

Journal of the International Phonetic Association

<http://journals.cambridge.org/IPA>

Additional services for *Journal of the International Phonetic Association*:

Email alerts: [Click here](#)

Subscriptions: [Click here](#)

Commercial reprints: [Click here](#)

Terms of use : [Click here](#)



Osaka and Kagoshima Japanese citation tone acoustics: A linguistic-tonetic comparative study

Shunichi Ishihara

Journal of the International Phonetic Association / Volume 42 / Issue 01 / April 2012, pp 1 - 21

DOI: 10.1017/S0025100311000478, Published online: 12 March 2012

Link to this article: http://journals.cambridge.org/abstract_S0025100311000478

How to cite this article:

Shunichi Ishihara (2012). Osaka and Kagoshima Japanese citation tone acoustics: A linguistic-tonetic comparative study. *Journal of the International Phonetic Association*, 42, pp 1-21
doi:10.1017/S0025100311000478

Request Permissions : [Click here](#)

Osaka and Kagoshima Japanese citation tone acoustics: A linguistic-tonetic comparative study

Shunichi Ishihara

College of Asia and the Pacific,
School of Culture, History and Language,
The Australian National University
shunichi.ishihara@anu.edu.au

The pitch realisations of the accentual systems in Osaka Japanese (OJ) and Kagoshima Japanese (KJ) have been auditorily described in detail, and analysed within various phonological frameworks. However, little linguistic-phonetic descriptive research has been undertaken on the accent types of Japanese dialects in such a way as to enable a cross-dialectal comparison of their acoustic realisation. In this study, linguistic-tonetic representations of OJ and KJ tonalities are derived from normalised acoustic representations for pitch patterns conventionally described as LH, LHL, LLH and LLLH. A comparison of these representations across the two dialects demonstrates some significant differences in the acoustic realisation of the H/L units. The implications of these observed differences for surface tonal representation of KJ within Autosegmental-Metrical theory are also explored.

1 Introduction

The accentuation patterns of Japanese dialects have been among the most vigorously studied areas in Japanese linguistics. This can be attributed to the long history of auditory transcriptions of the pitch patterns of words in various Japanese dialects (e.g. Ikeda 1951; Hirayama 1951, 1957; Kibe 2000). Much acoustic work has also been done on Standard Japanese (e.g. Fujisaki & Hirose 1984, Poser 1984, Kubozono 1993, Pierrehumbert & Beckman 1988), and some on varieties of Japanese (Ishihara 2004a on Kagoshima Japanese, Kori 1987 on Osaka Japanese, Matsuura 2008 on Nagasaki Japanese, and Nagano-Madsen 2004 on Kochi Japanese). However, knowledge of the linguistic-phonetic properties of Japanese dialects is still lacking. Linguistic-phonetic properties have both language-internal and cross-linguistic aspects. With regard to the former, linguistic-phonetic properties characterise ‘any and all distinctions which may ever be manipulated systematically in a particular language’ while for the latter, they comprise ‘[a]ny phonetic property . . . in terms of which utterances in Language X are systematically different from utterances in Language Y’ (Anderson 1978: 133–134).

In many Japanese descriptive works, H and L have been traditionally used to represent pitch (Hirayama 1951, 1957, 1960; Kibe 2000; the National Language Research Institute 1959, 1967–1975; the Society for the Study of Kyushu Dialects 1969). The aims of these works vary: to compare the accentual differences across regions; to compile an accent dictionary; to conduct a phonological analysis; to reconstruct the accentuation of proto-Japanese, etc. While phonetic representation using this H–L dichotomy may be suitable for the above purposes, it is unlikely that it is detailed enough to characterise differences in the acoustic realisation of a tonal feature across dialects. The use of H and L to represent accentual pitch is not suitable for linguistic-tonetic analysis, not only because it is not detailed enough, but also because it fails to express linguistically important between-speaker differences and it is not quantifiable. For linguistic-phonetic comparisons, therefore, representations are needed that can provide enough phonetic detail; that are able to capture the invariant properties of the various acoustic realisations of given linguistic information; and that can express linguistically important between-speaker differences (Rose 1993: 190). That is, linguistic-phonetic representations need to be able to show in detail to what quantifiable extent the productions of different speakers are linguistically the same.

Osaka Japanese (OJ) and Kagoshima Japanese (KJ), which typologically belong to different accent systems (Hirayama 1960, 1967; Uwano 1999), nevertheless share some pitch patterns. However, it can be inferred from previous phonetic studies on OJ and KJ tone patterns (Ikeda 1951, Ishihara 2004a, Kori 1987) that there appear to be some f_0 realisation differences in a number of pitch patterns between OJ and KJ. However, we do not know exactly how they are acoustically different.

Therefore, focusing on the shared LH, LHL, LLH and LLLH pitch patterns, the first aim of this paper is to specify the linguistic-phonetic properties of OJ and KJ tonalities in such a way as to permit inter-dialectal comparison. The second aim of this paper is to explore implications for surface tonal representations of KJ on the basis of observed linguistic-tonetic differences between OJ and KJ. This will be done in the framework of Autosegmental-Metrical (AM) theory (Pierrehumbert & Beckman 1988, Pierrehumbert & Hirschberg 1990).

In this study, linguistic-tonetic representations of OJ and KJ tonalities are derived from normalised acoustic representations for the LH, LHL, LLH and LLLH pitch patterns (see Section 3 for detailed explanation of normalisation). Linguistic-phonetic description should then allow us to identify the phonetic features which specify the sounds within a given language or variety, and those features which underlie sound contrasts between varieties (Ladefoged & Maddieson 1986; Rose 1987, 1991).

Different levels of representation are relevant to the current research. This is because our conventional model of language presupposes that an utterance, which can be acoustically represented (e.g. by its f_0 contour), is a realisation of its phonological form (whether as phonemes, systematic phonemes, morpho-phonemes or autosegmental H and L tones). The two research aims of this study explained above can be rephrased in terms of the different levels of representation involved. That is, for the first aim, we will show how one can derive linguistic-tonetic representations from speakers' utterances (acoustic representation). For the second aim, we will investigate how two levels of representation – namely linguistic-tonetic and phonological representations – are mapped onto each other.

The varieties referred to as OJ and KJ in this study are spoken in Osaka City and Kagoshima City, respectively (Figure 1), and all speech samples for this study were recorded in these cities.

1.1 Accentuation of Osaka and Kagoshima Japanese

Tables 1 and 2 contain some words exemplifying the differences in accentual contrast between OJ and KJ. The pitch pattern of words in OJ is determined: (i) by the presence or absence of a lexical accent; (ii) if a lexical accent is present, by its position, which dictates where the accentual pitch fall starts; and (iii) by whether the word is 'low-pitch beginning' or



Figure 1 Map of Japan showing location of Osaka and Kagoshima.

Table 1 Words exemplifying OJ accentual contrast.

Syllable number	High/Low	Accented	Unaccented
2	low-beginning	saru* [LHL] 'monkey'	hana [LH] 'nose'
2	high-beginning	ka*mi [HL] 'paper'	niwa [HH] 'garden'
3	low-beginning	kabu*to [LHL] 'helmet'	suzume [LLH] 'sparrow'
3	high-beginning	oto*ko [HHL] 'man'	sakura [HHH] 'cherry blossom'
4	low-beginning	nokogi*ri [LLHL] 'saw'	tsukemono [LLLH] 'pickles'
4	high-beginning	kamina*ri [HHHL] 'thunder'	niwatori [HHHH] 'chicken'

* = location of an accent.

Table 2 Words exemplifying KJ accentual contrast.

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

‘high-pitch beginning’ (Shibatani 1990, Sugito 1995). Those words which have a lexical accent – its phonetic realisation is a pitch fall – are called ‘accented’ words, and those which do not are called ‘unaccented’ words. So, for example, in Table 1, both *kabuto* [LHL] ‘helmet’ and *otoko* [HHL] ‘man’ are ‘accented’ words with a lexical accent on the second syllable. The former is a ‘low-pitch beginning’ word and the latter is a ‘high-pitch beginning’ word. On the other hand, *suzume* [LLH] ‘sparrow’ and *sakura* [HHH] ‘cherry blossom’ are both

‘unaccented’ words, the former belonging to the ‘low-pitch beginning’ group and the latter to the ‘high-pitch beginning’ group. A pitch rise may be observed between the penultimate and final syllable in unaccented ‘low-pitch beginning’ words, as can be seen in *suzume* [LLH] ‘sparrow’ (Sugito 1997).

As can be seen in Table 2, KJ is much simpler, exhibiting only a two-way accentual contrast: L^nHL vs. L^nH where $n \geq 0$ (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 (i.e. *hana* [HL] ‘nose’, *sakura* [LHL] ‘cherry blossom’ and *kagaribi* [LLHL] ‘watch fire’). 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 (i.e. *hana* [LH] ‘flower’, *usagi* [LLH] ‘rabbit’ and *kakimono* [LLLH] ‘document’). The L^nHL type is referred to as Type A and the L^nH type as Type B in this paper (after Hirayama 1960). The main contrast between KJ Type A and Type B is thus whether or not there is a fall in pitch at the end, although unlike Standard Japanese there is not a lexical contrast in terms of the location of an accent in KJ.

