F_0 models show Chinese speakers of Japanese insert intonational boundaries and drop pitch

Hiroko Hirano [1], *Keikichi Hirose* [2], *Goh Kawai* [3], *Wentao Gu* [4], *Nobuaki Minematsu* [5]

[1, 2, 5] University of Tokyo, [3] Hokkaido University, [4] Chinese University of Hong Kong

[1,2,5] {hiran,hirose,mine}@gavo.t.u-tokyo.ac.jp, [3] goh@kawai.com, [4] wtgu@ee.cuhk.edu.hk

Abstract

We used a command-response additive F_0 model to analyze F_0 patterns of Japanese spoken by native speakers of Mandarin Chinese. Compared to native speakers of Japanese, we found that Chinese speakers exhibit the following characteristics: (a) higher pitch, (b) more phrases, (c) *bunsetsu* decomposition, and (d) utterance-final plunging. These characteristics physically manifest themselves as: (a) higher baseline F_0, (b) more phrase commands, (c) more accent commands, and (d) negative commands. These characteristics may be subjectively perceived as: (a) tinnier speech (possible L1 marker but does not degrade communication), (b) disjoint phrases (requires mental consolidation), (c) choppy prosodic words (requires reconstruction), and (d) abrupt utterance termination (possibly misconstrued as emphatic or rude). We believe these difficulties arose from tonal and syllable-timed interference, which can be overcome by prosodic control and planning.

Index Terms: F_0 contour, command-response model, L2 learning, Japanese, accent, phrase.

1. Introduction

Chinese learners comprise over two-thirds of learners of Japanese as a second language. In our initial work, we compared prosodic patterns of Japanese declarative sentences uttered by native speakers and second-language learners whose native language is Chinese by observing rise-fall F_0 variations in each phrase and F_0 range variations over the entire utterance [1]. (For brevity, we will refer to native speakers of Tokyo-dialect Japanese as *Japanese speakers*, and non-native speakers of Japanese whose first language is Mandarin Chinese as *Chinese speakers*.) We continued our work in capturing essential differences between native and non-native prosody by quantitatively analyzing F_0 patterns using a F_0 contour generation process model that divides pitch into global sentence intonation and local word accent [2][3]. (We will refer to the F_0 contour generation process model as the *F_0 model*.) This paper extends our previous work by (a) adding more human subjects (15 males and 15 females each in Japanese and Chinese speaker groups), (b) showing gender differences among Japanese speakers, and, most importantly, (c) describing the way Chinese speakers insert intonational boundaries and drop pitch, and how this may negatively affect communication. The following sections describe the corpus (section 2), analysis methods (section 3), and results (section 4).

2. Materials

The corpus is comprised of two similar datasets collected under different circumstances.

2.1. Dataset 1

7 sentences were read aloud by 10 Japanese speakers and 10 Chinese speakers (5 males and 5 females in each group). The Chinese speakers were intermediate to advanced learners of Japanese as a second language, had been living in Japan between 8 and 18 months, and were students at the same Japanese language institute. Speakers read scripts with *furigana* (written pronunciation guides for readers unfamiliar with *kanji*) to avoid mispronunciations caused by lack of lexical knowledge; however, punctuation marks were removed to allow speakers to choose phrase boundaries themselves. Speakers read a fair amount of material besides the 7 sentences we analyzed -- the sentences we analyzed were spoken when the speakers became sufficiently comfortable with the recording task (after "warming up") -- thus we believe these utterances represent each speaker's speech style. Recordings were made in a soundproof booth, using a Sony TCD-D10PRO digital audio tape recorder (sampling rate 48 kHz, resolution 16 bits) and a Sony C-38B microphone. A native Japanese language instructor experienced in phonetic labeling segmented and labeled the utterances at the phone level by visually studying waveforms and spectrograms in *PRAAT* [4].

