An Acoustic–Phonetic Comparative Analysis of Osaka and Kagoshima Japanese Tonal Phenomena

Shunichi Ishihara
College of Asia and the Pacific, Australian National University, ACT, Australia
shunichi.ishihara@anu.edu.au

Abstract
This paper, first of all, demonstrates that the LHL and the L(L)H pitch patterns exhibit significantly different F0 realisations between Osaka Japanese (OJ) and Kagoshima Japanese (KJ) which typologically belong to different accent groups. Secondly, it is argued that the observed difference results from an acoustic–phonetic perspective. The present study is based on a normalisation technique for the two dialects in their F0 realisations using a normalisation technique based on 12 OJ and 14 KJ speakers.

Index Terms: Osaka Japanese, Kagoshima Japanese, F0, accent, normalisation

1. Introduction
The accentuation of Japanese dialects has been one of the most rigorously studied areas in Japanese linguistics [1, 2, 3]. This article adds to this long history of auditory based transcriptions of pitch patterns of words in various Japanese dialects [4, 5, 6]. Although various same pitch patterns are observed across many dialects, this does not necessarily mean that the same pitch pattern is auditorily described across dialects (i.e. HL) is acoustically–phonetically realised in the same manner.

In this study, first of all, it is demonstrated that the LH, LHL and LLH pitch patterns observed both in Osaka Japanese (OJ) and Kagoshima Japanese (KJ) significantly differ between these two dialects in their F0 realisations. This is done by an acoustic–phonetic transcriptions of F0 realisations of the LH, LHL and LLH pitch patterns by means of a normalisation technique with a multiple number of speakers. Secondly, it is argued that the observed acoustic–phonetic realisation difference between OJ and KJ is due to the phonological difference between them in their tonal representations. Finally, the implications observed difference would bring for the surface tonal representation of KJ are explored.

The pitch pattern of words in OJ 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 [7, 8]. 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 [HHH] ‘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’ [9].

KJ exhibits a two way accentual contrast; L*LHL and L*H*H*H where n = a positive integer or zero [10]. 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. 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. usagi LLH, ‘rabbit’ and kakinono LLLH, ‘document’). The L*H*L type is referred to as Type A and the L*L*H* type as Type B in this paper [10]. As can be seen from KJ’s Type A and Type B contrast, although unlike Standard Japanese (SJ), there is not a lexical contrast in terms of the location of an accent in KJ, the main contrast between Type A and Type B is whether there is a fall in pitch at the end or not.

Typologically, OJ and KJ belong to different accent systems (the Kyoto–Osaka system and the two–pattern system, respectively) [11, 7]. 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 in Japanese dialects. It has been reported that the phonological–typological differences regarding suprasegmental features observed in different languages or dialects are manifested in some aspects of their phonetics [12, 13, 14].

This study focuses on the LH pattern of disyllabic words and the LHL and LLH patterns of trisyllabic words, because a significant between–dialect difference is constantly observed at the beginning of these pitch patterns. The LH and LHL pitch patterns are unaccented low–pitch beginning words and Type B words, respectively for OJ and KJ. The LH pitch pattern is an accentuated low-beginning word and a Type A word for OJ and KJ, respectively.

2. Procedures
As the database, the ‘DVD version Speech Database of the Dialects all over Japan’ is used [15]. The recordings for this database were collected from 1989 to 1993. Besides other dialects, this database contains the recordings of a large amount of citation words and utterances together with syllabic sounds and numbers uttered by 86 OJ and 79 KJ native speakers of various generations. The speech samples of 12 OJ speakers consisting of 6 males and 6 females and 14 KJ speakers consisting of 7 males and 7 females are used for this study. These speakers belong to the three generations of Old (Old ≥ 60), Middle (60 > Middle ≥ 40), and Young (40 > Young ≥ 20). The average age (standard deviation) of the 12 OJ speakers is 57 (18.6) for...
All speech samples of the citation words were read once by each speaker in this database, and recorded on professional digital equipment. All sounds are digitised in 16 KHz.

Table 1 contains the target words which were selected for acoustically–phonetically describing the F0 realisation of the three target pitch patterns. The syllable structure of all target words is (C)V. Various voiced and voiceless consonants and vowels are included in the target words in order to neutralise their known intrinsic effects [16, 17].

After being segmentally annotated, all target tokens were analysed using the ESPS routine of the ‘Snack Sound Toolkit’ [18]. On the basis of the annotation, F0 was sampled at the onset and every 20% point for each vowel. The identification of the vowel 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’ [19, p7]. Some sort of glottal stop–like laryngeal tension was observed in many recordings of the falling pitch tokens (i.e. [LHL]) at their word final position.

3. Normalisation Results

As stated in §1, this study investigates acoustic–phonetic realisation differences of the LH, LHL and LLH pitch patterns between OJ and KJ. For this, it is essential to give quantified descriptions of the linguistic–acoustic properties of the three pitch patterns in question. 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’ [20, p7]. Therefore, the individual content needs to be factored out as much as possible to extract the linguistic–acoustic content of the target pitch patterns. The z–score normalisation technique, of which performance has been empirically attested [21, 22, 23], is used in this study to extract the acoustic correlates of the target pitch patterns [20]. The z–score normalisation procedure is: $F_{0\text{norm}} = \frac{F_0 - \bar{x}}{\text{sd}}$, where $F_0$ is a sampling point, $\bar{x}$ is the average F0 from all sampling points, and sd is the standard deviation around the mean of those points. The z–score normalisation parameters (x and sd) were obtained from various monosyllabic, disyllabic and trisyllabic words. In this particular normalisation procedure, each F0 observation is expressed as so many standard deviations above and below a speaker’s overall mean F0. Standard deviation (sd) is used as the unit for normalised F0 values because normalisation is based on the sd of the samples.

