A study of F0 declination in Japanese: Towards a discourse model of prosodic structure

Mieko Banno

University of California, Santa Barbara

ABSTRACT

This study investigates F0 declination as a global-level prosodic phenomenon, establishing a new discourse-based model of prosodic structure in Japanese. The model includes two levels of declination in a hierarchical order: utterance units and prosodic paragraphs, a higher level of declination consisting of embedded declinations. Comparing and contrasting three types of discourse -- read speech, conversation, and narrative -- this study provides an accurate description of declination, based on an all-points linear regression line procedure. The results reveal a significant occurrence of declination not only in read speech but also in spontaneous speech. In addition, all three types of discourse exhibit both levels of declination, i.e. utterance units and prosodic paragraphs, providing evidence for the discourse model proposed in this study.

1. INTRODUCTION

This study is intended to make theoretical and descriptive contributions to the study of declination by establishing a model of prosodic structure based on discourse-level data in Japanese.

In previous research, models have shown declination to be an important phenomenon in speech cross-linguistically, e.g., in Dutch (Cohen and 't Hart [1]), in English (Maeda [2], Cooper and Sorensen [3]), in Danish (Thorsen [4]), in Swedish (Bruce [5]), in French (Vaissière [6]), in Japanese (Fujisaki and Sudo [7], Pierrehumbert and Beckman [8]), and in many other languages. However, most of these models have not made much progress in characterizing declination in discourse data.

The primary theoretical objectives of this study are to provide a new account of declination phenomena in Japanese, and to establish a model for prosodic structure in discourse. Unlike most prior work, which has based models of prosody on local-level units produced out of context, this study attempts to account for declination phenomena based on recordings of utterances produced in natural discourse contexts. The discourse data in this study involve two different levels of declination, utterance units as the basic domain of declination, and prosodic paragraphs as a higher level of declination domain that consists of a lower level of declination domain. To account for the phenomenon of declination on different levels, establishing a discourse model is essential. The discourse model developed here is shown to account for declination at higher-level units of speech.

The primary descriptive objectives of this research are to provide empirical evidence for F0 declination in Japanese discourse, and to describe adequately the intonational structure of declination in different discourse types. The quantitative characteristics of declination may differ depending on different discourse types, such as read speech versus spontaneous speech, or conversational versus monologic speech. In other words, there may be structural and contextual differences in the way declination is used in different discourse types.

One of the emphases of this study is that the investigation of declination requires the use of different types of discourse data. This is because the existing models of declination have been established based on read speech, and there is a lack of consensus about whether declination actually occurs in spontaneous speech. Some descriptive studies on declination have reported lesser occurrence of declination in spontaneous speech, compared with read speech (Umeda [9], Lieberman et al. [10]). However, the conclusions have been drawn from only limited data and with limited methodology. Therefore, this study investigates different discourse types to determine how the exact nature and characteristics of declination in spontaneous speech differ from those of read speech.

In addition, there is a critical need for an adequate methodology for characterizing declination. A major early approach in the development of models of declination is a topline and/or baseline procedure, in which regression lines are visually fitted to F0 peaks and valleys in an intonation contour (Cohen and 't Hart [1], Maeda [2], Cooper and Sorensen [3], Thorsen [4], and many others). Unfortunately, this early approach lacks consensus from observer to observer and has theoretical contradictions. As many researchers claim, the study of declination is required to use quantitative instrumental measurements using an objective and replicable criterion like an all-points linear regression line procedure, which is based on a calculation of the exact F0 value of all points on an intonation contour (Lieberman et al. [10], Liberman and Pierrehumbert [11], Ladd [12]). For instance, Lieberman et al. [10] explicitly point out that an all-points linear regression line procedure allows the identification of regression lines in both spontaneous and read speech, while a quantitative understanding of declination cannot be obtained by a topline and/or baseline procedure. In order to provide an accurate description of declination phenomena in discourse, this study thus uses an all-points linear regression line procedure.

