CONTRAST MAXIMIZATION IN F_0 DECLINATION: JAPANESE SHIKI-ACCENT DIALECTS

Kenji Yoshida

Indiana University, Bloomington, USA
keyoshid@umail.iu.edu

ABSTRACT

Japanese dialects spoken in the Kinki and Shikoku regions have a lexical contrast of pitch register types called *shiki*. The shallow f_0 downtrend in the high-pitched shiki type is thought of as a manifestation of backdrop f_0 declination that is blind to tonal structure. Acoustic analysis of three Japanese dialects with 2- or 3-way contrast of shiki types, however, reveals that the degree of f_0 declination takes different values depending on the number and structure of contrastive shiki types. This can be regarded as dialect-specific ways of maximizing difference in f_0 trends in word-sized units for enhancing contrasts between phonological categories, suggesting that f_0 declination patterns in shiki dialects are encoded in the contrastive structure of their phonology.

Keywords: F_0 declination, Shiki-accent systems, contrast maximization

1. INTRODUCTION

Major changes have been brought about in the study of intonation after the Autosegmental-Metrical model was proposed [9]. One of them is the introduction of downstep and final lowering in modeling f_0 downtrend in an utterance [7], reducing the role of declination, i.e., "the gradual backdrop reduction in the pitch reference … that is blind to the phonological sequence of accents or tones" [10]. Studies suggest, however, that declination may have both speaker- and tone-specific variation [3]. Unfortunately, due to the existence of multiple factors contributing to f_0 downtrend, the declination component often has to be "distilled" from the results of fitting a f_0 production model including other parameters which induce lowering of f_0 [9, 10].[1]

Japanese dialects with lexical contrasts in pitch register, termed *shiki* [13] provide an interesting case where it is possible to isolate the declination component. Figure 1 shows one such case from the Kishiwada dialect near Osaka [5]. The unaccented words in H-beginning (Level) shiki have a high-pitched stretch. The f_0 downtrend observed therein has been interpreted as being due to the declination observable in the interpolation between word-initial and final H-tones [10]. The present study examines whether the degree of declination may differ across dialects with different shiki systems and discusses the implication of such systems for intonational phonology.

Figure 1: F_0 contours of Level- (●), Rising- (+) shiki and Accented (□) word types in Kishiwada dialect, an example of system-A (mean of 5 repetitions).

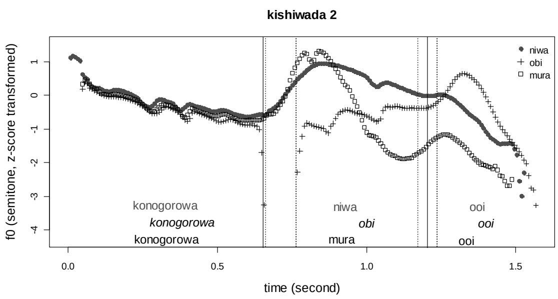


Figure 2: Level (●), Falling (+), Rising (□) and Accented (○) types in Ibukijima dialect, system-B.

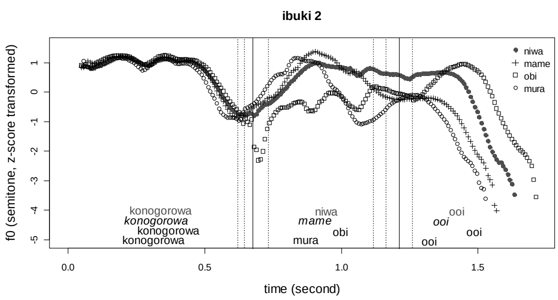
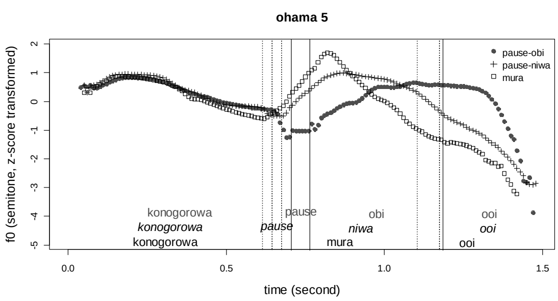


Figure 3: Non-falling (●), Falling (+) and Accented (□) types in Ohama dialect, system-C.



2. METHODS

2.1. Data from three shiki-accent dialects

The speech production data come from a larger corpus of six Japanese dialects with shiki-accent systems spoken in the Kinki and Shikoku regions [14]. The shiki-accent systems of the 10 speakers from the three dialects in (1) are examined.

