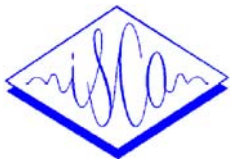

INTERSPEECH 2004 - ICSLP
8th International Conference on Spoken Language Processing
ICC Jeju, Jeju Island, Korea
October 4-8, 2004

Spread of High Tone in Akita Japanese

Kenji Yoshida

Department of Intercultural Communication
Shoin University, Atsugi-City, Kanagawa, Japan
kenjiyo@shoin-u.ac.jp

Abstract

The present study explores the phonetic implementation of accentual H(igh) tone of the Akita dialect of Japanese, specifically its timing control. Lexical accent in the dialect is implemented as an eminence in F0. The conspicuous feature is that F0 does not fall sharply as in Tokyo Japanese. Two hypotheses are examined concerning the extent of the duration of the pitch summit. One is that the H lasts for some fixed period of time. The other is that it lasts until some prosodic edge appears. An experimental study was performed for two native speakers of Akita, with a speaker of Tokyo dialect. The comparison elucidates the conspicuous characteristics of the phonetic implementation of the lexical accent differing from Tokyo Japanese. The result grossly supports the latter hypothesis. It suggests that the H tone in the dialect is phonetically implemented so that it lasts until the end of a Intonational Phrase.

10.21437/Interspeech.2004-83

1. Introduction

Intonational structure of world languages are currently one of the fields of attraction in phonology. The Japanese language (two major dialects, Tokyo and Osaka) is the one with which the leading theory has been constructed [1]. Therefore, dialectal variation of Japanese language is especially worth scrutinizing. The dialect put to examination here is the Akita dialect of Japanese, whose accent system is classified as a variety of Tokyo type, in view of the historical development of the accentual system of the Japanese language [2].

Akita Japanese has a system of lexical accent where the number of contrasting accent categories increases as the length of the word increases (n+1 types for n-mora word) [3]. Despite this similarity at the phonological level to Tokyo Japanese, the phonetic implementation is fairly different. An important characteristic is that the delay of the F0 peak [4]. The F0 peak for some instances is delayed beyond the magnitude normally observed in Tokyo Japanese. This is the reason why I posited the L*+H tonal specification for the broad focus speech of the dialect [3].

The other characteristic examined here is the F0 control after the L*+H. The sharp F0 fall observed in Tokyo Japanese is rather rare in Akita Japanese, instead, F0 falls gradually and a stretch of high pitched plateau is observed as in Figure 1. It is inappropriate to posit HL here. The question arises about the phonetic implementation after the H tone and its tonal specification. Two working hypothesis are possible.

A) The high plateau is controlled to last for a certain fixed interval (hereafter, *Fixed Duration Hypothesis*).
B) The high plateau is controlled to last until some prosodic edge comes, e.g. until the end of the intonational phrase (hereafter, *Edge Alignment Hypothesis*).

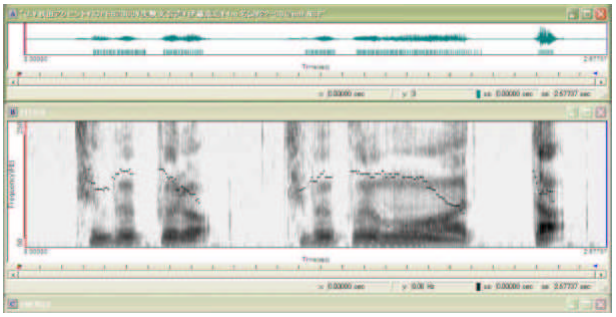

Figure.1 An example of H spread (tebukuro by NS)

We can find the cases where one of these hypotheses appear to apply in descriptive studies of Japanese dialects. There are dialects where ascending intonation is put on the accent and the F0 falls instantly, e.g., Narada [5], which appear to conform to the *Fixed Duration Hypothesis*. There are also dialects where the F0 does not fall until the end of 'bunsetsu' (grammatical unit largely corresponding to the Intonational Phrase), e.g., Aomori [6], which appear to conform to the *Edge Alignment Hypothesis*. But these studies are based on the description by auditory impression. There are no previous studies exploring the issue with numerical approach. The author designed an experiment examining the hypothesis. This is a preliminary report of the results.

