Japanese pitch accent: examination of final-accented and unaccented minimal pairs

Yukiko Sugiyama
University at Buffalo, SUNY

This paper presents partial results from ongoing research on the acoustic correlates of Japanese pitch accent. This research exploits bimoraic/disyllabic minimal pairs that are identical except for their accent types. While previous studies have consistently found a notable acoustic difference between final-accented and unaccented words in the following particle, it has been debated whether there is any difference within the words themselves. The current study used all 20 minimal pairs that have a relatively high familiarity rating in the computerized dictionary by Amano and Kondo (2000). Preliminary results from five talkers suggest that findings from previous research, which are based on a few pairs of words, hold for a larger set of words in general. That is, the absolute F0 maximum is higher and F0 difference between the first and second syllable is larger for final-accented words than unaccented words only when they are produced sentence-medially.

1. Introduction

Japanese is a mora-timed, pitch accent language in which words are marked for their accent.¹ In Japanese, words can be accented or unaccented. For words that have no accent (unaccented words), the first syllable has a low pitch and the remaining syllables have a high pitch.² For accented words, one syllable is marked for accent. If a word has

---

¹ I would like to thank Karin Michelson and Jim Sawusch for their advice and valuable discussions. I am obliged to Haruo Kubozono for his assistance in designing the production experiment. I am also grateful to Mitsu Shimojo for helping me recruit subjects. I would also like to thank Heike Lehnert-LeHouillier for her comments on an earlier version of this paper. This research is partly funded by the Mark Diamond Research Fund.

² Strictly speaking, syllables should be distinguished from moras as they play a different role in assigning pitch accent in Japanese (cf. McCawley 1978). However, since this research deals only with words in which one syllable consists of one mora, the distinction is not very crucial. In this paper, the term syllable will be consistently used.

Toronto Working Papers in Linguistics 26: 73–88
Copyright © 2006 Yukiko Sugiyama
accent on the first syllable, the syllable has a high pitch and all the following syllables have a low pitch. If a word has accent on the second or later syllable, the first syllable has a low pitch and the syllables from the second until the accented one all have a high pitch. After the accented syllable, all the remaining syllables have a low pitch. For example, the word /tatamioˆmote/ ‘tatami mat’ has accent on the fourth syllable (indicated by a caret ‘ˆ’). The first syllable ta has a low pitch and the second ta, third mi, and fourth o syllables have a high pitch. After the accented fourth syllable o, the remaining two syllables, mo and te, both have a low pitch. The overall pitch pattern of the word is therefore low-high-high-high-low (LHHHˆLL).

In the literature on Japanese pitch accent, it is still unresolved if words that have accent on their final syllable (final-accented words) and words that have no accent (unaccented words) are acoustically different. The differences between the two types are most noticeably manifested in the following particle when there is one. The particle has a low pitch when it follows a final-accented word, just like syllables have a low pitch after an accented syllable. By contrast, the particle has a high pitch when it follows an unaccented word. Thus, when the final-accented word /hanaˆ/ ‘flower’ is followed by the particle “ga”, a nominative marker, it has the accent type of /hanaˆga/ (LHˆL), whereas the unaccented word /hana/ ‘nose’ has the accent type of /hana ga/ (LHH).

The distinction between final-accented words and unaccented words was first mentioned by Yamada (1892). The final-accented word /hanaˆ/ was described as having a high pitch on the second syllable, and the accent type of the unaccented word /hana/ was described as being flat. Sakuma (1929) presented early experimental data that show the two accent types are acoustically distinct. Sakuma found that the two types of words are distinct in that final-accented words have a higher F0 peak than unaccented words on the final syllable. In order to illustrate this difference, Sakuma introduced another pitch level, middle (M). In this system, the final-accented word /hanaˆ/ has the accent type of LH, while its unaccented counterpart /hana/ has the accent type of LM. This notation brought about enthusiastic debates over the nature of Japanese accent, because it proposed that Japanese has three pitch levels, as opposed to the widely-accepted idea that Japanese has two pitch levels. Sakuma’s representation also suggests that the F0 difference between the first and second syllable determines whether the word is final-accented or unaccented. Since final-accented words have a high pitch and unaccented words have a middle pitch, the magnitude of the F0 rise from the first to the second syllable should be larger for final-accented words. While there is substantial literature on the phonological representations of final-accented and unaccented words in Japanese (e.g. Hattori 1954, Kindaichi 1947, McCawley 1968), the research in this study is mainly related to work that experimentally investigated the distinction between the two accent types.

