shows that, within a language, the realisation of PFC can be conditional upon lexical prosody. We further argue that this condition itself is language-specific, by comparing Japanese PFC with that in other +PFC languages. It is hoped that these results will further our understanding of the nature of focus intonation across languages.

Index Terms: Japanese, pitch accent, post-focus compression

1. Introduction

Post-focus compression (PFC) of fundamental frequency (f_0) range is a prosodic phenomenon commonly observed across languages. In general, it refers to the compression of f_0 range after prosodic focus, but exactly how it is phonetically realised varies across languages. In American English, for example, it is realised as compressed and lowered f_0 range for statements, and as compressed but raised f_0 range for questions (e.g. [1]). On the other hand, in utterances without much f_0 fluctuation, like a Mandarin statement that consists of only High tone words [2], PFC can manifest itself as mean f_0 lowering instead. Thus when examining whether a language exhibits PFC, functions that may interact with focus must be taken into account or the wrong conclusion about focus marking could be reached. To the best of our knowledge, the term was first explicitly used in [3] where she measured the excursion size of the Japanese Accentual Phrase (AP)-initial rise; elsewhere, Xu [4] used the term ‘suppressed’ to describe lowered and compressed f_0 range after focus in Mandarin.

Since [5] PFC has been increasingly studied from a typological perspective using largely comparable methods. Comparing the focus markers in Taiwan Mandarin vs. Beijing Mandarin, two mutually intelligible dialects, it was found that while PFC was present in Beijing Mandarin, it was absent in Taiwan Mandarin, which is in close contact with Taiwanese, a -PFC language. This led to the proposal that PFC is a phonetic feature easily lost through contact. Subsequently, a typological border between +PFC and -PFC languages started to emerge as studies in this line of research began to fill up the geographical map.

However, recent findings have cast doubt on the all-or-nothing assumption of PFC realisation. In particular, [6]–[8] reported that in Japanese the realisation of PFC required a preceding lexical pitch accent. The theoretical implication of these findings is that even if a language is typologically +PFC, PFC is not necessarily applied across the board and can be subject to certain conditions, which in Japanese is a preceding

accented word. Likewise, it has been reported that in Mandarin an early focus does not lower the Low tone significantly [4].

That said, much remains unknown about the details of the interaction between word (i.e. pitch accent condition) and sentential prosody. How is PFC realised in Japanese under different combinations of pitch accent conditions? How does sentence length come into play? How do we understand this interaction based on what we know about other languages? Here we present a production experiment to fill these gaps. We argue that PFC, as the difference between AP (~word)-internal maxima and minima (as in [6], [7]), is realised only after a pitch accent. In turn, we claim that although typologically PFC is a focus marker in Japanese (i.e. a +PFC language), it does not follow that PFC is applied across the board in this language. We illustrate our argument through a carefully collected corpus that controls for the accent condition of all words in the sentence, focus condition, and word length.

2. Methods

We conducted a production experiment in London with 13 native speakers from the Greater Tokyo area (Tokyo, Saitama, Kanagawa, Chiba). Of these speakers one turned out to be an English-Japanese parallel bilingual, one withdrew from the experiment after deciding the target sentences were difficult, and another completed the task but did not produce most of the contrasts we intended to elicit (e.g. focus); the data from these three speakers were thus discarded. Here we report data from the ten remaining speakers, five from each sex, aged between 24 and 36 (mean 30.3). All participants were remunerated a small sum for their time, and granted their written consent to being tested. This study has been approved by the UCL Research Ethics Committee [Ref. # SHaPSetXU002].

