The Perceptual Cues of a High Level Pitch-accent Pattern in Japanese: Pitch-accent Patterns and Duration

Tsutomu Sato

English Department
Meiji Gakuin University, Tokyo
tsato@ltr.meijigakuin.ac.jp

10.21437/Eurospeech.2003-597

Abstract

It has been pointed out that a head-high pitch accent pattern (hereafter HLL) of a loanword in Japanese tends to be flattened and pronounced as a high level pitch-accent pattern (phonetically represented here as HHH) by younger generation.

This paper attempts to clarify how Japanese who are in their twenties distinguish a high level pitch-accent pattern from a level pitch-accent pattern (hereafter LHH) as a form of perception experiment, using speech synthesis techniques.

In the first part of this paper, it will be shown how often HHH patterns appear among 10 Japanese college students' production experiment and the F0 configurations will be investigated. Then, the result of perception experiment indicates that HHH patterns are taken for LHH patterns in faster speech. Finally, it will be suggested that durational differences caused by pitch-accent patterns can be perceptual cues in telling the pitch-accent patterns apart.

1. Introduction

Some experts in certain working fields such as the music business and computers started to flatten a head-high pattern, HLL, of a loanword and pronounce the word with HHH, such as "o-dio" (audio) in the late 1970's The flattening is extensively used by relatively young people nowadays [1][2].

In the following section, the occurrences of HHH, instead of the realization of HLL, will be examined and the acoustic characteristics of HHH will be mentioned.

2. Production experiment

2.1. Data collection

The data to be analyzed in this experiment is a three-mora loanword from English called "shi-n" (scene) which is transcribed as having HLL. The fundamental frequency (F0) curve of the word will be compared with that of a three-mora Japanese word called "shiin" (the death), the pitch-accent pattern of which is a level pattern, i.e., LHH.

10 Japanese college students with the Tokyo pitch-accent system, five males and five females, took part in the data collection. The average age is 23.3 years old. The participant read 20 sentences, which included two sentences shown in (1):

(1) a. shi-n to oto ga hairu. (A scene and sounds come in.)
b. shiin to nazo ga wakaru. (The death and mystery are found out.)

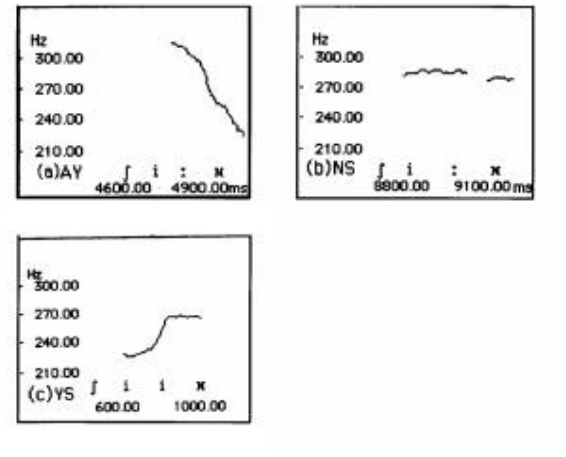
The sentences were shown one by one in random order on a computer display for 4.0 seconds as slow rate, then shown for 2.8 seconds as mid rate, and finally for 1.7 seconds as fast rate, with five repetitions. For presenting the sentences, a HyperCard stack written by [3] was used. In this way, 300 tokens framed in sentences (2 tokens x 3 speech rates x 5 repetitions x 10 participants) were recorded in total.

2.2. The Occurrences and characteristics of HHH

Five out of the ten participants flattened HLL and produced HHH. Four of them uttered the word with a HLL pattern. One participant used both patterns. The F0

curve of HHH differs from that of LHH in that the fundamental frequency value of the first mora of HHH is higher than that of LHH. The comparison of the three different patterns is shown in Figure 1:

Figure 1. F0 curves of "shi-n" by (a) AY, (b) NS, and "shiin" by (c) YS

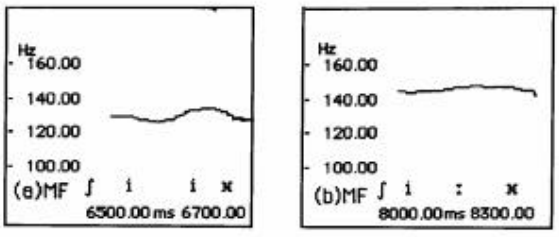


The abstraction and representation of the contours were made by a program called *Signalyze*. The three speakers are all female. The pitch-accent pattern uttered by (a) AY is HLL and the pattern can be considered a standard pattern in Tokyo Japanese [4]. On the other hand, the pattern shown in (b) NS is a new pattern, i.e., HHH. The F0 value of the first mora is, obviously, higher than that of (c) YS, i.e., LHH pattern of a Japanese word "shiin".