2.2. Dataset 2

7 sentences identical to Dataset 1 were spoken by 20 Japanese speakers and 20 Chinese speakers (10 males and 10 females in each group). The script format was identical to Dataset 1. The Chinese speakers included graduate, undergraduate, and language institute students. All were intermediate to advanced learners of Japanese, and had been living in Japan between 2 and 30 months. Recordings were made in either a soundproof booth or in a quiet classroom using a computer and a head-mounted microphone (sampling rate 16 kHz, resolution 16 bits). Participants viewed sentences as they appeared on the computer screen, and, guided by prompts, said them aloud. Subjects were allowed to play back their recordings, and to re-record as many times as they pleased. Phones were segmented automatically using *Julian* [5].

3. Methods

3.1. F_0 model

The F_0 model describes F_0 contours in the logarithmic scale as the sum of phrase components, accent components, and a baseline level (lnF_b, where F_b is the baseline frequency). The phrase and accent components are represented by phrase and accent commands, respectively, both of which having magnitude (shown on the vertically axis in Figure 1) and time location (shown on the horizontal axis in Figure 1).

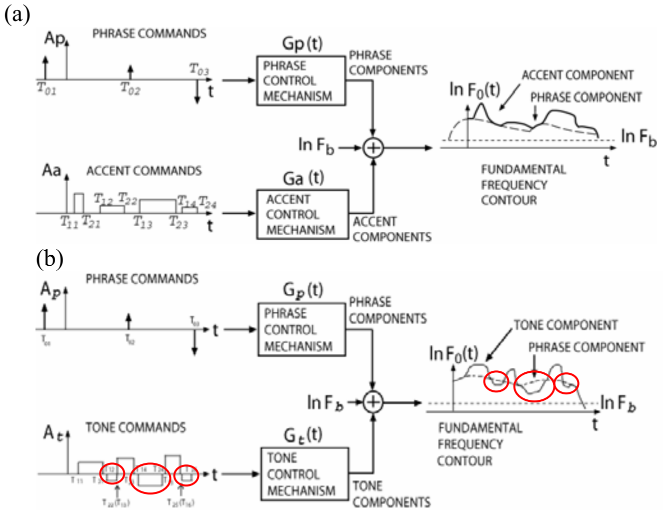

Figure 1: *F_0 contour generation command-response models for Japanese (above) and Chinese (below). Both languages include phrase commands that determine phrase components. Japanese uses accent commands for pitch accents. Accent commands in the Japanese model are always positive for Japanese speakers. Chinese uses tone commands for lexical tones. For example, rapidly falling lexical tones are described by combination of positive and negative tone commands.*

Phrase commands are responsible for the overall shape of sentence intonation. Accent commands characterize local F_0 changes. Both mechanisms are assumed to be critically-damped second-order linear systems. The commands correspond well with linguistic and paralinguistic speech phenomena. The top and bottom figures in Figure 1 show how the F_0 model describes Japanese and Chinese. Note that in Chinese, accent commands are replaced with tone commands representing lexical tones. Unlike accent commands (whose magnitudes are always positive for Japanese speakers), tone command magnitudes may be either positive or negative depending on the lexical tones.

3.2. Estimating F_0, phrase command magnitudes, and accent command magnitudes

420 utterances in Datasets 1 and 2 were downsampled to 10 kHz at 16-bit resolution. F_0 was estimated at 10 ms intervals by modified autocorrelation analysis of LPC residuals. Phrase and accent command magnitudes were first automatically estimated using analysis-by-synthesis [6], yielding F0 contours consisting of continuously-differentiable third-order polynomial segments, then manually corrected using linguistic information and phone alignment. Time constants of the phrase and accent control mechanisms were set at their default values 3/s and 20/s, respectively.

4. Results

Unless stated otherwise, the analyses below are based on the 420 utterances collected from Japanese and Chinese speakers (15 males and 15 females in each group saying 7 Japanese sentences each). JPM, JPF, CHM and CHF respectively stand for Japanese males, Japanese females, Chinese males and Chinese females.

