The Tone-Bound Declination Slope: Evidence from Japanese

Mariko Sugahara
University of Edinburgh
E-mail: sugahara@ling.ed.ac.uk

ABSTRACT

In this paper I show that there are two types of time-dependent downtrend in Tokyo Japanese. One of them is a global declination which unfolds across a wider domain, possibly an Intonational Phrase or an Utterance. The other is a "tone-bound" declination, which is a new finding in the intonational studies. I propose in my downtrend model that the tone-bound declination lasts until the tonal value of each tone is achieved and it is reset immediately after that achievement.

1. INTRODUCTION

It is known that time has the effect of lowering F0 in many languages such as in Dutch ([1]), in English ([4]), in Danish ([9]), and in Japanese ([6], [5]). Following the predecessors, I refer to the time-dependent downtrend of F0 as declination. The time-dependent declination is known for its global nature: the F0 value of tones gradually lowers (about -10 ~ -15 Hz per second) as the duration between those tones and the onset of an utterance increases. Given this nature, it is widely accepted that the time-dependent declination is due to subglottal air pressure decrease during the course of an utterance (Gelfer et al., 1983). With respect to speech perception of tonal values and prominence, the declination is formalized as a gradually declining slope of the *base line* of a pitch range unfolding over the whole utterance or across phonological phrase boundaries ([3]). According to them, tones are superimposed on the gradually declining slope without changing their relative tonal values. Henceforth, I refer to the time-dependent declination of this character as "global declination".

However, there is little literature on time-dependent downtrend that systematically investigated its more local aspects, for example effects of passage of time on a local domain such as between two neighboring tones. My experimental work presented in this chapter reveals that the effect of time-passage within the local domain of two neighboring tones is far greater than the downtrend expected by the global declination. I argue for an additional "tone-bound" declination, which I refer to as "local declination" from time to time. The adequate model of downtrend should be able to provide a formal representation of the local declination as well as the global declination. I propose that tones be superimposed on tone-bound declination slope, a slope that expands only two neighboring tones, T1 and T2, which is reset every time the F0 target of T2 is achieved. This tone-bound slope is superimposed on the global declination slope. In the following sections, I present the design of my declination experiment and its results.

2. THE EXPERIMENT

2.1. DESIGN OF THE TARGET SENTENCES

The target sentences contain a sequence of two target words, which are shown in (1). Those target words are unaccented, and syntactically form an immediate constituent, i.e. there is no large syntactic boundary at the left edge of the second word. The number of syllables of the initial word was varied from three to twelve syllables. Except for the shortest one consisting of only three syllables, compound words were used for the initial word. It is because compound formation makes it easier to increase the number of syllables within a single word. Also, presence of a compound initial word is considered to prevent those two words coalesced into a single Minor Phonological Phrase, a.k.a. Accentual Phrase and abbreviated as MiP ([7], [8]). That is, we expect Word1 and Word2 of the forms in (1) to be mapped onto separate Minor Phonological Phrases.

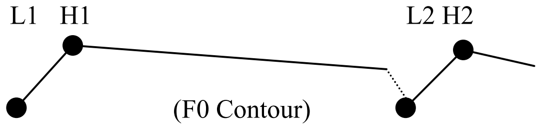
(1) Target Phrases

	[Word1	**]$_{\mathrm{MiP1}}$**	**[Word2]$_{\mathrm{MiP2}}$**
a,	*yama-no* barn-Gen		*umaya-no* barn-Gen
b.	*yamamura-no* (village name)-Gen		*umaya-no* ... barn-Gen …
c.	*yamanakamura-no* (village name)-Gen		*umaya-no* ... barn-Gen ...
d.	*yamanakagawamura-no* (village name)-Gen		*umaya-no* ... barn-Gen ...
e.	*minamiyamanakagawamura-no* (village name)-Gen		*umaya-no* ... barn-Gen ...

The left edge of a Minor Phonological Phrase in Japanese is marked with a sequence of LH edge tones: the L edge tone is associated with the initial syllable and the H edge tone is typically associated with the second syllable of the Minor Phonological Phrase. Therefore, we expect LH tones to be present both at the left edge of Word1 (MiP1) and at the left edge of Word2 (MiP2) as shown schematically in (2). From now on, I refer to the edge tones of Word1 (MiP1) as "L1" and "H1" and those of Word2 (MiP2) as "L2" and "H2".

 ISBN 1-876346-48-5

(2) [□… syllable number varied …□]$_{MiP1}$ [□ □ □ □]$_{MiP2}$



The core question asked in this experiment is whether the L2 and H2 edge tones of MiP2 (Word2) undergo time-dependent declination as the duration of MiP1 (Word1) increases.

