Kyoto–Osaka Japanese Citation Tone Acoustics: A Linguistic Tonetic Study

Shunichi Ishihara

Japan Centre (Faculty of Asian Studies), College of Asia and the Pacific
The Australian National University
Shunichi.Ishihara@anu.edu.au

Abstract

An acoustic–description of the contrastive accentual types of Kyoto–Osaka Japanese is provided on the basis of 12 informants (6 males and 6 females) using monosyllabic and disyllabic words. The linguistic–phonetic properties of the accentual contrast are specified from mean and standard deviation normalised F0.

1 Introduction

The aim of this paper is to give a quantified description of the linguistic–phonetic acoustic properties of the Kyoto–Osaka Japanese (KOJ) accentual contrast. There are some auditory descriptions of the accentual contrast of KOJ in the literature (Hirayama 1960), and various phonological treatments have been proposed for KOJ accentuation (McCawley 1986; Haraguchi 1977; Shibatani 1990). Besides acoustic studies of individual speakers (Sugito 1997; 1998), the suprasegmental features of KOJ have been studied from various angles, such as physiology of accent (Sugito and Hirose 1978) and accent perception using speech synthesis (Sugito 1998). However, there has to my knowledge been no study which specifies the linguistic–phonetic acoustic characteristics of KOJ by normalising acoustic data from a number of speakers. This study addresses this gap.

The pitch pattern of words in KOJ is determined: 1) by the presence or absence of a lexical accent, 2) if it is present, by its position, which dictates where the accentual pitch fall starts, and 3) by which group, that is either 'low–pitch beginning' or 'high–pitch beginning', a word belongs to (Shibatani 1990; Sugito 1996). Those words which have a lexical accent—of which phonetic realisation is a pitch fall—are called 'accented' words, and those which do not are called 'unaccented' words. That is, both *kabuto* [LHL] 'helmet' and *otoko* [HHL] 'man' are 'accented' words having a lexical accent on the second syllable. The former is a 'low–pitch beginning' word and the latter is a 'high–pitch beginning' word. Likewise, *suzume* [LLH] 'sparrow' and *sakura* [HHH] 'cherry blossom' are both 'unaccented' words. However, the former belongs to the 'low–pitch beginning' group and the latter to the 'high–pitch beginning' group. A pitch rise may be observed at the ultimate syllable in unaccented 'low–pitch beginning' words, as can be seen in *suzume* [LLH] 'sparrow' (Sugito 1997).

1.1 Conventions

In studies on KOJ accentuation, conventions combining H/L pitches and numerals have traditionally been used (i.e. H1, L0, L2, etc) to indicate the accentual type of a given word. The H/L pitches of the convention show which group, 'low–pitch beginning' or 'high–pitch beginning' a given word belongs to, and the numerals indicate the location of an accent (i.e. positive integers) or its absence (i.e. 0). Thus, H2 means that the word is 'high–pitch beginning' having a lexical accent on the second syllable. L0 means that the word is unaccented 'low–pitch beginning'. These phonological conventions are used throughout this paper.

Table 1: *Accentual Types*

1 σ				2 σ					
L0	L1	H0	H1	L0	L1	L2	H0	H1	H2
√	-	√	√	√	-	√	√	√	-

If one applies these conventions to monosyllabic and disyllabic words, the combinations given in Table 1 are logically possible. However, as shown in Table 1, several gaps exist in the logical combinations, such as no L1 for monosyllabic words and no L1 or H2 for disyllabic words. That is, in KOJ there are only three contrastive accentual types for monosyllabic words and four for disyllabic words.

1.2 Procedure

In this study, Japanese Speech Corpora of Major City Dialects (JCMD) vol 2 is used as a database (Tahara, Egawa, Sugito, and Itahashi 1993). This database contains recorded samples of 250 citation words and short phrases uttered by 20 native speakers of KOJ. These words and phrases were read once by each speaker, and recorded on professional digital equipment in 1991. Detailed recording conditions and elicitation procedures are not found for the JCMD database. All sounds in the JCMD database are digitised in 16 KHz. In this study, only monosyllabic and disyllabic words recorded by the 12 speakers who had completed their compulsory education at the time of recording were used. The remaining 8 speakers were not included as their language skills may not have been fully developed. The 12 speakers comprise 6 males and 6 females of three generations (cf. Ladefoged 1997, p140). The average age of these 12 speakers at the time of recording was 57 (18.6) for males and 46 (14.8) for females. In this paper, the speakers are referred to as {(O)ld/(M)iddle age/(Y)oung]}{(M)ale/(F)emale}{1/2}. For example, MF1 refers to the first (1) middle aged (M) female (F) speaker.

