Lexical Accent Status Affects Perceived Prominence of Intonational Peaks in Japanese

Takahito Shinya

Department of Linguistics
University of Massachusetts, Amherst
tshinya@linguist.umass.edu

Abstract

This study shows that lexical accent status affects perceived prominence of fundamental frequency (F0) peaks in Japanese. In Japanese, word accent type can be identified from two different sources: lexical accent status and phonetic F0 contour shape. This study examines whether listeners compensate for the accentual boost of an accented word based only on the word's lexical accent status, when no F0 contour information is available. A perceptual experiment was conducted in which participants judged the relative prominence between two F0 peaks. The experiment showed that for a given second F0 peak height, the first F0 peak height was higher when the first word was lexically accented than when it was lexically unaccented in order for the two words to be equal in perceived prominence. This suggests that the accentual boost of an accented word is subtracted in perception. However, it is also pointed out that another account based on a perceptual compensation for downstep is possible. It is concluded that lexical accent status as phonological knowledge affects perceived prominence of F0 peaks.

1. Introduction

It is known that perceived prominence of accents or F0 peaks in an utterance is affected by a number of factors. One of the most robust factors is F0 excursion size, the vertical distance between the peak of an accent and the baseline, the abstract lower bound of the speaker's pitch range ([2, 7, 11]). With other conditions being equal, greater excursion size evokes greater perceived prominence. Another factor that influences perceived prominence of F0 peaks is F0 declination, a gradual F0 downtrend over the course of an utterance. When perceiving the relative pitch height of two F0 peaks in an utterance, the second peak (P2) needs to be lower than the first peak (P1) in order for P2 to sound as high or prominent as P1. This effect is considered to be due to a perceptual compensation for F0 declination: listeners know that an F0 peak which occurs later in an utterance is lower than an earlier peak due to F0 declination and compensate for it in perception ([5]).

It should be noted that the two perceptual effects described above – F0 excursion size and F0 declination – could both be language-independent. These effects do not assume any language-specific perceptual processes or mechanisms. In this sense, they seem to be purely auditory in character. This raises a question, are there any language-specific phonetic or phonological effects? This study provides a positive answer to this question by examining the perception of F0 peaks in Japanese. It shows that lexical phonological knowledge of word accent type affects perception of intonational prominence.

In Japanese, each word is lexically either accented or unaccented. There is a general consensus among phonologists that this difference in lexical accent status is specified for each lexical item in the lexicon, either in the form of a diacritic ([3]) or in the form of a special lexical tone, i.e. H*+L ([6]).

The difference between an accented and an unaccented word is also reflected in phonetic F0 contour shape. An accented word is phonetically realized with a higher F0 peak (hence larger F0 excursion size) than an unaccented word in the same context ([4, 6]). This is called accentual boost.

In Shinya ([8, 9]), I experimentally showed that lexical accent affects perception of perceived prominence in Japanese. Listeners heard two F0 peaks whose F0 heights and lexical accent type were varied (both lexically and acoustically). I found that P2 had to be lower in an accented-unaccented (*au*) sequence than in an accented-accented (*aa*) sequence to sound equivalent to P1. In explaining this effect, I suggested a perceptual process whereby the accentual boost of an accented word is perceptually subtracted. I called this process *accentual boost subtraction*. Thus, in order for P1 and P2 to have an equal perceived prominence P2 needs to be higher in *aa* than in *au* because listeners subtract the accentual boost that is associated with P2 in *aa*.

Since Shinya's experimental words all had naturally realized F0 properties, we do not know which of the two types of information, lexical accent status and phonetic F0 contour shape, plays a role in accentual boost subtraction. The present study proposes to find an answer to this question. It aims to determine whether or not the perceptual compensation effect of accentual boost is observed based only on accent status as lexical knowledge. In the perceptual experiment reported below, we use words whose F0 properties are ambiguous between accented and unaccented types so that listeners can only use the lexical accent status of the words (= string of segments) to figure out their word accent types. The results show that the effect of accentual boost subtraction is observed on a word when its F0 contour shape does not signal its accent type. The experiment further shows that the same effect is seen even when a word exhibits the pitch pattern of the other accent type. These findings suggest that lexical accent status plays a significant role in perception of intonational prominence.

