THE EXPONENTIAL NATURE OF F0 TARGET TONE INTERPOLATION

Shunichi Ishihara

Japan Centre (Asian Studies) and Phonetics Laboratory (Linguistics, Arts),
The Australian National University.

ABSTRACT-The intonational phenomena of Kagoshima Japanese (KJ) will be investigated in this paper. More precisely, this paper has two aims. Ishihara (1998, 2000) reports that KJ's accentual oppositions: Type A and Type B, show a different F0 realisation range at word level. First of all, I will show that this difference between Type A and Type B appearing at word level is also maintained at phrase level. Secondly, I will show that KJ's HLLLLL(L) sequences—which are represented by the interpolation from a high to a low tone in Autosegmental-Metrical (AM) model (i.e. Pierrehumbert and Beckman, 1988)—exhibit exponential F0 curves, by acoustical-phonetically describing the sequences in question.

INTRODUCTION

KJ exhibits a two way accentual contrast; (L)^{0}HL and (L)^{0}H (cf. Hirayama, 1960). In the former pattern, only the penultimate syllable of a word has a high pitch, and every other syllable has a low pitch. In the latter pattern, only the last syllable of a word has a high pitch, and every other syllable before it has a low pitch. Following Hirayama (1960), (L)^{0}HL type is referred to as Type A and (L)^{0}H type as Type B. Table 1 shows some examples of this accentual contrast.

Syllable	Type A		Type B	
2	hana [HL]	"nose"	hana [LH]	"flower"
3	sakura [LHL]	"cherry blossom"	usagi [LLH]	"rabbit"
4	kagaribi [LLHL]	"bonfire"	kakimono [LLLH]	"document"

Table 1: The accentuation of KJ.

It has been reported that the F0 realisations of Type A and Type B are significantly different from each other at word level (i.e. Ishihara, 1998; 2000). Namely, Type A shows overall higher F0 realisation than Type B. Firstly, in this paper, the F0 realisations of KJ's accentual oppositions are going to be compared to see whether this difference appearing at the word level is maintained at the phrase level. Secondly, KJ's HLLLLL(L) sequences—which are represented by the interpolation from a high to a low tone in Autosegmental-Metrical (AM) model (i.e. Pierrehumbert and Beckman, 1988)—having different accentual types are acoustic-phonetically described to see how the interpolation from a high tone to a low tone is realised phonetically. Some implications will be discussed on the basis of these results.

EXPERIMENT PROCEDURE AND NORMALISATION

Four native speakers of KJ (two females: TY and YN and two males: TT and NK aged between 25 and 35 years of age) participated in this study. One corpus containing 18 noun phrases differing in syllable number—with approximately 15 dummy phrases which were scattered at random throughout the corpus—was prepared. The pitch configurations of these target noun phrases are presented in Table 2. The syllable structure of these target noun phrases is C$^{(j)}$V. The informants were asked to read this corpus 10 times in the frame given in Table 2 (e.g. Naomi-wa yoka sakana to itta gayo "Naomi said good fish"). The recording was conducted in a quiet room in Sydney with very low ambient noise, and the reading material was recorded onto high-quality normal position tapes using professional equipment. The raw material was digitised with Computerised Speech Laboratory (CSL) (sampling rate = 10000 Hz).

All phrases have the same LHL$^{(n)}$H(L) pitch configuration (n = 1, 2, 3, 4, 5, 6), but have different accentual types (Type A or Type B). In this paper, the phrases having a Type A + Type A pitch configuration are referred to as AA phrase, and the other combinations also follow the same convention. Another convention which needs explanation is that, for example, 3a7a stands for the

$HL^{(n)}$ sequence which is extracted from an AA phrase whose first component consists of 3 syllables and second component of 7 syllables (= LHL.LLLLLHL).

