

Modeling pitch errors of Japanese intonational phrases spoken by a native speaker of Chinese

Hiroko Hirano¹ and Goh Kawai²

¹Graduate School of International Media and Communication ²Institute of Language and Culture Studies
Hokkaido University, Sapporo, Japan

¹hiran@ilcs.hokudai.ac.jp ²goh@kawai.com

Abstract

We compared pitch patterns of Japanese intonational phrases spoken by a native speaker of Chinese.

Several types of pitch errors were distinguished according to pitch contour shapes within and across intonational phrases. These shapes differ categorically from one another—for instance, the learner uses flat contours while the native uses hump-like contours.

Comparisons of the learner and native speaker show that (a) the learner's intonational phrases are shorter, (b) the learner's pitch contour repertoire is smaller, (c) the learner has difficulty spreading pitch declinations over multiple syllables, and (d) segmental pronunciation errors cause pitch changes perceived by natives as intonational phrase boundaries.

1. Introduction

65 percent of the 100,000 foreign students in Japan are mainland Chinese and Taiwanese [1]. Acquiring spoken language is their paramount concern [2][3]. So far, pronunciation learning has focused on segmental skills [4][5][6].

Correctly expressing intonational phrases enhances understanding of the spoken message [7]. We have two long-term objectives regarding intonational phrases: (a) to improve instructor-led learning, and (b) to implement CALL systems.

In this paper, we analyzed pitch within and across intonational phrases spoken by a native speaker of Chinese, and compared them with a native speaker of Japanese. The remainder of this paper explains the speech data (section 2), pitch within intonational phrases (section 3), pitch across intonational phrases (section 4), and conclusion (section 5).

2. Speech data

2.1 Recording speech

A female adult Taiwanese gave a short talk about her graduate school research plans. The script for her talk had been written, edited and memorized for a presentation she gave six months prior to speech data collection. Recordings were made in a quiet conference room, using an Olympus DM-10 digital voice recorder (built-in monaural microphone, sampling rate 44.1 kHz, bitrate 64 kilobit/s, frequency response 300–7000 Hz,

WMA file format). The subject did not read a script; she had her talk memorized. The talk was transcribed and read aloud by a female native speaker. Total speech time was 446 seconds for the Chinese speaker and 402 seconds for the Japanese speaker.

2.2 Labeling intonational phrase boundaries

Punctuation marks and other typographic cues for intonational phrases were removed from the transcription, and given to three native Japanese language instructors who listened to the Chinese and Japanese talkers' recordings and marked the transcription wherever intonational phrase boundaries were perceived [8]. We accepted intonational phrase boundaries where two or more language instructors agreed. There were 397 intonational phrases for the Chinese talker and 197 for the Japanese talker. Defining agreement among labelers as:

$$a = \frac{p_a}{p_t}$$

a : agreement among labelers

p_a : number of intonational phrase boundaries agreed upon by two or more labelers

p_t : total number of unique locations labeled as intonational phrase boundaries by one or more labelers

Agreement was 80 percent for the Chinese talker and 92 percent for the Japanese talker. (Had we accepted intonation boundaries only when labelers were unanimous, agreement would have been 66 and 80 percent, respectively.)

The high agreement for the Japanese talker may indicate consistent, clear expressions of intonational phrase boundaries, while the moderately low agreement for the Chinese talker may indicate inconsistent, ambiguous acoustic cues.

Impressionistic listening supports the above interpretation. On the one hand, the Japanese talker used multiple acoustic cues (such as pitch changes, segmental lengthening and pause insertions) to intentionally mark intonational phrase boundaries. On the other hand, in many instances, the Chinese talker unwittingly used one or more acoustic cues that labelers perceived as phrasal markers. The varying number and degree of acoustic cue usage may have caused uncertainty among the labelers.

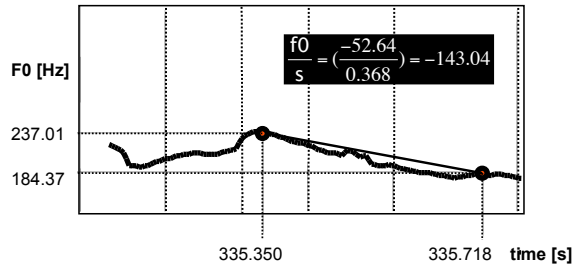


Figure 1. Pitch slope measurements. Within an intonational phrase, measure pitch drop between F_0 peaks over two accented mora, or the pitch peaks of the sole accented mora and the last mora. Pitch slope is the pitch drop over the time interval between measurement points.