(1) Shiki system investigated (number of speakers)

A **Level** vs. **Rising**: Kishiwada (n=3)
B **Level** vs. **Rising** vs. **Falling**: Ibukijima (n=4)
C **Falling** vs. **Non-falling** : Ohama (n=3)

The labels of the shiki types describe the global direction of pitch movement within the word [11, 13]. In system-A, f_0 constantly declines in Level-shiki type, whereas it rises in Rising (Figure 1) [5, 10]. System-B is the only 3-way contrast system attested [13]. F_0 declines very little in Level, whereas Falling has a steeper downtrend (Figure 2). System-C has two high-pitched shiki types that resemble Level and Falling in System-B, respectively (Figure 3) [11].

The speakers produced the four 2-mora words in Table 1 embedded in the frame sentence (2) five to six times. While shiki can coexist with pitch accent, we need to isolate the f_0 trends of shiki by making observations of unaccented words. An accented word *mura* was included for comparison. Excluding the utterances with disfluencies, a total of 214 tokens were obtained.

Table 1: Test words and the shiki-accent categories in the three shiki-accent systems. (L=Level; R=Rising; F=Falling; N=Non-falling; 0=unaccented; 1=1st mora accented; 2=2nd mora unaccented).

System	A	B	C
niwa 'garden'	L0	L0	F0
mame 'beans'	R2	F0	F0
obi 'belt'	R0	R0	N0
mura 'village'	L1	L1	F1

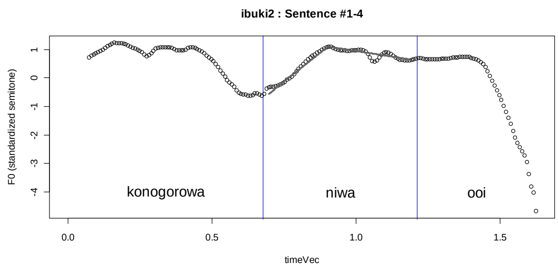
(2) *Konogorowa* (target) -*ga* *ooi*.
Nowadays -NOM. plenty
'Nowadays we have lots of (target)'

2.2. Acoustic analysis

Following Nolan [8], f_0 values were converted to semitones (Re. 100Hz) and z-score transformed for each speaker. Also, time normalization to the mean word duration across the repeated renditions was performed. These two procedures allow us to compare the rate of f_0 change across speakers and dialects.

Our interest is in comparing the degree of f_0 downtrend (or uptrend) across the shiki-accent categories and the three dialects. To obtain the slope of f_0 movement after the word-initial f_0 rise, two linear regression lines were fitted to the observed f_0 samples within the target word [10] as exemplified in Figure 4. The best two regression lines were determined by minimizing total squared deviation from the fitted function. The slope of the second regression line is taken as representing the f_0 trends characteristic of the shiki types.

Figure 4: An example of fitting two linear regression lines to the f_0 data within the target word. The two best-fit regression lines are overlaid in grey. (L0, Ibukijima).

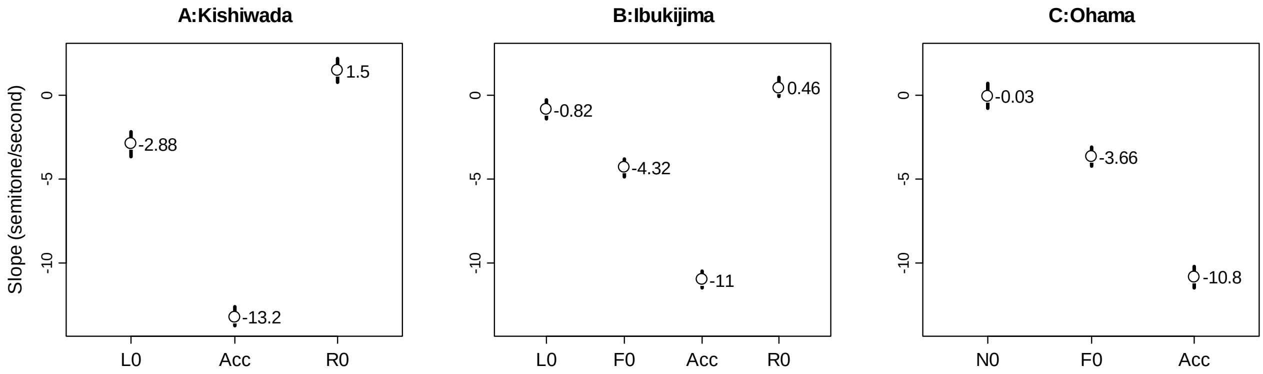


3. RESULTS

The results show that in each dialect, the shiki types are reliably distinguished by f_0 slopes alone. This was confirmed by Bayesian ANOVA [6], conducted for each of the three dialects separately. The slope of the second regression line was the predicted variable, whereas the shiki-accent category was the predictor variable, as well as the individual speaker to explain speaker-specific differences.

