Initial Pitch in Words Beginning with a CVV Syllable with a Long Vowel in Tokyo Japanese

KAMIYAMA Takeki
Laboratoire de Phonétique et Phonologie (UMR 7018) CNRS / Sorbonne Nouvelle, Paris, France
E-mail: takekik@phiz.c.u-tokyo.ac.jp

ABSTRACT

The study investigates fundamental frequency (F0) differences in the two first morae in Tokyo Japanese words beginning with a heavy syllable containing a long vowel. The two morae that constitute the syllable in question are usually considered to be both high in traditional descriptions. However, it is not well known how other factors, such as the initial onset consonant, or the presence of accent kernel later in the word, influence the melodic pattern. Analyses of a corpus composed of 22 four-mora words and phrases (CVCVCVCV and CVVCVCV) read by four speakers indicate that 1) F0 range is smaller in CVV than in CVCV syllables; 2) in CVV, maximum F0 is higher in accented words than in unaccented words, but the difference in F0 range is not statistically significant; 3) in CVV, minimum F0 is lower, and the range is larger when the initial consonant is voiced than otherwise.

1. INTRODUCTION

The variation of Japanese spoken in and around Tokyo is generally considered to be the standard one today.

The syllable structure of the language is as follows. Each syllable is either light or heavy. Light ones are composed of an optional onset consonant and a short vowel. Heavy ones consist of an optional onset consonant followed either (i) by a long vowel or a closing diphthong, or (ii) by a short vowel, and a moraic nasal or the first half of a geminate consonant. Light syllables thus have one mora, heavy ones two.

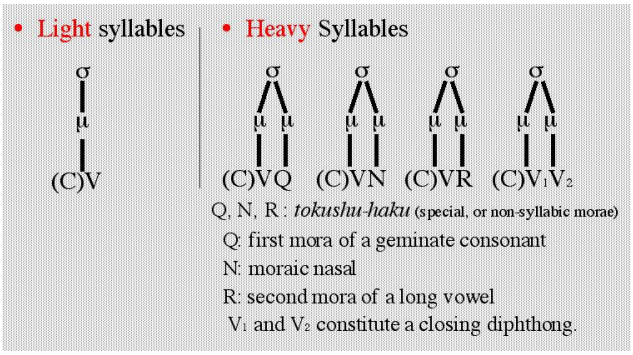

Figure 1: Syllable structure of Tokyo Japanese.

Tokyo Japanese is usually classified as a pitch accent language. Each lexical word is either accented or unaccented. If accented, the accent is realised as a fall in pitch after the accented or marked mora. In unaccented words, no marked fall in pitch is observed. As for the rise toward the accented mora, it can be predicted from the position of the accented mora: if the first mora is accented, the word begins with high pitch, and then passes on to a lower pitch; otherwise (including the case where the word is unaccented), a rise is observed from the low-pitched first mora toward the high-pitched second one. Once the pitch passes from high to low, after the accent kernel, it remains low through to the end of the word. Thus the presence/absence and the position of the accent kernel dictate the melody pattern of the word. The description above is in accordance with those made by most linguists. [1] [2] [3] [4] [5].

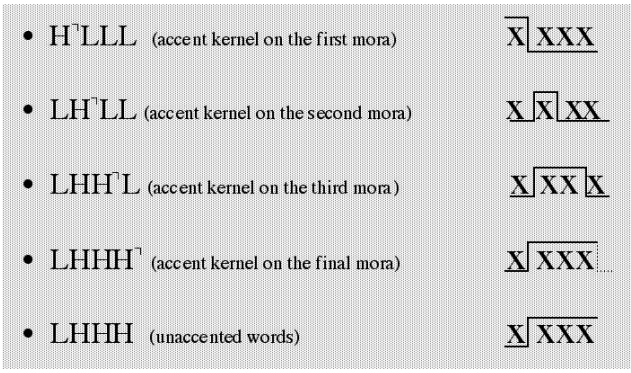

Figure 2: Possible accent patterns for four-mora words ("˥" represents the position of the accent kernel).