The monosyllabic and disyllabic words analysed in this paper are listed in Tables 2 and 3, respectively together with their accentual types. The syllable structure of both monosyllabic and disyllabic words is (C)V. As can be seen in Table 2, the initial consonants for the monosyllabic target words are voiceless, with the exception of ([m]). This is due to database limitations. This unbalanced situation may

Proceedings of the 11th Australian International Conference on Speech Science & Technology, ed. Paul Warren & Catherine I. Watson. ISBN 0 9581946 2 9
University of Auckland, New Zealand. December 6-8, 2006.

Accepted after full paper review

Table 2: *Monosyllabic target words and their accentual types*

ka	*mosquito*	H0	tɕi	*blood*	H0
ha	*leaf*	H1,H0	ke	*hair*	H1
te	*hand*	L0	me	*eye*	L0

intrinsically cause higher F0 at the onset of a rhyme (House and Fairbanks 1953; Lehiste and Peterson 1961). The vowels have various vocalic segments in terms of their height (Peterson and Barney 1952). As for the disyllabic target words shown in Table 3, voiced and voiceless consonants and various vowels are fairly evenly distributed.

Table 3: *Disyllabic target words and their accentual types.*

niɯa	*garden*	H0	kata	*shoulder*	L0, H1
hana	*nose*	H0	matsu	*pine tree*	L0, H0
kami	*paper*	H1	umi	*sea*	L0, H0
jama	*mountain*	H1	koko	*here*	L0, L2
naka	*inside*	L0	saɾu	*monkey*	L2
nani	*what*	L0	aki	*autumn*	L2
ame	*rain*	L2	mado	*window*	L2

As can be seen in Tables 2 and 3, inconsistencies exist between speakers with respect to the accentual types for some words (i.e. [ha], [kata], [matsu], [umi] and [koko]). In fact, these inconsistencies are due to the atypical behaviour of two speakers. All speakers recognise the monosyllabic word [ha] as /H1/, except for OM1 who perceives it as /H0/. Similarly, all speakers bar OF1 recognise the disyllabic words [kata], [matsu], [umi] and [koko] as /L0/. This kind of inconsistency has been reported particularly across generations (Nakai 1997; Sugito 1997). According to Sugito (1997: 210–212), the inconsistency between /H0/ and /L0/ (i.e. [matsu] and [umi]) is very rare. The inconsistent tokens uttered by OM1 and OF1, as well as several other tokens, were excluded from the analysis. Please refer to §2 for more about these tokens.

After being annotated, all target tokens were analysed using the ESPS routine of the *Snack Sound Toolkit* (Sjölander 2006). On the basis of the annotation, F0 was sampled at the onset and every 10% point for each rhyme for the monosyllabic target words, and at the onset and every 20% point for each rhyme for the disyllabic target words. The identification of the rhyme onset is determined straightforwardly from the audio speech waveforms and spectrograms; however, the offset of a word is more difficult to judge. It 'was adjudged to occur at the point where the glottal pulse train showed an obvious discontinuity in the regularity of increase of period' (Rose 1982, p7). Although it was not limited to the accented tokens, some sort of glottal stop–like laryngeal tension was observed in many recordings of the falling pitch tokens (i.e. /H1/ and /L2/).

2 Pitch Realisations of Accentual Types

As previously stated, there are three contrastive accentual types for KOJ monosyllabic words (/L0/, /H0/, /H1/) and four for disyllabic words (/L0/, /L2/, /H0/, /H1/). In this section, auditory descriptions of these contrastive accentual types are given using H/L pitch dichotomy. Tables 4 and 5 contain the pitch realisation(s) of each word across all informants for the monosyllabic and the disyllabic target words, respectively.

Table 4: *Pitch realisations of monosyllabic target words across informants, bold=excluded tokens.*

	ka	ha	te	tɕi	ke	me
	H0	H1,H0	L0	H0	H1	L0
OM1	H	**H**	LH	H	HL	L
OM2	H	HL	LH	H	HL	LH
OF1	H	HL	L	H	HL	L
OF2	H	HL	LH	H	HL	LH
MM1	H	HL	LH	H	HL	LH
MM2	H	HL	LH	H	HL	L
MF1	H	HL	LH	H	HL	LH
MF2	H	HL	LH	H	HL	LH
YM1	H	HL	L	H	HL	L
YM2	H	HL	LH	H	HL	LH
YF1	H	HL	L	H	HL	LH
YF2	H	HL	LH	H	HL	LH

As mentioned in §1.2, OM1 recognises [ha] as a /H0/ word, while the other speaker recognise it as /H1/. As a result, Table 4 shows [ha] realised as [H] for OM1 and as [HL] for other speakers.