	First component	Second Component	n
Type A + Type A (AA)	LHL	$L^{(n)}HL$	1, 2, 3, 4, 5
Type A + Type B (AB)	LHL	$L^{(n)}H$	2, 3, 4, 5
Type B + Type A (BA)	LH	$L^{(n)}HL$	1, 2, 3, 4, 5
Type B + Type B (BB)	LH	$L^{(n)}H$	2, 3, 4, 5

Frame					
	Naomi-wa	_____	to	itta	gayo.
	LLH L		L	HL	HL
	Name-TOP	_____	QTN.	say-past	SFP.
	Naomi said	_____.			

QTN = Quotation
SFP = Sentence final particle

Table 2: The pitch configuration of target noun phrases and carrier frame.

F0 was extracted from the $HL^{(n)}$ sequences using CSL's Automatic Pitch Extraction. F0 samples were taken at the onset, 50% and the offset of each syllable nucleus except for the initial high pitched syllable from which only maximum F0 value was sampled.

The logarithmic z-score normalisation—which Zhu (1999) reports the superiority of in F0 normalisation—was used in this study in order to exclude between-speaker differences and specify the invariant features (cf. Rose, 1987). The logarithmic z-score normalisation procedure is: $F0_{norm} = (F0_i - x) / SD$, where $F0_i$ is a sampling point, x is the average F0 from all sampling points, and SD is the standard deviation around the mean of those points, all of which are logarithmic terms. All statistical comparisons are conducted on the basis of logarithmic z-score normalised F0 values. Table 3 contains the normalisation parameters of the four informants.

Speaker	TY	YN	TT	NK
x	2.133	2.013	2.284	2.262
SD	0.053	0.043	0.073	0.043

Table 3: Normalisation parameters in log F0

COMPARISON BETWEEN TYPE A AND TYPE B

Those phrases having the same duration of $HL^{(n)}$ sequence were statistically compared at the maximum and minimum points in order to factor out the declination effect (Vaissière, 1983). These maximum and minimum points are considered to be associated with a target tone. Table 4 below contains all possible combinations of statistical comparisons. As with the HLLLLLL sequence, for example, 3a7a and 3a6b are comparable only at the minimum point.

$HL^{(n)}$	Combination	Location
HLLLLLL	3a7a vs 3a6b	Min
HLLLLL	3a6a vs 3a5b vs 2b7a vs 2b6b	Min, Max
HLLLL	3a5a vs 3a4b vs 2b6a vs 2b5b	Min, Max
HLLL	3a4a vs 3a3b vs 2b5a vs 2b4b	Min, Max
HLL	3a3a vs 2b4a vs 2b3b	Min, Max

Table 4: All possible combinations of statistical comparisons.

Table 5 below contains the average F0 minima and maxima of each phrase according to the group in comparison. If the difference observed between Type A and Type B at word level is carried over to phrase level, Type A should exhibit a higher value than Type B for both F0 minima and maxima. As with the F0 minima, it can be seen from Table 5 that those phrases having a Type A for the second component exhibit higher value than those having a Type B for the second component (i.e. 3a7a: -0.942 > 3a6b: -1.416). As with the average F0 maxima, those phrases having a Type A for the first component show higher value than those having a Type B for the first component (i.e. 3a6a: 2.214; 3a5b: 2.271 > 2b7a: 2.148; 2b6b: 2.122). Comparing the average F0 minima and maxima, the

difference appearing in Type A and Type B is larger in the F0 minima than the F0 maxima. Unpaired two-tail t-test, ANOVA with post-hoc Scheffe F-test were selectively conducted on the F0 maxima and minima depending on the number of group members. The results of these statistical comparisons—which were conducted according to Table 4—are listed in Table 6 for F0 minima and in Table 7 for F0 maxima.