The results from past work are not consistent as to whether final-accented and unaccented words themselves are different, and if they are, whether the accent contrast is maintained even when they are produced in isolation. In early literature on Japanese pitch accent, researchers argued that there is no difference between final-accented and unaccented words even when they are followed by a particle in a sentence frame (e.g. Kindaichi 1947, McCawley 1968). The only difference between the two accent types was

---

3 The particle itself does not have a specific pitch by itself. It is subject to the accent type of the preceding word.
thought to be whether the following particle has a high pitch or a low pitch. However, Poser (1984) found that final-accented words have a higher F0 maximum than unaccented words when they were produced in a sentence, although the difference was neutralized when they were spoken in isolation. There are also some studies that show that the two accent types are different even when they are spoken in isolation, at least for some talkers (Sugito 1979, 1983, Vance 1995).

Sugito (1968) recorded 75 disyllabic and bimoraic words (simply disyllabic words hereafter) to test the validity of Sakuma’s claim that final-accented words have a high pitch and unaccented words have a middle pitch on their final syllables. Out of 75 words, 35 words were unaccented and 40 words were final-accented. She found that F0 was more influenced by whether the initial consonant of the second syllable was voiceless or voiced than the accent types. In other words, regardless of the accent type, the F0 tended to be higher (rising) when a syllable contained a voiceless consonant than it consisted of only voiced segments. However, this tendency did not hold for all 75 words. Sugito concluded that there was no phonetic difference between final-accented words and unaccented words, as measured by the F0 difference between the first and second syllable. The difference between the two accent types was more clearly observable in the F0 contour than in the F0 difference between the first and second syllable. The F0 tended to stay flat for unaccented words while F0 showed a clear descending slope for the final-accented words. However, considering that segmental sequences were not matched between the two accent types and the F0 difference between the two accent types is normally small (less than 20Hz) to begin with (Kubozono 1993, Vance 1995), it is possible that the accent contrast was obscured because the segmental sequences were not controlled. Although many disyllabic words in Japanese have identical segmental sequences that differ only in the accent type, no such minimal accent pairs were included in Sugito’s study.

Contrary to what Sugito (1968) found, later studies by Sugito (1979, 1983) found that some talkers did produce disyllabic final-accented and unaccented words differently while other talkers did not. In these studies, five pairs of disyllabic words that differed only in the accent type of final-accented and unaccented were spoken by fourteen male talkers from Tokyo. The words were spoken in isolation and in three different sentences. However, not all minimal accent pairs were used for the subsequent acoustic analysis. Only one pair of /hana/ ‘flower’ and /hana/ ‘nose’ was analyzed on the grounds that the two words were the least subject to mispronunciation and the consonant in the second syllable was not voiceless, which enabled her to track F0 movements without interruptions. For this pair of words, the magnitude of F0 rise from the first to the second syllable, F0 maximum on the second syllable, and the F0 movement were measured. Sugito found that, for two talkers, the F0 difference between the first and second syllable was significantly larger for the final-accented word /hana/ ‘flower’ than the unaccented word /hana/ ‘nose’ even when they were spoken in isolation. The F0 maximum on the

---

4 The exact design of the experiment is not mentioned in Sugito (1968). Although not stated explicitly, it seems that words were spoken in isolation by one male talker.
5 Similar results are also found in Kawasaki (1983). This kind of consonantal effect on F0 has also been observed cross-linguistically (e.g. Lehiste and Peterson 1961)
6 In Sugito (1968), it is not stated explicitly whether this finding was based on words spoken in isolation or in a sentences. However, from the context, it seems that the finding is based on words spoken in isolation.
second syllable also had the tendency to be higher for the final-accented word for those talkers who made a clear distinction between the two accent types. Furthermore, when the F0 movements of /hana/ ‘flower’ and /hana/ ‘nose’ produced in the sentence “sore wa _ desu” ‘it is _’ were compared, the F0 curve shaped like a mountain for the final-accented word while the F0 was flat for the unaccented word. These findings were partially replicated by Vance (1995), who found that one out of four female talkers produced a measurable difference between /hana/ ‘flower’ and /hana/ ‘nose’ (Talker A). Her utterances were compared with those by a female talker who did not make a distinction between the two words (Talker B). The F0 maximum on the second syllable was reliably higher by 17 Hz for final-accented words than unaccented words for Talker A while it was only 6 Hz for Talker B (and not reliable). Vance also examined the size of the F0 rise from the F0 minimum on the first syllable to the F0 maximum on the first syllable. The F0 rise was again significantly larger for /hana/ for Talker A (mean = 43 for final-accented, mean = 31 for unaccented) while the F0 maximum difference was not significant for Talker B, who produced the words virtually the same (mean = 27 for final-accented, mean = 26 for unaccented). Unlike Sugito’s claim that the size of the F0 rise was more important in distinguishing the two accent types than the absolute values of F0 maximum, the difference between the two accent types was more clearly manifested in the F0 maximum on the second syllable than in the F0 rise in Vance.