The present study used four target pitch patterns common to both dialects: LH, LLH, LLLH and LHL. The first three pitch patterns represent unaccented low-pitch beginning words in OJ, and Type B words in KJ. The LHL pitch pattern is an accented low-pitch beginning word with an accent on the second syllable in OJ, and a Type A word in KJ.

1.2 Accent typology of Osaka and Kagoshima Japanese

It can be seen from the description given in the previous section that OJ and KJ belong to different accent systems. These are called the KYOTO–OSAKA system and the TWO-PATTERN system, respectively (Hirayama 1957, Shibatani 1990). As is also shown in that section, words can be classified into two groups in the Kyoto–Osaka system, unaccented and accented. If a word is accented, the location of an accent is lexically contrastive as in Standard Japanese (SJ). However, unlike SJ, words also contrast by virtue of starting with a high pitch or a low pitch. In the two-pattern system, all words exhibit only two-way accentual oppositions, and unlike SJ and OJ, the information as to where a pitch falls is not relevant to determining the meaning of a word.

In the spectrum of accentual complexity, the two-pattern system (i.e. KJ) has one of the simplest accentual systems and the Kyoto–Osaka system (i.e. OJ) has one of the most complex systems among the Japanese dialects.¹ This difference in accentual complexity between OJ and KJ was made apparent in Section 1.1. Note also that some scholars (i.e. Gussenhoven 2004, Yip 2002) regard Japanese as a tone language with impoverished paradigmatic and syntagmatic tonal contrast. As such, no absolute division exists between accent languages and tone languages (Yip 2002: 257).

1.3 Previous studies

There have been a large number of studies on the accentuation of OJ and KJ. The accentuation of both OJ and KJ words and phrases has been phonologically analysed by many scholars within various phonological frameworks (McCawley 1968, 1970, 1977; Haraguchi 1977; Shibatani 1979). Recently, experimental studies have been undertaken to investigate the realisation of word accents in phonologically large units such as phrases and sentences for OJ (Kori 1987) and KJ (Kubozono & Matsui 1996, Ishihara 2004a). Seven volumes of ‘Study of Japanese prosody and speech sounds’ compiled by Sugito (1994–1999) contain a series of detailed studies on the acoustics, phonetics, physiology and other features of OJ accentuation. Kibe’s (2000) descriptive study based on her own fieldwork and historical

¹ In this context, accentual complexity means the number of underlying accentual contrasts for a word of a given number of syllables (or moras). OJ has more contrastive accentual types than KJ for words consisting of three light syllables, for example.

records is a comprehensive synchronic and diachronic study of two-pattern Japanese dialects of Kyushu, the island on which Kagoshima City is located.

With regard to the target pitch patterns of the present study (LH, LHL, LLH and LLLH), some scholars have remarked on the different realisations between OJ and KJ. It has been reported that the low-pitch beginning words of OJ have a continuous pitch rise (Ikeda 1951) and are ‘acoustically characterised by an over-all slow [f_0] rise’ (Kori 1987: 44), while Ishihara (2004a) demonstrated that in KJ, f_0 gradually falls before the rise in the penultimate syllable of a Type A word or in the final syllable of a Type B word. Judging from these previous studies, there appear to be some acoustic realisation differences in the LH, LHL, LLH and LLLH pitch patterns between OJ and KJ. However, we do not know exactly how they are acoustically different due to the lack of linguistic-phonetic description of Japanese dialects. It is one of the aims of this paper to rectify this lack of knowledge.

1.4 Autosegmental-Metrical theory

Pierrehumbert & Beckman (1988) present evidence that the linking and spreading of the tones of the accent to other syllables – as is assumed in much autosegmental work on Japanese (Haraguchi 1977) – is not the best account for surface tone patterns in SJ. Instead, they argue that the surface f_0 contour is produced by phonetic interpolation between tone targets, including not only the lexically stipulated accent but also tones associated with higher levels of prosodic structure. With this idea, they introduced a new model for describing Japanese pitch accent by using a few tones per phrase, with interpolation between them. The approach which was adopted by Pierrehumbert & Beckman (1988) in order to analyse Japanese pitch accent (based on Liberman 1975, Bruce 1977, and Pierrehumbert 1980) is now called the Autosegmental-Metrical (AM) theory (Ladd 1996).

In AM theory, all relevant pitch is represented by a single H or L autosegment or by a combination of the two. However, it is not clear what H and L tones actually mean (i.e. pitch target? surface tone? f_0 target? etc). Nor is it made clear, given a specific f_0 contour, on what basis H and L tones are actually identified. Bruce (1977: 131–143) initially identified intonational tones on the basis of the turning points in the f_0 contour. Local maxima, then, correspond to high tones and local minima to low tones. In AM theory, however, tones do not always correspond to turning points, and turning points do not always reflect the surface phonetic realisation of tones (Pierrehumbert 1980). In terms of their surface phonetic realisation, phonological ‘High’ and ‘Low’ tones are not clearly defined in any work relating to the development of the AM theory (Ladd 1996: 104). As an answer to this problem, one might say that the high–low representation can be phonologically characterised in the ‘relative’ Jakobsonian sense (Anderson 1985: 116–139). That is, the first syllable, for example, is represented as L because it is relatively lower in f_0 than the maximum f_0 that is associated with the second syllable. The only condition is whether it is relatively high or relatively low, and the precise f_0 value does not matter.

Some implications that follow from observed differences between OJ and KJ will be explored with respect to the tonal representation of KJ using AM theory (Pierrehumbert & Beckman 1988, Pierrehumbert & Hirschberg 1990). Specifically, some indeterminacies with respect to tone assignment which result from the lack of acoustic specification of a target tone in AM theory will be pointed out while discussing the tonal representation of KJ.

2 Data and method

The data for analysis are taken from Sugito & Kakehi (2005), a DVD database of recordings of Japanese dialects collected from 1989 to 1993 and including a large number of citation forms of words and utterances, together with syllabic sounds and numbers. From the 86 OJ

Table 3 Corpus.

	OJ				KJ			
LH	kata 'shoulder'	matsu 'wait'	umi 'sea'	koko 'here'	iro 'colour'	kame 'turtle'	hato 'pigeon'	deru 'exit'
LHL	nezumi 'mouse'	suzume 'sparrow'	mukashi 'old'	–	kuruma 'car'	musume 'daughter'	asobu 'play'	–
LLH	ichigo 'strawberry'	kujira 'whale'	katana 'sword'	–	nigeru 'escape'	kagami 'mirror'	oyogu 'swim'	–
LLLH	imasara 'at this time'	karakuchi 'sharp'	koroshiya 'killer'	–	kudamono 'fruit'	yorokobu 'glad'	furiyamu 'stop raining'	–

and 79 KJ native speakers of different generations recorded, the speech samples of 12 OJ speakers (six males, six females), and 14 KJ speakers (seven males, seven females) are used for this study. These speakers belong to the three generations of Old (aged 60 years and over), Middle (aged between 59 and 40 years), and Young (aged between 39 and 20 years). This particular grouping was selected because an initial auditory-based analysis identified a discrepancy between these three generations and the generations younger than 20 years old (Sugito & Kakehi database contains recordings from primary and secondary students). This discrepancy is presumably due to the susceptibility of primary and secondary students to the accent of SJ. As far as the three generations aged 20 or older are concerned, pitch patterns are fairly consistent across the speakers, maintaining the pitch contrast reported in previous studies and accent dictionaries. In particular, the speakers selected for this study behave in the same way with respect to the pitch patterns of the target words. The average age of the 12 OJ speakers is 57 years ($sd = 18.6$) for males and 46 years ($sd = 14.8$) for females. That of the 14 KJ speakers is 41 years ($sd = 24.8$) for males and 47 years ($sd = 16.7$) for females.

Table 3 contains the target words selected for describing the f_0 realisation of the four target pitch patterns. A standard romanisation system for SJ is used to represent OJ and KJ target words, as there are no important segmental differences. The syllable structure of all target words is (C)V. The words were selected to control as much as possible for intrinsic vocalic and consonantal effects on f_0 (Peterson & Barney 1952, Lehiste & Peterson 1961). Thus, various voiced and voiceless consonants and vowels are included in the target words in order to control for their known intrinsic effects (Peterson & Barney 1952, Lehiste & Peterson 1961). In Section 5.2 below, OJ and KJ will be compared in terms of the f_0 realisation of LH, LHL, LLH and LLLH pitch patterns, particularly focusing on the initial syllable. As far as the vowel quality of the initial syllable is concerned, as can be seen from Table 3, OJ and KJ share the same distribution of the vowels (two /i/, four /u/, one /e/, two /o/ and four /a/ vowels), and thus the intrinsic effect arising from the vowel types is well controlled in this study.

All words were read once by each speaker, and recorded on professional digital equipment (16 kHz sampling rate, 16 bit quantisation).