As shown in Table 4, there appears to be two allotonic pitch realisations ([LH] and [L]) for the /L0/ words [te] and [me]. Unlike the case of [ha], the [LH] and the [L] are allotonic pitch realisations for the /L0/ accentual type. The JCMD contains recordings of several short phrases combining the target words with a particle. When [te] and [me] are extended with a particle (i.e. *ga* nominative case marker), the resultant two–syllable phrases (i.e. *te–ga* 'hand–NOM') consistently show a [LH] pitch contour for all speakers. It is clear from this that all speakers recognise these two words as /L0/, but that they have two allotonic realisations in the citation form. Some speakers use either [LH] or [L] consistently (i.e. OM2, OF1), while others use both of them, possibly in free variation (i.e. OM1, MM2). Therefore, the F0 realisations of the three accentual types (/L0/, /H0/ and /H1/) are acoustically–phonetically described in terms of [LH], [L], [H] and [HL] for citation monosyllables.

As in the case of the monosyllabic word [ha] uttered by OM1, mentioned in §1.2, OF1 recognises [kata], [matsu], [umi], and [koko] as different accentual types from the other speakers. This results in divergent pitch realisations for these disyllabic words (refer to Table 5).

Some inconsistencies can be observed from Table 5 with respect to the pitch realisations of [ame], [saɾu], [aki] and [mado]) which are phonologically recognised as /L2/ words by all speakers. The majority of speakers show a [LHL] pitch pattern for all or some of these words. However, YF1 consistently utters these words with a [LH] pitch pattern and YM2 uses a [LH] pitch only for [saɾu]. This also appears to be the case of dual allotonic pitch realisations ([LHL] and [LH]) of /L2/.

The pitch realisation that YF1 shows for the /L2/ words ([LH]) results in the merger of /L2/ and /L0/ disyllabic words because the pitch realisation of /L0/ disyllabic words is also [LH]. Although the pitch realisation of /L0/ and /L2/ disyllabic words may be the same in citation in L/H dichotomy, it is clear that YF1 phonologically distinguishes

Proceedings of the 11th Australian International Conference on Speech Science & Technology, ed. Paul Warren & Catherine I. Watson. ISBN 0 9581946 2 9
University of Auckland, New Zealand. December 6-8, 2006.

Accepted after full paper review

Table 5: *Pitch realisations of disyllabic target words across informants, x=unavailable, bold=excluded tokens.*

	niɯa	hana	kami	jama	kata	matsu	umi	koko	naka	nani	ame	saɾu	aki	mado
	H0	H0	H1	L0	L0,H1	L0,H0	L0,H0	L0,L2	L0	L0	L2	L2	L2	L2
OM1	HH	HH	HL	HL	LH	LH	LH	LH	LH	LH	LHL	LHL	LHL	LHL
OM2	HH	HH	HL	HL	LH	LH	LH	LH	LH	LH	LHL	LHL	LHL	LHL
OF1	HH	HH	HL	x	**HL**	**HH**	**HH**	**LHL**	LH	LH	LHL	LHL	LHL	LHL
OF2	HH	HH	HL	HL	LH	LH	LH	LH	x	x	LHL	LHL	LHL	LHL
MM1	HH	HH	HL	HL	LH	LH	LH	LH	LH	LH	LHL	LHL	LHL	LHL
MM2	HH	HH	HL	HL	LH	LH	LH	LH	LH	LH	LHL	LHL	LHL	LHL
MF1	HH	HH	HL	HL	LH	LH	LH	LH	LH	LH	LHL	LHL	LHL	LHL
MF2	HH	HH	HL	HL	LH	LH	LH	LH	LH	LH	LHL	LHL	LHL	LHL
YM1	HH	HH	HL	HL	LH	LH	LH	LH	LH	LH	LHL	LHL	LHL	LHL
YM2	HH	HH	HL	HL	LH	LH	LH	LH	LH	LH	LHL	**LH**	LHL	LHL
YF1	HH	HH	HL	HL	LH	LH	LH	LH	LH	LH	**LH**	**LH**	**LH**	**LH**
YF2	HH	HH	HL	LH	LH	LH	LH	LH	LH	LH	LHL	LHL	LHL	LHL

two lexical classes (/L0/ and /L2/). The phonological difference between /L0/ and /L2/ becomes phonetically apparent when something (i.e. *ga* nominative marker) is subsequently attached. When the nominative marker is attached to a /L0/ disyllabic word (i.e. *kata–ga* 'shoulder–NOM'), the phrase shows a [LLH] pitch configuration whereas when the nominative marker is attached to a /L2/ disyllabic word (i.e. *ame–ga* 'rain–NOM'), the phrase shows a [LHL] pitch configuration. YM2 also uses a [LH] pitch for the /L2/ word [saɾu] despite the fact that he shows a [LHL] pitch contour for the rest of the /L2/ disyllabic target words.

