

Focus prosody in Japanese reconsidered

Shinobu Mizuguchi & Koichi Tateishi*

Abstract. The purpose of the paper is to examine how focus in Japanese is realized and perceived in various accentual conditions. Japanese is a pitch accent language where lexical items are divided into two groups: Accented (A) and Unaccented (U). Pierrehumbert and Beckman (1988) first pointed out that the focus prosody of A words and U words are different. Much work has been done on narrow focus in A words following their work. Some previous studies (Pierrehumbert and Beckman (1988), Ishihara (2003) among others) suggest that, for A words, focus boosts the pitch of the accent peak and triggers either compression of the pitch range and/or Downstep in the post-focal domain. On the other hand, the previous studies are not clear on how focus on U words is implemented. The lack of accent and apparent lack of pitch compression in their studies made it harder to measure the effects of a focus.

In this paper, we would like to discuss (i) what types of cues make U words focused, and (ii) whether or not the effects of focus are strong enough to override other phrasing/pitch-manipulating effects such as Downstep and Downtrends. We have conducted a focus production experiment and a focus identification experiment which tightly controlled the focus status and accent status of word pairs (e.g. AA, AU, A[+F]U, AU[+F]) in the stimuli. We hypothesize that, similar to focus on A words, focused U words will be endowed with F0 Rise and Post-focal Fall. Also, if such acoustic marking is systematically used to mark focus on U words, we expect that listeners can identify focus with high accuracy.

Our results show that U words have the focal characteristics of F0 Rise and Post-focal Fall. These focal features, however, failed to show up constantly in all sequences considered. We can say that F0 cues and focus prominence are not correlated so straightforwardly in Japanese. As the effects of focus on pitch were not strong enough in our production experiment, we predicted that participants would have difficulties in perceiving focus. Overall, there is no significant difference between focus on A words and focus on U words in perception. However when the focus is preceded or followed by a U word, the identification ratio of the focus on U words is significantly lower than the focus on A words in the same context. F0 Rise and Post-focal Fall are the cues of focus listeners rely on, but, as the cues are not clearly distinguishable from cues to lexical accents, confusion will be caused. There is no clear-cut categorical cue to focus that we could find.

There are controversies over whether focus builds its own independent prosodic phrase (e.g. Sugahara (2003), Ishihara (2003)). If focus necessarily initiates a new Phonological (Major) Phrase, we would not expect such murky results as ours, so we should conclude that focus does not initiate a new prosodic category, which accords with Ishihara (2003) and others.

Keywords. focus; prosody; prosodic categories; accent; intonation; Japanese

* The authors hereby express their utmost gratitude to Tim Mahrt and Gabor Pintér, collaborators of experiments precursor to the current ones. Tim Mahrt also kindly proofread the whole manuscript, to which the authors would like to express gratitude as well. Thanks also go to participants of the series of experiments for their cooperation and patience, without which this study would not have been possible. The authors also would like to thank the participants of the meeting of the MIT Phonology Circle on Nov. 20, 2017 for their invaluable comments and suggestions to a presentation by Tateishi that helped us improve our ideas. The authors would also like to express their gratitude to visitors to our poster presentation at the 92nd meeting of the Linguistic Society of America for their questions and comments. Remaining errors are of course ours. This paper is in part supported by JSPS KAKENHI Grant-in-Aid for Scientific Research (C) Number JP15K02480. Authors: Shinobu Mizuguchi, Kobe University (mizuguti@kobe-u.ac.jp) & Koichi Tateishi, Kobe College (tateishi@mail.kobe-c.ac.jp).

1. Introduction. The purpose of this paper is to examine how focus in Japanese is realized (and perceived) in various accentual conditions. Although it is established that focus in Japanese affects realization of tonal pitch in some ways, the manner in which focus affects pitch has been a matter of controversy. Pierrehumbert and Beckman (1988) state, based on their description of F0 trends of Japanese, focus forces the initiation of a new major (phonological) phrase, in addition to a series of major phrases based on syntax (Selkirk and Tateishi (1991)) and other prosodic reasons. According to Pierrehumbert and Backman, this new major phrase resets effects of Downstep (Catathesis) and deepens the effect of Initial Lowering (and raising towards the phrasal H tone) than a simple Minor (accentual) phrase, as ordinary Major Phrase based on syntax will do. However, Ishihara (2003) argues that the effect of focus on pitch does not require reference to prosodic phrasing structure. In the present study, we conducted two experiments—one in production and one in perception of focus in various accentual conditions. We would like to argue in this paper that the latter is on the right track. In particular, we argue that the effects of focus on pitch in speech production is not as strong as initiating a new phonological category, which in part accords with Shinya’s (1999) important observation that focus can be downstepped. In our experiment, we found a case in which Post-Focal Lowering in the accented and unaccented environments are not significantly distinctive, suggesting that it may be rather a post-focal environment itself than Downstep that lowers the pitch post-focally. Also, we found that perception of focus can be significantly worse in some particular prosodic/accental environments, which we would not expect if the effects of focus is categorical.

Before going into the details of our experiments and results thereof, we provide a brief overview of Japanese accentual systems and intonation, along with reviews of the literature on the production and perception of prosodic focus in Japanese. After that, details of our experiments, results, and discussions are elaborated.

2. Backgrounds and previous studies. Japanese is a pitch accent language and its accent is characterized by a falling H*-L bi-tonal contour. Japanese lexical items are divided into two groups: Accented (A) and Unaccented (U). They are often tonally transcribed as %LH*L (accented) and %LH (unaccented), where %L is the Initial Low tone.

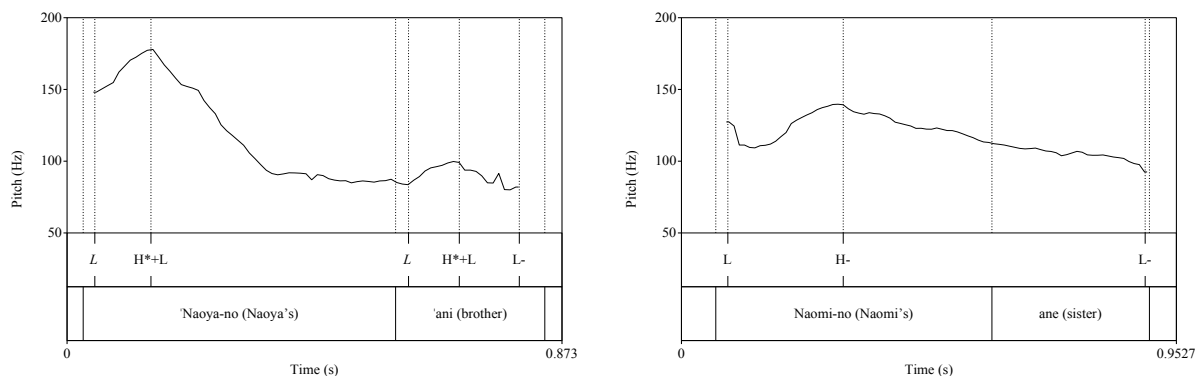


Figure 1. Sample accented (left) and unaccented (right) word sequences (examples adapted from Ishihara (2016))

Each lexical item can bear at most one accent.