Altogether 128 target sentences (2 sentence lengths \times 8 accent conditions \times 2 sentence types \times 4 focus conditions) were elicited (see Table 1). There were four possible focus conditions for each target sentence, namely Initial, Medial, Final, and Neutral; since all target sentences contained three words, Medial focus here was equivalent to penultimate focus in comparable studies. Each target sentence was repeated four times (i.e. 5 utterances). Focus was elicited with a leading question that contained a piece of inaccurate information in the word location of interest, and the speaker was to say the leading question and the target statement in pair in order to elicit narrow corrective focus. Neutral focus was elicited with a leading question that contained all accurate information (thus yielding no corrective focus). From the 6,400 utterances collected, 149 had to be discarded due to mis-production of accent condition. A total of 6,251 were retained for statistical analysis. For a more focused discussion, in this paper we report only the subset of results concerning statements.

Table 1: Target sentences used in this study (accented mora is marked with a following apostrophe).

		Word I	Word II	Word III
Short	A	<i>me'i-ga</i> May-NOM	<i>mo'mo</i> thigh	<i>-o mi'ta</i> -ACC saw
	U	<i>Mei-ga</i> Niece-NOM	<i>momo</i> peach	<i>-ni nita</i> -DAT resembled
Long	A	<i>mu'umin-ga</i> Moomin-NOM	<i>bu'dou</i> martial arts	<i>-o mi'ta</i> -ACC saw
	U	<i>noumin-ga</i> Farmer-NOM	<i>budou</i> grapes	<i>-ni nita</i> -DAT resembled

Although admittedly lab speech can be exaggerated and have extra boosted foci, we had to use meaningless sentences in order to ensure that all observed contrasts are minimal and free from microprosodic confounds like segmental perturbation and vowel-intrinsic f_0 . We also used nasal initial consonants as much as possible to unify the effect of segmental perturbation.

Recording took place in a sound-proofed room in University College London. Participants were seated in front of a computer, which displayed one question-statement pair at a time. The stimuli were presented in random order, and the repetitions were collected over four random occasions. Stimuli were presented in standard Japanese orthography (mixed use of *hiragana*, *katanana*, and *kanji*), with the focused item underlined and boldfaced. Participants were given oral instructions about the task, and time to practice before recording began. The raw sound data were first chunked into individual utterances, and then segmented by the mora on Praat [9]. A heavy syllable was segmented into two morae equal in duration. Vocal pulse markings were manually checked and rectified. The segmented data were then fed into ProsodyPro [10] to extract acoustical measurements. Raw f_0 values were converted to semitones with speaker-mean f_0 as reference. The resultant acoustic data were subsequently analyzed using Smoothing Spline ANOVA (SS ANOVA) [11], [12] in order to assess if the f_0 contours of different focus conditions differed from one another at different points in time.

3. Results

This experiment set out to explore the details of how word prosody (i.e. pitch accent condition) affects the phonetic realisation of PFC in Japanese. Each of the SS ANOVA plots below represents one target sentence, and the solid lines are the smoothing splines that best fit the 50 repetitions of each of the four focus conditions (i.e. Initial, Medial, Final, Neutral) from 10 speakers. The shading superimposed on the splines are 95% Bayesian confidence intervals. At any point along the x-axis, if two f_0 contours overlap, they are not significantly different from each other. The vertical dotted lines are word boundaries.

Fig. 1 shows the f_0 contours of the target sentences in which all three words are accented (henceforth AAA) whereas Fig. 2 shows the target sentences where all words are unaccented (UUU). Fig. 3 represents the UUA condition in which the first two words are unaccented while the last one is accented. For all figures the upper panel is the short sentence (8-mora long) whereas the lower panel is the longer version (11-mora long). Refer to Table 1 for details.