2.2. ELICITATION PROCEDURE

Each of the reading materials was presented to the speakers typed on a card using the Japanese writing system (i.e. Japanese *kanji/kana* orthography), and was embedded in a dialogue. The speakers and the experimenter (i.e. the author) played hypothetical roles in a conversational exchange. Four female Tokyo Japanese speakers participated in recording sessions: AS, MR, NK, and SK. Except for AS, who was at her late fifties, they were all at their late twenties. They were all born and raised in the Tokyo area, and had no speech impairment. Recordings were made at a sound proof studio in the University of Tokyo, Komaba, using a SONY DAT recorder. The recorded utterances were re-digitized in the sample rate of 22kHz. Their F0 was analyzed by PitchWorks on a Mac PowerBook G3.

2.3. MEASUREMENTS

Tokens without the Minor Phrase boundary between Word1 and Word2 were excluded because absence of the LH edge tones at the left edge of Word2 makes it impossible to ask the core question. The measurement points are the peak F0 of MiP1 which corresponds to the H1 edge tone of that phrase, the peak F0 of MiP2 which corresponds to the H2 edge tone of that phrase, and the F0 of the onset of MiP2 which corresponds to the L2 edge tone. At the same time, the duration between H1 and H2 and that between H1 and L2 were measured.

2.4. ANALYSES

In the following, I introduce analyses adopted in this paper to obtain the time-dependent declination rate of H2 and that of L2.

Though I did not ask speakers to vary their pitch ranges, it is still possible that there is a substantial variation in the F0 height of H1 (i.e. the H edge tone of MiP1). Also, it is likely that the F0 value of H1 has an effect on that of H2 and L2. Such effect of H1 on L2 and that on H2 need to be factored out to obtain the pure effect of duration on the F0 of L2 and H2. For that purpose, "residual analyses" were used following [5].

The effect of H1 on L2 and that of H1 on H2 were factored out by obtaining residual values from the regression analysis where L2 or H2 is the dependent variable and H1 is the predictor. An example of the relationship between H1 and H2 is shown in Figure 1.

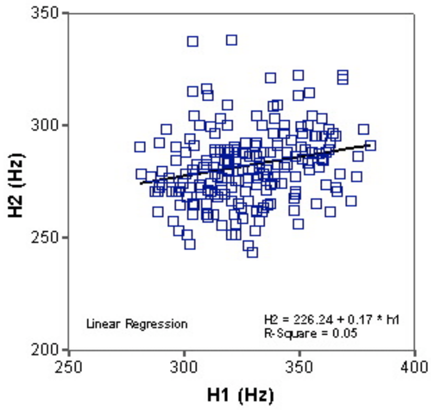


Figure 1: Relation between H1 and H2. Speaker AS.

Those residual values are, then, compared with the duration between H1 and L2 or H1 and H2 in the subsequent regression analyses, which I call "residual analyses" henceforth. If the variations among those residual values are well predicted by the syllable number variation (or the duration variation) between H1 and L2 or between H1 and H2, then we can conclude that the duration has an effect on the F0 of L2 or H2. More concretely, if the residual values decline as the duration between those two tones increases, then it is a good indication that L2 and H2 undergo time-dependent declination. The regression slope obtained from the residual analyses will be translated into time-dependent declination slope of the L2 and that of the H2 edge tone. In Section 3, I show results of the residual analysis.

3. RESULTS

3.1. DECLINATION OF H2

I first show results of the residual analyses in which the residual value of H2 is the dependent variable and the duration between H1 and H2 is the predictor. For all four speakers, negative regression slopes were obtained. Among three of the four speakers, the negative (declining) regression slope was significant ([AS: coefficient = -0.014, Std Error = 0.0032, t = -4.33, $p < 0.001$], [SK: coefficient = -0.018, Std Error = 0.0043, t = -4.27, $p < 0.001$], [NK, coefficient = -0.013, Std Error = 0.0022, t = -5.92, $p < 0.001$]). This result confirms the presence of the time-dependent declination slope across a Minor Phrase boundary. The only exception is MR. Though her H2 residual data points are also fit to a negative regression slope, the slope is not statistically significant (coefficient = -0.008, Std Error = 0.0054, t = -1.41, $p = 0.17$). An example of a scatter plot that represents the regression analysis (residual analysis) is shown in Figure 2.

ISBN 1-876346-48-5

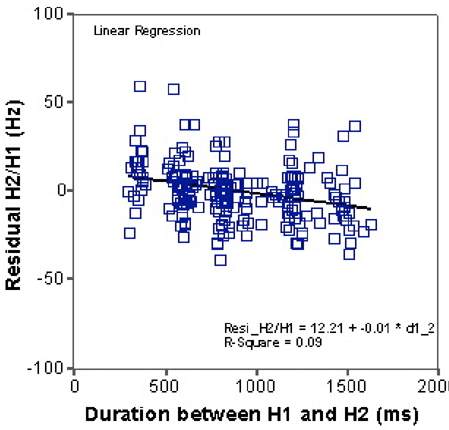


Figure 2: Residual values of H2 (the vertical axis) and duration between H1 and H2 (the horizontal axis).