2. Experiment

2.1. Method

2.1.1. *Stimuli*

The sentences in (1) were recorded by the author (male native speaker of Japanese, aged 32 years at the time of recording) in the sound attenuated booth of the Phonetics Laboratory at the University of Massachusetts, Amherst. The sentences contained three words: noun 1 (N1), noun 2 (N2) and intransitive verb. N1 and N2 were varied with respect to lexical accent type to give all of the four combinations of accented/unaccented N1 and N2 (an accented mora is indicated by an apostrophe). The verbs were all unaccented.

(1) a. accented-accented (*aa*)

Ina'mori-no	ani'yome-ga	inai.
Inamori-Gen	sister-in-law-Nom	not found

Inamori's sister-in-law is not found.

b. accented-unaccented (*au*)

Ina'mori-no	omiyage-ga	kieta.
Inamori-Gen	souvenir-Nom	disappeared

Inamori's souvenir disappeared.

c. unaccented-accented (*ua*)

Inamura-no	ani'yome-ga	inai.
Inamura-Gen	sister-in-law-Nom	not found

Inamura's sister-in-law is not found.

d. unaccented-unaccented (*uu*)

Inamura-no	omiyage-ga	kieta.
Inamura-Gen	souvenir-Nom	disappeared

Inamura's souvenir disappeared.

(Gen=genitive case marker, Nom = nominative case marker)

The recorded materials were stylized with *Praat* ([1]) using the pitch-scaling algorithm called "pitch synchronous overlap add" (PSOLA) such that the sentences had schematized F0 contours expressed only by the tones constituting the contours. After stylization, F0 properties of N1 were manipulated. For each N1 of the four experimental sentences, a 5-step continuum was created such that N1 showed canonical F0 properties for a 4-mora accented word at one endpoint and canonical properties for a 4-mora unaccented word at the other endpoint. The mean F0 values obtained from the seven naturally produced tokens were used for the canonical tonal values. The intermediate stimuli were prepared by changing the F0 height and alignment of N1's peak and the following two valleys (as denoted by V1 and V2 in Figure 1). The peak values (P1), which corresponded to H* accent tone for an accented word and to H phrasal tone for an unaccented word, were 179, 182, 186, 189, 193, and 193 Hz (from unaccented to accented). The values of the valleys were 122, 127.8, 133.6, 139.4, 145.2 and 151 Hz for V1, and 106, 108.6, 112.2, 113.8, 116.4 and 119 Hz for V2 (again from unaccented to accented). It turned out that peak height and its alignment was systematically related between the accented and the unaccented contours in such a way that the peak height was higher and aligned later in the accented contour than in the unaccented one. Therefore, these two properties were covaried. The step size was 3.4 Hz for F0 height and 3 ms for alignment. A systematic but reversed relationship between F0 height and alignment was seen in V1. The V1 height was lower and aligned earlier in the accented contour than in the unaccented one. The two properties were also covaried with the step size of 5.8 Hz for F0 height and 10 ms for alignment. The actual F0 contours used in the experiment are shown in Figure 1.

N2 was varied only in its F0 peak height when it was accented and in its F0 peak height and the following F0 peak (P3) when it was unaccented, with its overall F0 contour shape kept constant. It was also varied in 5 steps, with the step size 14 Hz. When N2 was unaccented, spreading of the H phrasal accent was seen ([10]), which made an F0 "shoulder" at the end of N2 (P3). The actual F0 values of N2 are shown in Table 1. The peak heights were identical between the accented and the unaccented N2. Note that both lexical accent status and F0 contour shape were available in N2. The 6 different F0 heights for N2 were combined with the 6 N1 stimuli, yielding 36 stimuli per continuum. The overall number of stimuli was 144 (36× 4 sentences in (1)).

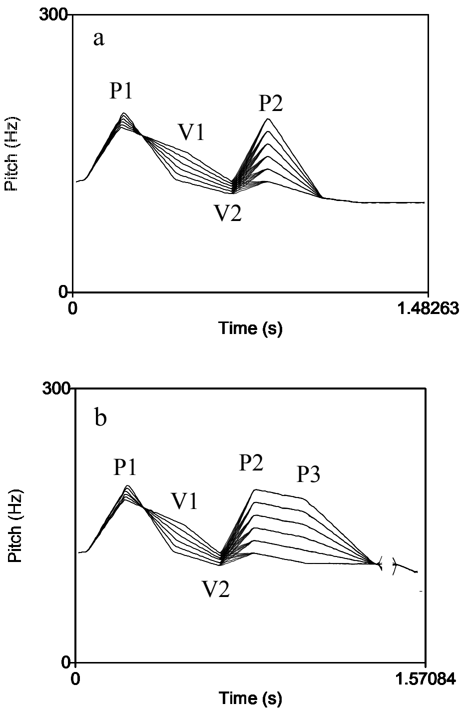

Figure 1 *F0 contours of the stimuli used in the experiment: P2 shows F0 properties for an accented word in (a) and for an unaccented word in (b).*