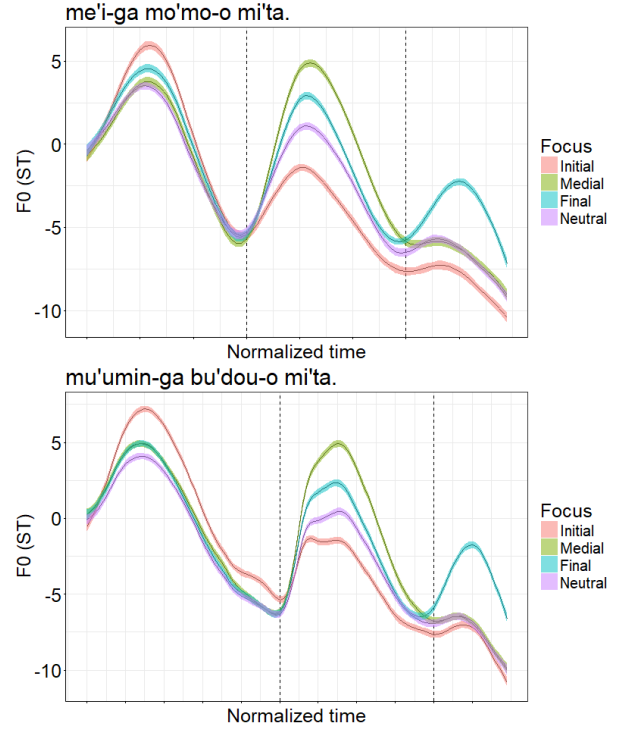


Figure 1: Time-normalised f_0 contours of an AAA sentence under different focus conditions.

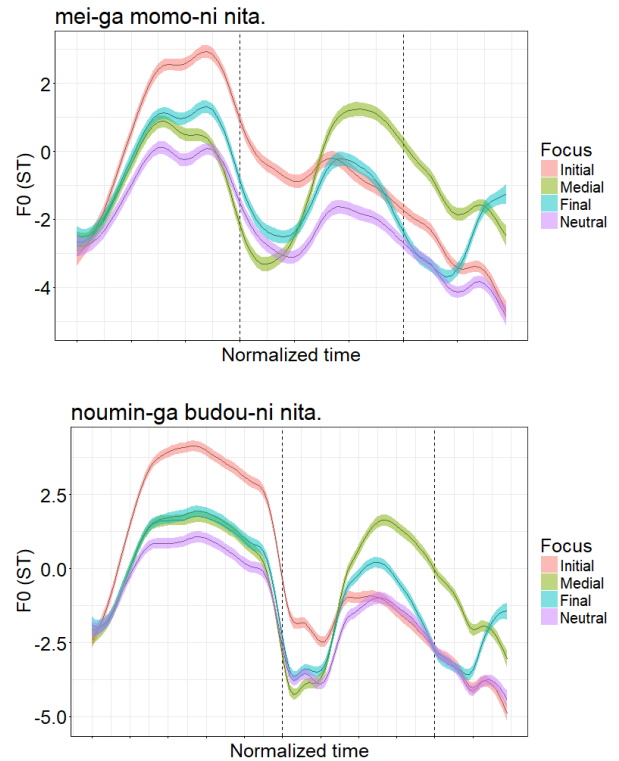


Figure 2: Time-normalised f_0 contours of a UUU sentence under different focus conditions.

Fig. 1 shows a typical case of PFC where the post-focus portion (second and third intervals) of the Initial Focus f_0

contours (pink) was significantly lower than that of its Neutral Focus counterpart (lilac). In line with previous studies (e.g. [6], [13]), there was also significant on-focus raising as indicated by a higher Initial Focus contour in the sentence-initial position (first interval from the left). Note that there was no PFC after Medial focus, like in Turkish [14] and Persian [15], as indicated by the overlapping Medial (lime) and Neutral (lilac) f_0 contours in the sentence-final position (third interval from the left). The same patterns were observed in both sentence length conditions.

The UUU condition represents the case where PFC is absent (Fig. 2). Contrary to the AAA condition above, here both Initial Focus (pink) and Medial Focus (lime) were significantly higher in f_0 than Neutral focus (lilac) in the post-focus region, as though there was on-focus raising. The two contours (i.e. Initial and Neutral) ultimately converged, after 5 morae (i.e. end of the utterance) in the upper panel and at the 8th mora (i.e. within the second interval) in the lower panel. Meanwhile, there was significant on-focus raising just like in the AAA condition.

