

# Intonation of Japanese sentences spoken by English speakers

Chiharu Tsurutani

School of Languages and Linguistics, Griffith University

C.Tsurutani@griffith.edu.au

## Abstract

This study investigated intonation of Japanese sentences spoken by Australian English speakers and the influence of their first language (L1) prosody on their intonation of Japanese sentences. The second language (L2) intonation is a complicated product of the L1 transfer at two levels of prosodic hierarchy: at word level and at phrase levels. L2 speech is hypothesized to retain the characteristics of L1, and to gain marked features of the target language only during the late stage of acquisition. Investigation of this hypothesis involved acoustic measurement of L2 speakers' intonation contours, and comparison of these contours with those of native speakers.

**Index Terms:** Japanese intonation, prosodic transfer, L2 speech

## 1. Introduction

Intonation is a feature of phrases or sentences expressed by pitch as the principle exponent [4]. Prosodic features, including pitch, are generally hard to teach and learn due to their abstractness. There are very few language classes that devote time specifically to intonation; nor are there many language textbooks that explain what kind of intonation patterns are to be assigned to sentences, apart from the description of final pitch patterns for interrogative and affirmative sentences. However, it is known that the acquisition of prosody has more impact on learners' intelligibility, as judged by native speakers, than does the acquisition of segmental features [1,10]. We need to elucidate the processes through which L2 learners progress before achieving the right intonation of the target language. This understanding of L2 intonation is crucial for L2 acquisition research and to produce the model for CAPTA (Computer Assisted Pronunciation Training and Assessment) programs. Prosodic errors refer to pitch and timing errors but, in this paper, we will focus on pitch as the first priority of our task, and take into consideration mainly linguistic factors that affect L2 prosodic structure. In order to analyze prosodic patterns assigned to L2 speech, we first need to illustrate the intonation pattern of the target language of this paper, that is, Japanese. After the prosodic pattern of Japanese language is clarified, the pattern of L2 speech will be analyzed and categorized for the purpose of creating L2 prosodic models.

## 2. Intonation pattern of Japanese sentences

When we observe the F0 contour of one sentence, we see that the pitch of a sentence may initially show a slight rise depending on the type of the word accent and then a gradual decrease towards the end of the phrase and sentence, regardless of the type of word accents in the sentence. This gradual, time-dependent down sloping of the fundamental frequency is called declination [6]. Pitch contour in Japanese sentences is largely influenced by the pitch pattern of words. Pitch decreases towards the end of the utterance while incorporating the accent patterns of words, resulting in a curve that looks like the Japanese character “ $\sim$ ” [he] [13]. To illustrate it, the following prosodic structure is used in this paper.

Utterance [ ]  
 Intonational [ ] [ ] [ ]  
 Phrase (IP)  
 Accentual [ ] [ ] [ ]  
 Phrase (AP)

Prosodic word *sankakuno yaneno mannakani okimasu*  
 (J\_ToBI model :1999)

IP=a pitch resetting and significant pause marks an IP boundary.

AP= a pitch rise and the subsequent gradual fall to a low marks AP. It often corresponds with noun, adjectival and verb phrases.

Figure 1: Japanese prosodic structure.

In Japanese, pitch accent is a lexical property of a word, and is not affected by the prosodic organization at the higher level. The pitch pattern of words (e.g., HL) cannot be reversed (LH) by intonation like it can be in English. To express accent patterns of Japanese words, the values high and low are commonly used. The following are examples of 4-mora words. The same two symbols do not denote exactly the same pitch since the pitch decreases over the period of time, due to declination. It is obvious that this binary notation per se is not sufficient to express the manifestation of prosodic organization higher than word level. Normally the first phrase of the sentence has the highest F0 and the following phrases subsequently have lower F0, unless the focus falls onto the particular phrase. The F0 of the focused phrase is higher than the F0 of other phrases and becomes as high as that of the first phrase. Alternatively, the first phrase starts with lower F0 than it typically does to make the focused phrase more prominent [5]. When words are combined to form an accentual phrase, the pitch patterns of original words are sometimes merged into one bigger pitch contour. This merge changes the pattern of the other words, or reduces the pitch of other words. The pitch patterns of accentual phrases are demonstrated in a two-word phrase in Figure 2. The abovementioned four types of pitch pattern are divided into two: accented (A), for the 2<sup>nd</sup> and 3<sup>rd</sup> types of patterns; and unaccented (U) for the 1<sup>st</sup> and 4<sup>th</sup> types of patterns.