	Step					
	1	2	3	4	5	6
N2 peak height *a* or *u*	120	134	148	162	176	190
P3	109	123	137	151	165	179

Table 1 *Heights of N2 peak and the F0 shoulder (Hz).*

2.1.2. *Participants*

Twenty-four undergraduate and graduate students were recruited at Sophia University in Tokyo (aged 19-32). All were native speakers of Japanese.

2.1.3. *Procedure and analysis*

Each listener sat at a computer terminal. They randomly heard each stimulus five times using the Multiple Forced Choice listening experiment object equipped with *Praat* ([1]) and were asked to judge which of N1 or N2 is given more "emphasis" by the speaker. The experiment was self-paced. There were 1.5 second inter-stimulus-intervals.

The data were analyzed following Gussenhoven et al. ([2]). For each listener, probit analysis was used to estimate the point where P1 and P2 have the same perceived prominence for each of the 6 N1 stimuli. These crossover point values were used as the measure of perceived prominence.

3. Results

The results of the probit analysis are represented in Figure 2. This graph shows P2 values that have the same perceived prominence as P1. The 5-step N1 continuum from the unaccented to the accented patterns is represented on the *x*-axis. Note that though each N1 stimulus is represented by P1 height, they also differ in other F0 properties such as the valleys that follow P1 and their tonal alignments. The left endpoint stimuli correspond to the F0 properties for the unaccented pattern and the right endpoint stimuli to those for the accented pattern. The function $y = x$ is provided as a reference line. This reference line can be used to help figure out the patterns that the relative prominence judgment exhibits. For example, if P1 and P2 need to have an equal F0 value to be perceived to be equally prominent, the P2 value should be on the reference line. Also, the distance between P2 and the reference line indicates how much lower P2 needs to be than P1 when they sound equal in perceived prominence.

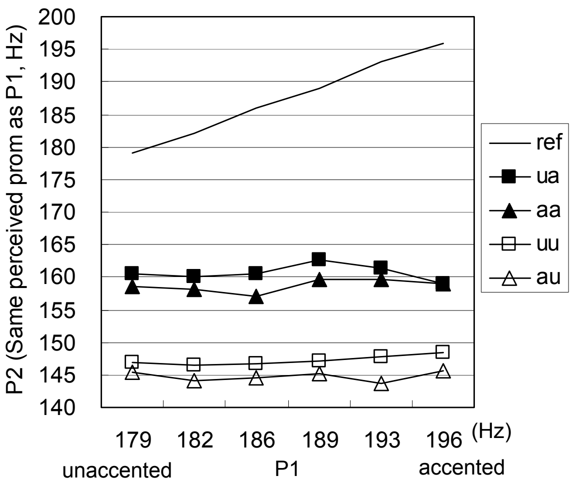

Figure 2 *Mean values of P2 giving the same perceived prominence as P1 for different N1 stimuli. "Ref" is shorthand for "reference", which represents the function* $y = x$.

One can readily see that all of the mean P2 values are well below the reference line $y = x$, which means that for P1 and P2 to be equal in perceived prominence, P2 has to be much lower in F0 than P1. In other words, we replicated the effect of perceptual compensation for F0 declination ([4]).

In addition to the F0 declination effect, we can make three other important observations. First, comparing cases in which only the lexical status of N1 is different, namely *ua* vs. *aa* and *uu* vs. *au*, the mean P2 values are consistently lower when N1 is accented than when it is unaccented in each of the pairs. In Figure 2, we can see that the filled triangles are lower than the filled squares and that the unfilled triangles are lower than the unfilled squares. In other words, *aa* needs larger P1-P2 differences than *ua*, and *au* needs larger P1-P2 differences than *uu* when P1 and P2 are perceived to be equal in prominence.

This pattern can be explained by the idea of accentual boost subtraction mentioned in §1. For a given accent type of N2, the F0 height of P1 needs to be higher when N1 is accented than when it is unaccented in order for P1 to be perceived to be equal to P2. This is so because the accentual boost on N1 is perceptually subtracted when it is accented but no perceptual subtraction occurs when it is unaccented. Consequently, a larger P1-P2 difference is needed when N1 is accented then when it is unaccented.