The rise in pitch at the beginning of the word, however, is realised differently when the first syllable is heavy. If the syllable in question is composed of a long vowel, the first two morae are described as both high [1] [6] [7]. According to Pierrehumbert and Beckman's acoustic analysis, "[…] inspection of fo contours leads us to believe that phrases beginning with unaccented long syllables also have H tone on the first mora […]. In a long syllable the fo peak is located near the middle, the second mora being entirely high. When the first two syllables are both short, the peak appears at the end of the second one." ([8]: pp. 127-128) Yokotani and Komatsu [9] conducted a series of perceptual experiments on this phenomenon. They used as stimuli isolated words with synthesised F0 patterns. Speakers of Tokyo Japanese were asked to judge whether the stimulus was "a little strange" or "not strange". The results indicate that "H-pronunciation" (with a high-pitched first mora) is accepted in words beginning with a CVR syllable (R: the second mora of a long vowel), but not in those beginning with CVCV, and that "H-pronunciation" is not only

ISBN 1-876346-48-5 © 2003 UAB

accepted but preferred to "L-pronunciation" (with a low-pitched first mora) in CVR syllables.

Based on these observations and findings, we focus on the F0 range (difference between the maximum and the minimum F0) in the first two morae, and aim to investigate whether there is any significant difference in F0 range 1) in words beginning with a CVCV syllable and those with a CVR syllable, 2) in accented and unaccented words beginning with a CVR syllable, 3) in words beginning with a CVR syllable with various onset consonants, 4) in words in a sentence-initial position and in a sentence-internal position.

2. CORPUS AND SPEAKERS

22 four-mora words and phrases were recorded. The first syllable (two first morae in the case of a CVR syllable) contains the vowel /o/. The words are shown in Table 1. Some words have different accent patterns (for the phrase "mosidewa", it may be pronounced with an accent kernel on the first mora, or on the third one). When the speaker pronounced the one we had not expected, we asked them to use the other accent pattern.

Heavy (unaccented)	Heavy (accented)	Light + light (unaccented)	Light + light (accented)
/to o kjo o/	/to o to˥ i/	/to ki do ki/	/to ko ro˥ de/
/do o kjo o/	/do o ke˥ si/	/do te ka ra/	/do te de˥ mo/
/so o me e/	/so o ka˥ si/		
/zo o da i/	/zo o ka˥ ti/		
/ro o bo ku/	/ro o so˥ ku/		
/mo o gju u/	/mo o ka˥ ru/	/mo gu ri te/	/mo si de˥ wa/
/o o mi ja/	/o o ki˥ i/	/o mo na ga/	/o ka ma˥ de/

Table 1: 22 isolated words and phrases ("˥" represents the position of the accent kernel).

The words and phrases were read first in isolation, then in embedded sentences: /sorewa … to iimasu/ ([we] call that …) and /… towa iimasen/ ([we] don't call [that] …).

We recorded four native speakers of Tokyo Japanese all born in Tokyo. They had all their primary and secondary education there. Their age ranged from 23 to 27 at the time of the recording. Speakers 1 and 4 are female; speakers 2 and 3 are male speakers.

Each speaker was asked to pronounce the words and phrases presented one after another on a screen. They were written in Chinese characters with *kana* letters. When they were finished with an item, the next one was presented. Once they had read all the items, we repeated the same procedure for two more times. In the words and phrases recorded, we put two dummy words at the beginning of the list, which we excluded from the analysis. We conducted the same procedures for the sentences.

3. DATA ANALYSIS

We measured the minimum and the maximum F0 using the phonetic analysis software *Praat* [10].

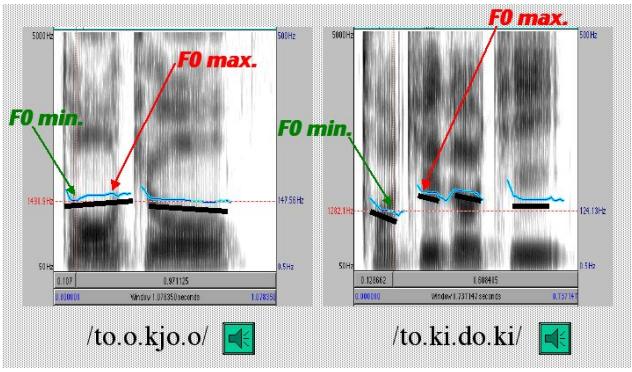

Figure 3: Minimum and maximum F0 in the first two morae.