Table 1. Pitch slope (k) descriptive statistics for non-native and native speakers. Native speaker shows wider range and variance.

k	non-native	native
max	0.45	8.16
mean	-106.51	-138.23
min	-309.08	-340.01
standard deviation	67.32	89.51

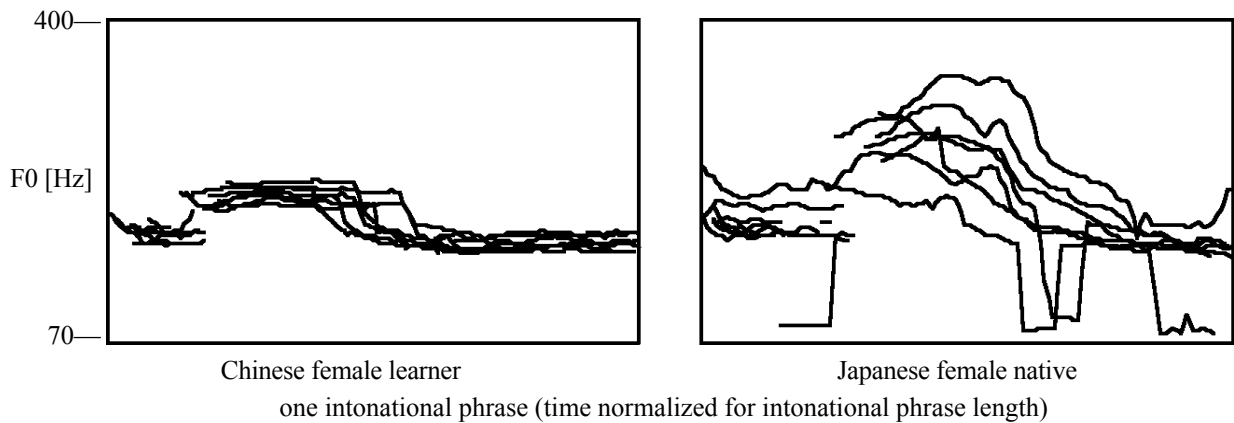


Figure 2. Pitch contours for the Japanese intonational phrase “Hokkaido {wo, no, ni, wa}” (total 7 phrases) spoken by a Chinese female learner (left) and a native Japanese female (right). Multiple utterances are shown overlapped, and time-normalized for intonational phrase length. Note reduced pitch range and uniform contour shape for learner. Native speaker shows wide pitch range and stylistic variation appropriate for context.

2.3 Selecting intonational phrases

Out of the 392 intonational phrases for the Chinese speaker, 108 consisted solely of words pronounced with the correct pitch accent. These intonational phrases corresponded to 95 intonational phrases for the Japanese speaker. (The Japanese speaker’s pitch accent was totally correct.) These were used for analyzing pitch within intonational phrases (section 3).

For analyses spanning adjacent intonational phrases (section 4), we chose 52 pairs from Chinese speaker, where labelers perceived boundaries even though there were no pauses. We were obliged to accept intonational phrases containing incorrect word accents because the Chinese talker made numerous errors. Confounding word accent and intonational phrase phenomena is unavoidable to some extent because non-native learners typically acquire both skills consecutively.

3. Pitch within intonational phrases

Using segment-level intonational phrase boundaries obtained in section 2, and with audiovisual reference to the speech files, waveforms and spectrograms, we hand-labeled phone and mora boundaries (the speech data files accompanying this paper are in Praat format [9]).

We measured the pitch slope for each intonational phrase as follows (an example is shown in Figure 1):

$$k = \frac{f_2 - f_1}{t_2 - t_1}$$

k : pitch slope

f_1, f_2 : F_0 peak values of the first and second accented mora, or of the sole accented mora and the last mora.

t_1, t_2 : Time points of the pitch peaks corresponding to f_1 and f_2 .