2.3. Speech rate effect

It is also found that the F0 range of LHH gets smaller as the speech rate increases. The resulting configuration is shown in Figure 2 (a):

Figure 2. F0 curves of (a) "shiin" and (b) "shi-n" by informant MF



As a comparison, the example of HHH by the same speaker (male) is given in (b). Although F0 peak on the second mora "i" can be still seen in (a), the configuration gets rather flattened.

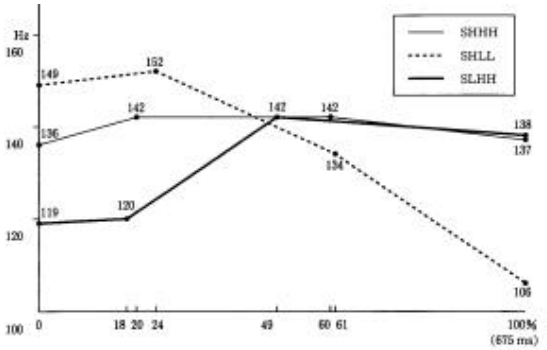
3. Perception Experiment

Then, the question arises of how HHH patterns can be detected by the young and differentiated from LHH in faster speech. This question will be pursued in this section.

3.1. F0 measurement and synthesis

After a careful investigation of pitch configurations of the data in 2.2, four measuring points were judged to be enough to make close-copy stylization of the contours [5]. The F0 measurements were done with all the points in the data and the average values were calculated. At the same time, the average duration of the second and third points in each data were also found. Based on the F0 and durational measurements of the points, the partial-correlation (PARCOR) analysis and synthesis of target contours were done by a computer program called *Onsei Rokubunken for Windows*. Three different patterns, i.e., HHH, HLL, and LHH, in three different speech rates, i.e., slow, mid, and fast, were synthesized; nine synthetic stimuli were made in total. The basic stylization of slow version is schematized in Figure 3:

Figure 3. Synthetic patterns of SHHH, SHLL, and SLHH



The first letter "S" shows the rate of speech, here slow. The average fundamental frequency values are vertically shown in Hertz (Hz) and the average durational values are horizontally shown in percentages (%). 675ms is the

average token duration of all the tokens in a slow mode. Similarly, the mid and fast versions of synthesis were made.

3.2. Participants

Seven participants with the Tokyo accent took part in this perception experiment. Five of them are in their twenties, three males and two females, whose average age is 22.8. Two other participants are older and joined as controls (male, 63 / male, 46).

Each participant was asked to choose which word, i.e., "shi-n" or "shiin" he or she hears through the headphones one by one. After the participant heard three example stimuli at the beginning, nine different stimuli were played five times in random order, and the participant ticked one of the two choices in the answer sheet. In this way, 45 stimuli in total were played to each participant.

3.3. Results

The number of responses to each pattern is shown in Table 1:

Table 1. The number of responses

	shi-n	shiin
SHHH	5	30
MHHH	3	32
FHHH	0	35
SHLL	35	0
MHLL	35	0
FHLL	32	3
SLHH	1	34
MLHH	0	35
FLHH	0	35

The LHH patterns were judged for "shiin" in all speech rate versions, excluding one case of SLHH. The HLL patterns were chosen for "shi-n" in slow and mid versions, but the answer "shiin" was chosen in three cases for FHLL. The number of choices of "shiin" is 30 for SHHH, 32 for MHHH, and 35 for FHHH, as the shaded areas with dots indicate. In other words, in fast versions, no responses were made to the choice "shi-n". In the mid and slow rates, eight responses were found, i.e., 5 to SHHH and 3 to MHHH, as were shown in the shaded areas with slashes.

To summarize, the HHH pattern was judged as "shiin", not "shi-n" in all the fast versions and many cases of slower versions. To put it another way, even younger participants took the pitch-accent pattern in citation form for "shiin" and could not identify the pattern as "shi-n".

4. Pitch-accent patterns and duration

So far, it has been found that the HHH pattern in its citation form is taken as a contour for not "shi-n" but "shiin", especially in faster versions of the pattern. Then, is there any phonetic cue to differentiate the HHH pattern from LHH pattern, other than the pitch-accent contrast?