2.1. ACCENT AND PROSODIC PHRASING. In a sequence of accented words, Downstep, or Catathesis (Pierrehumbert and Beckman (1988)), is observed, where the pitch peak of the second accented word is significantly lower than the preceding one. However, in a sequence of unaccented words, in lieu of Downstep, dephrasing is often observed which nullifies the effects of the Initial Lowering. Ito and Mester (2013) analyze the phenomena in OT terms, and their analysis predicts

the following patterns¹:

- (1) [[A][A]]: Initial Rise on both words, and Downstepping on the 2nd word
 [[A][U]]: Initial Rise on both words, and Downstepping on the 2nd word
 [U A]: Initial Rise on the first word
 [U U]: Initial Rise on the first word
 (A = accented, U = unaccented)
 (Based on Ito and Mester 2013: (37), 33)

The square brackets in (1) indicate the existence of a phonological phrase. According to Ito and Mester (2013), the maximal phonological phrase in the layer of phonological phrases coincides with the Major Phrase in McCawley’s terms (1968), and the minimal one corresponds to the Minor Phrase (ibid.). A Minor Phrase triggers Initial Lowering and a Major Phrase triggers both Initial Lowering and Downstep.

2.2. FOCUS, PITCH AND PHRASING. As in many other languages, the existence of focus affects the realization of pitch. Pierrehumbert and Beckman (1988) first pointed out that focus prosody of A words and U words are different, but their discussions were mainly on Downstep and compression of pitch range followed by Focal Pitch Raising, with A words. Much work has been done on the narrow focus on A words following their work, and the literature (cf. Sugahara (2003), Shinya et al. (2004), Ishihara (2003, 2011b, 2016), among others) found that the F0 cues of narrow focus on A words are characterized by F0 range expansion (F0 Rise) and post-focal F0 range compression (Post-focal Fall).

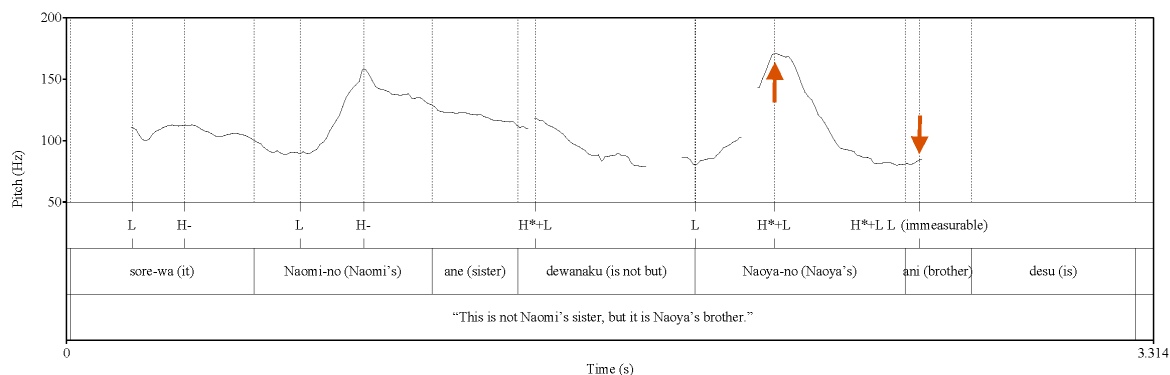


Figure 2. A Sample pitch track of focus on *Naoya* rising its pitch and compressing pitch on the following word

Not much work on focus prosody of U words is, however, done in the literature — partly because there is no accent or subsequent downstep in U words, though we do have a vague understanding that the pitch is raised when an U word is focused (Ishihara (2011a), Sugahara (2003)).

There are controversies over whether Focus builds its own independent prosodic phrase. Pierrehumbert and Beckman (1988) argue, based on their observation of focal pitches, that focus in Japanese initiates an independent Major Phrase (or Intonational Phrase). However, many other studies point to a different direction. Shinya (1999), cited in Ishihara (2003), points out that focus pitch itself can undergo Downstep. The same sort of observation has been done by Kubozono (2007), which points out that a *wh*-word, a typical focus associator, undergoes Downstep. These observations cast doubt on the hypothesis that focus initiates its own prosodic category. Also, Ishihara (2003) and Sugahara (2003) both point out that prosodic phrasing can also be observed in a domain where pitch range is compressed due to Post-focal Lowering. This also casts doubt

¹ However, as we observe later in this paper, actual realization of pitch does not necessarily accord with their analysis. Other factors, for example, speech rate and information structure, may affect it as well.

on the claim that focal pitch is a matter of prosodic phrasing.

3. Precursor: Experiment 1 (Mizuguchi, Tateishi and Mahrt (2016)). As a precursor to the current investigation, we conducted an experiment on the perception of contrastive focus. Our experiment uses the same methodology as Lee et al.’s (2015) cross-language study of the production of prosodic cues to focus and the perception of focus. Using corrective number sequences as in (2), Lee et al. measured prosodic cues such as duration, intensity and pitch in focal positions in various languages and examined how they correlate with focus.

- (2) A: Is Mary’s number 215-418-5623?
 B: No, the number is 215-417-5623.

Lee et al. also conducted a focus identification experiment and measured identification rates.

| | South Kyungsang Korean | Seoul Korean | Tokyo Japanese | Suzhou Wu | Standard French | Mandarin Chinese | American English |
|--------------------------------|---|--------------|----------------|-----------|-----------------|------------------|------------------|
| Production | median z-score values of focused digits | | | | | | |
| duration | 0.64 | 0.13 | 0.10 | 0.48 | 1.73 | 1.19 | 0.95 |
| intensity | -0.26 | 0.24 | -0.24 | 0.53 | 0.97 | 0.36 | 1.28 |
| pitch | 1.00 | 0.62 | 0.60 | 0.61 | 1.17 | 3.13 | 2.96 |
| Perception identification rate | 55.6% | 44.6% | - | - | - | 94.9% | 97.3% |




Table 1: Results of production and perception experiments adapted from Lee et al. 2015, where the z-scores reported for the acoustic cues indicate median values for focused after normalizing the values across all the values in an utterance²

Lee et al. found that focused digits, compared to unfocused digits, exhibit greater duration, intensity and pitch in Mandarin Chinese and American English, but not in Korean and Tokyo Japanese. This shows that languages can differ in how well their speakers communicate the location of contrastive focus by purely prosodic means of pitch, intensity and duration.