3.1. Within Dialect Comparison

Fig. 1 contains the mean normalised F0 curves of the three pitch patterns concerned plotted for OJ and KJ, respectively against the mean absolute duration (sec.) together with one standard deviation above and below the mean for OJ (a) and KJ (b), respectively.

Figure 1: Mean normalised F0 curves of all target pitch patterns (LH, LHL and LLH) plotted together against the mean absolute duration (sec.) together with one standard deviation above and below the mean for OJ (a) and KJ (b), respectively.

males and 46 (14.8) for females. That of the 14 KJ speakers is 41 (24.8) for males and 47 (16.7) for females.

Intrinsic z–score normalised F0 values of the LH, LHL and LLH patterns concerned plotted for OJ and KJ, respectively.

### Table 1: Target words.

<table>
<thead>
<tr>
<th></th>
<th>OJ</th>
<th>KJ</th>
</tr>
</thead>
<tbody>
<tr>
<td>LH</td>
<td>kata katsudzumi musu koushita - na - ni</td>
<td>iro kame musu asa navi owa ru</td>
</tr>
<tr>
<td>LHL</td>
<td>ichigo kijira kouta - -</td>
<td>kame musu kagami oyogu</td>
</tr>
<tr>
<td>LLH</td>
<td>- - -</td>
<td>- -</td>
</tr>
</tbody>
</table>

...
Intrinsic z−score normalised F0

Figure 2: Mean Normalised F0 curves of the LH pattern.

Figure 3: Mean Normalised F0 curves of the LHL pattern.

Figure 4: Mean Normalised F0 curves of the LLH pattern.

### 3.2. Between Dialect Comparison

Figs. 2, 3 and 4 contain the mean normalised F0 curves of the LH, LHL and LLH pitch patterns plotted separately between OJ and KJ against the equalised durations.

A notably consistent difference between OJ and KJ is the normalised F0 values associated with the first syllables (refer to the arrows of Figs 2, 3 and 4). The normalised F0 values of the first syllables lie between -1 and 0 sd for OJ, while they are around 0 sd for KJ. This observation about KJ conforms to Ishihara’s [23] report that normalised F0 declines around 0 sd in the initial low pitched syllable(s). The first syllables have significantly higher normalised F0 values (p ≤ 0.0001 at all sampling points) in KJ than OJ for all target pitch patterns.

Due to this difference between OJ and KJ, the LLH pitch pattern exhibits significantly different normalised F0 contours between OJ and KJ in that normalised F0 gradually rises from the first to the third syllable in OJ, while in KJ it gradually declines in the first two syllables before starting to rise in the last syllable. This rising F0 contour of low–pitch beginning unaccented words has previously been reported [24, 26].

### 4. Discussion

As introduced in §1, it is lexically determined whether a word starts with a high or a low pitch in OJ. Using the Autosegmental–Metrical Theory [26, 27], Pierrehumbert and Beckman [26] give the tonal representations of the low–pitch beginning accented (i.e. LHL) and unaccented (i.e. LLH) words as presented in Fig. 5. As can be seen in Fig. 5, a significant aspect of OJ is that both initial and final boundary tones are lexical tones and they are the properties of words.

Although it is not entirely clear about the tonal representations of KJ both at the lexical and surface levels, it is sensible to consider, if we posit an initial boundary tone for KJ, that it should be a non–lexical tone because KJ does have the same kind of lexical contrast as OJ for the initial boundary tone. With respect to the value of the initial tone, two possibilities which are presented in Fig. 6 have been proposed by Ishihara [23].

As given in Fig. 6a, following the phonological pitch configurations of Type A and Type B words (i.e. LLLLHL and LL−LLH), one might intuitively posit an L tone as the initial tone. However, a few points need to be addressed regarding this L tone.

First of all, the onset normalised F0 value of the initial low pitched syllable is around the mean (= 0), as can be seen in Figs. 2, 3 and 4. This means that the onset F0 value of the initial low pitched syllable is around the average F0 of the speakers’ F0 distribution. It is possible to interpret the F0 associated with the onset of the initial low pitched syllable as the most neutral F0 value that is around the mean F0 of a speaker’s F0 distribution. Considering this, as given in Fig. 6b, it might not be necessary to assign a particular tone as the initial tone because the F0 realisation of the initial low pitched syllable is predictable from the information regarding a speaker’s F0 distribution. A very similar interpretation can be seen in Dainora [28] in the intonation of English. Referring to the analysis on the intonation of English by Goldsmith [29] in which the intonation was analysed in terms of H, M and L, Dainora [28, p36] 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, M tone is always the first tone in the sequence and is not associated with an accented syllable.

If Fig. 6b is the case for KJ, it is straightforward to understand the F0 realisation difference between OJ and KJ regarding the initial low pitched syllables because OJ and KJ assign different tonal values (L vs. ∅) to the initial tone. Even if Fig. 6a is the case for KJ, there is a significant difference between OJ and KJ in that the initial L tone of OJ is lexical while that of KJ is
Figure 5: Tonal representation of low–pitch beginning accented (a) and unaccented (b) words.

Figure 6: Two possible tone values (a and b) of the initial tone are given using a 3 syllable word as an example.

5. Conclusions

This study has first of all demonstrated that the LHL and L(L)H pitch patterns of OJ have significantly higher F0 realisations than those of KJ at the initial low–pitched syllables. Secondly, it has been argued that this observed F0 realisation difference can be phonologically accounted for with respect to the different tonal representations. Finally, some implications to the tonal representations of KJ have also been discussed.

6. Acknowledgements

This research is financially supported by the College of Asia and the Pacific, the Australian National University. I thank three anonymous reviewers for giving valuable comments for this research.

7. References