The second observation is that the pattern pointed out is seen not only when N1 exhibits F0 properties that are ambiguous between the accented and the unaccented patterns but also when it clearly shows the acoustic characteristics for one of the accent types (except between *ua* and *aa* when P1 = 196 Hz). For example, *aa* shows lower P2 values than *ua* even at the unaccented endpoint of N1 stimulus (P1=179). This is a circumstance under which N1 in *aa* shows F0 properties for an unaccented word. The phonetic F0 shape of N1 gives cues for an unaccented word in both *aa* and *ua*. The difference observed between these two conditions suggests that lexical accent status is robust enough that it may override F0 contour shape in assessing intonational prominence.

The third observation concerns the differences between the cases where N2 is (both lexically and phonetically) accented (*ua* and *aa*, i.e. filled symbols in Figure 2) and those where it is unaccented (*au* and *uu*, i.e. unfilled symbols). Across the whole N 1 stimuli, the *ua* and the *aa* conditions, in which N2 is accented, show remarkably higher mean P2 values than the *uu* and the *au* conditions, in which N2 is unaccented. This means that a greater P1-P2 difference in F0 is needed when N2 is unaccented than when it is accented in order for P1 and P2 to sound equally prominent.

This third observation can again be explained by the process of accentual boost subtraction. Since listeners know accented words are realized higher than unaccented words on P2, they undo the accentual boost on P2 in perception and, consequently, P2 has to be higher in F0 when N2 is accented than when it is unaccented in order for P1 and P2 to have the same perceived prominence. This process accounts for the larger P1-P2 differences when N2 is unaccented than when it is accented.

Although the first and the third observations are both consistent with the idea of accentual boost subtraction, they differ from one another in their effect size. The size of the effect is much greater in the third observation than the first one. That is, the differences between cases in which the accent type of N2 differs (*ua* vs. *uu* and *aa* vs. *au*, or filled vs. unfilled symbols) are remarkably greater than the differences between cases in which the accent type of N1 differs (*ua* vs. *aa* and *uu* vs. *au*, or squares vs. triangles). If we recall that only the lexical status information is available on N1 but both lexical status and phonetic F0 contour shape are available on N2, this bigger difference can be seen as an effect of lexical accent status and F0 contour shape being combined.

The statistical analysis supports the observations described above. A three-way repeated-measures ANOVA was performed with accent type of N1 (AccN1), that of N2 (AccN2), and N1 stimulus step (Step) as the independent variables. The results are summarized in Table 2. As seen, only AccN1 and AccN2 are significant, and Step and the all of the interactions are not. From these results, we can reasonably conclude that (1) lexical accent status plays a role in perception of intonational prominence and that (2) the effect gets greater when lexical status is expressed with differences in phonetic F0 properties.

Factor	*F* value	*P* value
AccN1	$F(1,23) = 7.625$	**$p = 0.013$**
AccN2	$F(1,23) = 71.383$	**$p < 0.0001$**
Step	$F(5,115) = 0.420$	$p = 0.834$
AccN1*AccN2	$F(1,23) = 0.313$	$p = 0.581$
AccN1*Step	$F(5,115) = 0.221$	$p = 0.953$
AccN2*Step	$F(5,115) = 1.154$	$p = 0.336$
AccN1*AccN2*Step	$F(2,46) = 0.526$	$p = 0.756$

Table 2 *Results of an ANOVA. A p-value with bold type indicates that the difference is significant.*

4. Discussion

The most important result of this experiment is that lexical accent status as lexical knowledge affects perceived prominence of intonational peaks even when there is no F0 information, or contradictory F0 information. The experiment has shown that for a given P1 F0 value, P2 that has the same prominence as P1 needs to be lower in *aa* than in *ua* and in *au* than in *uu*. In other words, P1-P2 differences are greater when N1 is accented than when it is unaccented when the peaks sound equally prominent. These P1-P2 differences are brought about not by the difference in F0 properties of N1 but by the difference in its lexical accent status. For example, N1 in *aa* is identical to N1 in *ua* with respect to phonetic F0 contour shape, and therefore acoustic properties can never be responsible for the greater P1-P2 differences observed when N1 is accented. This result can be taken as evidence that listeners' knowledge of lexical accent status affects perceived prominence of intonational peaks.