The rate of the time-dependent declining slope of the residual values of H2 (i.e. the negative coefficient of the declination slope) is expressed in terms of *m s* (millisecond), and they are translated into the *sec* (second) terms. The rate per second is -14 Hz/sec for AS; -8 Hz/sec for MR; -18 Hz/sec for SK; and -13 Hz/sec for NK. This result is similar to the rate obtained by [5], i.e. about -10 Hz/sec, and confirms the presence of a declination slope unfolding across a phonological phrase boundary. This is not, however, the only observation made in this experiment. The additional finding is provided in Section 3.2.

3.2. DECLINATION OF L2

In addition to the global declination of the H2 edge tone between -8 and -18 Hz/sec, I found more substantial declination slope associated with the L2 edge tone of MiP2, which immediately follows the H1 edge tone of MiP1 and immediately precedes the H2 edge tone of MiP2. The declination rate of L2 is -26~-45 Hz/sec, which is far greater than the rate expected by the global declination factor only. I will interpret this finding as evidence for a more local declination slope on which each tone is superimposed and which is reset each time the target of that tone is achieved. The details of this "tone-bound" declination are presented below.

Residual values of L2 relative to H1 were made into the dependent variable and the duration between H1 and L2 was made into the predictor. Then a regression analysis was carried out. For all four speakers, negative regression slopes were again obtained as shown in Figure 3, and the coefficients of regression slopes are all significant ([AS: coefficient = -0.045, Std Error = 0.003, t = -15, p < 0.001], [MR: coefficient = 0.033, Std Error = 0.005, t = -7.3, p < 0.001], [SK: coefficient = 0.030, Std Error = 0.005, t = -5.6, p < 0.001], [NK: coefficient = 0.026, Std Error = 0.003, t = -9.16, t < 0.001]). Those coefficient values of the declining regression slopes are translated into the declination rates in terms of a second: 45 Hz/sec for AS; 33 Hz/sec for MR; 30 Hz/sec for SK; 26 Hz/sec for NK, and they are far greater than the declination rates obtained for H2.

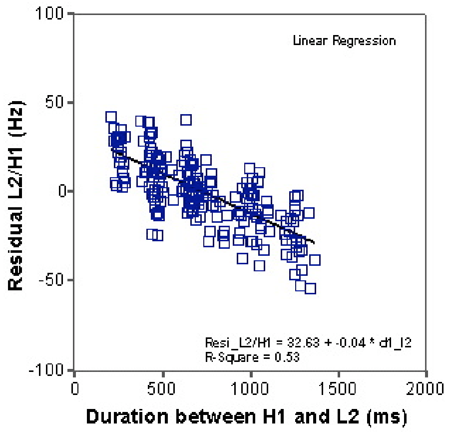


Figure 3: Residual values of L2 (the vertical axis) and duration between H1 and L2 (the horizontal axis).

4. DISCUSSION

The steep declining regression slopes of L2 indicate that the time-dependent downtrend of L2 is qualitatively different from the time-dependent declination of the following H2 edge tone. H2 only underwent a time-dependent declination of -8~-18 Hz/sec, which is consistent with the subglottal air pressure decrease account of [2]. However, the time-dependent downtrend of -26~-45 Hz/sec obtained for L2 is too large to be accounted for in that way. Also, it needs an account for why the H2 edge tone that follows L2 does not undergo such steep declination.

I propose that a pitch range undergo both global declination and a more local downtrend. The global downtrend unfolds over an Utterance or an Intonational Phrase while the latter, the local downtrend, lasts only a small interval and is constantly reset. This local downtrend is reset immediately after the F0 target of some tone is achieved. After the reset, the pitch range undergoes a new local downtrend until the target of the next tone is achieved. I am calling this local downtrend “tone-bound” declination. More specific proposals about this tone-bound declination in this model are the following, and are depicted in Figure 4.

- Both the top line and the lower line of a pitch range undergo a tone-bound declination.
- This tone-bound declination of the top line and the lower line is reset to the original point whenever the F0 target of some tone is achieved.
- The original points to which the local downtrend slope of the top line and the lower line are reset undergo a global declination.