In [6], the same ten speakers in 2.1 also produced two-mora token, "shin" meaning "pencil lead" in the same way as three-mora tokens, and the average duration of the token (TD), and the vowel "i" (VD) were calculated. In order to make the durational comparison among HHH, LHH, and HLL, the values of tokens and vowels of words "shi-n" and “shiin” were divided by those of "shin". These values are called "durational ratio". As a result, the token durational ratio (TDR) and vowel durational ratio (VDR) of HHH, LHH, and HLL in different speech rates are presented in Table 2:

Table 2. Durational ratios of HHH (left), LHH (middle), and HLL (right)

	Slow	Mid	Fast
TDR	1.45/1.47/1.39	1.58/1.49/1.40	1.50/1.41/1.32
VDR	3.09/2.95/2.57	3.02/2.76/2.45	2.64/2.58/2.27

Excluding the contrast of TDR in Slow, the durational ratios of HHH and LHH are larger than those of HLL. It is assumed in this study that the differences in the durational ratios were caused by the pitch-accent contrasts. In other words, level or rising patterns, i.e., HHH or LHH, might have caused their durational ratio dominance over a falling pattern, i.e., HLL.

A one-sided two-sample t-test with different variances (Welch's test) was conducted on the durational values of TD and VD of all the data. Significant differences were found in the following cases between HHH and HLL: P value=0.04 (72.4ms) in TD/Slow and P value=0.026 (56.5ms) in VD/Slow, P value=0.049 (48.6ms) in TD/Mid, and P value=0.029 (22.8ms) in TD/Fast (the values in braces are differences between their average durations). Significant differences were also found between HHH and LHH in TD/Fast (P value=0.008, 25.0ms) and between LHH and HLL in VD/Slow (P value=0.025, 39.7ms). The cases, whose P values are $0.05 \leqq P < 0.1$, are 0.05 in TD/Slow (48.0ms) between LHH and HLL, and 0.06 in VD/Mid (28.8ms) between HHH and HLL.

[7] and [8], comparing accented words and unaccented words, claimed that accent is not a primary factor or has only a minimal effect in affecting moraic or vowel duration in Japanese. This study, furthermore, makes it clear that there are durational rate differences among accented words with different pitch-accent patterns, at their word and vowel levels, although the statistic analysis for the values did not show significant differences in all the cases. In short, accented words might have their durational differences in accordance with pitch-accent patterns. [9] indicates that level or tail-high pitch-accent patterns cause longer duration of the final vowels in phrase-final nouns. The result of this study supports her findings, while the data here are confined to a limited number of three-mora nouns.

5. Conclusion

In this paper, it has been explored how Japanese people in their twenties perceive a new pitch-accent pattern, HHH, using synthetic speech. The result is that they associated the HHH pattern with a word "shiin" which typically takes a LHH pattern, even in most cases of slow versions, in addition to all the cases of fast versions. In the final section, durational ratio differences among pitch-accent patterns were pointed out and it was assumed that the differences can be perceptual cues in differentiating the pitch-accent contrasts.

For further study, a perception experiment on a larger scale, taking the durational ratio differences into consideration, is planned to be carried out with more participants, especially, older generations.

6. References

[1]Ishino, Hiroshi (1989) Gairaigo. In Tamamura Fumio (ed.) *Kooza Nihongo to Nihongo Kyooiku 6, Nihongo no Goi, Imi (Jyoo).* 97-120. Meiji Shoin.

[2]Inoue, Fumio (1992) Gyookai yoogo no Akusento: Senmonka Akusento no Seikaku. *Gengo*, 21-2, 34-39.

[3]Varden, J. Kevin (1998) *On High Vowel Devoicing in Standard Modern Japanese: Implication for Current Phonological Theory.* Doctoral Dissertation. University of Washington.

[4]NHK Hoosoo Bunka Kenkyuujo (ed.) (1998) Nihongo *Hatsuon Akusento Jiten. Shinban.* Nihon Hoosoo Shuppan Kyookai.

[5]'t Hart, Johan, Rene Collier, and Antonie Cohen (1990) *A Perceptual Study of Intonation: An Experimental-phonetic Approach to Speech Melody.* Cambridge University Press.

[6]Sato, Tsutomu and John A. Maidment (2000) The Acoustic Characteristics of Japanese Identical Vowel Sequences in Connected Speech. *The Proceedings of the 6th International Conference of Spoken Language Processing, Beijing, China.* Vol. 2, 632-635.

[7]Mitsuya, F. and M. Sugito (1978) A Study of the Accentual Effect on Segmental and Moraic Duration in Japanese. *Ann. Bull. RILP,* University of Tokyo, No12, 97-112.

[8]Beckman, Mary (1982) Effects of Accent on Vowel Duration in Japanese. *J. Acoust. Soc. Am. Suppl.* 1, Vol. 71, Spring, 3:35.

[9]Mori, Yoko (2001) Effect of Accentual Fall on Final Lengthening in Japanese. *Journal of the Phonetic Society of Japan*. Vol.5, 92-106.