For languages where Lee et al. found strong acoustic differentiation (i.e. American English and Mandarin Chinese), high accuracy was obtained in the perception study. However, for languages with weak prosodic marking, namely Seoul Korean and South Kyungsang Korean, low accuracy was obtained in the perception study. Lee et al. did not conduct an identification experiment on Japanese, because they could not recruit enough native speakers of the language. However, as with the two varieties of Korean, Japanese speakers did not acoustically mark focus and thus we would predict that Japanese speakers cannot perceive focus accurately. In this precursor experiment, we tested this hypothesis.

3.1 METHODS AND MATERIALS. In our identification experiment using Japanese, we used materials similar to Lee et al.’s. We used tokens as in (3).

- (3) An example of one of the Japanese materials (recorded by one of the authors, a male speaker of Tokyo Japanese)
 (As an answer to the question, “Is Yamada’s number 215-418-5623?”)
 A: Ie. 215-418-6623 des-u.
 no. COPULA-PRES
 “No. (It’s) 215-418-6623.”

² The perception study was not done for all of the languages investigated.

Twenty-two Japanese college students at two universities (20 F, 2M, mean age 20.45, SD 0.87)) were recruited for the experiment. They reported no auditory difficulties. The experiment was conducted in a quiet room where the audio stimuli were projected from a room speaker. Each participant heard 30 utterances, which were presented in a random order. The utterances contained one digit in the ten-digit sequence with focus. Participants were provided with an answer sheet that contained 30 sequences of 10 “x”s. Their task was to mark the position they heard as focused for each of the 30 utterances. For example, when the fifth number was contrastive in a token they heard, they were supposed to mark the fifth “x”. Before the experiment, participants were instructed about how focal pitch sounds with simple examples. After the instructions, they were given two practice trials. Each session lasted about thirty minutes.

3.2 RESULTS. Table 2 shows the results.³

| | Target Word Position (% accuracy) | | | | | | | | | |
|---------------|-----------------------------------|----|----|------------------------|----|----|------------------------|----|----|----|
| | 1 st phrase | | | 2 nd phrase | | | 3 rd phrase | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| L1 (Japanese) | 100 | 88 | 91 | 97 | 98 | 92 | 65 | 91 | 82 | 58 |

Table 2: Accuracy of native Japanese speakers in words bearing contrastive focus in the specified position of a ten digit number string (with the nine other numbers being already part of the discourse) in Japanese utterances (percentage values)

Participants did unexpectedly well, contra Lee et al.’s findings on production of Japanese focus.

The native speaker of Tokyo Japanese who recorded the materials pronounced a sequence of numbers in the following way:

- (4) [XXX][XXX][[XX][XX]]⁴
 (where X represents each number and [...] represents prosodic phrasing)

Due to the accent rules of Japanese, the first, fourth, seventh and ninth numbers, i.e. numbers at the beginning of a phonological phrase, are obligatorily unaccented in a sequence of numbers no matter what the lexical accent patterns of the numbers are, while, on the other hand, the second, fifth, and eighth numbers, i.e. the post-initial numbers, are always accented. This means that participants heard both accented and unaccented foci throughout the experiment. The high accuracy in most positions shows that listeners were able to correctly perceive focus regardless of the accent status of the word.

Also, the fact that participants had relative difficulties in the third (i.e. final) phrase may suggest that the effects of focus is not strong enough to override the effects of downtrends typical

³ We also recruited eighteen Japanese students (13 Males, 5 Females, mean age 20.5, SD 1) to have them identify focus in English in the same setting. The following table shows the results of identification experiments using English materials to Japanese speakers:

| | Target Word Position (% accuracy) | | | | | | | | | |
|--------------|-----------------------------------|-----|-----|------------------------|-----|----|------------------------|----|----|----|
| | 1 st phrase | | | 2 nd phrase | | | 3 rd phrase | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| L2 (English) | 100 | 100 | 100 | 100 | 100 | 98 | 100 | 98 | 96 | 94 |

The result accords with Lee et al.’s results with English. Even for non-native speakers, English focus is easy to detect.

⁴ Our preliminary pitch measurements of the numbers suggest that the three major prosodic constituents are Minor Phrases or Minimal Phonological Phrases in the sense of Ito and Mester (2013), as we observe clear Initial Lowering at the beginning of each phrase while later phrases tend to be compressed in pitch ranges.

of the end of the utterance and perhaps the preceding series of downstep (i.e. after accented words).

4. Experiment 2: Production of Focus. Given the result of Experiment 1, we have two research questions in mind. First, what types of cues makes an unaccented word focused? The previous studies suggest that, for A words, focus boosts the pitch of the accent peak and triggers either compression of the pitch range and/or Downstep in the post-focal domain. On the other hand, the previous studies are not clear on how focus on U words is implemented. The only exception we are aware of is Sugahara (2003). Based on the results of a series of experiments she conducted, Sugahara shows that, focused or not, a sequence of U words undergoes dephrasing — but only for some speakers/cases, and thus it is not totally clear what the effect of focus on U words is on pitch. Sugahara schematized her findings as follows:

- (5) Unaccented Cases for Sugahara
 [[U][U]] or [U U]: Without Focus we see both patterns. (Presence/Absence of Initial %LH)
 [[U[+F]][U]] → [U[+F] U]: for SOME speakers (no Initial %LH)

However, in Experiment 1, we observed that focus can be effectively detected even if the focused number is unaccented. This leads us to our first research question for Experiment 2: What acoustic cues signal focus for unaccented words? If, as Ishihara (2003) and others suggest, Post-focal Fall is caused by focus and if it is separated from Downstep, we would expect some sort of pitch lowering effects following a focused U word. However, if Post-focal Fall is a matter of accent, which Pierrehumbert and Beckman (1988) seem to suggest, we would not expect any Post-focal Fall following a focused U word.

Our second research question is: are the effects of focus strong enough to override other phrasing/pitch-manipulating effects such as Downstep and Downtrends? In Experiment 1, we found that the perception identification rate of focus lowers in the final prosodic phrase, where we would expect the influences of Downstep and Downtrends most strongly. Does that suggest that the effects of focus on pitch are not strong enough to cue the perception of a new Major Phrase?

With these research questions in mind, we conducted an experiment on the production of contrastive focus in Japanese in various accentual conditions.