The primary focus of Sugito and Vance was to examine how the contrast between final-accented and unaccented words was realized at the lexical level. There is also a series of studies from the perspective of intonational phonology. Although their ultimate goal is to understand the role of lexical accent in prosody, rather than the nature of lexical accent itself, their research has found some acoustic differences between final-accented and unaccented words. Kubozono (1993) used near minimal pairs of noun phrases that consist of an adjective and a noun, such as /umaˆi oimo/ ‘tasty potato’ and /amai oimo/ ‘sweet potato’ to examine the effect of accent in the first word. The adjective /umaˆi/ is accented on the second syllable and has the accent type of LHˆL. By contrast, the adjective /amai/ is unaccented and has the accent type of LHH. Near minimal pairs were constructed to measure if the F0 peak differed between accented high syllables and unaccented high syllables in the word medial position. It was found from a male talker that the peak was significantly higher by 7–13 Hz for accented adjectives than unaccented adjectives and Kubozono named this phenomenon “accentual boost”.

The F0 difference between accented syllables and unaccented syllables in prosody is accounted for in Pierrehumbert and Beckman (1988) as different types of accents having varying sizes of domains in an utterance. While many earlier theories on Japanese phonology claim that final-accented and unaccented words are indistinguishable within the word, Pierrehumbert and Beckman predicts that the two accent types would differ within the word. In short, accented syllables have an accent peak which they represent as a sequence of high-low “HL”, and second syllables of a phrase have a phrasal peak high tone which they represent as “H”, except for words that have accent on their initial syllable. The H of an accent peak is higher than H of a phrasal peak, as attested by their experimental data and Poser (1984). However, as Warner (1997) points out, Pierrehumbert and Beckman’s theory applies only to words that are at least three-syllable

---

Readers are referred to Pierrehumbert and Beckman (1988) for details.
long. Disyllabic words are problematic because final-accented words would have both an
accent peak and a phrasal peak on the second syllable, and they have no predictions for
the effect of having two peaks on top of each other. In order to test Pierrehumbert and
Beckman’s theory, therefore, words have to have at least three syllables. On the other
hand, monosyllabic and disyllabic words are the only choices if one is to completely
eliminate the influence of surrounding segments, as there are no pairs of final-accented
and unaccented words that differ only in the accent types for three syllable or longer
words in Japanese.

In most studies so far, final-accented and unaccented words were not examined
under conditions in which other factors that could influence phonetic details of those
words were strictly controlled. In the few studies that have examined disyllabic words
that differed only in their accent, the number of words used was very limited. The current
research used a larger corpus of disyllabic word pairs that differed only in their accent
(LH and LH pairs such as /hana/ ‘flower’ and /hana/ ‘nose’). A computerized Japanese
dictionary by Amano and Kondo (2000) was used to search for all pairs of disyllabic
words that consist of the same segments and differ only in the accent type. The words in
each pair were then checked for their familiarity and only words that were relatively high
in familiarity were selected. Each of these words was produced in isolation, as the focus
in a sentence, and as non-focus in a sentence. All factors other than the accent difference
between the words in each pair were controlled.

2. Production experiment

2.1 Method

2.1.1 Subjects

Twenty native speakers of Japanese (ten male and ten female) who grew up in
Tokyo or its neighboring areas (Kanagawa, Saitama near Tokyo, and Chiba) were
recruited at the University at Buffalo. None of them reported any history of a speech or
hearing disorder. Each participant received $6 on completion of the one-hour experiment.
In this paper, preliminary results collected from five talkers (two male and three female)
are reported. All of them were from Tokyo with ages ranging between 28 and 33 years
old.