Table 1: Pitch pattern of Japanese 4-mora words.

|              |                    |                           |
|--------------|--------------------|---------------------------|
| 1) Flat      | <i>ko-o-ko-o</i>   | LHHH~HHHH                 |
| 2) Head-high | <i>bi-n-bo-o</i>   | HLLL                      |
| 3) Mid-high  | <i>to-sho-ka-n</i> | LHLL                      |
|              | <i>ka-ra-ka-sa</i> | LHLLL=low pitched mora    |
| 4) Tail-high | <i>o-to-o-t-o</i>  | LHHH H= high pitched mora |

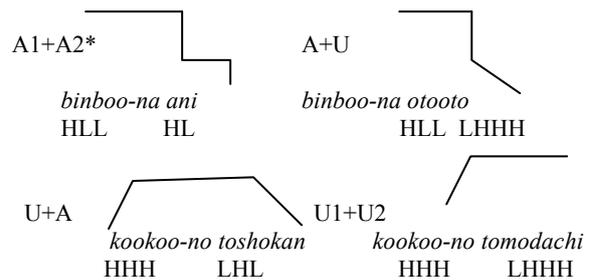


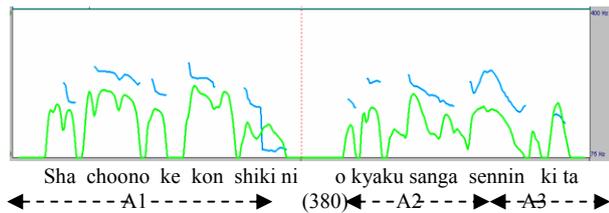
Figure 2: Formation of accentual phrases.

**\*A1+A2**

This staircase-like pitch contour is called a downstep, and the two words form two accentual phrases. However, when the first word receives focus and there is no initial rise observed on the second AP, two accented words could form one AP [12].

To illustrate the pitch contour above the accentual phrase level, we will use a sentence in regards to which L2 learners' performance will later be discussed.

The sentence "Shachoo no kekkonshiki ni okyakusan ga sennin kita" is expected to have the F0 contours at word level as below.



# lower lines indicate intensity. ( ) = the length of pause in millisecond  
Intonational /Accentual Phrase

Prosodic word  
Shachoo no / kekkonshiki ni/ okyakusan ga / sennin/ kita.  
LHH LHLLL LHHH HL HL  
U + A U A1 + A2

Figure 3: Native speaker's utterance.

### 3. Intonation patterns of Japanese assigned by L2 learners

More marked features of the target language are assumed to be harder to acquire than other features during prosodic transfer. Some researchers have a view that L2 learners simply use the intonation pattern of their L1 in the L2 sentence of the same meaning [2,9]. I hypothesize that the L2 intonation is a complicated product of L1 transfer at two levels of prosodic hierarchy: the word level and the phrase level, and retains the characteristics of L1 with marked features of the target language unacquired until the late stage of acquisition. In accordance with this hypothesis, L2 prosody was examined using Australian learners of Japanese. Typical learners' errors will be discussed at two different prosodic levels. Previous studies on non-native prosody measured the number of phrases, F0 differences of pitch fall, mean duration, and pitch range of a sentence spoken by native and non-native participants [7,8]. As expected, non-native speech had more phrases, more pauses, different pitch pattern, and greater length than did speech of natives. However, exact error patterns of pitch had not been illustrated in a systematic manner. This section reports on the quantitative analysis conducted on the pitch pattern of prosodic words and phrases in utterances of L2 speakers.