Although it has to my knowledge never been explicitly mentioned in the literature, it is evident from the JCMD that the phonetic contrast between /L0/ and /L2/ is disappearing amongst the informants excluded from this study (those who had not completed their secondary education at the time of recording). All of these young informants show the [LH] rather than the [LHL] pitch pattern for the /L2/ disyllabic target words. However, it is clear that they still maintain the phonological contrast between /L0/ and /L2/.

Since we do not have sufficient data to describe the acoustic–phonetic realisation of the /L2/ disyllabic words with a [LH] pitch contour (those LH tokens which are given in bold face in Tables 4), these tokens are excluded from analysis. Those tokens marked with 'x' are also excluded because they are uttered with an interrogative intonation.

3 Normalisation Results

As stated in §1, the aim of this paper is to give a quantified description of the linguistic–phonetic acoustic properties of KOJ accentual types. However, this is not an easy task because '[t]he acoustic properties of the radiated speech wave are a unique function of a speaker's vocal tract anatomy, and since speakers' vocal tracts differ, so will their acoustic output—even for phonetically the same sound' (Rose 1987, p7). Therefore, the individual content needs to be factored out as much as possible to extract the linguistic and accentual content. The z–score normalisation technique, of which performance has been empirically attested (Zhu 1994; Ishihara 2004), is used to extract the acoustic correlates of KOJ's accentual contrast (Disner 1980, p253; Rose 1991, p229). Z–score normalised values (z_i) can be obtained using the formula $z_i = (x_i - m_y)/s_y$, where x_i is a sampling point, and m_y and s_y are the arithmetic mean and standard deviation of x_i ($i = 1, 2, \cdots, n$), respectively. In this particular normalisation procedure, each F0 observation is expressed as so many standard deviations above and below a speaker's overall mean F0. Intrinsic normalisation parameters (m_y and s_y) were acquired from the words to be normalised. The normalisation parameters are given in Table 6 for each speaker. Standard deviation (sds) is used as the unit for normalised F0 values because normalisation is based on the sds of the samples.

Table 6: *Normalisation parameters in Hz.*

	m_y	s_y		m_y	s_y
OM1	151.1	25.8	MF1	208.3	33.3
OM2	160.3	29.8	MF2	205.0	31.9
OF1	149.4	19.7	YM1	86.5	10.5
OF2	211.7	47.8	YM2	112.9	16.4
MM1	111.7	25.6	YF1	217.7	20.6
MM2	155.3	29.9	YF2	211.8	30.0

3.1 Monosyllables

Fig. 1 contains the mean normalised F0 curves of KOJ's accentual contrast for citation monosyllabic words together with one standard deviation above and below the mean, plotted against mean absolute duration. Fig. 1 shows that the mean normalised F0 curves for the accentual types of monosyllables lie between ±2 sds above and below the mean.

The /H0/ accentual type ([H]) shows a very moderate F0 fall between 1 sds and 0 sds. The /H1/ accentual type ([HL]) shows a clear sharp fall across the entire F0 range of a speaker. Although the /H1/ and the /H0/ accentual types are phonetically transcribed as [HL] and [H], respectively, the former is realised significantly higher in F0 than the latter between the onset point and the 20% point ($p < 0.013$).

As auditorily described in §2, the /L0/ accentual type clearly exhibits two different F0 curves (low level and low ring curves) reflecting the two allotonic pitch realisations ([L] and [LH]), respectively. The low level [L] and the low rising F0 curves [LH] for the /L0/ accentual type have similar F0 values (≈ -0.5 sds) from the onset to around the

Proceedings of the 11th Australian International Conference on Speech Science & Technology, ed. Paul Warren & Catherine I. Watson. ISBN 0 9581946 2 9
University of Auckland, New Zealand. December 6-8, 2006.