HLLLLLL	3a7a	3a6b		
Min (SD)	-0.942 (0.277)	-1.416 (0.341)		
Max	-	-		
HLLLLL	3a6a	3a5b	2b7a	2b6b
Min	-0.890 (0.340)	-1.285 (0.332)	-0.796 (0.417)	-1.280 (0.430)
Max	2.214 (0.345)	2.271 (0.444)	2.148 (0.387)	2.122 (0.482)
HLLLL	3a5a	3a4b	2b6a	2b5b
Min	-0.892 (0.368)	-1.087 (0.285)	-0.736 (0.456)	-1.018 (0.369)
Max	2.290 (0.378)	2.343 (0.469)	2.044 (0.388)	2.029 (0.517)
HLLL	3a4a	3a3b	2b5a	2b4b
Min	-0.563 (0.338)	-0.842 (0.432)	-0.444 (0.437)	-0.804 (0.357)
Max	2.367 (0.327)	2.408 (0.378)	2.270 (0.403)	2.297 (0.411)
HLL	3a3a		2b4a	2b3b
Min	-0.159 (0.405)		-0.085 (0.316)	-0.529 (0.345)
Max	2.311 (0.550)		2.134 (0.357)	2.284 (0.381)

Table 5: Average F0 minima and maxima. SDs are presented in parenthesis.

HLLLLLL	DF (1, 77) = 6.773, p = 0.0001		
	3a7a > 3a6b		
HLLLLL	DF (3, 151) = 17.364, p = 0.0001		
	3a5b	2b7a	2b6b
3a6a	6.818**	N/A	6.722**
3a5b		10.454**	N/A
2b7a			10.360**
HLLLL	DF (3, 155) = 6.778, p = 0.0003		
	3a4b	2b6a	2b5b
3a5a	1.785	N/A	0.740
3a4b		5.854**	N/A
2b6a			3.767**
HLLL	DF (3, 153) = 9.307, p = 0.0001		
	3a3b	2b5a	2b4b
3a4a	3.198**	N/A	2.421*
3a3b		6.704**	N/A
2b5a			5.565**
HLL	DF (2, 114) = 17.487, p = 0.0001		
	2b4a	2b3b	
3a3a	N/A	10.607**	
2b4a		15.078**	

Table 6: Results of t-test, ANOVA and Scheffe F-test on F0 minima. * = significant at 90% and ** = significant at 95%. Five groups of comparisons according to Table 4. N/A = not applicable.

The statistical results presented in Table 6 confirm that the F0 minima of the phrases having a Type A as the second component is significantly higher than those having a Type B in the majority of comparisons (13/15). 12 out of the 14 F-values were significant. For example, according to Table 6, the F0 minima of 3a7a is significantly higher than that of 3a6b [DF (1, 77) = 6.773, p = 0.0001].

However, as far as the F0 maxima is concerned, the observation from Table 5 can not be statistically supported. According to Table 7, the differences appearing in the F0 maxima between those phrases having a Type A and a Type B as the first component is not statistically different in the majority of the comparisons (only 3/14 were significantly different).

HLLLLL	DF (3, 151) = 1.011, p = 0.3895		
	3a5b	2b7a	2b6b
3a6a	N/A	0.160	0.316
3a5b		0.558	0.835
2b7a			N/A
HLLLL	DF (3, 155) = 5.403, p = 0.0015		
	3a4b	2b6a	2b5b
3a5a	N/A	2.026	2.279*
3a4b		3.037**	3.348**
2b6a			N/A
HLLL	DF (3, 153) = 1.077, p = 0.3607		
	3a3b	2b5a	2b4b
3a4a	N/A	0.415	0.216
3a3b		0.856	0.555
2b5a			N/A
HLL	DF (2, 114) = 1.810, p = 0.1683		
	2b4a	2b3b	
3a3a	N/A	0.038	
2b4a		1.133	

Table 7: Results of ANOVA and Scheffe F-test on F0 maxima. * = significant at 90% and ** = significant at 95%. Four groups of comparisons according to Table 4. N/A = not applicable.

F0 REALISATIONS OF THE HLLLLL(L) SEQUENCES

In this section, the prosodic nature of the HLLLLL(L) sequences will be acoustic-phonetically described. In the AM model, the HLLLLL(L) sequence would be marked with a high initial tone and a final low tone, and the F0 is realised on the basis of the interpolation between these tones. Figure 1 below contains the mean normalised F0 curve plotted against the mean absolute duration with SDs. What can be observed from Figure 1 is that the interpolation between a high tone and a low tone does not seem to be realised as a simple linear F0 regression, instead it is realised as an exponential curve.