After being segmentally annotated, all target tokens were analysed using the *get_f0* function of the ‘Snack Sound Toolkit’ (Sjölander 2006). The *get_f0* function is an implementation of ‘the robust algorithm for pitch tracking (RAPT)’ proposed by Talkin (1995), which is widely used for estimating f_0 . On the basis of the segmental annotation, f_0 was sampled at the onset and every 20% point for each vowel. RAPT is very accurate in estimating f_0 , yet will never, of course, be totally free from errors (i.e. halving and doubling effects). The f_0 values of each token were therefore visually inspected, and those tokens which contained obvious f_0 tracking mistakes were re-analysed by setting different tracking thresholds. Identification of vowel onset or offset adjacent to an obstruent or a nasal is fairly straightforward from the audio speech waveforms and spectrograms; however, liquids and

glides make the identification of the onset and offset of a vowel more difficult. In these cases, segmentation was done by visually identifying discontinuities in the acoustic parameters, such as waveform shape, F-pattern, and amplitude. Phonation offset in many falling-pitch tokens (e.g. LHL) was either abrupt or creaky. This showed as glottal pulse striations with jitter at the end of the vocalic F-pattern, causing the algorithm to extract f_0 inaccurately. Thus, this portion was excluded from analysis.

3 Normalisation

The aim of this study is to give quantified descriptions for both dialects of the linguistic-acoustic properties of the four pitch patterns in question. However, extracting the specific phonetic properties of a given language or variety is not an easy task, as different speakers have different acoustic outputs for what is perceived to be the same linguistic information. This is 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: 343). Therefore, the individual content (in other words, between-speaker differences) arising from individual anatomical differences needs to be factored out as much as possible by a process of normalisation to extract the linguistic-acoustic content of the target pitch patterns.

However, not all between-speaker differences are associated with different vocal tract anatomy: speakers can differ LINGUISTICALLY as well (Rose 2002: 175–194). If such linguistic between-speaker differences occurred in the topology of a given language, it would be very difficult to spot them without normalisation, since they would be mixed up with the non-linguistic anatomical differences. Thus, normalisation enables us to specify an invariant acoustic property from different acoustic outputs and present it as a linguistic-phonetic representation while quantifying the linguistically important between-speaker differences. This representation, then, can be compared across languages and dialects. Particularly for cross-language/dialect comparisons, ‘[n]ormalisation is the only way, other than (the unlikely) recourse to bilingual speakers, of testing transcriptionally based hypotheses on the nature of linguistic-phonetic tonal variation’ (Rose 1987: 344).

It has been mooted in Section 1 above that the use of H and L as a phonetic representation is a poor way of specifying the tonetic properties of a given language/dialect, as this kind of broad representation fails to identify the linguistic-tonetic differences between OJ and KJ. One might consider that the mean f_0 curve obtained by simply averaging the raw data can be used as a linguistic-tonetic representation. Although such a representation might be able to provide a general idea of how f_0 values change in a particular word, it is a linguistically poor representation, not only because it does not express the invariant aspect of given linguistic information (its units would still be meaningless Hz), but also because it fails to express the linguistically important between-speaker differences.

The z -score normalisation technique, the performance of which has been empirically attested for f_0 (Rose 1991, Zhu 1994, Ishihara 2004a), is used in this study. The z -score normalisation procedure is: $f_{0\text{norm}} = (f_{0i} - m) / s$, where f_{0i} is a sampling point, and m and s are the arithmetic mean and standard deviation of f_{0i} ($i = 1, 2, \dots, n$), respectively. The z -score normalisation parameters (m and s) were obtained for each speaker from a set of mono-to-four-syllabic words. This set, which includes the target words listed in Table 3, is given in the appendix. The procedure which was used to sample f_0 values from these words is the same as that explained in Section 2 above. In this particular normalisation procedure, each f_0 observation is expressed as so many standard deviations above and below a speaker’s overall mean f_0 ($= m$). Thus, standard deviation (sd) is used as the unit for normalised f_0 in this study.

Table 4 Z-score normalisation parameters (Hz).

OJ				KJ			
Sex	Speaker	Mean	sd	Sex	Speaker	Mean	sd
Male	OJM1	149.6	24.7	Male	KJM1	150.2	28.1
	OJM2	162.1	30.9		KJM2	111.7	16.3
	OJM3	115.5	26.5		KJM3	112.8	21.8
	OJM4	155.7	29.5		KJM4	124.6	13.8
	OJM5	86.8	10.2		KJM5	146.4	13.5
	OJM6	112.5	16.1		KJM6	134.7	22.6
Female	–	–	–	Female	KJM7	135.3	11.4
	OJF1	148.7	19.7		KJF1	182.8	24.3
	OJF2	211.1	45.4		KJF2	194.6	23.0
	OJF3	209.9	40.6		KJF3	192.3	22.6
	OJF4	199.2	32.5		KJF4	197.2	20.2
	OJF5	222.1	23.8		KJF5	205.7	43.6
	OJF6	216.8	34.9		KJF6	217.9	22.8
	–	–	–		KJF7	216.1	24.9

The z-score normalisation parameters are given for each speaker in Table 4. As can be seen in this table, there are considerable differences in these parameter values across the speakers. For example, the range of the mean values is about 75 Hz (86.8 Hz–162.1 Hz) for the OJ male speakers while it is about 39 Hz (111.7 Hz–150.2 Hz) for the KJ male speakers (i.e. about half that of the OJ male speakers). As for the female speakers, the range of the mean values for the OJ females is approximately 73 Hz (148.7 Hz–222.1 Hz) whereas that of the KJ females is about 35 Hz (182.8 Hz–217.9 Hz), again about half that of the OJ females. As explained above, these differences are treated as being due to individual anatomical differences, and as factors to be removed in order to represent the invariant aspects of KJ and OJ tonalities by means of normalisation. This assumes of course that there are no differences between the dialects in f0 range.

4 Sample size and margin of error

The sample size of the current study (26 = 12 KJ speakers + 14 OK speakers) satisfies the minimum of three people for each sex for quantified phonetic work suggested by Ladefoged (1997: 140). However, the number of speakers needed depends on the variance of the data, the margin of error and the desired confidence, thus it needs to be statistically decided. Since we know the sample size of this study ($n = 26$) and the mean of the speakers' standard deviations ($s = 24.8$, calculated from Table 4), the margin of error (E) of 9.5 Hz was estimated with 95% confidence ($z = 1.96$).² An error of 9.5 Hz is equivalent to $sd = 0.38$. Hence, when we compare the f0 realisation differences between OJ and KJ, any differences which fall within this margin of error ($sd = 0.38$) will not be considered.

² $n = (z * s / E)^2$ thus, $E = \sqrt{(z * s)^2 / n} = \sqrt{(1.96 * 24.8)^2 / 26}$ (Devore & Peck 1997: 281).

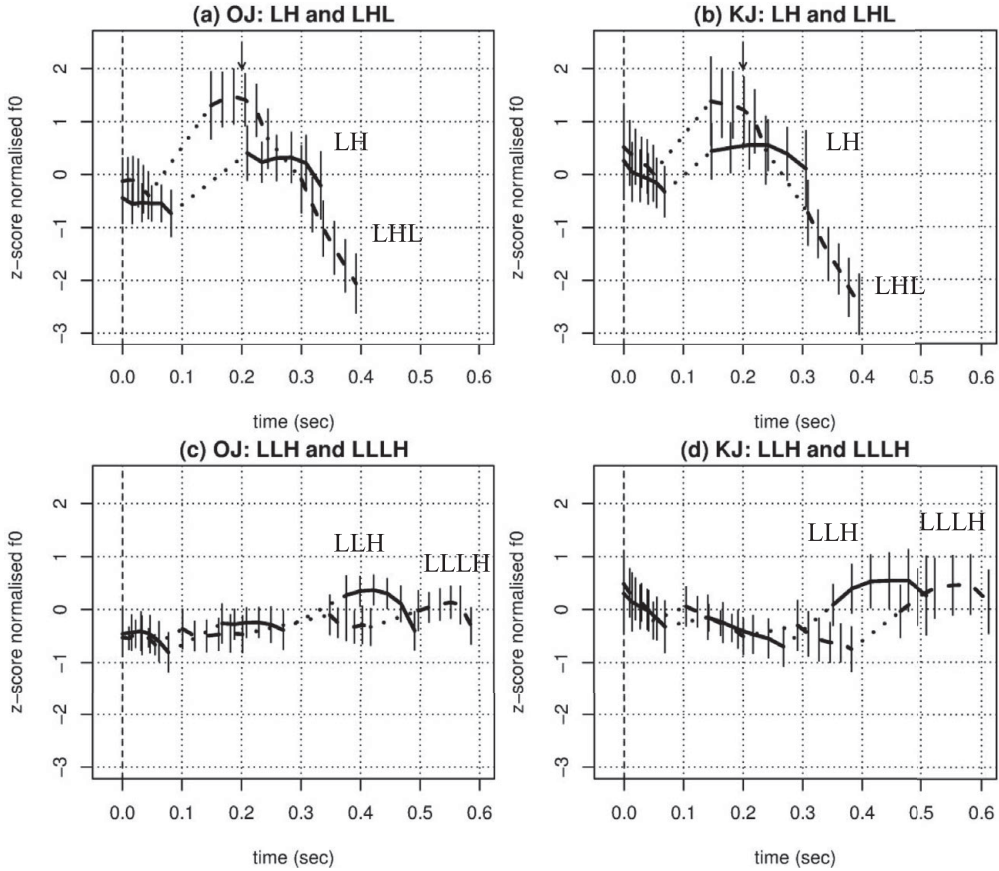


Figure 2 Mean normalised f_0 curves for all target pitch patterns plotted against mean absolute duration. Panels (a) and (c) show OJ; panels (b) and (d) show KJ. Vertical lines indicate one sd above and below the mean. Solid lines are for LH and LLH; dashed lines for LHL and LLLH. Dotted lines show putative intervocalic consonantal f_0 . Arrows indicate the f_0 corresponding to the high-pitch syllable of the LHL pitch pattern. Note that it is realised higher in normalised f_0 than the high-pitch syllable of the LH pitch pattern.