#### 3.1. Method

##### 3.1.1 Participants

The participants in this study were 19 native Australian English speakers who have no other languages spoken at their home (one speaker's recording was not clear and was excluded). They were second year university students who had studied Japanese for 200 hours in class. This level of learners was

chosen because complete beginners have difficulty in reading a sentence with a certain level of fluency. Participants' ages ranged from 17 to 22 years. The recording was extracted from a computer exercise in which participants read Japanese sentences.

##### 3.1.2 Materials

The sentences are all 23-24 mora length and consists of five prosodic words. Sentences 1- 3 have the same type of the first AP <Noun-GEN, Noun -PP/ACC>, and sentences 4 and 5 have the same AP type <Noun-PP>. Pitch was not indicated on the display screen, although after their first trial participants were able to hear a correct pronunciation. Most words had a flat pitch pattern, which is difficult for English speakers to produce. A space was left between prosodic words on the display screen (*wakachigaki*) for ease of reading and a technical reason regarding speech recognition by machines.

Table 2: Japanese sentences

1. *Shachoo-no kekkonshiki-ni okyakusan-ga sennin kita.*  
LHH LHLLL / LHHH / HL HL  
President's wedding- to/ guests / one thousand came
2. *Obasan-no yoosu-o byooin-de kitee kudasai.*  
LHHH HHH / HHH / LHH HHHH  
Aunt's condition / the hospital-at / please ask.
3. *Watashi-no kookoo-de isshoni shashin-o torimashoo.*  
LHHH HHH / LHH / LHH LHHHL  
My high school-at /together photos let's take.
4. *Shuumatsu-kara futari-no hito-to shigoto-o suruyoteedesu.*  
HHHHH / LHHH LHL / LHHH LHHHHL  
Weekend-from / two people-with / work (I) plan to do.
5. *Yuubinkyoku-de hyakuen-no kitte-o gomai katta.*  
HHLLL / LHHH LHH / LH LH  
The post office-at/ 100 yen stamps / five(pieces) bought.  
/ indicates accentual phrase boundary.

##### 3.1.3 Procedure

The participants read five Japanese sentences that appeared on the computer screen. The sentences appeared one at a time and the participants were allowed to retry more than once until they were satisfied with their pronunciation. The utterances of 19 participants (5 males, 14 females) were phonetically transcribed and marked by three experienced Japanese language teachers according to the speaker's performance.

### 3.2. Pitch range / pitch contour

A study on Chinese learners of Japanese reported a narrower pitch range in the learners' performance due to the conscious suppression of the characteristics of the tone language [7]. In the current study, the pitch range varied from speaker to speaker and quite often the learners' L2 pitch ranges were wider than those of natives, although there were a few extremely monotonous speech segments. Presumably the pitch range is related to learners' L1 background. The pitch range of English, a stress accent language, is wider than that of Japanese, which is a pitch accent language [3]. In the current study, there was no difference in pitch range between native and L2 speakers for female speakers (see Table 2). In general, women have a wider pitch range than men. The smaller pitch range in male L2 speakers in comparison to native males is hard to explain as the males' L2 performance scores were relatively high, and the smaller pitch range does not necessarily mean lack of confidence. Further investigation regarding this point is required.

Table 2 : Averaged pitch range in Hz: Natives and L2 learners.  
( ) = No. of speakers

|        | Native   | L2 learners |
|--------|----------|-------------|
| Female | 126 (5)  | 126 (14)    |
| Male   | 95.2 (5) | 52 (5)      |

The pitch contour of the whole sentence was captured by measuring pitch at the highest point of intensity in each syllable and plotting this point within each speaker's pitch range. This normalized pitch contour was used to compare the pitch contour of native speakers' and learners' utterances in Sentence 1. Ten native speakers' utterances showed the first peak at "choo-no", and the focused word "sen" had the second highest peak. "Shi" in "shiki" and "ki" in "kita" did not have their pitch measured as [i] was devoiced. In IP3 the downstep shows a straight fall of pitch after "sen".