Starting from an uncontroversial vague priors (*Normal*(μ=0, σ=10000)), the Markov Chain Monte Carlo sample of posterior (Burn-in step=2000, Thin-step=20) yielded 6000 credible combinations of the overall baseline of f_0 slopes and the deflections from it due to the two predictors. The analysis converged well. Linear combination of the posterior samples provides the best estimates of the slope for each shiki-accent category for each dialect, given the data. The estimates are summarized by the mean and upper and lower bounds of the 95% HDI (highest density interval) as in Figure 5. All the differences between shiki-accent categories are credibly non-zero.[2]

Figure 5: Estimated slopes (in z-score transformed f_0) obtained with Bayesian ANOVA. The open circles indicate the means and the lines extending from them show the 95% HDIs of the posterior samples. L0=Level-shiki, unaccented; R0=Rising-shiki, unaccented; F0=Falling-shiki, unaccented; N0=Non-falling-shiki, unaccented; Acc=Accented.



Next we examine the differences and similarities between the three shiki-accent systems. First, for all three dialects, the accentual f_0 fall is much steeper than all the shiki types, which is reasonable given the defining characteristic of pitch accent of these dialects, i.e., steep, local pitch fall, as in the majority of Japanese dialects [13].

Second, in System-A, L0 (Level-shiki, unaccented) has a mean f_0 downtrend of –2.88 st./sec. In System-B, by contrast, L0 has a substantially shallower downtrend (–0.82 st./sec.), whereas F0 (Falling-shiki, unaccented) has a downtrend steeper than L0 in System-A (–4.32 st./sec.), as the label of the shiki type (Falling) suggests. In System-C, F0 has a downtrend that is comparable to, though a little shallower than that of System-B (–3.66 st./sec.).

Third, R0 (Rising-shiki, unaccented) in System-A is characterized with an upward trend of f_0 (1.5 st./sec.). Similarly in System-B, the slope of f_0 movement tends to be rising but less steep than that of System-A (0.46 st./sec.). This is because the sharp f_0 rise that characterizes Rising-shiki is often delayed and realized in the next word (Figure 2, also reported in Kori [5]). Yet the slope alone reliably differentiates R0 from the other categories for both System-A and B, although the difference in pitch range (low for R0) is already sufficient to differentiate it from the others.

Finally, N0 (Non-falling-shiki, unaccented) in System-C is characterized by the f_0 trend that is around 0 st./sec, as the label (Non-falling) suggests. While this is close to that of R0 in System-B, N0 typically has an earlier rise and is realized in high pitch range, thus is quite different from R0 in System-A and B in its f_0 contour [11].

4. DISCUSSION AND CONCLUSION

The results of the acoustic analysis reveal that the shiki types of the three dialects investigated show different patterns of contrasts in their f_0 slopes. This offers important insights into the way how the contrast between phonological categories is implemented.

Of particular interest is the difference in the slope of L0 between the System-A and B. As Table 1 suggests, the shiki-accent categories are in regular correspondence among Japanese dialects, which traces back to the word prosodic system of the Early Middle Kyoto Japanese [4]. Specifically, L0 is found in largely the same group of words in both System-A and B, as is true for the majority of other shiki-accent dialects [12]. Thus, it is likely that phonetic realization of L0 has developed differently in System-A and B from the same phonetic form. The most plausible explanation of the different L0 slope values would be the number of contrasting shiki categories, two in System-A and three in System-B. That is, the very shallow f_0 downtrend of L0 in System-B must have resulted due to the systematic pressure of having one more contrastive shiki category, F0.

System-C provides another such instance. N0 corresponds to R0 in most of other shiki-accent dialects [11, 12], suggesting that it is likely that N0 has developed from R0. As discussed earlier, however, N0 is typically realized in high pitch range, as is F0 in this dialect. In such a situation, it is likely that the direction of f_0 carries more weight in differentiating the two shiki types in System-C. The downtrend of F0 in this system that is steeper than L0 of System-A may have resulted due to the pressure of having N0.