4.1. STIMULI AND METHOD. The stimuli consist of two words with similar segmental characteristics (cf. (6)), covering the four possible accentual sequences: UU, UA, AU, and AA (cf. (7)).

- (6) The Structure of Stimuli
 [Phrase [Mod XXXX][Head YYYYY]]-desu
- (7) AA: /a'oi 'uni 'desu/ “It is a blue sea urchin.”
 blue sea urchin be-PRES
 /a'oyama-no 'ani 'desu/ “It is my brother in Aoyama.”
 (place name) brother be-PRES
- AU: /a'oi ume 'desu/ “It is a blue plum.”
 plum
 /a'oyama-no ane 'desu/ “It is my sister in Aoyama.”
 sister
- UA: /amai 'uni 'desu/ “It is sweet sea urchin.”
 sweet
 /oomiya-no 'ani 'desu/ “It is my brother in Omiya.”
 (place name)
- UU: /amai ume 'desu/ “It is a sweet plum.”
 /oomiya-no ane 'desu/ “It is my sister in Omiya.”

We included a [+Focus] condition where either the first word in the sequence was focused,

the second word in the sequence was focused, or neither word was focused. This yields 12 stimuli conditions: UU, U[+F]U, UU[+F], UA, U[+F]A, UA[+F], AU, A[+F]U, AU[+F], AA, A[+F]A, and AA[+F]. The context for focus was given as answers to a *wh*-question. (cf. (8)).

- (8) 1. What is it? Type (gives the broad focus context.)
 2. What Y was it? Type (gives the narrow focus context on Mod.)
 3. X's what was it? Type (gives the narrow focus context on Head.)

Five Tokyo-Japanese speakers (F2, M3) participated in the experiment. Recordings were conducted in a sound-attenuated booth at Kobe University and were saved to a laptop computer at 44.1kHz with 16 bits per sample. During the recordings, stimuli were presented to speakers through PowerPoint slides in a pseudorandomized order. Participants, watching a computer display, first silently read a question that gives the context. After that, they read aloud the stimulus sentences as answers to the contextual questions. 24 stimuli and 24 distractors were produced per subject.

For the analysis, we follow Ishihara (2016), and take the normalized F0-means of the six measurement points: the 1st F0-minimum, the F0-maximum, and the 2nd F0-minimum of Word 1 (L1-1, H1, L1-2) and the same measures for Word 2 (L2-1, H2, L2-2). The data from all five subjects were normalized to factor out pitch range differences among speakers, using the formula in (9), where x is an actual value in each subject's data, R_1 is the F0 peak, and R_2 is the F0 valley of the whole utterance.

- (9) Normalization formula: $y = x - R_2 / R_1 - R_2$

4.2. RESULTS. Figures 3 through 6 show the measurements of mean normalized F0 in each of the four two-word accentual conditions (AA, UA, AU, and UU respectively). For the sequence AA (Figure 3), focus does not exhibit as large a boost in pitch as we expected for AA[+F].

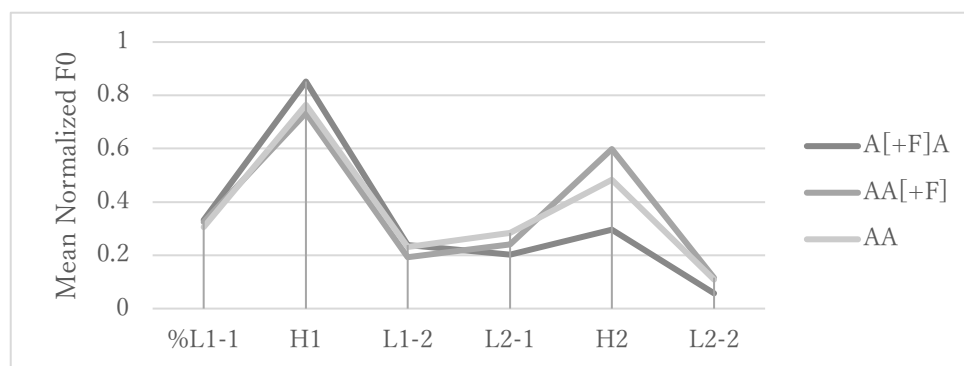


Figure 3. Mean normalized F0 in AA sequences

From the UA sequence plot (Figure 4), focus appears to induce Post-focal Fall.

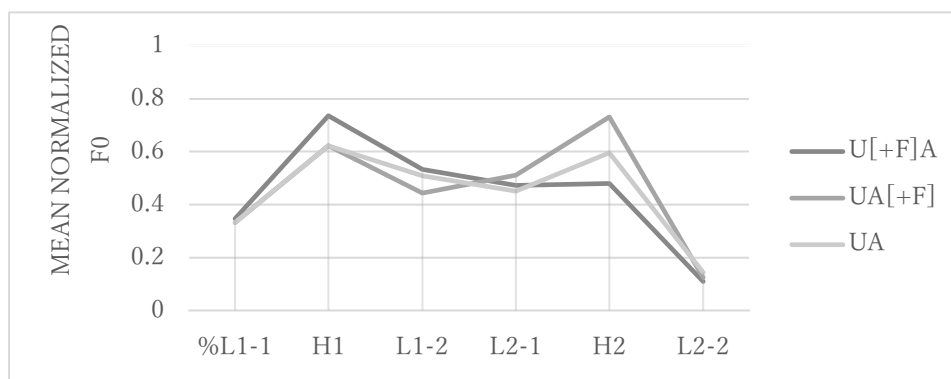


Figure 4. Mean normalized F0 in UA sequences

In the AU sequence (figure 5), one notable fact is that, in AU[+F], the focused U word appears to be downstepped.

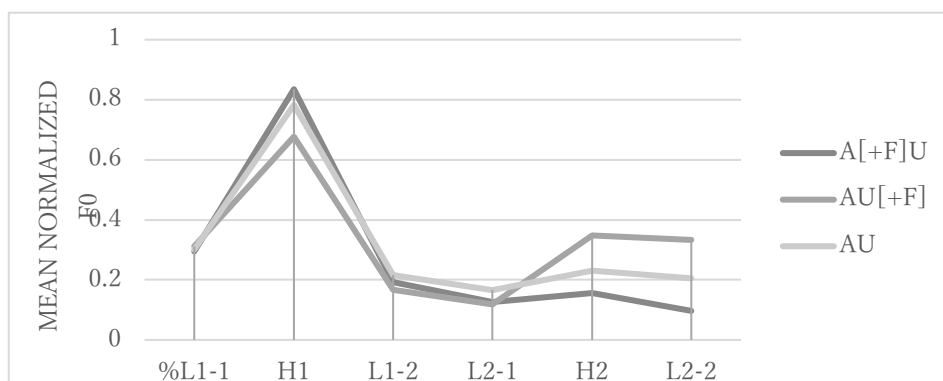


Figure 5. Mean normalized F0 in AU sequences

Finally, in the UU sequence (figure 6), it appears that a focus on the first word causes lowering of pitch on the second word.