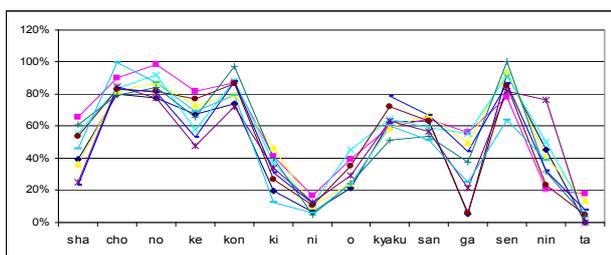


Figure 4: 10 Native speakers' pitch contour of Sentence 1.

The three IPs formed three moderate humps in native speakers' utterances. In L2 utterances, the shape of IP drastically differed from that of natives and the sentence focus on "sen" in "sennin" was not observed, nor was the pitch fall after "sen". Instead the regular occurrence of an accented syllable was evident. The pitch contours of learners who had pitch error on the first and third prosodic words ("sha" and "kyaku") showed three or four peaks that appeared in regular interval, as displayed below.

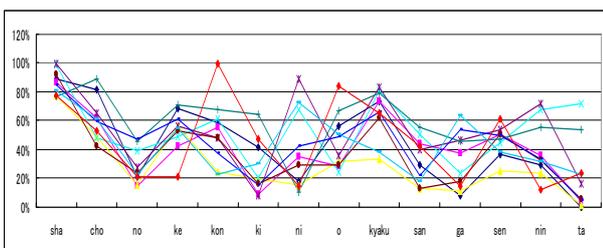


Figure 5: 10 learners who had pitch error on the first and third prosodic words ("sha" and "kyaku").

The rest of the learners, including the two top learners, showed individual differences in their pattern and did not form patterns as a group. Assigning a high pitch on a syllable at the beginning of each phrase was observed throughout all five sentences.

### 3.3. Pitch tracking

The pitch contour of the entire sentence shows individual differences even among native speakers. However, native speakers' speech show general agreement in the location of the syllable to which the highest and lowest pitch in the accentual phrase are assigned. Thus, the highest pitch point in each phrase and their proportional heights compared with the beginning and the end of the utterance was measured. The

value of pitch was extracted at five locations from Sentence 1 "Sha (3) choo no (2) kekkonshiki ni (4) okyakusanga sen (1) nin kita (5)" and plotted as below.

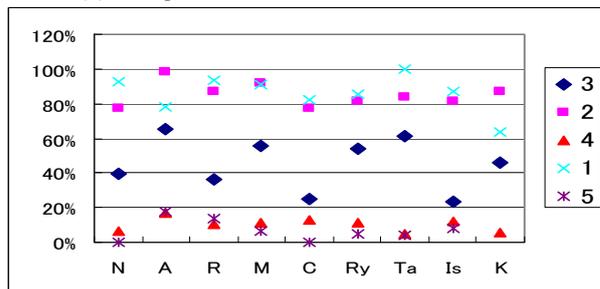


Figure 6: Native speakers' pitch hierarchy.

The Y axis shows that the relative value of pitch was highest in location 1 and lowest in location 5. As expected, native speakers uniformly presented the same pitch hierarchy.

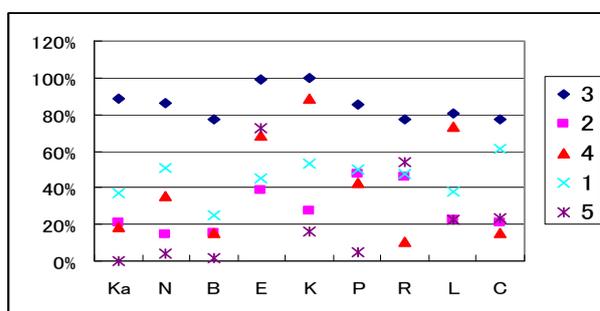


Figure 7: Learners' pitch hierarchy.