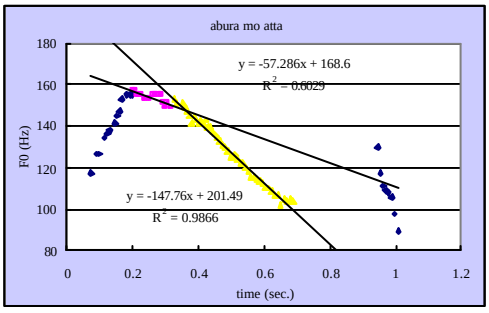

Figure.2 An example of the determination of the end of H plateau by the Angled Linear Regression Analysis

2. Procedure

2.1. Subjects

The subjects are two native speakers of Akita Prefecture, in the northwestern part of the main island of Japan. Both are male and over 70 years old who has spent all their lives in their native town. They will be referred to here as TA (for Tashiro) and NS (for NishiSenboku). NS was the most conservative and typical speaker in the previous studies by the present author [3, 4, 7]. The speech data from a speaker of Tokyo Japanese was also examined as a baseline. He was born and raised in Tokyo (Nakano-ward) and will be referred to here as TO (for Tokyo).

Subjects (1)

TA, Tashiro-Town, 80 years old, male, forest service
NS, Nishisenboku-Town, 70 years old, male, farmer
TO, Nakano, Tokyo, 54 years old, male, math teacher

2.2. Speech Materials

The test words were 128 nouns of 3 and 4 mora length. I will report here the result of thirty 4 mora noun as shown in (2). All the words were pronounced twice with the carrier sentence '-- mo atta (There is also --)'. The target noun coupled with the particle 'no', constitutes an Intonational Phrase. Some words belong to the unaccented class, and accordingly, no F0 excursion due to accent was observed. Discarding these tokens and the cases where measurement was impossible (due to breathy voice), we got 41 observations for subject TA, 44 for NS, and 43 for TO.

List of test words (2)

asagao 'morning glory', ashioto 'footstep',
amazake 'drink made from the sake sediment',
uguisu 'Japanese bush warbler',
uzumaki 'gyre', umeboshi 'salted plum',
oto'oto 'younger brother', oyayubi 'thumb',
kanemochi 'the rich', kamakiri 'mantis',
kaminari 'thunder', kyo'odai 'brother',
kusuriya 'pharmacy', ko'omori 'bat',
sakamichi 'slope', shi'itake 'a kind of mushroom',
sho'ogatsu 'beginning of a year',
sentaku 'laundry', tebukuro 'mitt',
tomodachi 'friend', niwatori 'chicken',
ninjin 'carrot', nokogiri 'sew',
matsutake 'a kind of mashroom',
mizu'umi 'lake', mitsubachi 'honey bee',
murasaki 'violet',
mochitsuki 'pestling of the race cake',
mono'oki 'closet', ro'osoku 'candle'

The speech materials were recorded at 20 KHz sampling rate and 16 bit quantization directly to a portable PC at the fieldwork location. They were classified into 5 accent categories (0, 1, 2, 3, and 4). The numbers stand for the location of the lexical accent. The classifications were made by auditory impression. The data of isolated form were used as an additional clue of the classification.

2.3. Measurements

All the measurements were made on Multispeech Signal Analysis Workstation 3700 (Ver. 2.6.2) by Kay Elemetrics. First, timing measurements were made at the following acoustic landmarks. 1) the start of the accented mora, 2) F0 peak 3) the end of the Iintonational Phrase (end of the particle 'mo'). In order to calculate the duration of the H plateau, the onset of highest F0 was chosen for F0 peak if there are several consecutive points of the same value.