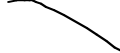
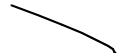
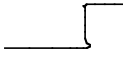
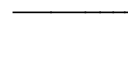
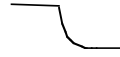
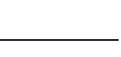

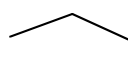
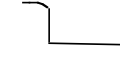

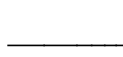
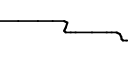
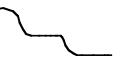
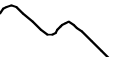
zone	Beginning of intonational phrase.	Up to and including accented mora.	Immediately following accented mora.	End of intonational phrase.
appropriate pitch movement	Pitch should rise gradually in this zone.	Pitch should continue to rise for most of this zone, followed by a gradual drop.	Pitch should decline gradually.	Pitch should continue to decline gradually.
correct contour shape				
incorrect contour shapes				
				
				
				

Figure 3. Intonational phrase zones, appropriate pitch movement, correct and incorrect contour shapes. Each column shows contour shapes appearing in consecutive zones in intonational phrases. Incorrect contour shapes were analyzed from intonation phrases where pitch accents were produced correctly.

While the native speaker almost always incorporated more than one accented mora in intonational phrases, the non-native learner produced only 10 such phrases; 98 phrases consisted of one accented mora. In these cases, for convenience, we substituted the final unaccented mora in the intonational phrase for pitch slope calculations. The smallest 5 percent of k were removed to compensate for F0 estimation errors. Table 1 shows descriptive statistics.

Comparisons of the learner and native speaker show that (a) the learner's intonational phrases are shorter in duration, (b) the learner has a limited repertoire of pitch contours, and (c) the learner has difficulty spreading pitch declinations over multiple syllables. Figure 2 shows a set of representative examples. Note the reduced pitch range and uniform contour shape for the learner. The native speaker shows considerably wider pitch range and stylistic variations appropriate for context.

We identified several types of pitch errors distinguishable by pitch contour shapes in four consecutive zones: (1) the beginning of the intonational phrase, (2) up to the end of the accented mora, (3) immediately after the accented mora, and (4) up to the end of the intonational phrase. Each zone exhibits several characteristic shapes. These shapes differ categorically from one another—for instance, the learner uses flat contours while the native uses hump-like contours. Figure 3 contrasts correct and incorrect shapes for each zone. Note that the learner does not make every conceivable error, but is rather limited in her selection, while at the same time producing shapes that are distinctively non-native.

4. Pitch across intonational phrases

We observed that the native speaker used multiple acoustic cues such as pitch change, segment lengthening and pause insertions to redundantly signal phrase boundaries. In order to focus on the non-native learner's

pitch phenomena, we chose adjacent intonational phrases where labelers perceived an intonational phrase boundary although there was no silence in between (52 boundaries out of 392 total). We then compared these 52 phrase pairs with corresponding intonational phrases produced by the native speaker.

Our analyses indicate that short-term changes in pitch are, while important, not sufficient conditions for an intonational phrase boundary. Pitch changes across boundaries are better understood in a wider context—observing pitch movements in the preceding intonational phrase's zones 2, 3 and 4, and the succeeding intonational phrase's zones 1 and 2 allowed us to cluster pitch contours (Figure 5. We identified three types of preceding and succeeding pitch shapes (logically there should be nine combinations but we found no instance of one type; hence Figure 4 shows eight types).

With the notable exception of the top-center type, seven types have pitch movements that are either abrupt, unsmooth, at different slopes, or in opposite directions. We noticed that the absolute value of pitch change itself is not necessarily decisive; after a flat pitch contour, tiny pitch changes can be perceived as intonational phrase boundaries.

The top-center type is the correct shape for a single intonational phrase, except the succeeding intonational phrase contains inappropriate U-shaped, downward-upward pitch movements. In other words, had the learner controlled her pitch in the succeeding phrase's zone 2, no boundary would have been perceived.

The learner occasionally produced incorrect pitch accent. Lexical items with the wrong pitch caused the labelers to perceive pitch boundaries although morpho-syntactically there clearly was none. Lexical pitch and intonational phrases need to be taught together in the Japanese language classroom.