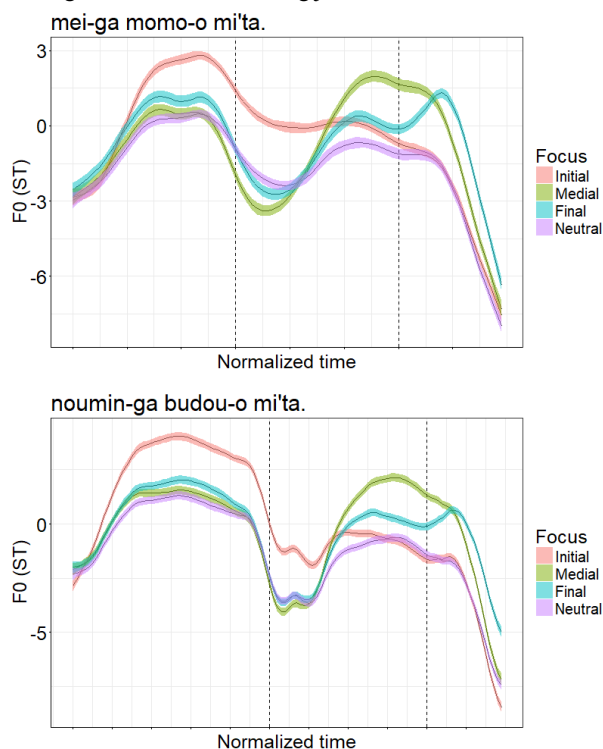


Figure 3: Time-normalised f_0 contours of a UUA sentence under different focus conditions.

The UUA condition (Fig. 3) shows that in the post-focus region, the Initial Focus and Neutral Focus contours converged at the accent peak (i.e. sharp turning point in the third interval) in the upper panel, in contrast to the UUU condition (Fig. 2 upper panel) where convergence only occurred at the end of the utterance. In the lower panel, like in Fig. 2, the convergence of Initial and Neutral foci occurred at the 8th mora.

4. Discussion

Several points can be concluded from the above results. Firstly, PFC was observed only after a pitch accent (cf. Fig. 1), in line with previous findings [6], [7]. Where there was no preceding accent, post-focus f_0 remained higher than Neutral Focus (cf. Fig. 2). Secondly, where PFC was absent in the UUU condition,

it took as many as five morae for the Initial and Neutral contours to converge (Fig. 2). Thirdly, where narrow focus and Neutral Focus diverged, a pitch accent in the post-focus region had the effect of drawing the two contours together (Fig. 3).

The first observation that there is PFC when there is a preceding pitch accent aligns Japanese with other +PFC languages. This observation was revealed by our experimental design that exhausted all logical combinations of accent condition in the target sentences. Our findings echoed those of previous studies that used different target stimuli and experimental procedures. What is more interesting, however, is the fact that PFC was absent in the UUU condition, confirming earlier reports [6], [7] that measured PFC in the ‘standard’ way (i.e. f_0 maxima less minima within a given post-focus domain). Just because Japanese is a +PFC language it does not mean that PFC must be applied across the board; there can be within-language variation with regards to the phonetic realisation of PFC. As far as Japanese is concerned, PFC is conditionally realised, after a lexical pitch accent.

Is the lack of PFC after an unaccented focus due to the absence of sharp turning points in a Japanese unaccented word? This possibility is unlikely. This is because even in an all-T1 (high level tone) utterance in Beijing Mandarin where f_0 should remain relatively constant, PFC is realised as post-focus lowering whereas in Taiwan Mandarin the same is not true. The interaction between word prosody and PFC is therefore language-, or more precisely dialect-specific, and not necessarily due to the shape of lexical prosody. In fact, as the context that permits PFC is also what gives rise to downstep in Japanese (i.e. a preceding pitch accent, see e.g. [16]), which is an integral part of its lexical prosody, one cannot rule out the possibility that PFC is also a strategy that Japanese speakers use to maintain lexical contrast in sentential prosody.