5 Normalisation results

Figure 2 contains the mean of all speakers' normalised f_0 curves of the four pitch patterns, separately plotted for OJ (Figures 2a and 2c) and KJ (Figures 2b and 2d). Normalised f_0 is plotted against mean absolute duration in order to preserve the f_0 derivative. The range of f_0 values is shown by vertical bars indicating one standard deviation above and below the mean. The temporal axis of Figure 2 indicates the average duration of words with each pitch pattern. This was obtained by summing the mean segment durations in the relevant target words across all speakers. For example, for the LH pitch pattern of OJ (see Figure 2a), the mean duration of the pitch pattern of 0.33 s is the sum of the mean duration of the first vowel = 0.09 s, that of the intervocalic consonant = 0.12 s, and that of the second vowel = 0.12 s.

The normalised f_0 contours presented in Figure 2 can be considered as a type of linguistic-phonetic representation of KJ and OJ accentual patterns. These linguistic-phonetic representations can now be used for both intra-dialectal comparisons – for example, to

investigate the acoustic realisation differences between different accentual types within a dialect – and for inter-dialectal comparisons – for example, to investigate the realisation differences in the LH pitch pattern across different dialects. Both types of comparison will now be exemplified.

5.1 Within-dialect comparison

Figure 2 shows that the mean f_0 curves for the target pitch patterns lie between $sd = c. -2.5$ and $sd = c. 1.5$. Compared to other sampling points, a relatively large amount of between-speaker variation ($sd = c. 0.7-0.8$) is found at the onset and offset points of each vowel for some syllables. This can be seen in the vertical standard deviation bars in Figure 2. This is due to the perturbatory effect of the syllable-initial consonant, and the final creaky phonation which was referred to at the end of Section 2 above. Otherwise, between-speaker variation is within the range of $sd = 0.4-0.6$, which is comparable with previous studies (see Ishihara 2004a for KJ, Rose 1993 and Zhu 1994 for Shanghai).

In both OJ and KJ in Figure 2, it can be clearly seen that the high-pitch syllable of the LHL pattern, indicated by the arrows (panels a and b), is realised significantly higher in f_0 than that of the LH pattern (unpaired two-tailed t -tests, $p < .0001$ at the highest point). This is a well-reported phenomenon associated with accented words in OJ (Kori 1987), and Type A words in KJ (Kubozono & Matsui 1996, Ishihara 2004a). The same phenomenon has been found in Standard Japanese, where it is called ‘accentually induced f_0 boost’ (Kubozono 1993: 85). The implication of this is that pitch-accent is not merely tonal, but involves an additional parameter of prominence, as implied by the term ‘accent’.

5.2 Between-dialect comparisons

In Figure 3, the mean f_0 curves are plotted separately according to the pitch patterns across the dialects (sd bars are omitted for clarity). The x-axis of Figure 3 is equalised duration (%).

A consistent difference between OJ and KJ is observed in the normalised f_0 values associated with the L on the first syllable, indicated by arrows in Figure 3. The f_0 values of the first syllable lie between $sd = -1$ and $sd = 0$ for OJ, while they are around $sd = 0$ for KJ. This observation for KJ is consistent with Ishihara’s (2004a) finding that f_0 declines around $sd = 0$ in KJ initial L syllables. Unpaired two-tailed t -tests indicate that the first syllable has significantly higher f_0 values ($p < .0001$ at all sampling points) in KJ than OJ for all target pitch patterns. Note too that the difference between KJ and OJ at the first syllable is larger than the margin of error ($sd = 0.38$) estimated in Section 4 above at all sampling points. In the case of the LLLH pitch pattern (Figure 3d), for example, the difference between OJ and KJ in f_0 value at the onset point is $sd = 1.1$ (OJ: $sd = -0.52$; KJ: $sd = 0.58$).

Due to this difference between OJ and KJ, the LLH and LLLH pitch patterns exhibit significantly different f_0 contours between the two dialects in that f_0 gradually rises from the first to the final syllable in OJ, while in KJ it gradually declines from the first to the penultimate syllable before rising in the final syllable. Due to the gradually falling contour observed in the L syllables of LLH and LLLH in KJ, the average maximum normalised value of the KJ first L syllable ($sd = 0.45$ for LLH and $sd = 0.53$ for LLLH) is significantly higher (unpaired two-tailed t -tests, $p < .0001$) than the average minimum normalised value of the penultimate L syllable ($sd = -0.72$ for LLH and $sd = -0.80$ for LLLH). Whereas for OJ, as a result of the gradually ascending contour, the average minimum normalised value of the first syllable ($sd = -0.85$ for LLH and $sd = -0.90$ for LLLH) is significantly lower ($p < .0001$) than the average maximum normalised value of the penultimate L syllable ($sd = -0.10$ for LLH and $sd = 0.02$ for LLLH).

The normalised f_0 curves of OJ shown in Figure 3 agree with Ikeda’s (1951) auditory-based transcription of a rising pitch contour for low-pitch beginning words. However, the f_0

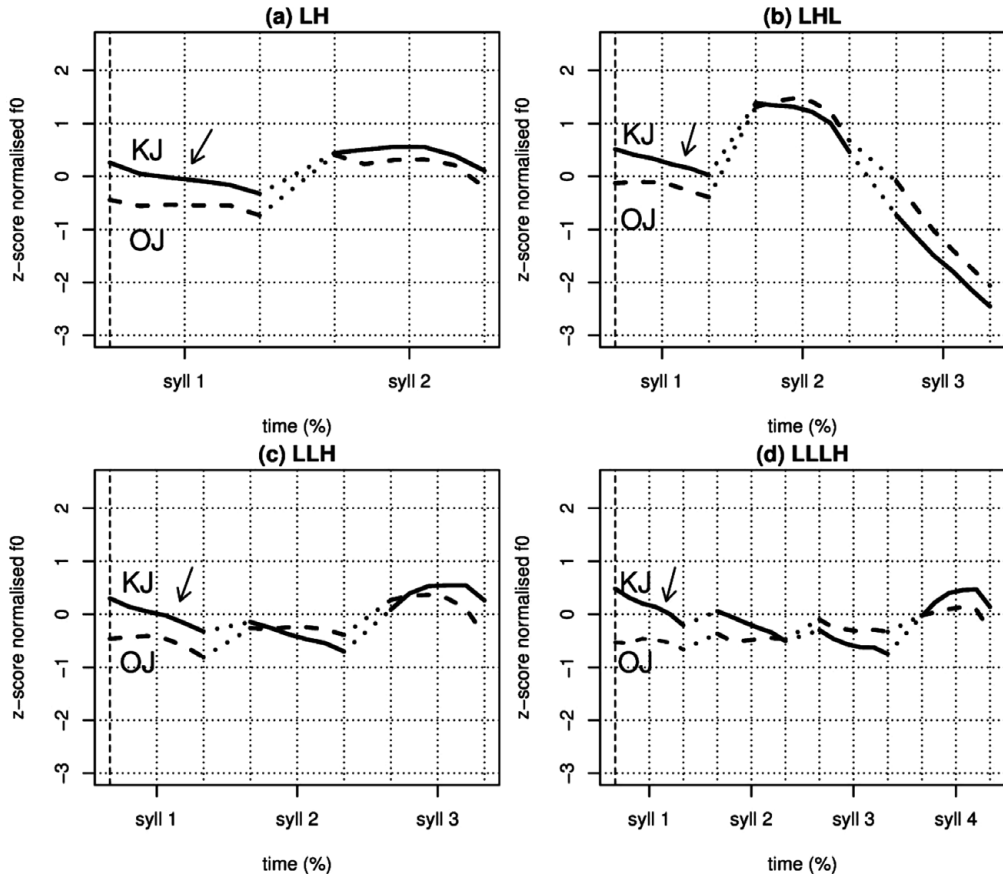


Figure 3 Between-dialect comparison of mean normalised f0 curves of the LH (panel (a)), LHL (b), LLH (c) and LLLH (d) patterns. Dashed lines = OJ; solid lines = KJ; syll = syllable. Arrows point to the f0 on the first syllable of KJ. Dotted lines show putative intervocalic consonantal f0.

curves given in Figure 3 provide more detailed information about the pitch contour which may not be captured by transcription.