The low value at location 1 indicates a failure to assign the sentence focus on "sen" in "sennin". A lower value at location 2 than at location 3 indicates that learners failed to produce the pitch pattern LHH on "shachoo". The high value at location 4 or 5 indicates pitch raising at the phrase final position. In this way, the hierarchical order indicates correctness of intonation contour.

### 3.4. Number and location of incorrect HL pitch pattern

Each sentence contained two to four words with flat pitch pattern. The occurrences of errors in all five sentences were counted. In Table 3, the percentage of speakers who made errors in pitch pattern is displayed according to the syllable structure of the word and position in a sentence.

Table 3: Syllable structure of flat pitch pattern words that had HL pitch pattern errors (%).

| syll structure | Error rate | syll structure | Error rate |
|----------------|------------|----------------|------------|
| CVCCV          | 0.10       | CVCV           | 0.75       |
| CVVCV          | 0.17       | C'VVCVN        | 0.77       |
| C'VCVVN        | 0.30       | VC'VCV         | 0.80       |
| CVGVCV         | 0.38       | C'VCVV         | 0.87       |
| C'VCVC         | 0.44       | CVCVCV         | 0.88       |
| VCC'V          | 0.50       | CVVCVV         | 0.89       |
| CVGVV          | 0.50       | CVVCV          | 0.92       |
| CVCCV          | 0.60       | CVVCV          | 0.92       |
| VCVCVN         | 0.61       | C'VVCVCV       | 0.94       |

Words that had a long vowel at the beginning elicited higher rates of error than did other word types. The position of these words in a sentence is also shown below. AP boundary(■) and pitch raising before AP boundary are presented together with five prosodic word (PW)s.

Table 4: HL pitch errors and pitch raising in the sentence (%).

|    | PW1  | R    | PW2  | R    | PW3  | R    | PW4  | PW5 |
|----|------|------|------|------|------|------|------|-----|
| S1 | 0.87 | ---- | HL   | 0.53 | 0.8  | 0.3  | HL   | HL  |
| S2 | 0.61 | ---- | 0.92 | 0.7  | 0.77 | 0.23 | 0.92 | HL  |
| S3 | 0.17 | ---- | 0.89 | 0.38 | 0.5  | 0.27 | 0.44 | HL  |
| S4 | 0.94 | 0.5  | 0.38 | ---- | 0.75 | 0.5  | 0.88 | HL  |
| S5 | 0.5  | 0.78 | 0.3  | ---- | 0.6  | 0.5  | 0.5  | 0.1 |

R=pitch raising, HL= pitch pattern of the words were originally HL  
 Long vowels at word initial syllable are assumed to attract stress-type accent. A high rate of error in this type of word was observed for all prosodic words. Pitch raising was observed only before AP boundaries and found even in the speech of participants with high scores. As declination is a universal phenomenon in the utterance of statement, pitch raising at the end of the sentence was rare. Although pitch raising is sometimes heard in casual speech in the young generation in recent times, it never occurs in statements read by native speakers.

#### 4. Summary

This study attempted to provide data on L2 intonation using Japanese sentences read by L2 English speakers. The following errors were observed as typical prosodic errors.

<Word level>

- Pitch raising at word (phrase)-final position
- Frequent assignment of HL pitch pattern and the resultant regular occurrence of accented syllables

<Accentual Phrase level>

- Forming more phrases than there should be

<Intonational Phrase level>

- No sentence focus
- No downsteps
- Monotonous intonation

L1 transfer from the English intonation pattern was apparent in word (phrase)-final pitch raising and the assignment of HL pitch pattern. Pitch raising at the last foot of the sentence is a phenomenon known as tonic syllable in English. The intonation contour of Japanese was heavily affected by the pitch patterns of prosodic words. In L2 speech the shape of AP was formed around one accented syllable that was accompanied by a sharp drop of pitch, while there were more gentle humps within an AP in native speakers' utterances due to flat pitch patterns. Flat pitch pattern, which is a marked prosodic feature, seems to be the key to acquiring the correct intonation.