2. DATA AND METHODOLOGY

The data for this study consist of audio-recorded Japanese discourse in three different types:

Data I: **read speech**, reading text aloud, based on a Japanese newspaper article taken from the Nikkei Daily News

Data II: **high-interactional discourse**, taken from one-to-one or multi-party conversations

Data III: **low-interactional discourse**, taken from narratives of personal experience in a one-to-one interview

These three discourse types have been selected so that the phenomenon of declination can be accounted for in two different dimensions: read speech versus spontaneous speech, and within spontaneous speech, high interactional versus low interactional.

The speakers are 24 native Japanese speakers (12 male and 12 female) who use Tokyo intonation. Each of these speakers produced all thee different discourse types so that the data allow investigation of speech production by the same speaker in contrasting discourse types. Portions of approximately 10 successive global prosodic units (i.e. either utterance units or prosodic paragraphs) for each speaker from each discourse type were selected for analysis. The corpus for F0 analysis consists of a total of 770 units, approximately 170 minutes of speech.

The all-points linear regression lines and the regression results of all 770 units were extracted from the measurements of F0 contours. Based on the regression results, different domains of declination (utterance units, prosodic paragraphs, and embedded declinations within the domain of prosodic paragraphs) were analyzed. Each slope coefficient (Hz/s) was checked for the level of significance; if the regression lines reached a significantly linear correlation at $p < 0.05$, the declination contour was assessed as having achieved linearity.

3. PROPOSED MODEL

The proposed model of prosodic structure in discourse is designed to establish global levels of prosodic units based on a quantitative analysis of F0 contour. One theoretically important aspect of this model is to provide new accounts for prosodic structure in discourse-level data, which will contribute to establish a higher level of prosodic constituency. The discourse model developed in this research describes prosodic structure in terms of the following hierarchy: *local prosodic units*, which exist at a local level within the domain of declination and have their own pitch contours; *utterance units*, which exist at a global level and are the domain of declination; and *prosodic paragraphs*, which constitute a higher level of declination involving multiple embedded declinations. In short, this discourse-based model involves one local level and two different global levels, that of utterance units and that of prosodic paragraphs. At the global level of prosodic structure, three domains of declination are postulated: utterance units, prosodic paragraphs, and embedded declinations. These levels are shown below in (1).

(1) Proposed model of prosodic structure

Global level	Prosodic paragraphs	[/$_{emb}$	/$_{emb}$	/$_{emb}$	]
	Utterance units	[	][	][	]
Local level	Local prosodic units	[( )( )( )(		)(	)]

The texts used in this research include long stretches of speech in which syntactic constituents of varying lengths are involved. Because of this characteristic of texts, the phenomenon of 'declination within declination' has been observed in the data. This phenomenon exists when the speaker uses a noticeably long utterance, which is longer than a complex-clause pair (i.e., a dependent clause plus a main clause, or two conjoined clauses). Such long utterances show a more global level of declination involving multiple embedded declination lines within its domain.

This domain may represent a higher level of prosodic unit, similar to the prosodic paragraph, as noted by Lehiste [13], Silverman [14], Pierrehumbert and Hirschberg [15]. These studies report that prosodic cues indicate paragraph structure and sequential F0 downdrift across sentences in the paragraph. In this study, F0 declination by instrumental measurements shows explicitly a higher level of declination of prosodic paragraphs. The domain of prosodic paragraphs, which involve multiple embedded declination lines within their domain, contrasts with that of utterance units, which are associated with one linear declining slope. The size of prosodic paragraphs and location of their boundaries may vary, and may not correspond to the text-analytic category of paragraphs such as syntactic structure or discourse structure (e.g., topic structure).

4. F0 ANALYSIS OF DECLINATION

The results of F0 analysis of declination show how the phenomenon of declination occurs in spontaneous speech and how its characteristics differ from those of read speech. First, the results of quantitative measurements using an all-points linear regression line procedure are given in terms of the linearity assessments of declination in each discourse types. Next, the slope coefficients of declination are examined. Furthermore, comparing slope coefficients, the quantitative characteristics of declination in two different levels, utterance units and prosodic paragraphs, are presented. Finally, the relationship between F0 declination and duration are analyzed.