6 Discussion

Having identified differences between OJ and KJ in the acoustic realisation of their LH, LHL, LLH and LLLH pitch patterns, this section investigates how linguistic-tonetic and phonological representations are mapped onto each other. In particular, we explore some implications of the identified differences between these two dialects in their normalised acoustics with respect to the surface tonal representation of KJ using AM theory.

It was pointed out in Section 1.4 above that it is not clear in the AM theory as to what H and L tones actually mean. For example, does L tone mean that ‘f0 or pitch stays low until countermanded at some predetermined point’ or ‘moves to a low value’, or something else? Relating to this question, Pierrehumbert & Beckman (1988: 175) explain that ‘some of the tones have a specified duration, whereas all others are treated as points’, but they

are not explicit about which tones have a specified duration or which are treated as a point. Nevertheless, once target tones are determined, they are linearly interpolated as a subsequent process (Pierrehumbert & Beckman 1988: 175).

Tonal representation in OJ is addressed first.

6.1 Osaka Japanese tonal representation

As introduced in Section 1.1 above, whether an OJ word starts with a high or a low pitch is determined lexically. The AM tonal representations of OJ low-pitch beginning accented (i.e. LHL) and unaccented (i.e. LLLH) words are shown in Figure 4. Both lexical and surface representations (from Pierrehumbert & Beckman 1988: 229) are shown together with the linguistic-tonetic representation – the corresponding mean *z*-score normalised *f*₀ contours derived in this study. The lexical representations are given in the first row, and surface representations are given in the second. Figure 4 thus shows how the acoustic-phonetic reality relates to the phonological constructs. Note that the representations of an accented word (LHL) are given in the panels on the left of Figure 4 and those of an unaccented word (LLLH) are in the panels on the right.

6.1.1 LHL accented word

The different levels of the representations for the LHL accented word are given in the left hand side panels of Figure 4. In the lexical representation of the accented word (Figure 4a), the pitch accent is modelled by a sequence of HL tones linked to the accented syllable. The L tone, which models the low pitch at the beginning of the word, is a property of the word and is therefore linked to it, rather than to the word's syllables (Pierrehumbert & Beckman 1988: 229). At the surface, this L is linked to the word-initial syllable.

Although the words in Figure 4 were uttered as single words in a citation manner, they are not free from the tones associated with the prosodic levels higher than words. Thus, it is necessary to consider that these words are also associated with some tones which are linked to higher prosodic levels (e.g. utterance). Since all target words were uttered in a citation manner as a statement, it is reasonable to posit a final boundary L tone linked to the final syllable of the words (see Figures 4c and 4d). This final boundary L tone is possibly a property of utterance causing a final *f*₀ lowering (Pierrehumbert & Beckman 1988: 72–75).

It is now possible to specify how the phonological tones might map onto the normalised acoustics. As far as the L tones are concerned, both word-initial and pitch accent tones are mapped onto a normalised value of *sd* = *c.* 0, but the word-initial L is realised over the whole of the first syllable rhyme, whereas the pitch accent L is a point value at the beginning of the final syllable rhyme. The utterance-final L can also be considered as a point value, but associated with a value of *sd* = *c.* –2 (or 2 standard deviations below the previous L tone) at the end of the final syllable rhyme. The H of the pitch accent tone is realised at *sd* = *c.* 1.5 and is prolonged over the second syllable rhyme. The perturbations at the beginning and end of the rhymes can be modelled with a smoothing function.

6.1.2 LLLH unaccented word

The different levels of the representations for the LLLH unaccented word are given in the panels on the right of Figure 4. Since the LLLH word does not have an accent, its lexical representation (Figure 4b) consists of two word-level tones: an initial L, which accounts for the low-pitch onset as in the accented word, and a word-final H. In the surface representation, these tones are linked to the first and last syllables of the word, the last syllable being also linked to an utterance-final L, as in the accented word. As far as the mapping of these tones onto the normalised acoustics is concerned, the initial L is realised with a normalised *f*₀ value of *sd* = *c.* –0.5, prolonged over the first syllable rhyme. The lower realisation of the word-initial L in the unaccented word suggests that the higher value for word-initial L

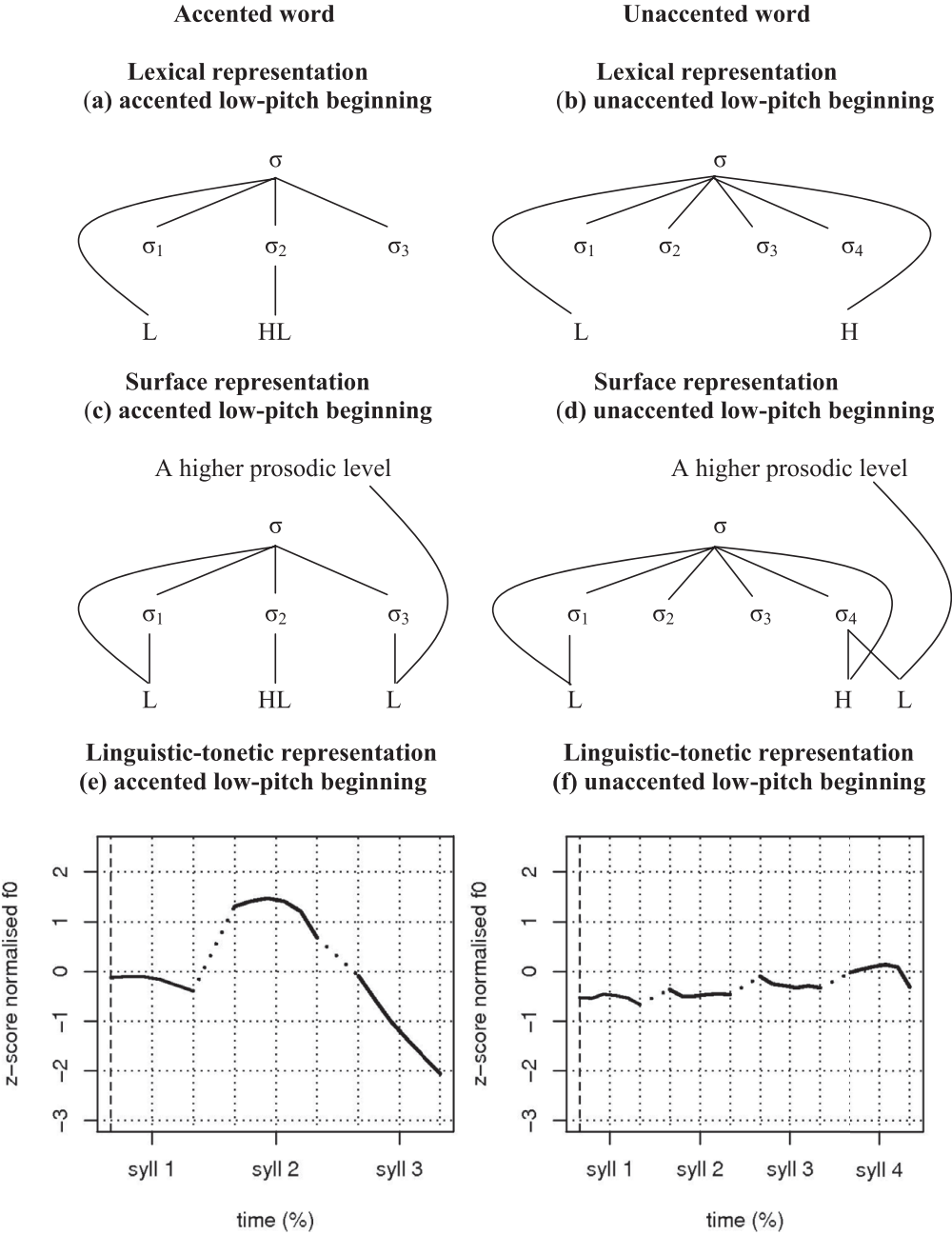


Figure 4 Tonal representations in OJ low-pitch beginning words. Right panels = unaccented words, left panels = accented words (accent is on the second syllable). Top row = lexical AM representation; middle row = AM surface representation; bottom row = linguistic tonetic representation (mean z-score normalised f0 contours).

($sd = c. 0$) in the accented word is a result of assimilation to the following H, and this should therefore be accounted for by rule.

The word-final H is realised with a normalised f_0 value of $sd = c. 0$. The big difference in the realisations of the H in the accented and unaccented words can be seen as resulting from either the effect of accentual boost on H when it occurs as part of the pitch accent, or the lowering effect of the utterance-final L on the unaccented word-final H, or both.

The gradually rising f_0 after the initial syllable can be seen as an interpolation between the values on the initial and final syllable. The rising f_0 on the final syllable possibly indicates that its word-final H is realised as a point value; the falling perturbation at the end of the rhyme is a typical phonatory offset effect. Although the f_0 on the second and third syllables appears fairly level (and is therefore suggestive of prolonged values), this is probably due to the effect of the initial consonantal perturbations.