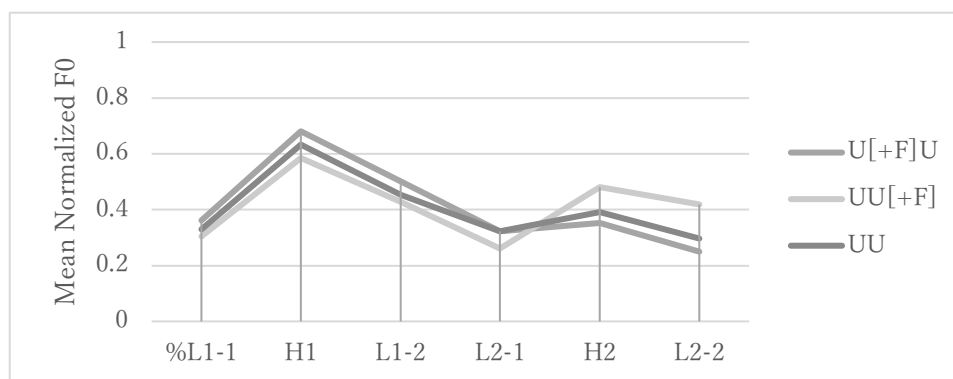


Figure 6. Mean normalized F0 in UU sequences

Based on the measurements of normalized F0, the differences are examined using Wilcoxon tests about the Initial F0 Rise of W1 and W2, and about the Post-focal Fall of W2 in comparison with the broad focus cases. From Table 3, we can say that F-0 Rise is observed overall, except in the U[+F]A and the AA[+F] sequences. Post-focal Fall is observed only in the A words after focus, i.e. in the A[+F]A and U[+F]A sequences, whereas it is not observed in the U words (i.e. A[+F]U and U[+F]U).

| | F0 Rise | Post-focal Fall | | F0 Rise | Post-focal Fall |
|--------|--------------|-----------------|--------|--------------|-----------------|
| A[+F]A | ** (p=0.008) | ** (p=0.007) | AA[+F] | ns (p=0.09) | - |
| A[+F]U | ** (p=0.008) | ns (p=0.26) | AU[+F] | ** (p=0.008) | - |
| U[+F]A | ns (p=0.859) | ** (p=0.008) | UA[+F] | ** (p=0.008) | - |
| U[+F]U | ** (p=0.008) | ns (p=0.314) | UU[+F] | ** (p=0.008) | - |

Table 3: Presence/absence of characteristics of focus (with p-values from Wilcoxon Tests)

4.3. DISCUSSION. We had two research questions when we conducted Experiment 2: (i) What types of cues make an unaccented word focused? (ii) Are the effects of focus strong enough to override other phrasing/pitch-manipulating effects such as Downstep and Downtrends? Let us discuss research question (i) first. From Figures 3-6 and Table 3, we can claim that F0 Rise is characteristic of focused U words, with the exception of the U[+F]A sequence.

Additionally, Post-focal Fall is observed in the sequences A[+F]A and U[+F]A, but not in A[+F]U and U[+F]U. There are two important patterns from this observation: First, Post-focal Fall is realized on post-focal A words, but not in post-focal U words. Second, both A words and U words trigger Post-focal Fall. The literature on Japanese focus shows that F0 Rise and Post-focal Fall are characteristic of the focus on A words, and our results suggest that U words also have these focal characteristics. We need to explain why F0 Rise is not observed in all focal contexts, though.

As for research question (ii), the AA[+F] sequence fails to show the characteristic F0 Rise of focus, and the first A word is higher than the second A word, i.e. the AA[+F] sequence undergoes Downstep, as observed in Figure 7. This observation of ours supports the claims by Shinya (1999) and Kubozono (2007). Downstep in the AA[+F] sequence casts doubt on the hypothesis that focus initiates its own prosodic category.

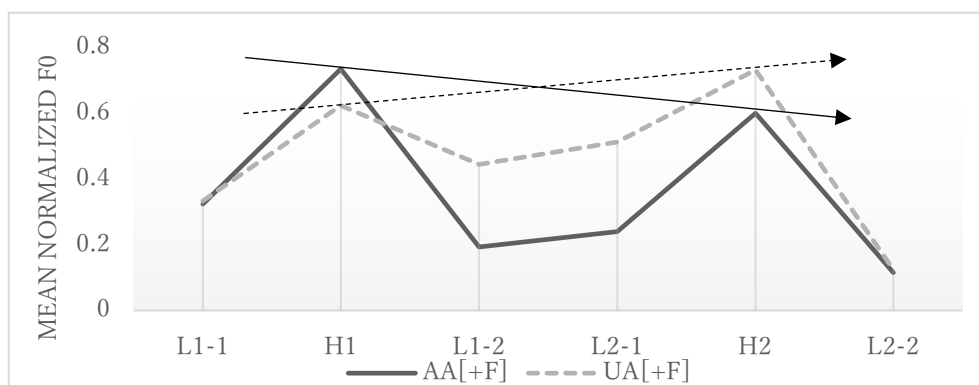


Figure 7. Mean normalized F0 of AA[+F] and UA[+F] sequences

In the UA[+F] sequence, on the other hand, the opposite is the case; A[+F] is higher in pitch than the preceding A word, and no downstep is observed, as in Figure 7. This appears to support the claim by Pierrehumbert and Beckman (1988) that focus initiates its own prosodic category.

As the prosodic features characteristic of focus failed to show up constantly in all sequences considered, we can say that the F0 cues and the focus prominence are not correlated so straightforwardly in Japanese. Given these results and the results of Experiment 1, we cannot claim the effects of focus on pitch is as strong as the focus initiates another phonological (major) phrase overriding other pitch-manipulating effects such as Downstep and Downtrends.

In experiment 1 we found that, in general, Japanese speakers acoustically mark focus and Japanese listeners are sensitive to the acoustic marking of focus, but we did not control the accent status of the target word or adjacent words. In experiment 2, we more tightly controlled the stimuli to examine the effect of focus on phrase-final and post-focal material for accented and unaccented words. The results show clear acoustic marking in some situations but not in others. What remains to be examined is whether listeners are able to hear focus in these various contexts. We examine this question in Experiment 3.