Accepted after full paper review

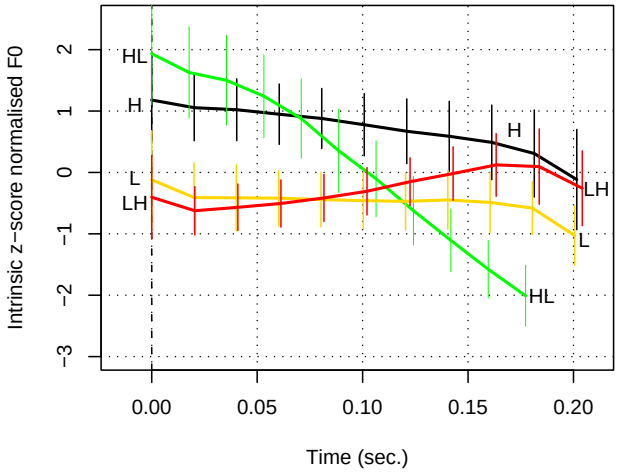

Figure 1: *Mean normalised F0 curves for the accentual types (H0, L0, H1) of KOJ citation monosyllables, plotted against mean absolute duration. Vertical bars indicate one standard deviation above and below the mean.*

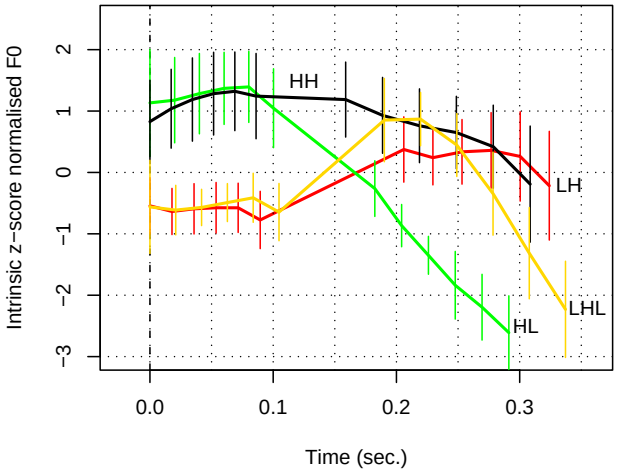

Figure 2: *Mean normalised F0 curves for the accentual types (H0, L0, H1 and L2) of KOJ citation disyllables, plotted against mean absolute duration. Vertical bars indicate one standard deviation above and below the mean.*

50% point. However, after that, the former stays more or less level while the latter begins to rise up to the offset of around 0 sds. The F0 values of the [LH] type are significantly higher from the 80% point to the offset point ($p \leq 0.018$) than those of the [L] type. Therefore, it appears both auditorily and acoustically plausible to hypothesise two allotonic realisations for the /L0/ type in monosyllables.

The moderate F0 falling contour ([H]) of the /H0/ and the low rising F0 contour ([LH]) of the /L0/ gradually merge towards the offset ($\approx$ 0 sds), resulting in no significant difference between them at the 90% point and the offset point ($p \geq 0.31$). All non–falling F0 curves ([H], [LH] and [L]) show a sudden drop in F0 of about 0.3 sds at the offset which is perhaps due to the offset laryngeal tension.

A relatively large amount of variation ($\approx$ 0.7-0.8 sds) is found at the onset point of all accentual types and at the offset point of some accentual types. This is due to the perturbatory effect of the initial consonant and the final laryngeal tension which was referred to in §1.2.

As far as the mean duration is concerned, the [HL] curve (0.175 sec) is shorter in duration than the rest, which share more or less the same length ([H]: 0.201 sec; [LH]: 0.204 sec; [L]: 0.200 sec).

3.2 Disyllables

Fig. 2 contains the mean normalised F0 curves of KOJ's accentual contrast for citation disyllabic words together with one standard deviation above and below the mean, plotted against mean absolute duration. Fig. 2 shows that the mean normalised curves for the accentual types of disyllables lie between about 1.3 sds and -2.5 sds.

As can be seen in Figs. 1 and 2, the relative distributional relationship between the [HH], [LH] and [HL] types of disyllables is almost identical to that between the [H], [LH] and [HL] types of monosyllables on the F0 plane. The /H0/ accentual type ([HH]) shows a fairly level F0 contour in the first syllable, except for the minor F0 rise at the beginning. In the second syllable, F0 gradually decreases towards the offset. As for the F0 realisation of the /H1/ type ([HL]), F0 stays almost level in the first syllable, and then starts falling from the offset of the first syllable, continuing to fall until the offset of the second syllable. The first syllable of the /H0/ type ([HH]) and that of the /H1/ type ([HL]) are very similar in their F0 values. A statistical comparison shows no difference between the /H1/ and the /H0/ accentual types in the first syllable ($p > 0.18$).