2.1.2 Materials

The recording materials were twenty pairs of disyllabic and bimoraic words. In
each pair of words, one word had the accent type of final-accented and the other had the
accent type of unaccented. The words were chosen to examine whether the two accent
types were acoustically distinct while controlling other factors that could influence the
production of the words. Several steps were taken to obtain the twenty pairs of words (40
words) from the computerized Japanese dictionary (Amano and Kondo 2000). First, all
pairs of words that were identical except the accent type between final-accent and
unaccented patterns were identified (e.g. /hasi/ ‘bridge’ and /hasi/ ‘edge’). This search
yielded more than 100 pairs. However, it turned out that so many words could be
pronounced with more than one accent type, as judged by ten native Japanese speakers from Tokyo (six male, four female, ages between 18–29 years old). For example, the word /hata/ ‘field’ could be pronounced with all three possible accent types that a disyllabic word could have: initial-accent, final-accent, and no accent (unaccented). The acceptability ratings for all possible accent types were also listed on a 5-point scale. These ratings could have been used to select only words that had the highest rating for the accent type needed. However, in order to avoid any potential confounding factors in interpreting the production data, words with more than one accent type were excluded. Only words that were unambiguously final-accented or unaccented words were included, resulting in 55 pairs of words. Next, these 110 words were checked for word familiarity. In Amano and Kondo (2000), at least 32 native Japanese speakers rated familiarity of each word on a 7-point scale with 7 indicating the highest familiarity. Since pitch accent is a lexical property in Japanese, and not fixed on a particular syllable like Czech and Polish, it was important to make sure that the words were familiar to the participants so that they knew how to pronounce them. There were three kinds of ratings for three modes of presentations: 1) auditory presentation, 2) visual presentation, and 3) simultaneous auditory and visual presentation. The 19 pairs of words that had a familiarity rating of 5.0 or higher for all three modes of presentations were used in the production study. To this list, the pair of words /kakï/ ‘fence’ and /kaki/ ‘persimmon’ was added for comparison with previous studies. Thus, 20 pairs of words were used in total.

2.1.3 Procedure

The 40 words were written on index cards in ordinary Japanese orthography with a mixture of Chinese characters and the hiragana syllabary. The pronunciations of all Chinese characters were written above them in hiragana because most Chinese characters can be read in more than one way. The words were produced in three environments: in isolation and two kinds of frame sentences. The frame sentences were written on another card so that the participants could see them while they produced each word. Since it was found in Pierrehumbert and Beckman (1988) that words in focus tend to have a larger F0 movements than words not in focus, two frame sentences were constructed to control this pragmatic effect. For the focus frame, the sentence used was “kare wa _ to itta” ‘he said _’, in which the word in focus was the test word. For the non-focus frame, the sentence used was “dare ga _ to ittano?” ‘who said _?’, in which “dare ga” ‘who’ is the focus of the sentence. See Appendix for details.

Each block consisted of producing the 40 words in isolation, the focus frame, and non-focus frame, respectively. After completing a set of three blocks once, the participants produced another set of three blocks in the same order as the first time to record two repetitions of each word spoken in each environment. The order in which the 40 words were presented to the participants was randomized in each block. The order in

---

8 The number of words that had more than one possible accent type was larger in Amano and Kondo (2000) than in Nihon Hooso Kyokai (1998). This is probably due to the fact that while the former is more descriptive in nature while the latter is more prescriptive.

9 Although not all familiarity ratings for /kaki/ ‘fence’ reached 5.0, it still is a well-known word. The lowest rating was from the visual presentation (4.75). All three ratings were higher than 5.0 for the counterpart, /kakï/ ‘persimmon’.

78
which each participant performed the three blocks was randomized, even though the order was consistent within each participant. In this paper, 20 pairs of words produced by five talkers were analyzed, resulting in 1200 tokens total. The experiment was a three variable factorial design of Accent type (final-accented vs. unaccented) × Environment (isolation, sentence focus, sentence non-focus) × Gender (male vs. female). The utterances were recorded to disk on a computer at the sampling rate of 44.1 kHz.