The definition of the minimum F0 (min F0) and the Maximum F0 (max F0) is as follows:

Min F0: the minimum F0 during the vowel of the first mora (or the minimum toward the beginning of a heavy syllable).

Max F0: the maximum F0 during the first two morae.

There are cases where F0 is difficult to measure; in some speakers' speech the vowel in the low-pitched first mora, especially when surrounded by voiceless consonants, is not intense, regular, or long enough for the software to detect F0 correctly. Even though it is possible to calculate it manually by measuring the period or by using a narrow band spectrogram, those cases were excluded from the data in favour of an invariable measurement.

We then calculated the F0 range using the following formula:

$$(\text{max F0} - \text{min F0}) / \text{min F0} * 100\ (\%) \qquad (1)$$

We also measured the length of the long vowel /o/ in CVR syllables as an indication of the speaking rate.

4. RESULTS

4.0. DURATION OF THE LONG VOWEL /o/ IN CVR SYLLABLES

The mean duration of the long vowel /o/ in CVR syllables was 281, 233, 180, and 235 milliseconds for the four speakers.

4.1. CVCV AND CVR SYLLABLES

Table 2 below shows the mean F0 range for each speaker.

Speaker	1	2	3	4
(C)VR	+10.40	+10.23	+15.99	+11.30
(C)VCV	+34.10	+35.33	+27.04	+25.84
t-test	t63=11.49 **s**	t64=18.83 **s**	t59=6.59 **s**	t64=9.39 **s**

Table 2: Mean F0 range (%) in the first two morae. The letter "s" means that the difference is significant. The convention holds also for the following two tables.

The result of Student's t-test indicate that this difference is highly significant (p<0.0001) for every speaker. We find thus a remarkable difference in F0 range in initial (C)VCV syllables and (C)VR syllables: the difference between min F0 and max F0 is smaller in (C)VR syllables.

4.2. CVR SYLLABLES IN UNACCENTED AND ACCENTED WORDS

The results shown in Table 3 indicate that the presence of accent kernel later in the word (or the phrase) leads to a greater range in the initial (C)VR syllable, but that the difference between accented and unaccented words is not significant at the 5% confidence level.

Speaker	1	2	3	4
unaccented	+8.34	+9.97	+14.59	+10.38
accented	+12.79	+10.49	+17.38	+12.23
t-test	t40=1.70	t40=0.44	t40=1.46	t40=1.00

Table 3: Mean F 0 range (%) in the first (C)VR syllable.

Min F0 (Table 4) shows a similar tendency: it is higher in accented words than in unaccented ones for all the speakers, but the difference is significant for only one of them (speaker 2: t40=2.60, p<0.05).

Speaker	1	2	3	4
unaccented	212.81	132.62	109,86	229.29
accented	226.86	139.29	112.00	238.71
t-test	t40=1.54	t40=2.60 **s**	t40=1.08	t40=1.94

Table 4: Min F0 (Hz) in the first (C)VR syllable.

On the other hand, max F0 in accented words is significantly higher (p<0.05 for all) than in unaccented ones in all the speakers' pronunciation (Table 5).

Speaker	1	2	3	4
unaccented	228.52	145.76	125.62	252.29
accented	249.38	153.76	131.24	267.29
t-test	t40=4.36 **s**	t40=2.98 **s**	t40=3.11 **s**	t40=4.78 **s**

Table 5: Max F0 (Hz) in the first (C)VR syllable.

Thus, both min F0 and max F0 are higher in accented words, but the difference is not always significant for min F0.

4.3. CVR SYLLABLES BEGINNING WITH VARIOUS CONSONANTS

The F0 range in CVR syllables is shown in Table 6. As long as the opposition of voicing is concerned (t-d, s-z), voiced consonants lead to a larger F0 range. Note that F0 values were measured during the vowel and not during the consonant. The difference is significant for all the speakers. When it comes to sonorants and zero consonant (i.e. absence of onset), the result of an ANOVA shows a global effect only for speaker 1 (F2, 15=4.90, p<0.05), and it is hard to describe any general tendencies.