4.1. Regression Results of Global Prosodic Units

There are various patterns of the declination in differently sized global prosodic units throughout the three discourse types; however, one of the striking findings of this study is that the slope coefficients nevertheless reached a significantly linear correlation at $p < 0.05$. The linearity assessments of the slope coefficients in the data are given in Table 1.

Table 1: Linearity of assessments of regression lines.

| | P< 0.05 | | P< 0.1 | | P> 0.1 | | Total |
	N	%	N	%	N	%	N
Data I	243	100.0	0	0.0	0	0.0	243
Data II	258	98.8	2	0.8	1	0.4	261
Data III	247	99.2	0	0.0	2	0.8	249
Total	748	99.3	2	0.3	3	0.4	753

Of 770 global prosodic units, 17 upward contours (inclination) have been observed: 9 units in Data II and 8 units in Data III. These units are excluded from the linearity assessments. Table 1 shows that of the total of 753 units, 748 units (99.3.%) achieved a significantly linear relationship in declining contour at $p < 0.05$ in all three discourse types. Only 2 units (0.3 %) reached $p < 0.1$, and a linear relationship could not be established in only 3 units (0.4 %) in the entire corpus. As shown above, declination occurs consistently in the data throughout all three discourse types. The results of the quantitative measurements using the all-points linear regression line procedure show F0 declination not only in read speech but also in spontaneous speech.

4.2 Declination and Slope Coefficients

The results of the quantitative measurements based on the all-points linear regression line procedure also show the variability of slope coefficients in each discourse type. Figure 1 shows a significant occurrence of declination in all discourse types; however, the frequency of the cluster that shows the slope coefficients varies in each type.
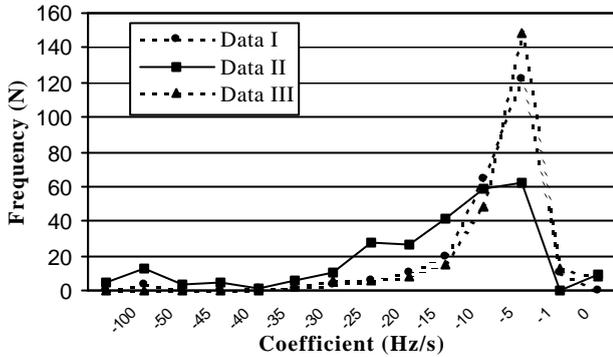

Figure 1: Frequency of declination in three discourse types.

Figure 1 shows that throughout the three discourse types, the declining slopes tend to cluster between -1 Hz/s and -5 Hz/s: 50.2 % in Data I, 23.0 % in Data II, and 58.0% in Data III. The distribution for the slope rate in Data II, however, differs from that of the other two discourse types. There are declining slopes of varying higher coefficients in Data II than the other types. In contrast, Data I and III tend to cluster in similar ways.

4.3 Two Levels of Declination: Utterance Units vs. Prosodic Paragraphs

The discourse-level data in this study show a higher level of structure, that of prosodic paragraphs. While an utterance unit involves a single linear declining slope in its domain, a prosodic paragraph consists of multiple declinations embedded in the domain. An example of prosodic paragraphs is given in Figure 2. In this higher-level domain, three embedded declination lines occur in succession.
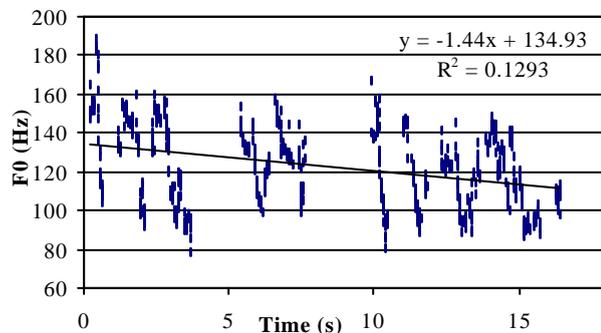

Figure 2: An example of multiple embedded declination lines in the higher-level of declination, prosodic paragraph.