The sluggish convergence of narrow focus and Neutral Focus in Fig. 2 is reminiscent of the Neutral Tone in Mandarin. Although the phonetic realisation of the Mandarin Neutral Tone appears to be highly variable, there is evidence showing that it had an invariant target implemented with weak articulatory strength [17], [18]. This weak strength was argued to be associated with the sluggish approximation to articulatory targets. The fact that post-focus f_0 contour in the UUU condition took as many as five morae to converge with Neutral Focus may likely be attributable to a weak articulatory strength too, although such an account will need to be verified using electromyography or at least analysis-by-synthesis with strength being a factor, like in [17]; until the pitch target is reached, the surface f_0 contour remains similar to the previous target due to inertia (i.e. continuation of the on-focus target). In turn, the pitch accent that draws together f_0 contours in the post-focus region (cf. Fig. 3) may possibly be seen as a sign of greater articulatory strength. An alternative to the present conditional realisation account of PFC would be to maintain that it is an all-or-nothing feature. As a matter of fact, whether there is PFC in the UUU condition depends on how PFC is measured. In both of [6], [7], f_0 range was not found to be compressed after an unaccented focus if PFC was measured as the difference between f_0 maxima and minima within that word. As [8] describes it, where the focus is unaccented ‘the pitch contour exhibits a high plateau following the focal f_0 rise’ which, as illustrated in Fig. 2, resembles post-focus ‘raising’. In this ‘raised’ plateau there is observable but compressed AP-initial rise, which was taken by [3] as evidence of PFC. If measured as the excursion size from the beginning of the word (or AP) to the plateau [3], Japanese would be another neat example of

+PFC with across-the-board realisation but will lead to a language-specific definition of how PFC is measured.

Nevertheless, there are three reasons why our account is more attractive. Firstly, in most studies investigating PFC, this phonetic feature was measured as the difference between the f_0 maxima and minima of a given post-focus domain; restricting the measurement of PFC in the first half of the first post-focus ‘AP’ (as in [3]) would render Japanese incomparable with other languages. Relatedly, measuring PFC as the excursion size from the beginning of the word is at odds with the notion of focus trizone [19], as the beginning of the first post-focus word is also the end of the on-focus domain. According to [19], narrow focus controls the f_0 range of the on-focus, pre-focus and post-focus regions, each with a different strategy. Restricting PFC to the initial rise that spans the first half of the first word after focus would mean this phonetic feature is no longer purely post-focus but is looking across both the on-focus and the post-focus regions — confusing for those to whom the focus trizone notion matters and confounding a post-focus feature with the effect of on-focus raising. Thirdly, attributing f_0 movements in unaccented utterances to carryover effects means resorting more to phonetic and mechanistic notions and less to abstract ones, which is a desirable outcome of the division of labour between phonetics and phonology.

5. Conclusions

This paper reports a production study of Japanese focus prosody that comprehensively controlled for accent condition, focus condition and word length. Our findings generally agreed with previous research, that there is on-focus enhancement and post-focus reduction in various forms. However, based on our data, we propose that PFC in Japanese is only realised after a pitch accent, and does not include cases of compressed initial rise in an exclusively unaccented utterance. We also observed that the raised contour after an unaccented focus gradually converged with the neutral focus contour, and posit that these two focus conditions have same underlying pitch target, although the post-focus contour approximates this target at a very low articulatory strength.

6. Acknowledgements

This study was partially supported by the UCL Overseas Research Scholarship awarded to AL. Preliminary findings were presented at the 1st International Conference on Prosodic Studies (ICPS 1), Tianjin in 2015. Thanks are due to Shinichiro Ishihara (Lund) and Bronwen Evans (UCL) who provided constructive comments on an early version of this paper.