Although the F0 contour of the /L0/ type ([LH]) rises overall, the F0 stays more or less level with some minor ups and downs within the first and second syllables. The /L2/ accentual type shows a rising–falling F0 contour which reflects the [LHL] pitch realisation. The F0 values of the first syllable are very similar to those of the /L0/ accentual type ([LH]). Statistically speaking, no difference was found between the /H0/ and the /H2/ accentual types in the first syllable ($p > 0.05$). A large F0 fall starts at an early point of the second syllable (around 20% into the second syllable) in the /L2/ type. The maximum F0 value in the second syllable appears to be higher in the /L2/ ([LHL]) than the /L0/ accentual type ([LH]). A statistical comparison between the /L0/ and the /L2/ types at their maximum values confirmed ($p < 0.001$) that the /L2/ (average maximum F0 value is 1.105 sds) reaches a significantly higher F0 than the /L0/ (average maximum F0 value is 0.688 sds).

Based on the mean duration values, the durations of the [HH], [HL], [LH] and [LHL] types can be expressed by their inequality ([HL]: 0.29 sec < [HH]: 0.31 sec < [LH]: 0.32 sec < [LHL]: 0.34 sec).

4 Discussion

In Fig. 3, mean normalised F0 realisations of KOJ's accentual types for monosyllables and disyllables are plotted against equalised duration. The dashed lines are for the monosyllables and the dotted lines are for the disyllables.

As can be seen in Fig. 3, the F0 realisations of the /H0/, /H1/ and /L0/ types are almost identical for monosyllables (1, 3 and 6 of Fig. 3) and disyllables (2, 4 and 7 of Fig. 3) when plotted against equalised duration. However, there are still some minor differences. As for the /H1/ type ([HL]), the F0 does not fall in the monosyllables (3) as much as the disyllables (4) (the gap is about 0.63 sds

Proceedings of the 11th Australian International Conference on Speech Science & Technology, ed. Paul Warren & Catherine I. Watson. ISBN 0 9581946 2 9
University of Auckland, New Zealand. December 6-8, 2006.

Accepted after full paper review

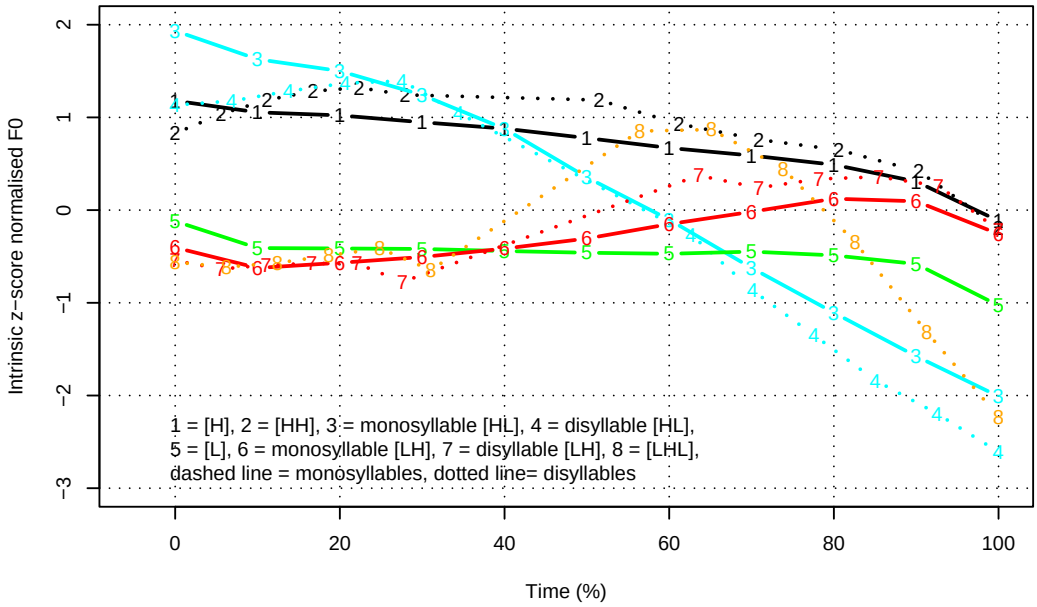

Figure 3: *Mean normalised F0 curves for the accentual types (H0, L0, H1 and L2) of KOJ citation monosyllables and disyllables plotted against equalised duration (%).*

at the offset point). This difference was statistically confirmed ($p < 0.001$) in terms of minimum F0 values, with the average minimum F0 for monosyllables and disyllables: -2.04 sds and -2.61 sds, respectively. Similarly, for the /L0/ type ([LH]), the F0 reaches approximately 0.25 sds higher in the monosyllables (6) than the disyllables (7). Again, this difference is statistically significant ($p < 0.04$) in terms of maximum F0 values,with the average maximum F0 for monosyllables and disyllables being 0.34 sds and 0.68 sds, respectively. A plausible account for these realisation differences is that the F0 realisations of /H1/ and /L0/ were 'undershot' in monosyllables due to their shorter duration than disyllables (Daniloff, Schuckers, and Lawrence 1980, p317; Clark and Yallop 1990, p119).