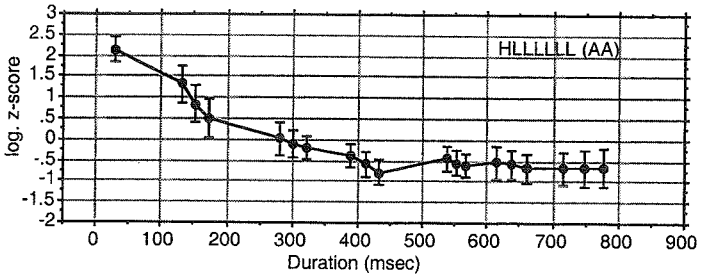

Figure 1: Mean normalised F0 curve of HLLLLL (AA) with one SD above and below it.

In order to show more clearly that the interpolation from a high tone to a low tone is realised as an exponential curve, a two-degree polynomial was fitted to all observed normalised F0 values for each

phrase. In Figure 2, all observed normalised F0 values collected from the HLLLLLL (AA) sequences are plotted against time with a two-degree polynomial curve. Moreover, in Figure 3, the two-degree polynomial curves of AA, AB, BA, and BB phrases are presented together. The relevant statistical information for Figure 3 is provided in Table 8.

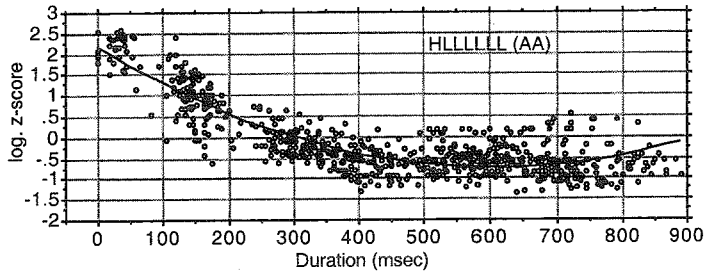

Figure 2: All observed normalised F0 values plotted against absolute duration with a two-degree polynomial curve. $y = 2.165 - 0.010x + 7.837E - 6x^2$, $R^2 = 0.770$.

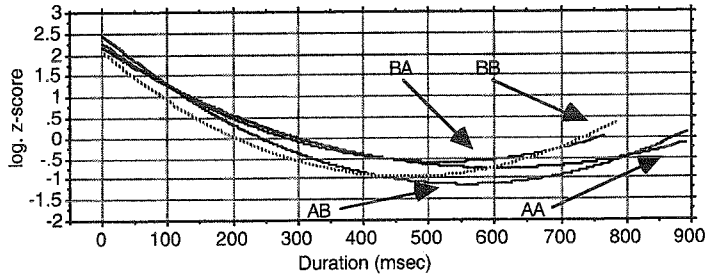

Figure 3: Two-degree polynomial curves for the HLLLLL(L) sequences collected from AA, AB, BA and BB phrases plotted all together.

Type	Formula, R-square	Analysis of Variance
AA	$y = 2.165 - 0.010x + 7.837E - 6x^2$, $R^2 = 0.770$	DF (2, 735) = 1232.689, p = 0.0001
AB	$y = 2.462 - 0.013x + 1.152E - 5x^2$, $R^2 = 0.756$	DF (2, 751) = 1161.018, p = 0.0001
BA	$y = 2.271 - 0.011x + 1.053E - 5x^2$, $R^2 = 0.679$	DF (2, 603) = 638.773, p = 0.0001
BB	$y = 2.070 - 0.013x + 1.367E - 5x^2$, $R^2 = 0.684$	DF (2, 637) = 689.046, p = 0.0001

Table 8: The relevant statistics for Figure 3.