2.1.4 Measurements

Three acoustic measurements were made for words produced in a sentence and two measurements were made for words produced in isolation. The three parameters were 1) F0 maximum of the second syllable, 2) F0 difference between the first and second syllable, which was obtained by subtracting the F0 minimum of the first syllable from the F0 maximum of the second syllable, and 3) F0 difference between the second syllable and the following particle, which was obtained by subtracting the F0 minimum of the following particle from the F0 maximum of the second syllable. Only the first two values were obtained for words produced in isolation for the obvious reason that F0 difference between the second syllable and the following particle is not measurable when there is no following word. All acoustic analyses were done with Praat (Boersma and Weenink 2005). Each token was annotated and a Praat script was run to extract F0 maxima and minima. Then, the outputs from the script were manually checked. When an utterance contained a lot of breathiness or a talker’s voice became creaky, Praat did not always detect an F0. These utterances were marked when the sound files were being annotated. After F0 maxima and minima were automatically extracted, the values for the marked utterances were manually assessed for accuracy.

2.2 Results

All talkers produced two repetitions of forty words (twenty final-accented words and twenty unaccented words) in three environments (isolation, focus frame, and non-focus frame) to examine the effect of accent and the influence of environment in which these words were spoken. For each token, the absolute F0 maximum of the second syllable, the magnitude of F0 difference between the first and second syllable (F0 rise), and the magnitude of F0 difference between the second syllable and the following particle were measured. The F0 rise was calculated by subtracting the F0 minimum of the first syllable from the F0 maximum of the second syllable. The F0 difference between the second syllable and the following particle was calculated by subtracting the F0 minimum of the particle from the F0 maximum of the second syllable. The values obtained from the two repetitions were averaged except for when one value out of two repetition was missing. Missing values occurred when a syllable was completely voiceless. When the first syllable of the target word was completely voiceless, the F0 difference between the first and second syllable was not measurable. Similarly, the F0 maximum was not measurable when the second syllable of the target word was completely voiceless. When a word was missing one out of two values, the value from the remaining token was taken as the mean of the word. When a word was missing values from both tokens, this resulted in a missing cell. Since the frequency of complete devoicing was variable within and
across talkers, the number of missing cells was variable. Out of 120 cells (40 words in three environments), the number of missing cells ranged from 2 to 29 for the F0 maximum among the five talkers, from 19 to 44 for the F0 difference between the first and second syllable, and from 2 to 31 for the F0 difference between the second syllable and the following particle. Once a talker’s mean F0 maximum of the second syllable, F0 difference between the first and second syllable, F0 difference between the second syllable and the following particle were obtained for each word in each environment, the means for the F0 maximum on the second syllable, the F0 difference between the first and second syllable, and the F0 difference between the second and the following particle were calculated for final-accented words and unaccented words in each environment.

In the following, F0 maximum of the second syllable, F0 difference between the first and second syllable, and F0 difference between the second syllable and the following particle are discussed in order. Because the results presented in this paper are based on data from only five talkers, no statistical analyses were performed. Considering that the five talkers consists of two males and three females and males and females have different F0s, statistical analyses would not be informative.

2.2.1 The F0 maximum of the second syllable

Table 1 shows the absolute F0 maximum of the second syllable. For words spoken in a sentence (both focus and non-focus frames), final-accented words had a higher F0 maximum than unaccented words for all talkers. This difference was very small in the non-focus frame for Talker 4 (only 2 Hz difference). The result that final-accented words have a higher F0 peak than unaccented words is consistent with most previous studies (e.g. Pierrehumbert and Beckman 1988, Vance 1995). As to the effect of focus and non-focus frames, it seems that the F0 maximum is higher in the focus frame than in the non-focus frame. If this tendency turns out to be significant when data from more talkers are analyzed, the result would be consistent with Pierrehumbert and Beckman (1988). It is not clear if the contrast between final-accented and unaccented words becomes more distinctive in one frame than the other. The absolute difference in F0 maximum between final-accented and unaccented words was larger in the focus frame than in the non-focus frame for Talker 4 (15 Hz in the focus frame, 2 Hz in the non-focus frame), but the difference was larger in the non-focus frame for three talkers (Talkers 2, 3 and 5). The difference was the same for Talker 1 (7 Hz). The five talkers showed inconsistent patterns.