ISBN 1-876346-48-5 © 2003 UAB

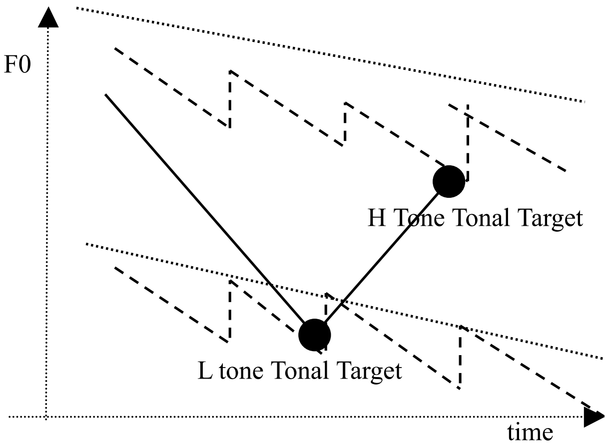


Figure 4: The Tone-Bound Declination Model. Dotted lines represent points to which the local downtrend slope of the top line and that of the lower line are reset, and they undergo a global declination. Dashed lines represent the local (tone-bound) declination of the top line and the lower line of a pitch range. Filled dots represent tonal targets within a given pitch range, and solid lines represent F0 trajectories that connect successive two tonal target to each other.

Now imagine that there are three tones and two Minor Phrases: a H1 edge tone associated with the second syllable of MiP1, a L2 edge tone associated with the initial syllable of MiP2 and a H2 edge tone associated with the second syllable of MiP2. Also imagine that there are three different linguistic forms carrying those tones and Minor Phrases. One of those forms has a relatively shorter duration, say 300 ms, between H1 and L2; the second form has an intermediate duration, say 600 ms, between those two tones; and the other form has the longest duration, say 1000 ms, between them. I call the first the "shortest form", the second the "intermediate form" and the last the "longest form". For the shortest form, the local downtrend slope that starts immediately after the realization of the H1 edge tone lasts only 300 ms until the L2 edge tone appears. As a result, the L2 edge tone is realized relatively higher. On the other hand, the local downtrend slope of the intermediate form and the longest form lasts for 600 ms and 1000 ms respectively. As a result, the L2 edge tones of those two forms are realized relatively lower. Between those two forms, their L2 tone values are different: the L2 tone of the longest form is lower than that of the intermediate form. When it comes to the H2 edge tone that immediately follows the L2 edge tone, it is not affected by the local downtrend. This is because the local downtrend is reset to the "original" point once the L2 edge tone is realized. That original point to which the local downtrend slopes are reset undergoes a global declination. As a result, H2 edge tone only undergoes a global declination.

According to this model, tones which undergo the more substantial declination (i.e. the tone-bound declination) are not limited to L edge tones. As long as duration between T1 and the immediately following T2 are variable whatever those tones are, T2 is expected to undergo the tone-bound declination as the duration between T1 and T2 increases. This prediction was tested in [8], and I found that even the H* pitch accent of an accented Minor Phrase undergoes a substantial declination of –26 ~ -98 Hz/sec as the duration between H* and the immediately preceding H edge tone increased. This finding further supports the presence of the tone-bound declination. For future studies, it is worthwhile testing whether the tone-bound declination is a language-specific or a universal phenomenon.

ACKNOWLEDGEMENTS

This paper is supported in part by the National Science Foundation grant BCS0004038 to Elisabeth Selkirk, University of Massachusetts at Amherst. Also, this paper is a shorter version of Chapter 2 of my PhD dissertation. I am grateful to my dissertation committee members: Elisabeth Selkirk, John Kingston, Charles Clifton and Joseph Pater for their precious comments.

REFERENCES

[1] R. Collier and J. t'Hart, "A Grammar of Pitch Movements in Dutch Intonation," *IPO Annual Reports*, vol. 6, pp. 17-21, 1975.

[2] C. Gelfer, K. S. Harris, R. Collier and T. Baer, "Is Declination Activity Controlled?" *Vocal Fold Physiology: Biomechanics, Acoustics and Phonatory Control*, I. R. Titze and R. C. Scherer (eds.) Denver, The Denver Center for the Performing Acts, 1983.

[3] C. Gussenhoven and T. Rietveld, "Fundamental Frequency Declination in Dutch: Testing Three Hypotheses," *Journal of Phonetics,* vol. 16(3), pp. 355-369, 1988.

[4] J. Pierrehumbert, *The Phonology and Phonetics of English Intonation.* Ph.D dissertation, MIT, Cambridge, 1980.

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

[6] W. Poser, *The Phonetics and Phonology of Tone an dIntonation in Japanese*, Ph.D dissertation, MIT, Cambridge, 1984.

[7] E. Selkirk, T. Shinya and M. Sugahara, "Degree of Initial Lowering in Japanese as a Reflex of Prosodic Structure organization," *Proceedings of ICPhS 2003* (this volume), 2003.

[8] M. Sugahara, *Downtrends and Post-FOCUS Intonation in Tokyo Japanese*, Ph.D dissertation, University of Massachusetts, Amherst, 2003.

[9] N. Thorsen (Grønnum), "Intonation Contours and Stress Group Patterns in Declarative Sentences of Varying Length in Asc Danish," *ARIPUC* vol. 14, 1980.