The result described in the previous paragraph clearly raises a problem for models of the perception of intonational prominence that do not take any language-specific factors into account. One such model is proposed by Gussenhoven et al. ([2]). In this model, three primary factors are assumed to affect the perceived prominence of an F0 peak. The first is the position of the F0 peak in the utterance. An F0 peak occurring later in an utterance evokes more perceived prominence than the same F0 peak occurring earlier in the utterance – the perceptual compensation for F0 declination ([5]). The second factor is F0 excursion size. Larger F0 excursion size leads to larger perceived prominence. Finally, Gussenhoven et al.'s model assumes a certain interaction between two peaks occurring in an utterance: when P1 is increased, the amount of P2 that needs to be increased in order to maintain the same prominence as P1 is smaller when P1 is high than when it is low. In other words, the difference between P1 and P2 increases as a function of P1 height.

Note, however, that Gussenhoven et al.'s model does not assume any mechanisms or processes that are language-specific. The experiment provides evidence that knowledge of lexical accent status plays a significant role in prominence perception. Gussenhoven et al.'s model needs to be modified to handle the language-specific effect found in the experiment.

Recall that in order to explain our first observation, I proposed the idea of accentual boost subtraction ([8, 9]): the accentual boost of an accented word is perceptually subtracted in assessing perceived prominence. However, it is also possible to account for the same pattern – greater P1-P2 differences when N1 is accented in its lexical accent status than when it is unaccented – by assuming a perceptual compensation for downstep ([8]). It is well known that the F0 height of a word is realized distinctively lower when the word is preceded by an accented word than by an unaccented word ([4, 6]). A comparable perceptual process which compensates for this production process can be posited. In this process, a word's F0 peak height that is lowered due to downstep is perceptually boosted such that the production effect is undone in perception.

With the perceptual compensation for downstep, our first observation can be explained as follows. P1-P2 differences are greater when N1 is lexically accented than when it is unaccented because when N1 is accented, downstep appearing on the following word is perceptually undone. As a result, when P1 and P2 have equal prominence, P2 is lower in F0 when N1 is accented than when it is unaccented. Thus, P1-P2 difference will be larger when N1 is accented than when it is unaccented.

5. Conclusions

This study shows that lexical accent status is crucially relevant to the perceived prominence of intonational peaks. The perceptual experiment shows that when P2 has the same perceived prominence as P1, P2 is significantly lower in F0 when N1 is accented than when it is unaccented. This result is crucially obtained when N1's accent type is differentiated with respect to lexical accent status only. We saw two possible accounts of this pattern: accentual boost subtraction and a perceptual compensation for downstep. Which process is responsible for the obtained pattern remains to be determined by future study. What is important here is that language-specific phonological knowledge influences perception of intonational prominence.

6. References

[1] Boersma, P. and Weenink, D. 1992-2005. *Praat: Doing Phonetics by Computer*. University of Amsterdam.

[2] Gussenhoven, C., Repp, B. H., Rietveld, A. C. M., Rump, H. H. and Terken, J. 1997. "The perceptual prominence of fundamental frequency peaks," *JASA,* 102, 3009-3021.

[3] Haraguchi, S. 1977. *The Tone Pattern of Japanese: An Autosegmental Theory of Tonology.* Tokyo: Kaitakusya.

[4] Kubozono, H. 1993. *The Organization of Japanese Prosody.* Tokyo: Kurosio Publishers.

[5] Pierrehumbert, J. 1979. "The perception of fundamental frequency declination," *JASA*, 66, 363-369.

[6] Pierrehumbert, J. and Beckman, M. 1988. *Japanese Tone Structure.* Cambridge, Mass.: MIT Press.

[7] Rietveld, A. C. M. and Gussenhoven C. 1985. "On the relation between pitch excursion size and prominence," *Journal of Phonetics* 13, 299-308.

[8] Shinya, T. 2005. "Lexical accent effects on the perception of fundamental frequency peaks in Japanese," presented at the 150th ASA Meeting, Minneapolis, Minnesota.

[9] Shinya, T. in progress. *The Role of Lexical Accent in Perception of Intonational Prominence: Evidence from Japanese and English.* Doctoral Dissertation, University of Massachusetts, Amherst.

[10] Sugahara, M. 2003. *Downtrend and Post-FOCUS Intonation in Tokyo Japanese*. Doctoral Dissertation, University of Massachusetts, Amherst.

[11] Terken, J. 1991. "Fundamental frequency and perceived prominence of accented syllables," *JASA* 89, 1768-1776.