6.2 Kagoshima Japanese tonal representation

Having explained above how the phonological tones are mapped onto the normalised acoustics in OJ, we now start looking into KJ. More precisely, what we attempt here is to investigate what sort of phonological tones (or autosegmental tones) are necessary to represent KJ at the surface level based on the linguistic-phonetic representations of KJ. This will be discussed using the Type B LLLH pitch pattern as an example.

As described in Section 5.2 above, one of the significant aspects of KJ is a gradual FALL on the initial L syllables. This gradual fall followed by a rise on the final syllable is a typical f_0 contour for Type B words. For OJ LLLH unaccented words, the gradually RISING f_0 after the initial syllable was modelled by interpolating between the L tone associated with the first syllable and the H tone associated with the final syllable. Similarly, this gradual fall of KJ from the initial syllable to the penultimate syllable can also be seen as an interpolation between the tone associated with the initial syllable and that associated with the penultimate syllable. Likewise, the rise in f_0 between the penultimate and the final syllables can be modelled by interpolating the tones associated with the penultimate and the final syllables, respectively. Since we do not know the values (e.g. L or H) of these three tones relevant to Type B LLLH pitch pattern, we tentatively call these tones T1, T2 and T3, as shown in Figure 5, in which a tentative surface representation (first row) of KJ's LLLH pitch pattern is given together with its normalised acoustic representation (second row).³

Now that we can see how these phonological tones (T1, T2 and T3) are mapped onto the normalised acoustics, we can discuss what tone values (e.g. H or L) are appropriate for these tones. In this paper we focus only on T1 and T2, our main question being: What are the tone values for T1 and T2? As shown in Figure 5, the tone values of T1 and T2 will be discussed by referring to LLLH Type B as an example. The tonal value of T2 will be discussed first.

6.2.1 Kagoshima Japanese tonal representation: T2

As can be seen from Figure 5, T2 is linked to the penultimate syllable, and its normalised value ($sd = c. -0.5$) is the lowest of all the syllables. Although the f_0 on the penultimate syllable continues to fall (indicating that T2 is realised as a point value), this fall may be due to a perturbatory effect of the consonant.

Previously, Ishihara (2000) investigated the f_0 behaviour of the (L)HL'H(L) sequence ($6 \geq n \geq 1$) of KJ adjective + noun phrases (i.e. LHL.LLLLH, where a period stands for a word boundary) and reported that the f_0 minima between the high-pitch syllables – which correspond to T2 – varies as a function of n in an exponential manner. This exponential decay

³ As for OJ in Section 6.1, the final syllable in KJ should also be linked to a final boundary L tone. However, the final boundary L tone is not explicitly included in Figure 5 as we focus here on the initial and the penultimate syllables.

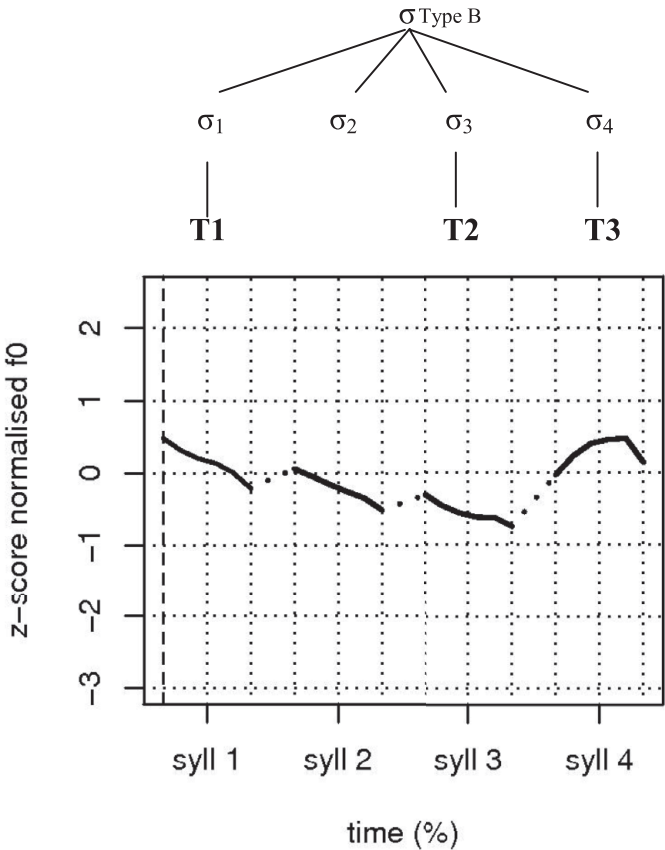


Figure 5 Tonal representations in KJ LLLH Type B word. Top row = surface representation with the tentative tones of T1, T2 & T3; bottom row = linguistic-tonetic representation (mean z-score normalised f_0 contour).

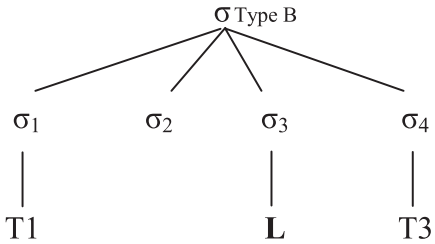


Figure 6 The tone value for T2.

is a typical characteristic of a target tone in a falling contour (Pierrehumbert & Beckman 1988: 35–46). Thus, this exponential f_0 decay indicates that there is a target tone (T2) linked to the penultimate syllable for Type B words. As the syllable that T2 is linked with has the lowest normalised f_0 value out of all syllables, an L tone is the appropriate tone value for T2, as represented in Figure 6.

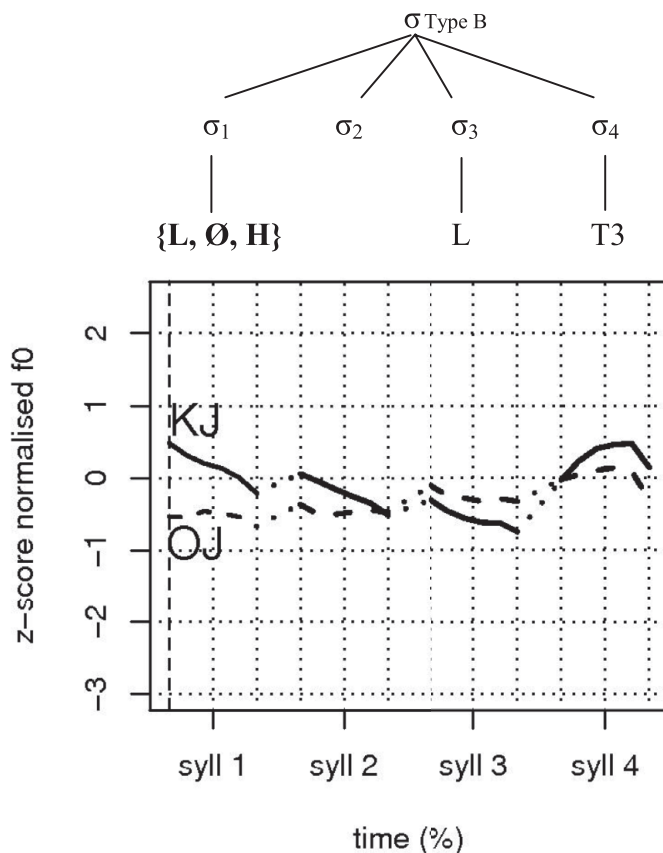


Figure 7 Three possible target tone values {L, Ø, H} for T1 together with the mean z-score normalised f0 contour of the KJ LLLH pitch pattern (solid line). The mean z-score normalised f0 contour of the OJ LLLH (dashed line) is also given for the sake of comparison.

6.2.2 Kagoshima Japanese tonal representation: T1

As we can see in Figure 5, the word-initial T1 is realised with a normalised f0 value of $sd = c. 0$. Its realisation value is constant regardless of the length of the Type B words (see Figures 3c and 3d). Like T2, the f0 on the initial syllable continues to fall (indicating that T1 is realised as a point value); this could also be due to a perturbatory effect of the consonant.

As for the tone value of T1, there appear to be three possibilities (L, Ø and H), as presented in Figure 7. Following the traditional auditory-based pitch transcriptions of KJ words (i.e. LLLLHL and LLLH), one might intuitively posit an L tone for T1. However, a few points must be addressed regarding this issue.

As has been reported, the normalised f0 values of the initial L syllable are around the mean ($sd = 0$) for Type B words. This means that the f0 values of the initial L syllable are around the average f0 of the speakers' f0 distribution. It is possible to interpret the f0 associated with the initial L syllable as the most neutral f0 value that is near the mean of a speaker's f0 distribution. Considering this, it may not be necessary to assign a particular tone for T1 (Ø) as the f0 realisation of the initial L syllable is predictable from the information

regarding a speaker's f_0 distribution.⁴ A similar interpretation can be seen in Dainora (2001) regarding the intonation of English. In reference to an analysis of the intonation of English by Goldsmith (1978), in which the intonation was analysed in terms of H, M and L, Dainora (2001: 36) mentions that 'the M does not represent a tone at all, but, rather, represents the neutral frequency used from the beginning of an utterance to the first pitch accent'. In Goldsmith's analysis, the M tone is always the first tone in the sequence and is not associated with an accented syllable. In the AM theory, as noted earlier, it is not clear on what basis low and high tones are actually identified in the f_0 contour.