Speech at the intonational phrase level involves advanced prosodic knowledge, such as the subtle difference of meaning created by pitch and by sentence focus. This prosodic knowledge could sometimes vary even among native speakers and will be difficult to learn for L2 learners. However, the following advice could help improve learners' intonation: 1) Numbers and adverbs tend to attract focus in a sentence. 2)

When words with HL pitch pattern occur one after the other within an IP, the second word has a lower pitch; this pattern is called a downstep. 3) Pitch declines within an AP and does not rise at the end of the phrase. While the pitch accent of words is crucial factor in forming the correct intonation pattern in Japanese, the incorrect pitch contours at phrase level might have less impact on overall performance. Occasional pitch raising at phrase-final position on a particle could be accepted by native speakers as it is heard in some emphatic patterns in Japanese, but not after every phrase.

The measurement of the entire pitch contour of a sentence by itself does not seem to be practical for assessment. However, the pitch pattern of Japanese words is unaffected by prosodic organization at higher levels. In fact, fixed pitch patterns are easier to learn than plastic (movable) pitch patterns such as those found in English [11]. Normalized pitch values at critical locations provide good indicators of correct intonation and would be a useful tool for assessment of intonation. This indication of pitch values could also reduce the risk of L1 prosodic transfer. Although the focus of this study was on the measurement of pitch contour, studies that measure duration and other related segmental errors will be necessary in order to illustrate L2 prosody as a whole. Further, the development of assessment tools that incorporate the findings of such studies will be expected in future research.

#### 5. References

- [1] Anderson-Hsieh, J. et al. "The relationship between native speakers' judgements of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure" *Language learning* 42: 4, 529-555, 1992,
- [2] Ayusawa, T. "L1 prosodic transfer in learners of Japanese", in Research report on the prosodic features of Japanese, Ministry of Education D1 Group. 100-110, 1993.
- [3] Beckman, M. and Pierrehumbert, J. "Intonational structure in Japanese and English", *Phonology Yearbook*. 155-300, 1986.
- [4] Cruttenden, A, *Intonation* Cambridge University Press, 1986.
- [5] Fujisaki, H. "Modeling the generation process of F0 contours as manifestation of linguistic and paralinguistic information" Symposium "Intonation: Models and parameters" 12<sup>th</sup> ICPHS, 19-24, 1999.
- [6] Guessenhoven, C. *The phonology of tone and intonation*. Dordrecht: Kluwer Academic publisher, 2004.
- [7] Hirano, H. and Kawai, G. "Pitch patterns of intonational phrases and intonational phrase groups in native and non-native speech" *Interspeech* 2005.
- [8] Jun and Oh: 2000 Acquisition of second language intonation *Interspeech* 2000.
- [9] Saito, H. "Nuclear-stress placement by Japanese learners of English - Transfer from Japanese" in (eds.) Kawaguchi, et al. *Prosody and Syntax, Cross-linguistic perspectives* 125-140. 2006.
- [10] Sato, H. "Production of prosody from phonetic point of view (Prosody no seisei)", 58-65, *Gekkan Gengo*. 1992.
- [11] Rasier, L. and Hiligsmann, P. "Prosodic transfer from L1 to L2." Theoretical and methodological issues. *Nouveaux cahiers de linguistique francaise* 28, 41-66, 2007.
- [12] Venditti, J.J., et al. "Prominence marking in the Japanese intonation system." In Miyagawa et al.(eds.) *Handbook of Japanese Linguistics*. Oxford University Press. 2008.
- [13] Yao, R. "L1 prosodic transfer in Japanese predicate spoken by Chinese speakers" in Research report on the prosodic features of Japanese, Ministry of Education D1 Group. 103-110, 1993.