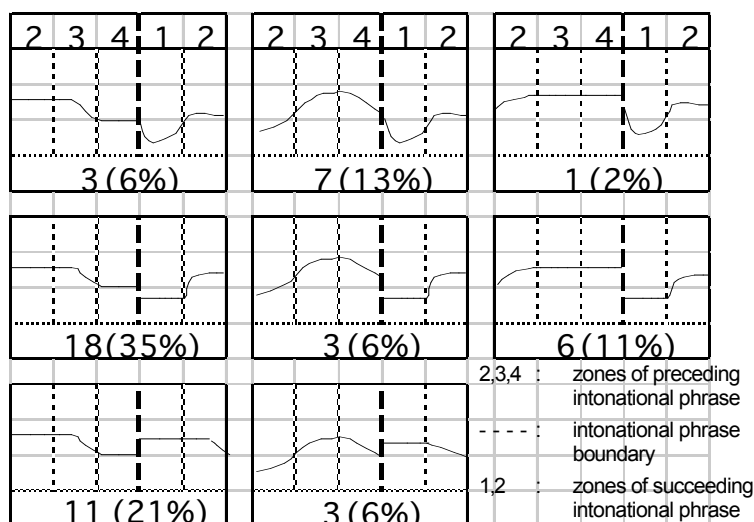


Figure 4. Pitch contours of adjacent, pause-less intonational phrases produced by the non-native learner. The vertical broken line is the phrase boundary. To the left of the boundary are zones 2, 3 and 4 of the preceding intonational phrase. Contours common to the preceding phrase are arranged by column. To the right of the boundary are zones 1 and 2 of the succeeding intonational phrase. Contours common to the succeeding phrase are arranged by row. A total of 52 pairs were observed and clustered by hand. Numbers indicate the frequency of each contour combination.

L1 (Chinese) syllabic tones may interfere with intonation production. As shown in Figure 2 left, the learner tends to use level pitch contours that step up or down depending on L2 pitch accent. The learner may be using consecutive flat tones (L1 tone 1) at two pitch heights to say high and low L2 mora. Pronunciation teaching techniques emphasizing the categorical dichotomy of high and low pitch accents may encourage this behavior. Learners need to learn how to adjust pitch throughout the intonational phrase to create hump-like pitch contours. Graphical representation of pitch contours may help acquire this skill.

5. Conclusion

We compared pitch patterns of Japanese intonational phrases spoken by a native speaker of Chinese. Several types of pitch errors were distinguished according to pitch contour shapes within and across intonational phrases.

We intend to improve our categorical model through analyses of more Chinese learners. Once we have a baseline set of pitch-error models, we will use them in instructor-led Japanese language learning, and develop CALL systems that automatically classify pitch contours.

References

- [1] http://www.mext.go.jp/b_menu/houdou/15/11/03111102.htm (accessed 10 May 2004)
- [2] Society for Teaching Japanese as a Foreign Language, ed. (1991) "Course design in Japanese language institutions" Bonjinsha, Tokyo
- [3] Sato, T. (1998) "Needs assessment of Japanese language learners in Korea and Taiwan" *Tohoku University Linguistic Science*, 2
- [4] Taniguchi, F. (1991) "Status and issues in Japanese pronunciation learning" *Proc. Japanese spoken language prosody*
- [5] Toki, S. (1986) "Textbooks analyzed with respect to spoken language learning" *Japanese Language Learning*, 59, 24-37
- [6] Kawano, T. (1998) "How spoken language skills are explained in the introduction sections of Japanese language textbooks" *Japanese Language Teaching Methods*, 5/1
- [7] Sugito, M. (1996) "Japanese language and Japanese language learning series, vol. 2, Phonetics and phonology, part 1" Meiji Shoin, Tokyo
- [8] Crystal, D. (1997) "A dictionary of linguistics and phonetics, 4th ed." Blackwell
- [9] Boersma, P. and Weenik, D. (2004) "Praat: doing phonetics by computer" <http://www.praat.org/> (accessed 10 May 2004) (Note: Version 4.2 was used in this study. The speech data files accompanying this paper are in Praat format, and contain speech files and their transcriptions.)