5. Experiment 3: Perception of Focus. From Experiment 2, we have some predictions about when we think listeners will be able to hear focus and when they won't, and conducted a focus identification experiment. If the effects on focus in terms of prosody are categorical, we expect that listeners can identify focus easily.

5.1. METHODS AND PREDICTIONS. In Experiment 3, we tested the perception of focus in the same twelve stimuli contexts as in Experiment 2: UU, U[+F]U, UU[+F], UA, U[+F]A, UA[+F], AU, A[+F]U, AU[+F], AA, A[+F]A, and AA[+F]. We also examined the perception of broad focus by providing stimuli produced under broad focus and providing a 'broad focus' response option. We expect that listeners will choose the 'broad focus' response option both when an utterance was

produced with broad focus and when the acoustic cues to focus are unclear. Twenty-three Japanese college students (F12, M11, Mean Age 19.65, SD 4.0) were recruited as participants. They reported no auditory difficulties. Before the experiment, they received instruction on the relation of *wh*-questions and focus, including broad focus, based on (8). In the experiment, they only heard answers without texts, and they were asked to mark which position in the target sequence was focused on the answer sheet. They were instructed to mark broad focus as “The Whole Utterance”. Each participant listened to 96 tokens of utterances: 48 with the twelve accentual and focal patterns (4 per pattern), 24 with broad focus, which were collected in Experiment 2, and 24 distractors (tokens with numbers used in Experiment 1). The experiments were conducted in a quiet room at Kobe University, and the audio stimuli were presented to listeners through a room speaker. Before the actual test, we presented three practice trials to listeners to familiarize them with the procedure. Each session lasted about thirty minutes.

We predicted that participants have difficulties perceiving focus, as the effects of focus on pitch were not strong enough in Experiment 2. Also, as the experiment requires listeners to choose among the three choices, narrow focus on the first word, one on the second word, and broad focus, we predicted that the task would be harder than the number-selecting task in Experiment 1.

5.2. RESULTS. Table 4 exhibits the identification rates of broad and narrow focus; columns correspond to focus positions within an utterance, and rows represent the focus identification rates by percentage values. The overall identification rate was 78.8% for narrow focus on A words, 65.88% for narrow focus on U words, and 38.6% for broad focus. In order to test whether the numerical difference among identification rates was statistically supported, we conducted a one-way ANOVA between focus type and identification rates, and the result was $F(2, 275)=8.618, p=0.0002$. A post-hoc multiple comparison analysis showed a significant difference between broad focus and narrow focus on A words ($p<0.001$) as well as between broad focus and narrow focus on U words ($p=0.004$), but not between narrow focus on A words and narrow focus on U words ($p=0.29$).

| Focus type | Max(%) | Min(%) | Mean(%) | SD |
|-------------------|--------|--------|---------|------|
| Narrow focus on A | 91.4 | 68.5 | 78.8 | 16.1 |
| Narrow focus on U | 76.1 | 57.6 | 65.8 | 4.6 |
| Broad focus | 42.4 | 15.3 | 38.6 | 5.4 |

Table 4: Focus identification rates (percentage values)

Though the overall identification rates are not significantly different between A[+F] and U[+F], a closer look shows a different picture. Table 5 is a confusion matrix of narrow focus and broad focus perception and we can see the identification rates of narrow focus vary among focus positions.

Figure 8 is the identification rates of narrow focus in the sequences of XA, XU, AX, and UX, where X is A[+F] and U[+F]. We conducted a T-test on focus identification rates between A[+F] and U[+F], and found that the difference was significant between the two focus types in the XU sequence ($t(44) = 3.13, p = 0.003$) and the UX sequence ($t(44) = 4.99, p<0.001$), but the difference is non-significant in the XA sequence ($t(44) = 1.87, p=0.07$) and in AX sequence ($t(44) = 1.79, p=0.08$). This result shows that focused A words preceding or following U words are more easily perceived than focused U words in the same sequence.

We also conducted a T-test on focus identification rates in AA[+F] and UA[+F] (cf. Figure 7), to see whether Downtrend or Uptrend affects the identification rate. We found out that the difference is non-significant ($t(44) = 1.78, p=0.08$). This means that Downtrend or Uptrend observed in Figure 7 does not affect the perception of narrow focus on A words.

| Perceived | Target | | | Perceived | Target | | |
|-----------|--------|------|--------|-----------|--------|------|--------|
| | A[+F]A | AA | AA[+F] | | A[+F]U | AU | AU[+F] |
| A[+F]A | 68.5 | 38.0 | 9.8 | A[+F]U | 77.2 | 51.1 | 16.3 |
| AA | 21.7 | 15.3 | 12.0 | AU | 22.8 | 31.5 | 18.5 |
| AA[+F] | 9.8 | 46.7 | 78.2 | AU[+F] | 0.0 | 17.4 | 65.2 |
| Perceived | Target | | | Perceived | Target | | |
| | U[+F]A | UA | UA[+F] | | U[+F]U | UU | UU[+F] |
| U[+F]A | 57.6 | 23.9 | 4.3 | U[+F]U | 64.1 | 37.0 | 10.9 |
| UA | 23.9 | 42.4 | 4.3 | UU | 26.1 | 21.7 | 13.0 |
| UA[+F] | 18.5 | 33.7 | 91.4 | UU[+F] | 9.8 | 41.3 | 76.1 |

Table 5: Confusion matrix of narrow and broad focus (percentage values)



Figure 8. Focus identification rate in four sequences (percentage values)

6. General Discussion. We had two research questions when we started this study; (i) What type of cue makes an unaccented word focused? (ii) Are the effects of focus strong enough to override other phrasing/pitch-manipulating effects such as Downstep and Downtrends? We have conducted production and perception experiments and have found out that, even though the focus on U words appears to be harder to detect than the one on A words (cf. Table 4), the difference in perception is not significant on ANOVA. Table 3 shows that the production cues of focus of U words, along with A words, are F0 Rise and Post-focal Fall. We could claim that as far as the prosodic cues of focus are concerned, there is no difference between A words and U words. The story is not this simple, however. Table 5 and Figure 8 show that perception rates vary depending on the focus positions. We need to consider what makes perception rates of focus higher in more detail. Table 8 elaborates on the results shown in Table 3. Only XA and XU can show Post-focal Fall by definition; however, XA, but not XU, is accompanied by Post-focal Fall in Table 6. This means Post-focal Fall does not affect focus perception in the XU sequence. Since focus perception is significantly different between A[+F]U and U[+F]U, we suspect this difference comes not from Post-focal Fall but from F0 Rise. This view is supported by the UX sequence, which does not show Post-focal Fall by definition; the focus perception is significantly different between UA[+F] and UU[+F] and we can say this difference is brought by F0 Rise.