As an alternative value for T1, basing judgement solely on the falling f_0 contour that Type B words exhibit on the initial L syllables, it is certainly possible to posit an H tone (particularly with respect to the concept of f_0 maxima and minima in the 'relative' Jakobsonian sense, Anderson 1985: 116–139). However, positing an H tone for T1 is highly counter-intuitive from a perceptual point of view (personal communication, Prof. Haruo Kubozono, National Institute for Japanese Language and Linguistics, Japan) as the initial syllable in question is perceived and transcribed as a low-pitch syllable by many descriptive linguists (Hirayama 1957, 1960; Kibe 2000). At this stage, as far as the f_0 realisation of citation words is concerned, it is unclear what value we should assign to T1 because all candidates given in Figure 7 appear to be plausible in one way or another. This point requires further investigation using units phonologically longer than words.

Nevertheless, regardless of the tone value of T1, the falling f_0 contour of Type B words can be modelled well by an interpolation between T1, which is linked to the initial syllable and of which normalised value is $sd = c. 0$, and the penultimate syllable L tone, which has normalised value of $sd = c. -0.5$.

Thus, again, whichever of the three possibilities presented in Figure 7 represents the tone value of T1, the f_0 realisation difference between OJ and KJ observed in the initial syllable of the pitch patterns concerned can be phonologically or phonetically accounted for. If the tonal value of T1 is either $\emptyset$ or H for KJ, it is straightforward to phonologically understand the f_0 realisation difference between OJ and KJ regarding the initial low-pitch syllables because OJ and KJ assign different tonal values to the initial tone (OJ: L vs. KJ: $\emptyset$ or H). Although it is necessary to look into the f_0 behaviour of KJ in larger phonological units such as sentences and utterances to determine the tone value of T1 (see Ishihara 2004b), it is sensible to consider that if we posit an initial boundary tone for KJ at T1, it should be a non-lexical tone because KJ does not have the same kind of lexical contrast as OJ for the initial boundary tone. Therefore, even if an L tone is proposed for KJ at T1, there is a significant difference between OJ and KJ in that the initial L tone of OJ is lexical while that of KJ is not. It is well known that a tone can be realised differently in f_0 depending on whether or not it is a lexical tone, even if the same value is associated with it (Pierrehumbert & Beckman 1988, Kubozono 1993, Ishihara 2004a). That is, the f_0 realisation difference observed at the initial syllable between OJ and KJ can be either explained by phonologically positing different tones for the first syllable (OJ: L vs. KJ: $\emptyset$ or H) or by phonetically implementing different realisation rules for the initial L tone.

⁴ One of the reviewers pointed out that if $\emptyset$ can be given for T1 since the normalised f_0 is $sd = c. 0$ at the first syllable, the same reasoning can be applied to the final syllable of OJ LLLH pitch pattern of which f_0 is also $sd = c. 0$ (see Figure 4f), and consequently $\emptyset$ can be assigned to the final syllable as well. However, as explained in Section 6.1 above, H tone is assigned to the final syllable in question instead of $\emptyset$ because the f_0 realisation of the final syllable is subject to the f_0 lowering effect of the final boundary L tone and/or the declination effect, thus the H tone was realised lower.

7 Conclusions

At the beginning of this paper, we argued that the previous auditory descriptions of OJ and KJ tonalities using H and L are not adequate for linguistic-tonetic purposes because they are not phonetically detailed enough, they fail to express linguistically important between-speaker differences, and they are not quantifiable. As a result, the previous auditory descriptions of OJ and KJ tonalities fail to capture the linguistic-tonetic differences between the two dialects that have been identified in this study for the LH, LHL, LLH and LLLH pitch patterns.

Besides this auditory-based representation, the current study was concerned with three levels of representation: acoustic, linguistic-tonetic and AM representations. Acoustic representation (i.e. the f_0 contour of an utterance) is the most surface level representation, whereas AM representation is a phonological representation. Linguistic-tonetic representation is situated somewhere in-between. This study showed, first of all, how linguistic-tonetic representations can be obtained from speakers' utterances (which can be acoustically represented). More precisely, linguistic-tonetic representations of OJ and KJ tonalities were derived from normalised acoustic representations for the LH, LHL, LLH and LLLH pitch patterns. These linguistic-tonetic acoustic representations can provide sufficient phonetic details for characterising differences in the acoustic realisation of a tonal feature; they are able to capture the invariant property of the various acoustic realisations of given linguistic information; and they can express linguistically important between-speaker differences. Thus, the linguistic-phonetic representations that were established in this way for OJ and KJ can be used to make acoustic comparisons, not only between these two dialects, but also with other varieties of Japanese or other languages.

By comparing the linguistic-phonetic representations of OJ and KJ for the LH, LHL, LLH and LLLH pitch patterns, this study has first of all demonstrated that these pitch patterns of OJ have significantly lower f_0 realisations than those of KJ at the initial L syllables. More precisely, the LHL pitch pattern starts at $sd = c. 0$ and the LH, LLH and LLLH patterns start at $sd = c. -0.5$ in OJ, while they start higher in KJ by $sd = 0.5-1$. An overall significant difference in contour shape can also be observed between OJ and KJ in that the former exhibits a gradual rise in the LLH and LLLH pitch patterns while the latter shows a gradual fall before a rise between the last two syllables.

Having established the linguistic-tonetic representations of OJ and KJ for the LH, LHL, LLH and LLLH pitch patterns, we have first demonstrated how OJ phonological representations are mapped onto its linguistic-phonetic representations. Then, based on the linguistic-tonetic representations of KJ, we discussed the phonological representation of KJ at the surface level using LLLH Type B words as an example. Through the discussion it was demonstrated that since it has not been clearly defined what H and L tones actually mean in the AM theory, the mapping between a linguistic-tonetic representation and its phonological representation (or AM representation) is not straightforward. In particular, the possibility of $\emptyset$ tone as the initial tone was discussed, focusing on the fact that the normalised f_0 values stay at $sd = c. 0$ in the first syllable in KJ.

Acknowledgments

This paper is a revised and extended version of a paper presented at Interspeech 2008 in Brisbane, Australia. This research was financially supported by the College of Asia and the Pacific, the Australian National University. The author is very grateful to Dr Phil Rose for his valuable comments and extensive discussions of earlier versions of this paper. Thanks are also due to the editors and three anonymous reviewers for their detailed comments. This paper is dedicated to Dr Phil Rose on the occasion of his retirement from the Australian National University.

Appendix. Words used for normalisation parameters

OJ				KJ			
Syllable #	pitch	word	English	Syllable #	pitch	word	English
1	H	ka	mosquito	1	HL	e	painting
1	H	chi	blood	1	HL	chi	blood
1	HL	ha	leaf	1	HL	ha	leaf
1	HL	ke	hair	1	HL	hi	day
1	LH	te	hand	1	H	te	hand
1	LH	me	eye	1	H	ki	tree
2	HH	niwa	garden	1	H	su	nest
2	HH	hana	nose	1	H	ha	teeth
2	HL	kami	paper	2	HL	kaze	wind
2	HL	yama	mountain	2	HL	oto	sound
2	LH	kata	shoulder	2	HL	hana	nose
2	LH	matsu	pine tree	2	HL	neru	sleep
2	LH	koko	here	2	LH	iro	colour
2	LH	naka	inside	2	LH	kami	hair
2	LH	nani	what	2	LH	hato	pigeon
3	LHL	saru	monkey	2	LH	deru	leave
3	LHL	mado	window	3	LHL	kuruma	car
3	HHH	kimono	cloth	3	LHL	musume	daughter
3	HHH	kemuri	smoke	3	LHL	asobu	play
3	HHH	sakura	cherry blossom	3	LHL	owaru	finish
3	HLL	otoko	man	3	LLH	kotoba	words
3	HLL	kokoro	heart	3	LLH	kagami	mirror
3	HLL	namida	tear	3	LLH	oyogu	swim
3	LLH	nezumi	mouse	3	LLH	nigeru	escape
3	LLH	suzume	sparrow	4	LLHL	hajimeru	start
3	LLH	mukashi	old	4	LLHL	nakiyamu	stop crying
3	LHL	ichigo	strawberry	4	LLHL	kasamaru	overlap
3	LHL	kujira	whale	4	LLLH	kudamono	fruit
3	LHL	katana	sword	4	LLLH	yorokobu	happy
3	HHL	atama	head	4	LLLH	irogami	coloured paper
3	HHL	kagami	mirror				
3	HHL	fukuro	bag				
4	LLLH	imasara	at this time				
4	LLLH	karakuchi	sharp				
4	LLLH	koroshiya	killer				

References

- Anderson, Stephen R. 1978. Tone features. In Victoria Fromkin (ed.), *Tone: A linguistic survey*, 133–175. New York: Academic Press.
- Anderson, Stephen R. 1985. *Phonology in the twentieth century: Theories of rules and theories of representations*. Chicago: University of Chicago Press.
- Bruce, Gösta. 1977. *Swedish word accents in sentence perspective*. Lund: Gleerup.
- Dainora, Audra. 2001. *An empirically based probabilistic model of intonation in American English*. Ph.D. dissertation, University of Chicago.