Secondly, the end of the H plateau must be decided. This is, however, not a straightforward task because H plateau should not strictly be the stretch of the same F0; it may rather be a gradual F0 fall yielding the natural F0 declination. In the present study, following procedure was adopted. All the raw F0 values and its timing are pooled from F0 peak to the adjacent bottoming of F0 and fed to the *'Angled Linear Regression Analysis'*. I used the program offered on the website of Professor Shigenobu Aoki [8].This is an protocol searching for the best point where a bending F0 trajectory is divided into two separate lines. It gives the point where the square sum of the residual of two regression analysis is the lowest. I take it as the end of the H plateau. Similar method was used for deciding the F0 turning point for Dominican Spanish [9]. An example of the analysis is given in Fig 2.

2.4. Normalizing the data

The raw data were then fed to the following normalization procedure for the comparison. What we want to examine here is the relative location of *the end of the H plateau* (HE) against the referential duration, i.e., *the time from the start of the accented mora* (Acc) to *the end of the intonational phrase* (IPE). It is better to represent the proximity to the IP edge as some relative term for the statistical comparison, so that sentence to sentence variation of duration can be canceled out. Therefore the total duration of the referential duration was divided by *the number of the mora within it (Acc to IPE* (mora)). The time from the end of the H plateau to the end of the intonational phrase is then divided by the value obtained by the preceding calculation. The result can be regarded as *the proximity of the HE to IPE normalized in terms of the mora unit* (Pr). This measure is given by the simple equation in (3).

$$\mathrm{Pr} = \frac{\mathrm{HE\,to\,IPE(ms)}}{\mathrm{Acc\,to\,IPE(ms)} \,/\, \mathrm{Acc\,to\,IPE(mora)}} \quad (3)$$

The other unit for the analysis is *the duration of H plateau* (HDur). This is *the duration of the high plateau*, i.e., the time from F0 peak (PK) to the PE, normalized by the same referential duration as in (3), which is given by the following equation.

$$\mathrm{HDur} = \frac{\mathrm{PK\,to\,IPE(ms)}}{\mathrm{Acc\,to\,IPE(ms)} \,/\, \mathrm{Acc\,to\,IPE(mora)}} \quad (4)$$

These two values enable us to assess the two hypotheses given in Introduction. If the *Fixed Duration Hypothesis* is correct, HDur remains constant irrespective of the Pr. In contrast, if the *Edge Alignment Hypothesis* is correct, HDur vary so as to keep the Pr small enough to align with IPE.

3. Results and discussion

3.1. Overall results and comparison of the speakers

We begin with the overall results of each subject. Figure 3-5 show the scatterplots of HDur against Pr of three subjects.

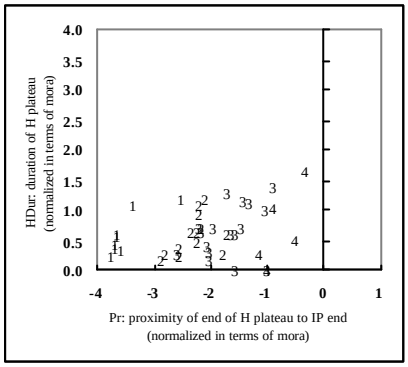

Figure.3 HDur against Pr, Tokyo(TO).

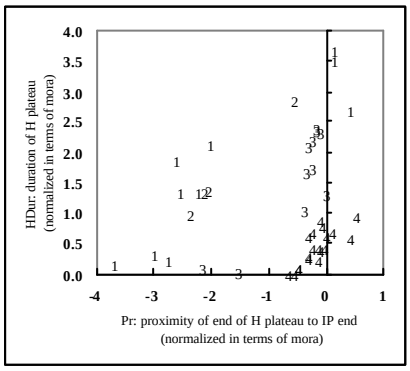

Figure.4 HDur against Pr, Nishisenboku (NS).

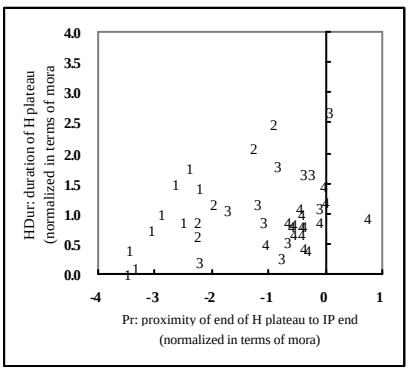

Figure.5 HDur against Pr, Tashiro (TA).