In this utterance, the multiple clauses it contains are divided into three embedded declinations in the higher-level of declination. Figure 2 shows that the declination of this prosodic paragraph has a slope of -1.44 Hz/s, which achieved a significantly linear correlation, at p< 0.001, over the course of the utterance. As seen in the figure, there are unfilled pauses between each embedded declination. Despite these pauses, the linear relationship of the declining slope was not interrupted.

The frequency of occurrence of utterance units versus prosodic paragraphs across the three different discourse types is given in Table 2. The results show that 166 out of 770 units (21.6 %) are prosodic paragraphs, and that a higher percentage of prosodic paragraphs is observed in Data III (28.4 %), and in Data I (25.5%), than in Data II (11.5 %). This suggests that prosodic paragraphs are more likely to occur when speech is monologic and texts contain longer units.

Table 2: Frequency distribution of declination in utterance units vs. prosodic paragraphs.

	Data I		Data II		Data III		Total	
	N	%	N	%	N	%	N	%
Utterance	181	74.0	229	85.1	174	67.7	584	75.8
Paragraph	62	25.5	31	11.5	73	28.4	166	21.6
Upward	0	0.0	9	3.3	8	3.1	17	2.2
P> 0.1	0	0.0	1	0.4	2	0.8	3	0.4
Total	243	31.6	270	35.1	257	33.4	770	100.0

There are significant differences in the quantitative characteristics of declination between utterance units and prosodic paragraphs. Table 3 shows that the overall mean slope coefficient is greater in utterance units than in prosodic paragraphs, -12.92 Hz/s and -3.21 Hz/s, respectively. Also, the overall mean duration is shorter in utterance units than in prosodic paragraphs, 6.3 seconds and 11.8 seconds, respectively. Statistical analysis (two-way ANOVA) reveals that these differences in slope and duration are significant (p<0.01).

Table 3: Mean slope coefficients (Hz/s) and duration (s) of declination in utterance units vs. prosodic paragraphs.

	Data I		Data II		Data III		Total Mean	
	Hz/s	s	Hz/s	s	Hz/s	s	Hz/s	s
Utterance	-8.78	5.7	-21.21	5.5	-7.05	8.0	-12.92	6.3
Paragraph	-2.64	9.6	-4.86	11.5	-3.00	13.9	-3.21	11.8

Table 3 also shows discourse-type differences: the greatest mean slope coefficient was found in Data II in both utterance units and prosodic paragraphs, -21.21 Hz/s and -4.86 Hz/s, respectively. The statistical analysis reveals a significant difference between utterance units in Data II and the other two types (p<0.01), whereas the differences across types for prosodic paragraphs are not significant. This indicates that a greater range of slopes is used in utterance units in high-interactional discourse than in the other types. On the other hand, a comparison of duration shows that in Data III, both utterance units and prosodic paragraphs have longer durations, which is statistically significant (p<0.01). These facts indicate that the shorter units tend to show steeper slopes, as in Data II, and that the longer units tend to show gentler slopes, as in Data III.

4.4 F0 Declination and Duration

The issue of the relationship between declination and the duration of utterance has been discussed in previous studies. Some researchers found that the pitch range of declination differs depending on the duration of utterance, in which the declining slope was much steeper in short utterances than in long utterances ('t Hart [16], Cooper and Sorensen [3]). Figure 4 shows the distribution of all global prosodic units of declination, a total of 753 units, presenting the correlation between declining slope and duration in each discourse type. It yields the correlation between slope coefficients and duration in the plotted pattern: the shorter the duration, the steeper the slope, and vice versa. It also shows that the steeper slope coefficients cluster around duration less than 4 seconds.
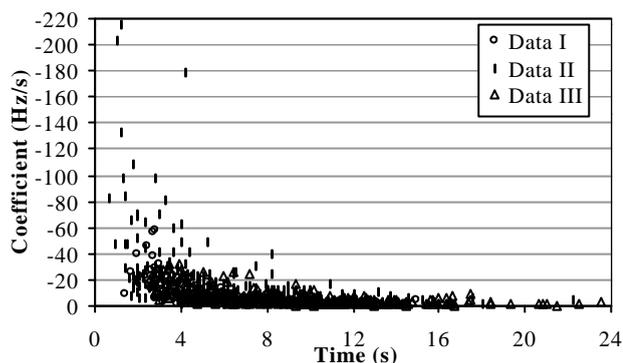

Figure 4: Frequency distribution of slope coefficients plotted against duration.