7. References

- [1] S. J. Eady and W. E. Cooper, “Speech intonation and focus location in matched statements and questions,” *J. Acoust. Soc. Am.*, vol. 80, no. 2, pp. 402–415, Aug. 1986.
- [2] Y. Xu, “Speech melody as articulatorily implemented communicative functions,” *Speech Commun.*, vol. 46, no. 3–4, pp. 220–251, 2005.
- [3] M. Sugahara, “Downtrends and post-focus intonation in Tokyo Japanese,” University of Massachusetts, Amherst, Amherst, MA, 2003.
- [4] Y. Xu, “Effects of tone and focus on the formation and alignment of F0 contours,” *J. Phon.*, vol. 27, no. 1, pp. 55–105, 1999.
- [5] Y. Xu, S.-W. Chen, and B. Wang, “Prosodic focus with and without post-focus compression: A typological divide within

- the same language family?,” *Linguist. Rev.*, vol. 29, no. 1, pp. 131–147, 2012.
- [6] A. Lee and Y. Xu, “Revisiting focus prosody in Japanese,” in *Proceedings of the 6th International Conference on Speech Prosody (SP2012)*, 2012, pp. 274–277.
- [7] S. Ishihara, “Focus prosody in Tokyo Japanese wh-questions with lexically unaccented wh-phrases,” in *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS 2011)*, 2011, pp. 946–949.
- [8] S. Ishihara, “Syntax-phonology interface,” in *Handbook of Japanese Phonetics and Phonology*, H. Kubozono, Ed. Berlin: Mouton de Gruyter, 2015.
- [9] P. P. G. Boersma and V. J. J. P. van Heuven, “Speak and unSpeak with PRAAT,” *Glott Int.*, vol. 5, no. 9/10, pp. 341–347, 2001.
- [10] Y. Xu, “ProsodyPro: A tool for large-scale systematic prosody analysis,” in *Proceedings of Tools and Resources for the Analysis of Speech Prosody (TRASP 2013)*, 2013, pp. 7–10.
- [11] L. S. Davidson, “Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance,” *J. Acoust. Soc. Am.*, vol. 120, no. 1, pp. 407–415, 2006.
- [12] C. Gu, “Smoothing spline ANOVA models: R package gss,” *J. Stat. Softw.*, vol. 58, no. 5, pp. 1–25, 2014.
- [13] J. B. Pierrehumbert and M. E. Beckman, *Japanese Tone Structure*. Cambridge, MA: Massachusetts Institute of Technology, 1988.
- [14] C. Ipek, “Phonetic realization of focus with no on-focus pitch range expansion in Turkish,” in *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS 2011)*, 2011, pp. 140–143.
- [15] M. Taheri Ardali and Y. Xu, “Phonetic realization of prosodic focus in Persian,” in *Proceedings of the 6th International Conference on Speech Prosody (SP2012)*, 2012, pp. 326–329.
- [16] M. E. Beckman and J. B. Pierrehumbert, “Intonational structure in Japanese and English,” *Phonol. Yearb.*, vol. 3, pp. 255–309, 1986.
- [17] S. Prom-on, F. Liu, and Y. Xu, “Post-low bouncing in Mandarin Chinese: Acoustic analysis and computational modeling,” *J. Acoust. Soc. Am.*, vol. 132, no. 1, pp. 421–432, 2012.
- [18] Y. Chen and Y. Xu, “Production of weak elements in speech: Evidence from F0 patterns of neutral tone in Standard Chinese,” *Phonetica*, vol. 63, no. 1, pp. 47–75, 2006.
- [19] Y. Xu, C. X. Xu, and X. Sun, “On the temporal domain of focus,” in *Proceedings of the 2nd International Conference on Speech Prosody (SP2004)*, 2004, pp. 81–84.