First, we examine the data of TO (Figure 3). H plateau do not extend over the 1.5 mora length. The ends of H plateau are located around the range of -4 to -3 about Accent Category 1, and -3 to -2 about Accent Category 2, etc. This means that F0 falls fairly soon after the F0 eminence by the H tone, in accordance with the HL tonal specification of Tokyo Japanese. Some cases where the F0 fall is delayed in a great degree are the result from the carryover of F0 eminence due to the vowel devoicing [10]. Secondly, we look at the data of NS (Figure 4). Many of the data scatter along the Y axis at 0 (IP edge), suggesting the alignment to the IP edge. The duration of H plateau can be lengthened to achieve this. There are, however, also some data which scatters far from the IP edge. These are similar to those of Tokyo Japanese. There seems to be two modes of phonetic category for implementation of H plateau [11, 12, argued in the next section]. Thirdly, we look at the data of TA (Figure 5). Timing control of H plateau is somewhere in between of TO and NS. HDur is longer than TO, however unlike that in NS, there is no clear alignment to the IP edge. TA lies in the middle of TO and NS.

The difference observed above was confirmed statistically with a 2 way ANOVA (Table 1) with the HDur as the dependent variable, and Speakers and Accent Categories as explanatory variables. Simple effects of Speaker are highly significant for both Pr and HDur. Post hoc test of multiple comparison reveals significant differences between all speakers ($p<.01$), confirming that speakers exert different timing control on the extent of H plateau

		SumSquare	*d.f.*	F value	p	
Pr	Speaker	17.6	2	22.8	0.000	**
	Category	84.0	3	72.5	0.000	**
	Spk*Cat	4.6	6	2.0	0.072	n.s.
HDur	Speaker	8.4	2	10.8	0.000	**
	Category	4.9	3	4.2	0.007	**
	Spk*Cat	8.1	6	3.5	0.004	**

Table 1 Result of ANOVA

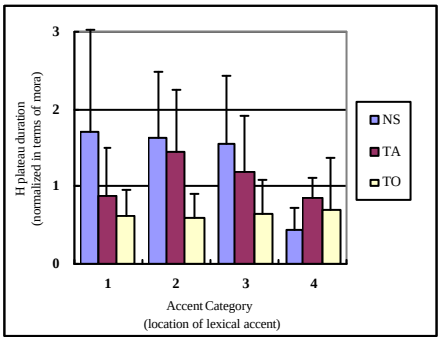

Figure.6 HDur by Speakers and Accent Categories

The detailed results of the HDur are in Figure 6 (mean and standard deviation.). The durations of H plateau of TO are invariable, whereas those of NS vary drastically. However, since there are the tokens with the H plateau similar to Tokyo

Japanese, the average duration does not extend over to 2 mora length. TA lies in between those two speakers. The detailed results of Pr are in Figure 7. The proximity to IP edge systematically changes for TO; a. reasonable result with the existence of HL tonal specification in Tokyo Japanese. In contrast, the proximity does not show linear change for NS. The end of H plateau is nearer to the IP edge. TA again lies in between those two.