Table 1: The absolute F0 maximum (in Hz) for final-accented and unaccented words (standard errors in parentheses)

<table>
<thead>
<tr>
<th></th>
<th>Focus</th>
<th>Non-focus</th>
<th>Isolation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Accented</td>
<td>Unaccented</td>
<td>Accented</td>
</tr>
<tr>
<td>T1</td>
<td>M 132 (2)</td>
<td>125 (2)</td>
<td>121 (2)</td>
</tr>
<tr>
<td>T2</td>
<td>M 148 (3)</td>
<td>133 (2)</td>
<td>155 (3)</td>
</tr>
<tr>
<td>T3</td>
<td>F 283 (8)</td>
<td>255 (5)</td>
<td>243 (6)</td>
</tr>
<tr>
<td>T4</td>
<td>F 271 (5)</td>
<td>256 (5)</td>
<td>236 (3)</td>
</tr>
<tr>
<td>T5</td>
<td>F 256 (6)</td>
<td>239 (5)</td>
<td>244 (5)</td>
</tr>
</tbody>
</table>

Note. Accented = final-accented; T = talker; M = male; F = female.
The distinction between the two accent types was not observed for words spoken in isolation, except Talker 5, who had a higher F0 maximum for final-accented words than unaccented words by 16 Hz. In Vance (1995), the F0 maximum was 17 Hz higher for the final-accented word (/hana^\prime/ ‘flower’) than the unaccented word (/hana/ ‘nose’) for the female talker who made a reliable distinction between the two accent types. For talker 4, unaccented words had a higher F0 maximum than final-accented words. Indeed, the F0 maximum was higher for final-accented words than unaccented words for 14 out of 20 pairs for this talker. Overall, the results for F0 maximum seem consistent with the results from previous research in that final-accented words have a higher F0 maximum than unaccented words when produced in a sentence but the distinction appears to be largely neutralized when the words are produced in isolation (e.g. Poser 1984, Sugito 1983).

2.2.2 The F0 difference between the first and second syllable

For each talker, the F0 difference between the first and second syllable was also examined in each environment for final-accented and unaccented words. In the past, Sugito (1979, 1983) and Vance (1995) analyzed the F0 maximum of the second syllable and the F0 difference between the first and second syllable. Sugito found that the final-accented word /hana^\prime/ had a higher F0 maximum and a larger F0 difference between the first and second syllable than the unaccented word /hana/. Although statistical analysis was only made for the F0 difference between the first and second syllable and not for the F0 maximum on the second syllable, Sugito argued that the size of F0 difference between the first and second syllable determines the accent type of the word. Vance (1995) used the same pair of words as Sugito but found the opposite result from a single talker. Both the F0 maximum and F0 difference were reliable for the talker but the F0 maximum showed a clearer separation between the final-accented and unaccented words than the F0 difference between the first and second syllable.

Table 2 shows the F0 difference between the first and second syllable found in the current research. The F0 difference was larger for final-accented words than unaccented words for all talkers when they were produced in a sentence. For both accent types, it seems that the F0 difference is larger when the word is the sentence focus than when it is not. Comparing the result for F0 maximum of the second syllable and F0 difference between the first and second syllable, it is the case that the F0 maximum is more stable, as indicated by the smaller standard errors for the F0 maximum than for the F0 difference. However, since the number of talkers is limited and no statistical analyses were performed, it is not possible to determine whether the absolute F0 maximum or the F0 difference is more reliable as an acoustic correlate of pitch accent. For words spoken in isolation, there was no consistent pattern between the two accent types. Some talkers produced a larger F0 difference for final-accented words than unaccented words while other talkers produced a larger F0 difference for unaccented words than final-accented words.
Table 2: F0 difference (in Hz) between the first and the second syllable (standard errors in parentheses)

<table>
<thead>
<tr>
<th></th>
<th>Focus</th>
<th></th>
<th>Non-focus</th>
<th></th>
<th>Isolation</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Accented</td>
<td>Unaccented</td>
<td>Accented</td>
<td>Unaccented</td>
</tr>
<tr>
<td>T1</td>
<td>M</td>
<td>39 (4)</td>
<td>32 (4)</td>
<td>25 (4)</td>
<td>18 (4)</td>
<td>22 (2)</td>
</tr>
<tr>
<td>T2</td>
<td>M</td>
<td>27 (4)</td>
<td>17 (2)</td>
<td>37 (3)</td>
<td>16 (3)</td>
<td>15 (2)</td>
</tr>
<tr>
<td>T3</td>
<td>F</td>
<td>91 (14)</td>
<td>62 (11)</td>
<td>36 (7)</td>
<td>13 (3)</td>
<td>48 (9)</td>
</tr>
<tr>
<td>T4</td>
<td>F</td>
<td>75 (8)</td>
<td>53 (6)</td>
<td>21 (3)</td>
<td>16 (3)</td>
<td>73 (6)</td>
</tr>
<tr>
<td>T5</td>
<td>F</td>
<td>63 (8)</td>
<td>46 (7)</td>
<td>54 (7)</td>
<td>29 (5)</td>
<td>55 (6)</td>
</tr>
</tbody>
</table>

Note. Accented = final-accented; T = talker; M = male; F = female.