Speaker	1	2	3	4
/t/	+6.53	+6.40	+12.30	+4.66
/d/	+13.16	+14.03	+17.98	+13.14
t-test (t-d) t10=	**4.88 s**	**5.03 s**	**3.11 s**	**3.85 s**
/s/	+3.38	+7.14	+7.76	+3.06
/z/	+12.45	+10.86	+13.22	+13.91
t-test (s-z) t10=	**3.98 s**	**3.15 s**	**2.30 s**	**6.74 s**
/m/	+7.00	+10.00	+17.04	+12.53
/r/	+10.06	+13.79	+21.01	+14.43
/ø/	+21.71	+9.41	+22.60	+17.39
ANOVA (m-r-ø) F2,15=	**4.90 s**	**2.99**	**2.02**	**2.40**

Table 6: F0 range (%) in words beginning with various consonants. The letter "s" means that the difference is significant (t-test), or that there is a significant global effect (ANOVA). The convention holds also for the following two tables.

Min F0 also shows a significant difference in the voiced/voiceless pairs, the former being lower than the latter, but again, the general tendency on the sonorants and zero consonant is not clear.

Speaker	1	2	3	4
/t/	252.50	145.33	123.67	253.50
/d/	208.50	133.00	107.67	224.67
t-test (t-d) t10=	**5.68 s**	**3.78 s**	**6.56 s**	**7.99 s**
/s/	237.17	141.33	119.17	257.50
/z/	207.50	133.33	109.67	221.67
t-test (s-z) t10=	**3.90 s**	**2.78 s**	**4.49 s**	**5.81 s**
/m/	198.33	123.17	108.83	228.00
/r/	222.33	133.83	105.50	229.67
/ø/	200.17	141.67	107.17	223.00
ANOVA (m-r-ø) F2,15=	**7.06 s**	**11.74 s**	**2.25**	**0.85**

Table 7: Min F0 (Hz) in CVR words beginning with various consonants.

Max F0 demonstrates a similar tendency to that of min F0. Note, however, the significantly lower figure for /m/ in the recordings of speakers 1 and 2.

Speaker	1	2	3	4
/t/	259.33	154.67	140.00	265.33
/d/	240.33	151.67	125.33	254.17
t-test (t-d) t10=	**2.27 s**	**0.66**	**5.30 s**	**1.78**
/s/	240.83	151.50	128.50	265.33
/z/	233.17	147.83	124.17	252.50
t-test (s-z) t10=	**0.84**	**0.85**	**1.02**	**1.83**
/m/	212.17	135.50	127.33	256.67
/r/	244.67	152.33	127.67	263.00
/ø/	242.00	154.83	131.33	261.50
ANOVA(m-r-ø) F2,15=	**13.16 s**	**10.94 s**	**1.08**	**0.35**

Table 8: Max F0 (Hz) in CVR words beginning with various consonants.

Thus, voiced obstruants /d/ and /z/ cause the F0 range to be larger, and min F0 to be lower. As for max F0, a similar tendency is observed, but the difference does not reach the level of significance. As for sonorants and zero consonant, no clear tendencies are observed.

ISBN 1-876346-48-5 © 2003 UAB

4.4. POSITION OF THE CVR WORDS IN SENTENCES

Table 9 indicates the F0 range, min F0 and max F0 in CVR syllables at sentence-initial and sentence-internal positions. Even though we find a general tendency that F0 range is larger in sentence-internal position (except for speaker 3), the difference is significant only for speaker 1. We also notice that both min F0 and max F0 are higher in initial position, but the difference reaches the 5% confidence level only in some cases (max F0 for speaker 1 and min F0 for speaker 4).

Initial	**1**	**2**	**3**	**4**
F0 range (%)	**+6.04**	**+15.70**	**+16.02**	**+10.66**
Min F0 (Hz)	213.76	144.12	114.74	253.50
Max F0 (Hz)	226.40	166.45	132.81	279.67
Internal	**1**	**2**	**3**	**4**
F0 range (%)	**+10.02**	**+17.65**	**+15.95**	**+12.82**
Min F0 (Hz)	205.67	141.45	113.81	244.12
Max F0 (Hz)	225.62	165.95	131.36	274.31

Table 9: F0 range, min F0 and max F0 in CVR syllables in sentence-initial and sentence-internal positions.