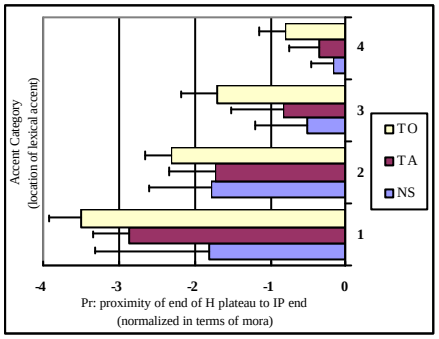

Figure.7 Pr by Speakers and Accent Categories

3.2. F0 implementation of H plateau

The result above confirms that F0 implementation of NS is in clear contrast with TO; TA is intermediate. There seems to be two kinds of timing control in Akita Japanese. Timing control of NS conforms to *Edge Alignment Hypothesis*. It is appropriate to posit phrasal H anchored to the end of IP. Unlike in Tokyo Japanese, timing of F0 fall after the eminence of accentual H is highly variable, thus not reliable cue of accent location. This is in accordance with the highly variable F0 peak in Akita Japanese [4]. Alignment of L relative to the accented mora shows better stability [7]. Those results in combination support the conclusion that L*+H would be a more likely tonal specification [7]. The Timing control of TA conforms neither hypothesis A or B. F0 fall starts before the IP edge, but it occurs significantly later than in Tokyo Japanese, but without any stable duration. This suggests that there exists no specification of F0 fall (no L tone), but there is no phrasal H either. The consequence is shallower and variable F0 falls toward the adjacent IP.

The difference between NS and TA is suggestive given the discussion of the origin of ascending intonation in the northern dialect of Japan. Uwano [12] discuss that the ascending intonation at the lexical accent is historically derived from descending type. Therefore, it is reasonable that in some context the original feature of descending intonation appears. The phonetic implementation of TA appears to be this instance, reflecting the development from the Tokyo type to Akita type (NS). Some tokens of NS and TA are not different from those of TO. These instances are presumably under the influence of standardization, but some equivocal instances of NS may have resulted from the phonetic implementation similar to those of TA.

4. Conclusions and future directions

The result obtained confirms the following conclusions.
1) Akita Japanese has spread of accentual H tone greater than Tokyo Japanese, suggestive of the lack of L tone.
2) One of the speakers shows the alignment to the IP end, suggestive of the existence of phrasal H.
The results also suggest an implication about the process of phonological change (slide of the lexical accent toward the following mora); it appears more fruitful to focus on the timing of L rather than H.

5. Acknowledgements

I am deeply grateful to the subjects for their cooperation. I also thank to Yosuke Igarashi for recommending the 'Angled Linear Regression Analysis'. This study is supported by the Grants-in-Aid for Scientific Research of Japanese Ministry of Education, Culture, Sports, Science and Technology, No. 14710297.

6. References

[1] Pierrehumbert, J. B. and Beckman, M., *Japanese Tone Structure*, MIT Press, Cambridge, 1988.
[2] Akinaga, K,, *Accent dictionary of Japanese*, 3rd ed., Sanseido, Tokyo, 2002.
[3] Yoshida, K., "The delay of the timing of F0 minima in Akita Japanese", poster presented at Conference of Laboratory Phonology 9, Illinois, 2004.
[4] Yoshida, K., "Timing control of the F0 peak in Nishisenboku dialect (in Japanese)", *Proc. of the 16th meeting of the Phonet. Soc. of Japan,*: 165-170, 2003.
[5] Uwano, Z., "Nihongo no akusento (Accent of Japanese language)", Iwanami Koza Nihongo 10, dialects, Iwanami Publishers, Tokyo, 1977.
[6] Kato, M., Mitsui, H., Onishi, T., and Shimura, F., "A report of dialect survey in Tsugaru region, Aomori)", *Bulletin of Research Institute of Japanese Culture, Suppl*., 25, 578-594, 1988.
[7] Yoshida, K, "Tonal Specification of Two Falling Pitch Patterns in Akita Japanese", *Proc. of the 15th ICPhS, Barcelona*, 1293-1296, 2003.
[8] Aoki, S., "Java Script for Statistical Analysis", http://aoki2.si.gunma-u.ac.jp/JavaScript/corr3.html,
[9] Willis, E., Prenuclear low tone alignment in Dominican Spanish, *Proc. of the 15th ISPhS, Barcelona*: 2941-2944, 2003.
[10] Kitahara, M., and Amano, S., Perception of pitch accent categories in Tokyo Japanese, *J. of Ling. Soc. of Japan*, 120:1-34, 2001.
[11] Pierrehumbert, J. B., "Phonetic diversity, statistical learning, and acquisition of phonology", *Lang.and Speech*, 46(2-3):115-154, 2003.
[12] Uwano, Z., "On the ascending kernel", *The bulletin of the Phonetic Society of Japan*, 199:1-13, 1992.