Figure 1 shows spectrograms for the final-accented word /namiˆ/ ‘wave’ produced by a female talker in the focus frame. The line with small dots that lay on top of spectrograms tracks F0 over time. The F0 rises gradually toward the accented syllable /miˆ/ and peaks just before the end of the syllable. Although F0 tracking is briefly interrupted at the voiceless stop of the following particle /to/, judging from the descending slope of F0 that appears after the voiceless stop, the decline of F0 after the peak seems gradual. For the final-accented word, the overall F0 movement at the target word and the following particle results in a profile that resembles a mountain. By contrast, F0 movement is relatively flat for unaccented words. Figure 2 shows spectrograms for the word /nam i/ ‘mediocre’ produced by the same female talker as Figure 1 in the focus frame. This word is the unaccented counterpart of the word /namiˆ/. There is no particular F0 rising toward the second syllable and no mountain-shaped peak on the second syllable. The F0 also stays relatively flat at the following particle. From the comparison of these two figures, the difference between the two accent types are observable in the F0 peak as well as the F0 difference between the first and second syllable, although data from more talkers are needed to make detailed analyses. It is worth noting in passing that completely voiceless words were also observed for words spoken in a sentence. Spectrograms for the completely voiceless word, /situˆ/ ‘room’ spoken by a male in the focus frame are shown in Figure 3.10 Some research shows that consecutive devoicing (devoicing more than one vowel in sequential syllables) is avoided in Japanese (e.g. Kondo 1997). However, consecutive devoicing was not uncommon in the five talkers’ utterances.

---

10 As mentioned by a few participants in the production study, this word is rarely used by itself. It is usually used with another noun to make a compound noun, such as /jikken-situ/ ‘experiment room’.
Figure 1: Final-accented word /nami/ 'wave' produced by a female talker in the focus frame. The line that lay on top of the spectrograms shows F0.
Figure 2: Unaccented word /nami/ 'mediocre' produced by the same female talker as Figure 1 in the focus frame. The line that lay on top of the spectrograms shows F0.
Figure 3: Completely voiceless final-accented word /situ^/ ‘room’ produced by a male talker in the focus frame. The line that lay on top of the spectrograms shows F0.
2.2.3 The F0 difference between the second syllable and the following particle

Finally, Table 3 shows the F0 differences between the second syllable and the following particle. Results on Japanese pitch accent have consistently shown that F0 of the particle is lower when it follows a final-accented word than when it follows an unaccented word. This measurement served as a reference to check if results from other measurements were in the right ballpark. As expected, all talkers had a larger F0 decline from the second syllable to the following particle for final-accented words than unaccented words. This contrast is also seen in Figures 1 and 2. As in the F0 maximum data, the F0 difference data showed a very small effect for Talker 4.

Table 3: F0 difference (in Hz) between the second syllable and the following particle (standard errors in parentheses)

<table>
<thead>
<tr>
<th>Talker</th>
<th>Focus</th>
<th>Non-focus</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Accented</td>
<td>Unaccented</td>
</tr>
<tr>
<td>T1 M</td>
<td>48 (2)</td>
<td>33 (3)</td>
</tr>
<tr>
<td>T2 M</td>
<td>54 (8)</td>
<td>12 (4)</td>
</tr>
<tr>
<td>T3 F</td>
<td>94 (9)</td>
<td>35 (10)</td>
</tr>
<tr>
<td>T4 F</td>
<td>37 (6)</td>
<td>17 (6)</td>
</tr>
<tr>
<td>T5 F</td>
<td>71 (9)</td>
<td>25 (6)</td>
</tr>
</tbody>
</table>

Note: Accented = final-accented; T = talker; M = male; F = female.