| Phonetic context where focus appears | | Production cue (p values from Wilcoxon test) | | Perception rate (p values from T test) |
|--------------------------------------|--------|---|-----------------|---|
| | | F0 Rise | Post-focal Fall | |
| XA | A[+F]A | ** 0.008 | ** 0.007 | ns |
| | U[+F]A | ns | ** 0.008 | |
| XU | A[+F]U | ** 0.008 | ns | *** 0.003 |
| | U[+F]U | ** 0.008 | ns | |
| AX | AA[+F] | ns | - | ns |
| | AU[+F] | ** 0.008 | - | |
| UX | UA[+F] | ** 0.008 | - | **** < 0.001 |
| | UU[+F] | ** 0.008 | - | |

Table 6: Focus prosodic cues and perception rates in four sequence types

If we are on the right track, we would expect a significant difference in focus identification between tokens with F0 Rise and tokens without it. A look at the XA and the AX sequence, however, shows this is not borne out; there is no significant difference in these sequences. We have to conclude that neither Post-focal Fall nor F0 Rise affects the perception of narrow focus in Japanese. We need another explanation for the difference in perception of focus between A words and U words.

Figure 9 shows mean normalized F0 contours of focus in four types of word sequences. Since U words lack an accent in Japanese, the only possible prosodic cue for focus would be Initial Lowering and phrasal H, except in U[+F]A. However, raising pitch on the second mora of a U word naturally raises the following pitch (cf. top left and top right figure of Figure 9), which may cause confusion in the identification of focus. Focus on A words are better in identification rates, but harder in the initial position (cf. Table 5). This may be due to the fact that even a non-focused A word at the initial position causes Downstep of the following word (cf. bottom left figure of Figure 9), making it harder to distinguish with a real focus. This fact makes us wonder whether listeners are really listening to focus.

Recall Experiment 1 and its result given in Table 2. We have observed that the number identification is bad in the seventh position in the third prosodic phrase. This position invokes the U[+F]A accent pattern and its pitch contour is similar to the one in top right figure of Figure 9. Listeners often misunderstand that the focus falls not on U words but on A words in this sequence.

We certainly observe focal effects on pitch, but they are not strong enough. Therefore, we would like to claim (a) significant Initial Lowering (and the following rise) fails to show up in certain focal contexts (Experiment 2), so that (b) Post-Focal Lowering is not so strong that some participants of Experiment 3 had trouble distinguishing a simple accented word and a focal one. (c) It is sometimes hard to distinguish effects of a focal U word (such as Initial Rise and the Lowering of the following tone) and mere Downtrends (Experiment 2). Recall that identification rates are low with broad focus (Table 5). This fact and the facts of narrow focus cases may show that it is simply hard to distinguish between a mere accent and focus, and, probably, a mere Initial Rise and focus. F0 Rise and Post-focal Fall are the cues of focus listeners rely on. As the cues are not clearly distinguishable from cues to lexical accents, confusion will be caused. There is no clear-cut categorical cue to focus that we could find.

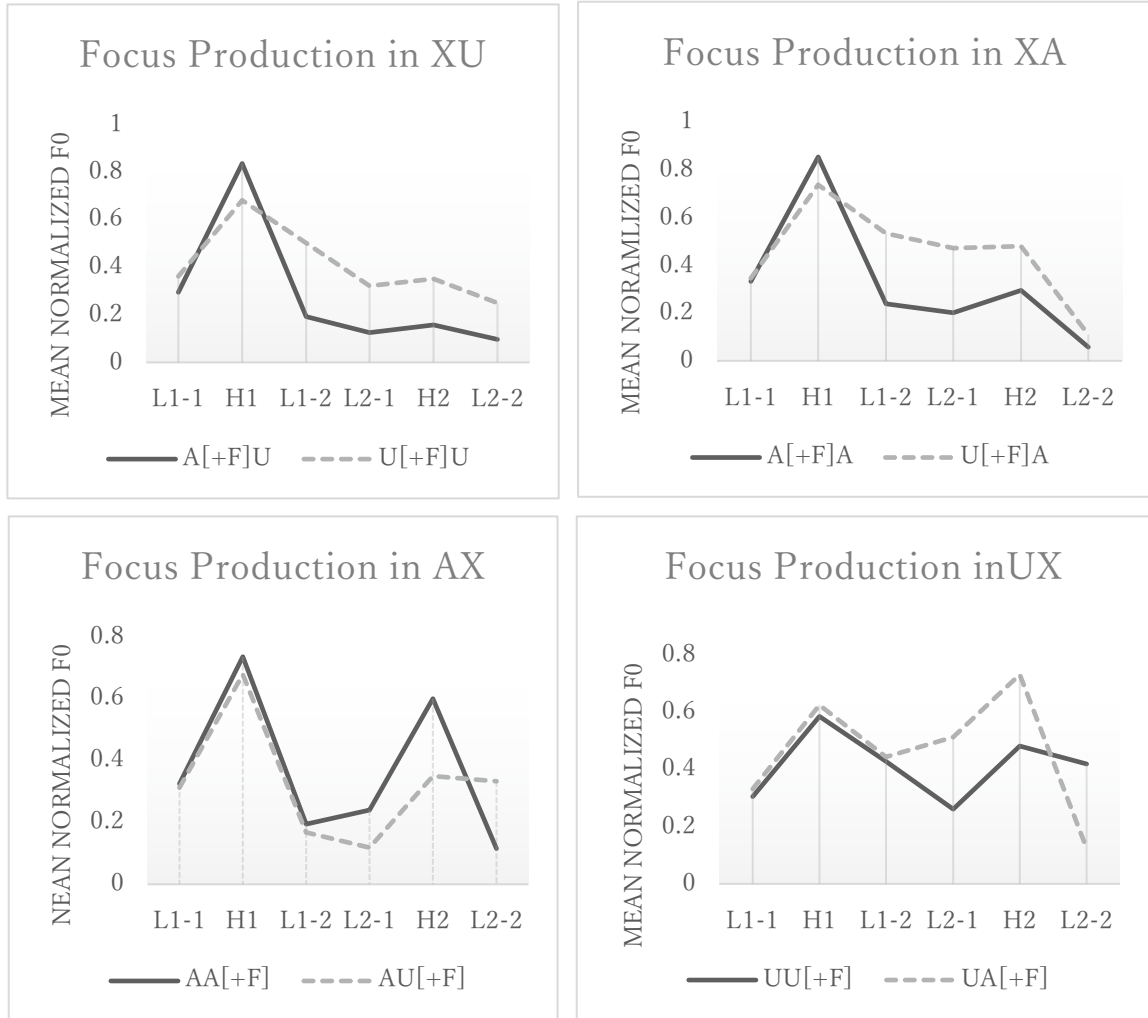


Figure 9. Mean normalized F0 of narrow focus in four sequence types

As for Research Question 2, we have observed that focus induced Downstep and Upstep, depending on the positions focus appears (cf. Figure 7) in Section 4. In addition, in Section 5, we found that there is no significant difference in perception between focus with Downstep and the one with Upstep. In a word, effects of focus on pitch are not always realized as they should be as claimed in previous studies, and actual utterances may even lack them. If focus necessarily initiates a new Phonological (Major) Phrase, we would not expect such murky results, so we should conclude that focus does not initiate a new prosodic category, which basically accords with Ishihara (2003) and others.