The highest R^2 value is 0.770 (AA), and the lowest one is 0.679 SDs (BA). The R^2 values of those phrases having a Type A for the first component appear to be higher than those having a Type B (AA: 0.770; AB: 0.756 > BA: 0.679; BB: 0.689). Although there are some differences in R^2 values which show how well the polynomial curve captures the nature of observed normalised F0 values, the high probability value (p = 0.0001) clearly indicates that these polynomial curves are useful for prediction. Furthermore, these R^2 values are far better than those calculated from simple linear lines [AA: 0.569; AB: 0.415; BA: 0.478; BB: 0.381].

DISCUSSION

In this paper, I have statistically shown that the realisational difference in F0 appearing between Type A and Type B is observed at phrase level as well as word level. However, the difference was only

confirmed statistically at the F0 minimum point, not at the F0 maximum point. This could be because of the syntactic structure of the carrier sentence, where a syntactic boundary exists immediately before a target phrase. That is, the initial part of a target phrase is anchored with a tone which is associated with a higher intonational level. Therefore, the realisation of the initial peak value could have been influenced by this tone so that the underlying difference between Type A and Type B was neutralised.

I have also demonstrated that the interpolation from a high tone to a low tone is acoustic-phonetically realised as an exponential curve in KJ. This exponential nature of the two target-tone interpolation could be accounted for in terms of production, particularly the bio-mechanical nature of muscular tissues (Fujisaki, 1983). The difference in F0 range caused by accentual types can be also observed amongst the two-degree polynomial curves presented in Figure 3. A large number of different frameworks have been developed for the annotation or coding of speech corpora (PROSA, TEI, ToBI, IPO, and more). At the prosodic level, what different schemes appear to have in common is that they seek for some sorts of F0 stylised approximation (Pijper, 1983) of what they claim to be the apparently arbitrarily pitch fluctuations found in natural speech. In the process of F0 stylisation, target points are interpolated with a linear function in most encoding schemes. However, as I have shown, the realisation of the interpolation between two target tones does not have a simple linear function in KJ. Therefore, stylised approximation using an exponential function seems to be appropriate in the case of KJ. In the field of speech synthesis, this sort of fundamental information is important because it is generally believed that prosody, particularly fundamental frequency, is the key to the naturalness (i.e. Carlson and Granström, 1997).

ACKNOWLEDGMENTS

The author would like to thank Dr Phil Rose and two anonymous reviewers for their valuable comments.

REFERENCES

Carlson, R. and Granström, B. (1997) *Speech Synthesis.* In Hardcastle, W. J. and Laver, J. (eds.), The Handbook of Phonetic Science, p. 768-788.

Fujisaki, H. (1983) *Dynamic Characteristics of Voice Fundamental Frequency in Speech and Singing.* In MacNeilage, P. (ed.), The Production of Speech, 39-55.

Hirayama, T. (1960) *Zenkoku akusento jiten.* Tokyo: Tokyodo.

Ishihara, S. (1998) *An Acoustic-Phonetic Description of Word Tone in Kagoshima Japanese.* Proceedings of the 5th International Conference on Spoken Language Processing 3, p. 595-598

Ishihara, S. (2000) *An Acoustic-Phonetic Descriptive Analysis of Pitch Realisations in Kagoshima Japanese.* In Henderson, J (ed.), Proceedings of the 1999 Conference of the Australian Linguistics Society. http://www.arts.uwa.edu.au/LingWWW/als99/proceedings.

Pierrehumbert, J. and Beckman, M. (1988) *Japanese Tone Structure.* Cambridge, Massachusetts, London, England: The MIT Press.

Pijper, D. R. de. (1983) *Modelling British English Intonation.* Dordrecht, Cinnaminson: Foris Publications.

Rose, P. (1987) *Considerations in the Normalisation of the Fundamental Frequency of Linguistic Tone.* Speech Communication 6, 343-351.

Vaissière, J. (1983) *Language Independent Prosodic Features.* In Cutler, A. and Ladd, D. R. (eds.), Prosody: Models and Measurements, p. 53-66. Berlin: Spring-Verlag.

Zhu, X. (1999) *Shanghai Tonetics.* Lincom Studies in Asian Linguistics 32. Lincom Europe.