2.3 Discussion & tentative conclusions

This paper presented preliminary results from the research on acoustic correlates of Japanese pitch accent that exploits disyllabic accent pairs of final-accented and unaccented words. In this research, the exhaustive set of disyllabic word pairs that met certain criteria was used to examine if final-accented words are different from unaccented words. All words used were either unambiguously final-accented or unaccented and had relatively high familiarity ratings. The reports in this paper were restricted to be descriptive in nature pending the results from additional talkers. Nevertheless, the production data from five talkers suggest that there are some acoustic differences between final-accented words and unaccented words. However, the difference appeared only when they were produced in a sentence followed by a particle. That is, final-accented words had a higher absolute F0 maximum on the second syllable than unaccented words. The F0 difference between the first and second syllable was also larger for final-accented words than unaccented words. These contrasts between the two accent types were not present when the words were spoken in isolation. These findings are generally consistent with findings from previous research (e.g. Pierrehumbert and Beckman 1988, Poser 1984, Sugito 1983, Vance 1995). Although Sugito (1983) and Vance (1995) found that some talkers produced final-accented words and unaccented words significantly differently in isolation, a majority of their participants did not make such a distinction. While previous studies that used disyllabic accent pairs focused on a few pairs, the current study shows that the contrast between the two accent types holds more generally. The comparison of F0 difference between the first and second syllable in the focus and non-focus frames suggests that the difference is enhanced when the target
word is the sentence focus, regardless of accent type. However, more data is required to assess the robustness of this pragmatic effect. In sum, the data from five talkers indicate that the lexical contrast of final-accent and unaccented in Japanese is carried by prosody in this language.

**APPENDIX  Words used in the production study**

The phonemic representations for each pair are indicated between slashes. The accent patterns are indicated in parentheses on the top. The words in Chinese characters and hiragana are the actual orthography used to present to the talkers.

<table>
<thead>
<tr>
<th>Final-accented (LH^)</th>
<th>Unaccented (LH)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. /aki/ 飽き ‘tiresomeness’</td>
<td>空き ‘vacancy’</td>
</tr>
<tr>
<td>2. /haji/ 恥 ‘shame’</td>
<td>はじ ‘edge’</td>
</tr>
<tr>
<td>3. /hana/ 花 ‘flower’</td>
<td>鼻 ‘nose’</td>
</tr>
<tr>
<td>4. /hane/ 跳ね ‘jump’</td>
<td>羽 ‘feather’</td>
</tr>
<tr>
<td>5. /hare/ 晴れ ‘clear weather’</td>
<td>腫れ ‘swelling’</td>
</tr>
<tr>
<td>6. /hasi/ 橋 ‘bridge’</td>
<td>端 ‘edge’</td>
</tr>
<tr>
<td>7. /hati/ 八 ‘eight’</td>
<td>蜂 ‘bee’</td>
</tr>
<tr>
<td>8. /kaki/ 塊 ‘fence’</td>
<td>柿 ‘persimmon’</td>
</tr>
<tr>
<td>9. /mame/ 豆 ‘bean’</td>
<td>まめ ‘hardworking’</td>
</tr>
<tr>
<td>10. /moti/ 持ち ‘durability’</td>
<td>餅 ‘rice cake’</td>
</tr>
<tr>
<td>11. /nami/ 波 ‘wave’</td>
<td>並 ‘mediocre’</td>
</tr>
<tr>
<td>12. /nori/ 海苔 ‘seaweed’</td>
<td>乗り ‘ride’</td>
</tr>
<tr>
<td>13. /osu/ 雄 ‘male’</td>
<td>お酢 ‘vinegar’</td>
</tr>
<tr>
<td>14. /sita/ 舌 ‘tongue’</td>
<td>下 ‘below’</td>
</tr>
<tr>
<td>15. /situ/ 室 ‘room’</td>
<td>質 ‘quality’</td>
</tr>
<tr>
<td>16. /take/ 丈 ‘length’</td>
<td>竹 ‘bamboo’</td>
</tr>
<tr>
<td>17. /tama/ 玉 ‘ball’</td>
<td>たま ‘infrequent’</td>
</tr>
<tr>
<td>18. /tori/ 取り ‘share (noun)’</td>
<td>鳥 ‘bird’</td>
</tr>
<tr>
<td>19. /tume/ 詰め ‘stuffing’</td>
<td>爪 ‘nail’</td>
</tr>
<tr>
<td>20. /yuki/ 雪 ‘snow’</td>
<td>行き ‘going’</td>
</tr>
</tbody>
</table>

**APPENDIX  Frame sentences**

**Focus frame**  彼 は _______ と 言った。

\textit{kare wa to \textit{iti-ta}}

\textit{He said _____ .}

**Non-focus frame** 誰 が _______ と 言ったの?

\textit{dare ga to \textit{iti-ta-no}}

\textit{Who said _____ ?}
References