Thus, declination domains exhibit a correlation between slope coefficient and duration. These results support the claim that the declination slope coefficient increases when the duration of utterance decreases.

5. SUMMARY

In this study, a hierarchical model for prosodic structure was established to analyze declination in Japanese based on the F0 analysis of three different discourse types using an all-points linear regression line procedure. The model yielded convincing results identifying and characterizing declination at global levels of prosodic structure. The regression results demonstrated a significant occurrence of declination in all three discourse types, providing convincing evidence that declination phenomena in Japanese are substantial prosodic events not only in read speech but also in spontaneous speech. In addition, the F0 analysis provided evidence for the discourse model of prosodic structure having hierarchically organized global levels of declination, i.e. utterance units and prosodic paragraphs. These levels differ in that declination lines on the latter level have relatively greater length and gentler slope. Their different properties especially those of prosodic paragraphs as a higher level of declination helped account for the variability of declination in discourse. A frequency analysis of regression lines revealed that the shorter the duration of declination, the steeper the slope. The evidence made available by this method thus provides a solid foundation for a discourse-based model of prosodic structure.

5. REFERENCES

1. Cohen, A., and 't Hart J. "On the anatomy of intonation." *Lingua* 19, 177-192, 1967.

2. Maeda, S. "A characterization of American English intonation." Ph.D.dissertation, MIT. 1976.

3. Cooper, W. , and Sorensen, J. *Syntax and speech.* Cambridge, MA: Harvard University Press. 1981.

4. Thorsen, N. "On the variability in F0 patterning and the function of F0 timing in languages where pitch cuesstress." *Phonetica,* 39, 302-316, 1982.

5. Bruce, G. "Textual aspects of prosody in Swedish." *Phonetica*, 39, 288-301, 1982.

6. Vaissière, J. "Contribution a la synthèse par regles du Francais." Ph.D.dissertation, Université de Langues et Lettres de Grenoble. 1971.

7. Fujisaki, H., and Sudo, H. "Synthesis by rule of prosodic features of connected Japanese." *Proceedings of 7th International Congress on Acoustics*, 3, 133-136, 1971.

8. Pierrehumbert, J., and Beckman M. *Japanese tone structure.* Cambridge, MA: MIT Press. 1988.

9. Umeda, N. " "F0 declination" is situation dependent." *Journal of Phonetics*, 10, 279-290, 1982.

10. Lieberman, P., Katz, W., Jongman, A., Zimmerman, R., and Miller, M. "Measures of the sentence intonation of read and spontaneous speech in American English." *Journal of the Acoustical Society of America,* 77 (2), 649-657, 1984.

11. Liberman, M., and Pierrehumbert, J. "Intonational invariance under changes in pitch range and length." In Aronoff, M., and Oerhle, R., *Language sound structure.* Cambridge, MA: MIT Press. 1984.

12. Ladd, D. Robert. "On the theoretical status of the baseline in modeling intonation." *Language and speech*, 36(4), 435-451, 1993.

13. Lehiste, I. "The phonetic structure of paragraphs." In Cohen, A., and Nooteboom, S., *Structure and processing in speech perception*, 195-203. Springer-Verlag. 1975.

14. Silverman, K. "The structure and processing of fundamental frequency contours." Ph.D. dissertation, University of Cambridge. 1987.

15. Pierrehumbert, J., and Hirschberg, J. "The meaning of intonation contours in the interpretation of discourse." In Pollack, M., and Cohen, P., *Plans and intentions in communication and discourse*, 271-311. Cambridge, MA: MIT Press. 1990.

16. 't Hart, J. "Explorations in automatic stylization of F0 curves." *IPO Annual Progress Report*, 14, 61-65, 1979.