A large difference can be seen in the F0 realisation of the /H1/ accentual type between the monosyllables (3) and the disyllables (4). The F0 starts about 0.74 sds higher in the monosyllables than the disyllables, but this difference no longer exists by the 30% point. However, this difference in the F0 realisation of the /H1/ between monosyllables and disyllables is not statistically significant ($p = 0.11$) in terms of the maximum F0 value (average maximum F0 values for monosyllables and disyllables are 2.00 sds and 1.66 sds, respectively) due to a large standard deviation.

Various tonal languages show apparent allotony as a function of phonological structure/length. Provided that the F0 realisation of a given tone on a long syllable (i.e. VV) is a default full–scale realisation, usually two allotonic realisation patterns are possible on a short syllable (i.e. V) (Hombert, Ohala, and Ewan 1979). One possible allotonic pattern is that a tonal contour can be truncated due to the shortness of the syllable. The other pattern is that the default full–scale contour is 'time–warped' so that the whole contour shape can be distributed over the shorter duration. When the F0 realisations of the the /H0/, the /H1/ and the /L0/ accentual types are plotted together for monosyllables and disyllables, it is clear that KOJ belongs to the latter pattern.

It was reported in §3.1 that the /H1/ accentual type ([HL]) starts higher in F0 than the /H0/ accentual type ([HH]) at an early stage. The main phonological difference between the /H1/ and the /H0/ accentual types is that the former has a lexical accent ([+accent]) while the latter does not ([-accent]). It has been reported in standard Japanese (SJ) that with everything else being equal an accented syllable is realised higher in F0 than an unaccented syllable (Kubozono 1993). Since the actual target words used for the monosyllabic /H1/ and /H0/ accentual types are phonetically cohesive (/H1/: [ha] and [ke] vs /H0/: [ka] and [tɕi]), as far as monosyllables are concerned, it appears that the lexical accent has an F0 raising effect in KOJ as well. However, the same phenomenon was not observed in the disyllabic /H1/ and the /H0/ accentual types. However, unlike this pair, the /L2/ and /L0/ pair—the other disyllabic words contrastive in terms of [±accent]—showed in §3.2 that the maximum F0 value of the /L2/ type (8) was significantly higher in the second syllable than that of the /L0/ type (7). This difference was statistically confirmed ($p < 0.001$). However, when it comes to the disyllabic /L2/ and /L0/ pair, it is not clear whether or not the F0 realisation difference was caused by the lexical accent. It is possible that the /L0/ accentual type is realised lower than the /L2/ accentual type in the second syllable due to the boundary intonational tone which induces F0 lowering (Pierrehumbert and Beckman 1988). Therefore, it is not clear at this stage whether a lexical accent causes any sort of F0 rise in KOJ (cf. Sugito 1998).

It is clear from Fig. 3 that there are at least two F0 onset points: the high point around 1–2 sds for the onset of the /H0/ and the /H1/ accentual types and the low point around -0.5 sds for the onset of the /L0/ and the /L2/. These two points are considered to be a function of the phonological groups of 'high–pitch beginning' and 'low–pitch beginning'. Three offset points can be identified in Fig. 3: the highest point around 0 sds for the offset of the /H0/ and the [LH] type, the mid point around -1 sds for the [L] type

Proceedings of the 11th Australian International Conference on Speech Science & Technology, ed. Paul Warren & Catherine I. Watson. ISBN 0 9581946 2 9
University of Auckland, New Zealand. December 6-8, 2006.

Accepted after full paper review

of the /L0/, and the lowest point around -2.2 sds for the /H1/. As briefly mentioned above, these onset and offset points are acoustically achieved regardless of the phonological length of a word.

Although they are not comparable in the strict sense, the /H1/ accentual type ([HL]) is realised shorter than the other accentual types, as described in §3.1 and §3.2, in both monosyllables and disyllables. This point conforms to what Sugito reports (Sugito 1997, p317). The short duration of a high–falling tone compared to a level/rising tone has been reported in the tonal literature as well (Ohala and Ewan 1973; Sundberg 1973).