5. CONCLUDING REMARKS

The results in 4.1. confirm the descriptions presented in the introduction, according to which the two morae in a word-initial (C)VR syllable are pronounced both high.

The findings in 4.2. will be explained by the anticipation of a fall in pitch due to the accent kernel located later in the word or in the phrase. It has been already observed that F0 is higher before an accentual fall ("accentual boost") [5].

The observations in 4.3. is related to the well-known fact that the F0 curve starts lower, and continues to be lower during the vowel, in syllables with an initial voiced stop than in those with the corresponding voiceless stop. The phenomenon seems to have consequences on the development of tones, as described initially by Haudricourt [11] (see also Hombert, Ohala and Ewan's observations [12]).

This paper has focused on the phonetic realisations of the first two morae, which are considered to be phonologically LH. This realisation seems to result from the physiological difficulty in rising F0 during a long vowel.

We have also confirmed the influence of co-intrinsic F0 value. There are also other factors to be considered such as the influence of attitudes, emotions (some characteristics are briefly described by Saito [6] (pp. 129-130) and modalities, which seems to have an overall effects on the sentence. It would be interesting to investigate how much the prosodic pattern resists under the influence of modalities, attitudes and to determine the weight of each component (Vaissière, [13]).

ACKNOWLEDGEMENTS

I would like to express my gratitude to my supervisor Jacqueline Vaissière for her guidance and to Cédric Gendrot for his comments on my first draft.

REFERENCES

[1] S. Kawakami, *Nihongo akusento-hou* (An Introduction to Japanese Accent), Gaku shobo shuppan, 1973.

[2] S. Hattori, *Onseigaku* (Phonetics). Tokyo, Iwanami shoten, 1984.

[3] Z. Uwano, "Nihongo no akusento" (Accent in Japanese), in *Kouza Nihongo to nihongo-kyouiku 2: Nihongo no onsei, on'in* (volume 1) (Series: Japanese language and the teaching of Japanese 2: Japanese phonetics and phonology (volume 1)), pp. 178-205, Tokyo, Meiji Shoin, 1989.

[4] M. Shibatani, *The Languages of Japan*, Cambridge, Cambridge University Press, 1990.

[5] H. Kubozono, *The Organization of Japanese Prosody*, Tokyo, Kuroshio shuppan, 1993.

[6] Y. Saito, *Nihongo onseigaku nyuumon* (Introduction to Japanese phonetics), Tokyo, Sanseido, 1997.

[7] S. Tanaka and H. Kubozono, *Nihongo no hatsuon kyoushitsu* (Introduction to Japanese Pronunciation), Tokyo, Kuroshio Publishers, 1999.

[8] J. B. Pierrehumbert and M. E. Beckman, *Japanese Tone Structure*, Cambridge, MA/London, England, The MIT Press, 1988.

[9] T. Yokotani and M. Komatsu, "Toukyou hougen ni okeru boin-rensa to Initial Lowering hu-tekiyou-sei -- chikaku-jikken houkoku" (Vowel hiatus and the inapplicability of Initial Lowering in Tokyo dialect -- perceptual experiment report). *Shogengo no on-in koozoo to on-in riron ni kansuru sougouteki kenkyuu* (Integral studies on phonological structures and theories in different languages). S. Haraguchi, Ed, pp. 131-153, 2002.

[10] Paul Boersma and David Weenink, *Praat* (http://www.praat.org), 1992-2002

[11] A.G. Haudricourt, "De l'origine des tons en vietnamien", *Journal Asiatique* 242, pp. 69-82, 1954.

[12] J.-M. Hombert, J. J. Ohala and W. G. Ewan, "Phonetic Explanations for the Development of Tones", *Language* 55-1, pp. 37-58, 1979.

[13] J. Vaissière, "Language-Independent Prosodic Features", in *Prosody: Models and Measurements*, A. Cutler and D.R. Ladd, Ed, pp. 53-66, 1983.