- Devore, Jay & Roxy Peck. 1997. *Statistics: The exploration and analysis of data*. Pacific Grove, CA: Duxbury Press.
- Fujisaki, Hiroya & Keikichi Hirose. 1984. Analysis of voice fundamental frequency contours for declaration sentences of Japanese. *Journal of the Acoustical Society of Japan (E)* 5, 233–242.
- Goldsmith, John. 1978. English as a tone language. *Communication and Cognition* 11, 453–476.
- Gussenhoven, Carlos. 2004. *The phonology of tone and intonation*. Cambridge: Cambridge University Press.
- Haraguchi, Shosuke. 1977. *The tone pattern of Japanese: An autosegmental theory of tonology*. Tokyo: Kaitakusha.
- Hardcastle, William J. & John Laver (eds.). 1997. *The handbook of phonetic sciences*. Oxford: Blackwell.
- Hirayama, Teruo. 1951. *Kyushu hogen onchono kenkyu* [Tonal studies of Kyushu dialects]. Tokyo: Gakkaino shihyosha.
- Hirayama, Teruo. 1957. *Nihongo onchono kenkyu* [Tonal studies of Japanese]. Tokyo: Meiji shoin.
- Hirayama, Teruo. 1960. *Zenkoku akusento jiten* [Dictionary of Japanese accents]. Tokyo: Tokyodo shuppan.
- Hirayama, Teruo. 1967. *Kokugo no onsei* [The sounds of the national language]. Tokyo: Iwanami shoten.
- Ikeda, Kaname. 1951. Kinki akusentono taikai [The accentual system of the Kinki dialect]. In Kishio Terakawa (ed.), *Kokugo akusento ronso* [A collection of essays on Japanese accents], 97–187. Tokyo: Hosei daigaku shuppanyoku.
- Ishihara, Shunichi. 2000. Linguistic-tonetic differences in target-tone realisation: Standard Japanese vs. Kagoshima Japanese. *The 8th Australian International Conference on Speech Science and Technology, Canberra, Australia*, 224–229.
- Ishihara, Shunichi. 2004a. *An acoustic-phonetic descriptive analysis of Kagoshima Japanese tonal phenomena*. Ph.D. dissertation, Australian National University.
- Ishihara, Shunichi. 2004b. The surface tonal representations of Kagoshima Japanese accentual contrast. *The 10th Australian International Conference on Speech Science and Technology and Australian Language Technology Workshop, Sydney, Australia* (CD-ROM).
- Kibe, Yoko. 2000. *Seinanbu kyushu niki akusentono kenkyu* [A study of the two-pattern accent system in South-East Kyushu]. Tokyo: Bensei shuppan.
- Kori, Shiro. 1987. The tonal behavior of Osaka Japanese: An interim report. *Ohio State University Working Papers in Linguistics* 36, 31–61.
- Kubozono, Haruo. 1993. *The organization of Japanese prosody*. Tokyo: Kuroshio Publishers.
- Kubozono, Haruo & Michinao Matsui. 1996. Nihongo akusentojakkato intoneshon. In the Phonetic Society of Japan (ed.), *The 10th General Meeting of the Phonetic Society of Japan*, 101–106. Tokyo: The Phonetic Society of Japan.
- Ladd, D. Robert. 1996. *Intonational phonology*. Cambridge: Cambridge University Press.
- Ladefoged, Peter. 1997. Instrumental techniques for fieldwork. In William J. Hardcastle & John Laver (eds.), *The handbook of phonetic science*, 137–166. Oxford: Blackwell.
- Ladefoged, Peter & Ian Maddieson. 1986. Some of the sounds of the world's languages: Preliminary version. *UCLA Working Papers in Phonetics* 64.
- Lehiste, Ilse & Gordon E. Peterson. 1961. Some basic considerations in the analysis of intonation. *Journal of the Acoustical Society of America* 33, 368–379.
- Liberman, Mark. 1975. *The intonational system of English*. Ph.D. dissertation, MIT.
- Matsuura, Toshio. 2008. *Nagasaki hogenniokeru goonchono oninron* [The tonology of the Nagasaki dialect]. Ph.D. dissertation, Kyushu University, Japan.
- McCawley, James D. 1968. *The phonological component of a grammar of Japanese*. The Hague: Mouton.
- McCawley, James D. 1970. Some tonal systems that come close to being pitch accent system but don't quite make it. *The Sixth Regional Meeting, Chicago Linguistic Society (CLS)* 6, 526–531.
- McCawley, James D. 1977. Accent in Japanese. In Larry Hyman (ed.), *Studies in stress and accent* (Southern California Occasional Papers in Linguistics 4), 261–302. Los Angeles, CA: University of Southern California.
- Nagano-Madsen, Yasuko. 2004. Phonetic characteristics of the Kochi dialect of Japanese: A comparison of the Yusu-hara and Kochi-city dialects. *Africa & Asia* 4, 90–114.

- National Language Research Institute. 1959. *Nihon hogen kijutsuteki kenkyu* [A descriptive study of Japanese dialects]. Tokyo: Okurasho insatsukyoku.
- National Language Research Institute. 1967–1975. *Nihon gengo chizu* [A linguistic map of Japan]. Tokyo: Okurasho insatsukyoku.
- Peterson, Gordon E. & Harold L. Barney. 1952. Control methods used in a study of vowels. *Journal of the Acoustical Society of America* 24, 175–184.
- Pierrehumbert, Janet. 1980. *The phonology and phonetics of English intonation*. Ph.D. dissertation, MIT.
- Pierrehumbert, Janet & Mary Beckman. 1988. *Japanese tone structure*. Cambridge, MA: MIT Press.
- Pierrehumbert, Janet & Julia Hirschberg. 1990. The meaning of intonation contours in the interpretation of discourse. In Philip R. Cohen, Jerry Morgan & Martha E. Pollack (eds.), *Intentions in communication*, 271–311. Cambridge, MA: MIT Press.
- Poser, William J. 1984. *The phonetics and phonology of tone and intonation in Japanese*. Ph.D. dissertation, MIT.
- Rose, Phil. 1987. Considerations in the normalisation of the fundamental frequency of linguistic tone. *Speech Communication* 6, 343–351.
- Rose, Phil. 1991. How effective are long term mean and standard deviation as normalisation parameters for tonal fundamental frequency? *Speech Communication* 10, 229–247.
- Rose, Phil. 1993. A linguistic-phonetic acoustic analysis of Shanghai tones. *Australian Journal of Linguistics* 13(2), 185–220.
- Rose, Phil. 2002. *Forensic speaker identification*. London: Taylor & Francis.
- Shibatani, Masayoshi. 1979. Review of Haraguchi 1977. *Language* 55, 928–936.
- Shibatani, Masayoshi. 1990. *The languages of Japan*. Cambridge: Cambridge University Press.
- Sjölander, Kåre. 2006. The Snack sound toolkit (version 2.2.10). <http://www.speech.kth.se/snack/> (retrieved 30 June 2011).
- Society for the Study of Kyushu Dialects. 1969. *Kyushu hogen kisoteki kenkyu* [A fundamental study of Kyushu dialects]. Tokyo: Kazamashobo.
- Sugito, Miyoko (ed.). 1994–1999. *Nihongo onsei no kenkyu* [Study of Japanese prosody and speech sounds], vols. 1–7. Tokyo: Izumishoin.
- Sugito, Miyoko. 1995. *Sound dictionary of accent: Osaka and Tokyo Japanese*. Tokyo: Maruzen. [CD ROM]
- Sugito, Miyoko. 1997. *Onseihaikiwa kataru: Nihongo onseino kenkyu* [A study of Japanese sounds], vol. 4. Tokyo: Izumishoin.
- Sugito, Miyoko & Kazuhiko Kakehi. 2005. DVD version speech database of the dialects all over Japan. *Journal of the Phonetic Society of Japan* 9(3), 14–22.
- Talkin, David. 1995. A robust algorithm for pitch tracking (RAPT). In Bastiaan Kleijn & Kuldip K. Paliwal (eds.), *Speech coding and synthesis*, 495–518. New York: Elsevier.
- Uwano, Zendo. 1999. Classification of Japanese accent systems. In Shigeki Kaji (ed.), *Symposium Cross-Linguistic Studies of Tonal Phenomena, Tonogenesis, Typology, and Related Topics, Tokyo University of Foreign Studies*, 151–178. Tokyo: Research Institute for Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies.
- Yip, Moira J. W. 2002. *Tone*. Cambridge: Cambridge University Press.
- Zhu, Xiaonong. 1994. *Shanghai tonetics*. Ph.D. dissertation, Australian National University.