5 Summary

In this paper, the linguistic–phonetic properties of KOJ accentual types were presented from mean and standard deviation z–score normalised F0 data for citation monosyllables and disyllables using six male and six female speakers. The linguistic–phonetic descriptions presented in this paper can be used to compare with those of other Japanese dialects in order to identify the acoustic–phonetic realisation differences/similarities across Japanese dialects.

6 Acknowledgments

I thank two anonymous reviewers for giving valuable comments for this research.

References

Clark, J. and C. Yallop (1990). *An Introduction to Phonetics and Phonology*. Oxford: Blackwell Publishers.

Daniloff, R., G. Schuckers, and F. Lawrence (1980). *The Physiology of Speech and Hearing*. New Jersey: Prentice–Hall.

Disner, S. (1980). Evaluation of vowel normalization procedures. *Journal of Acoustical Society of America 67*(1), 253–261.

Haraguchi, S. (1977). *The Tone Pattern of Japanese: An Autosegmental Theory of Tonology*. Tokyo: Kaitakusha.

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

Hombert, J., J. Ohala, and W. Ewan (1979). Phonetic explanations for the development of tones. *Languages 55*(1), 37–58.

House, A. S. and G. Fairbanks (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. *Journal of Acoustical Society of America 25*, 105–113.

Ishihara, S. (2004). *An Acoustic–Phonetic Descriptive Analysis of Kagoshima Japanese Tonal Phenomena*. Ph. D. thesis, the Australian National University.

Kubozono, H. (1993). *The Organization of Japanese Prosody*. Tokyo: Kuroshio Publishers.

Ladefoged, P. (1997). Instrumental techniques for fieldwork. In W. J. Hardcastle and J. Laver (Eds.), *The Handbook of Phonetic Science*, pp. 137–166. Oxford: Blackwell Publishers.

Lehiste, I. and G. E. Peterson (1961). Some basic considerations in the analysis of intonation. *Journal of Acoustical Society of America 33*, 368–379.

McCawley, J. (1986). *Phonological Component of a Grammar of Japanese*. The Hague: Mouton.

Nakai, Y. (1997). Individual differences of word-accent in the Kyoto dialect. *Mathematical Linguistics 20*(7), 18–29.

Ohala, J. and W. Ewan (1973). Speed of pitch. *Journal of Acoustical Society of America 53*, 345.

Peterson, G. E. and H. L. Barney (1952). Control methods used in a study of vowels. *Journal of Acoustical Society of America 24*, 175–184.

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

Rose, P. (1982). Acoustic characteristics of the Shanghai–Zhenhai syllable types. *Papers in South–East Asian Linguistics 8: Tonation, Pacific Linguistic Series A 26*, 1–53.

Rose, P. (1987). Considerations in the normalisation of the fundamental frequency of linguistic tone. *Speech Communication 6*, 343–351.

Rose, P. (1991). How effective are long term mean and standard deviation as normalisation parameters for tonal fundamental frequency? *Speech Communication 10*, 229–247.

Shibatani, M. (1990). *The Languages of Japan*. Cambridge: Cambridge University Press.

Sjölander, K. (2006). *The Snack Sound Toolkit*. http://www.speech.kth.se/snack/.

Sugito, M. (1996). *Osaka/Tokyo akusento onsei jiten* [Sound Dictionary of Accent: Osaka and Tokyo Japanese]*, CD-ROM*. Tokyo: Maruzen.

Sugito, M. (1997). *Onsei hakeiwa kataru* [Speech waveforms speak], Volume 4 of *Nihongo onseino kenkyu* [Studies on Japanese Speech]. Tokyo: Izumishoin.

Sugito, M. (1998). *Hana to hana* ['Flower' and 'nose'], Volume 5 of *Nihongo onseino kenkyu* [Studies on Japanese Speech]. Tokyo: Izumishoin.

Sugito, M. and H. Hirose (1978). An electromyographic study of the Kinki accent. *Annual bulletin Research Institute of logopedics and phoniatrics, University of Tokyo 12*, 33–51.

Sundberg, J. (1973). Data on maximum speed of pitch changes. *Quarterly Progress and Status Reports, Speech Transmission Laboratory, Stockholm, 1973/4*, 39–47.

Tahara, H., K. Egawa, M. Sugito, and S. Itahashi (Eds.) (1993). *Japanese Speech Corpora of Major City Dialects*, Volume 2. Ministry of Education, Science and Culture.

Zhu, X. (1994). *Shanghai Tonetics*. Ph. D. thesis, the Australian National University.

Proceedings of the 11th Australian International Conference on Speech Science & Technology, ed. Paul Warren & Catherine I. Watson. ISBN 0 9581946 2 9
University of Auckland, New Zealand. December 6-8, 2006.

Accepted after full paper review