As reviewed in the introduction, L0 in System-A has been interpreted as reflection of backdrop f_0 declination. Phonologically, it is a high-pitched stretch (an interpolation between word-initial and final H-tones) [10]. However, the present results suggest that the phonetic realization of L0 can vary depending on the number of contrasting shiki categories. System-C provides another case of variation in phonetic realization that depends on the structure of the contrast system. This situation may be explained by making an appeal to the theory of adaptive dispersion, which argues that "preferred phoneme and feature inventories reflect … the listener-oriented selection criterion of auditory distinctiveness" [2]. In the present case, the degrees of f_0 downtrend reveal the phonetic variation that enhances auditory distinctiveness between the shiki categories, which is conditioned by the number and structure of contrasting shiki categories in each system.

While the dispersion theory has been mostly tested for segmental systems, the present study is not the sole case demonstrating applicability of this theory in intonational phonology. Yoruba, which has a 3-way tonal contrast (H, M and L), reveals a case of tone-specific pattern of declination: the sequence of H or M tones shows very small declination, whereas L tone sequence shows substantially larger declination [1]. This appears readily interpretable as a consequence of the speakers' effort to maintain the contrast between the three tones.[3]

In conclusion, the present study has shown that there can be dialect-specific ways of maximizing difference in f_0 down- or up-trends in word-sized units for enhancing the contrasts between the shiki types. Since the shiki in Japanese dialects conveys lexical contrasts, categoricality of the differences in slopes is unquestionable. Furthermore, the acoustic analysis of 2- and 3-item sequences with the same design as the present study shows that the shiki-specific downtrends are co-existent with, thus independent of the downstep effects [14]. While further investigation is necessary to see whether the differences in f_0 trends can and should be represented with locally-specified tonal targets, it seems likely that degrees of f_0 declination can be a property for lexical contrast thus may not simply be a backdrop effect that is blind to tonal structure.

5. ACKNOWLEDGEMENTS

I thank Yosuke Igarashi for suggestion about f_0 standardization and time normalization methods, and Aaron Albin and the two anonymous reviewers for helpful comments. This work is supported by NSF grant (#BCS-0921129).

6. REFERENCES

[1] Connell, B., Ladd, R. 1990. Aspects of pitch realization in Yoruba. *Phonology* 7, 1-29.

[2] Diehl, R., Lindblom, B. 2004. Explaining the structure of feature and phoneme inventories: The role of auditory distinctiveness. In Greenberg, S., et al. (eds.), *Speech Processing in the Auditory System*. New York: Springer, 101-162.

[3] Gussenhoven, C. 2004. *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.

[4] Kindaichi, H. 1937. Pitch accent system in the Heian dynasty reconstructed through the comparison of the present dialects. *Hoogen* 7(6), 1-43.

[5] Kori, S. 1987. The tonal behavior of Osaka Japanese: an interim report. *Ohio State University Working Papers in Linguistics* 36, 31-61.

[6] Kruschke, J. 2011. *Doing Bayesian Data Analysis: A Tutorial with R and BUGS*. New York: Academic Press.

[7] Liberman, M., Pierrehumbert, J. 1984. Intonation invariance under changes in pitch range and length. In Aronoff, M., Oehrle, R. (eds.), *Language Sound Structure*. Cambridge: MIT Press, 157-233.

[8] Nolan, F. 2003. Intonational equivalence: An experimental evaluation of pitch scales. *Proc. 15th ICPhS* Barcelona, 771-774.

[9] Pierrehumbert, J. 1980. *The Phonology and Phonetics of English Intonation*. Doctoral dissertation, MIT.

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

[11] Sato, E. 1986. On the pitch accent of Takase dialect. *Yamate Kokubun Ronkoo* 7, 1-26.

[12] Uwano, Z. 1985. Genealogical relationships and the geographical distribution of the accents in mainland Japan 1. *Trans. Japan Academy* 40(3), 215-250.

[13] Uwano, Z. 1989. Accent in Japanese. In Sugito, M. (ed.), *Kooza Nihongo-to Nihongo Kyooiku Vol. 2*. Tokyo: Meiji Shoin, 178-205.

[14] Yoshida, K. In preparation. *The Phonology and Phonetics of Shiki-Accent Systems in Mid-western Japanese Dialects*. Doctoral dissertation, Indiana University.

[1] Throughout this paper, "f_0" is used for the fundamental frequency, whereas the upper case "F0" is reserved for a shiki-accent category, "Falling-shiki, unaccented".

[2] For all the differences among the shiki-accent categories, mean differences range from 3.5 to 14.7; HDIs of differences do not include zero, ranging from 2.7 to 15.6; 100% of the samples are larger than zero.

[3] Connell and Ladd [1] suggested that the larger declination in L tone sequence may be due to the effect of final lowering. But this does not explain why the effect is not found in H and M tone sequences.