One may point out that some prosodic feature's dependence on some prosodic category does not necessarily mean its realization is also categorical. However, note that this is why we eliminated other contextual factors like new/given information status throughout all the experiments. Also, note that we did both production and perception experiments, and we could not find categorical results in neither.

7. Conclusion. Our results may suggest several interpretive possibilities, among which we are not in a position to choose. First, it may suggest that a view of Fujisaki and Hirose (1984) is on a right track. Fujisaki and Hirose view intonation as superimposition of several informational strata, accentual, phrasal and so on. We could imagine focus could only be another information to be overlaid. Or, we could imagine adopting Shinya's (2009) view that effects of accent and focus involve phonological and phonetic ones, which may "nullify" the phonological

categorization in some phonetic contexts. We could also imagine adopting a series of proposals by Liberman and his colleagues (Lee et al. (2015), Lee et al. (2016), among others) that phonological and phonetic effects of focus on prosody is not universal and that each language implements focus as language-particular regulations. It is a matter of future research which direction is the most appropriate.

Another point to note will be that the focus that we manipulated is a narrow focus. We have not discussed cases of focus where the categorical results have been already obtained, namely the prosodic effects of a “real” focus associator such as *wh*-phrases in Japanese as Ishihara (2003, 2016, 2017) discussed. In the case of *wh*, focus is marked morphologically, prosodically, and possibly syntactically as well. On the other hand, the focus we discussed in this paper is only prosodically (and pragmatically/semantically) marked. It is still not clear whether such differences affect the prosodic effects of focus.

A third remaining problem is that we have no explanation why broad focus is harder to identify. In particular, when the initial word is accented, the identification rate is very low. Although we do not deny that, as participants were instructed to identify focus, their response may have been biased towards narrow focus, the low identification rate may suggest that listeners are actually using accent as the sole cue of focus. In the recourse, we probably need to consider that, in Japanese, the base line of utterance lowers due to Downstep/ Downtrend (cf. Pierrehumber and Beckman 1988, Shinya 2009), and that the pitch height is not directly linked to prominence. We will leave this for future study.

References

- Fujisaki, Hiroya and Keikichi Hirose. 1984. Analysis of voice and fundamental contours for declarative sentences in Japanese. *Journal of the Acoustical Society of Japanese* 5-4. 233-242.
- Ishihara, Shinichiro. 2003. *Intonation and interface conditions*. Cambridge, MA: MIT dissertation.
- Ishihara, Shinichiro. 2011a. Focus prosody in Tokyo Japanese Wh-Questions with lexically unaccented Wh-phrases. *ICPhS XVII*. 946-949.
- Ishihara, Shinichiro. 2011b. Japanese focus prosody revisited: Freeing focus from prosodic phrasing. *Lingua* 121. 1870-1889. <http://dx.doi.org/10.1016/j.lingua.2011.06.008>
- Ishihara, Shinichiro. 2016. Japanese downstep revisited. *Natural Language and Linguistic Theory* 34. 1389-1443. <http://dx.doi.org/10.1007/s11049-015-9322-8>.
- Itô, Junko and Armin Mester. 2013. Prosodic subcategories in Japanese. *Lingua* 124. 20-40. <http://dx.doi.org/10.1016/j.lingua.2012.08.016>.
- Kubozono, Haruo. 2007. Focus and intonation in Japanese: Does focus trigger pitch reset? *Workshop on Prosody, Syntax, and Information Structure (WPSI)* 2. pp.1-27.
- Lee, Yong-Cheol, Bei Wang, Sisi Chen, Martine Adda-Decker, Angélique Amelot, Satoshi Nambu, and Mark Liberman. 2015. A crosslinguistic study of prosodic focus. *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference*. 4754-4758.
- Lee, Yong-Cheol, Ting Wang and Mark Liberman. 2016. Production and perception of tone 3 focus in Mandarin Chinese. *Frontiers in Psychology* 7. 1-13. <http://dx.doi.org/10.3389/fpsyg.2016.015058>.
- McCawley, James D. 1968. *The phonological component of a grammar of Japanese*. The Hague: Mouton.
- Mizuguchi, Shinobu, Koichi Tateishi and Tim Mahrt. 2016. Perception of contrastive focus by L2 learners.” A paper presented at *Tone and Intonation in Europe (TIE) 2016*, University of Kent.
- Pierrehumbert, Janet B. and Mary E. Beckman. 1988. *Japanese tone structure*. Cambridge, Mass.: The MIT Press.

- Pintér, Gabor, Shinobu Mizuguchi and Koichi Tateishi. 2014. Perception of prosodic prominence and boundaries by L1 and L2 speakers of English. *Proceedings of InterSpeech 2014*. 544-547.
- Selkirk, Elisabeth and Koichi Tateishi. 1988. Constraints on minor phrase formation in Japanese. *CLS* 24. 316-336.
- Selkirk, Elisabeth and Koichi Tateishi. 1991. "Syntax and downstep in Japanese." C. Georgopoulos and R. Ishihara (eds.) *Interdisciplinary approaches to language: Essays in Honor of S.-Y. Kuroda*. Dordrecht: Kluwer. 519-543.
- Shinya, Takahiro. 1999. Eigo to nihongo ni okeru fōkasu ni yoru daunsteppu no soshi to cho-on-undo no chōgo [The blocking of downstep by focus and articulatory overlap in English and Japanese] *Sophia Linguistics Society* 14. 35-51.
- Shinya, Takahiro. 2009. *The role of lexical constraint in the perception of intonational prominence in Japanese*. Amherst, MA: University of Massachusetts dissertation.
- Shinya, Takahiro, Elisabeth Selkirk, and Shigeto Kawahara. 2004. Rhythmic boot and recursive minor phrase in Japanese. *International Conference on Speech Prosody* 2. 183-186.
- Sugahara, Mariko. 2003. *Downtrends and post-focus intonation in Tokyo Japanese*. Amherst, MA: University of Massachusetts dissertation.