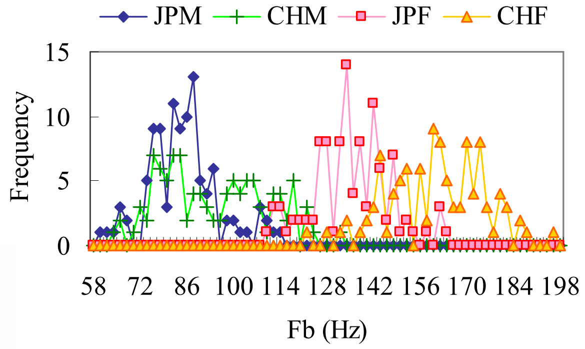

Figure 2: *F_b histograms. Chinese males and females have higher F_b than their Japanese counterparts. Given no reason to suspect physiological differences between same-gender Japanese and Chinese speakers, we speculate there are cultural or linguistic differences that encourage higher F_b for Chinese speakers.*

Table 1: *Mean and standard deviation of F_b in Hz. n=105 for each column. Chinese means are larger than Japanese for both males and females. Chinese variance is larger than Japanese.*

	JPM	JPF	CHM	CHF
mean	84.18	134.38	92.96	159.80
SD	10.70	11.65	15.77	14.33

4.1. F_b (baseline frequency)

Figure 2 shows F_b histograms. Chinese males and females have higher F_b than their Japanese counterparts.

Table 1 lists F_b means and standard deviations. Chinese mean F_b is larger than Japanese for both males and females (Welch's *t*-tests show JPM *vs* CHM: $t(183)=4.72$, $p<.01$, JPF *vs* CHF: $t(200)=14.1$, $p<.01$). Chinese speakers' F_b vary more than Japanese (*F*-tests show JPM *vs* CHM: $F(105,105)$ $248.6/114.4=2.17$, $p<.01$, JPF *vs* CHF: $F(105,105)=205.4/135.7=1.51$, $p<.01$).

4.2. Number of phrase and accent commands

Figure 3 shows histograms of the number of phrase and accent commands found in Japanese speech. Descriptive statistics are shown on Table 2.

A *bunsetsu* is a syntactic unit defined as a content word with *n* succeeding function words ($n \geq 0$). Figure 4 compares the number of *bunsetsu* with the number of phrase and accent commands. In some Chinese utterances (particularly sentences 1, 5, 6, and 7), phrase commands are produced at every *bunsetsu* boundary. Such over-generation results in about double the number of *bunsetsu* found in Japanese utterances (the ratio is greatest in sentences 3 and 4).

While the number of accent commands is roughly the same as the number of *bunsetsu* for Japanese speech, Chinese speech has more accent commands. Figure 5 shows that in the majority of Japanese and Chinese speech there is exactly one accent command per *bunsetsu*; however Chinese speech may have up to 4. By contrast, neighboring *bunsetsu* sharing an accent command (effectively lowering the accent-command-

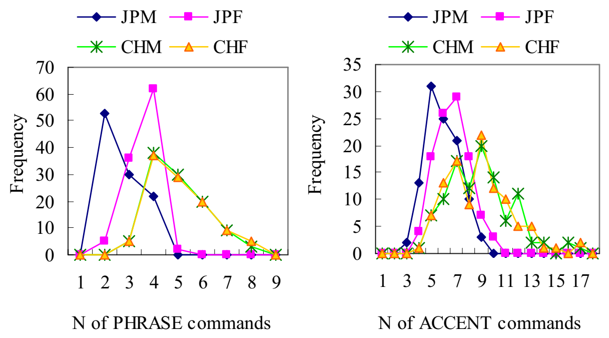

Figure 3: *Histograms of the number of phrase and accent commands found in Japanese speech.*

Table 2. *Descriptive statistics of the number of phrase and accent commands found in Japanese speech. Japanese females produce more phrase and accent commands than Japanese males. Chinese males and females show no difference among them, but together, produce more phrase and accent commands than Japanese females*

	phrase commands				accent commands			
	JP		CH		JP		CH	
	M	F	M	F	M	F	M	F
sample N	105	105	105	105	105	105	105	105
mean	2.7	3.6	5.0	5.0	5.9	6.7	8.9	8.8
SD	0.6	0.4	1.5	1.4	1.3	1.4	2.6	2.6
F-test	t = 2.17**		t = 1.51		t = 1.05		t = 1.01	
t-test	t = 8.92** df = 196		t = 2.94 df =208		t = 4.32** df =208		t = 0.29 df =208	

**: significance at *p*<.01

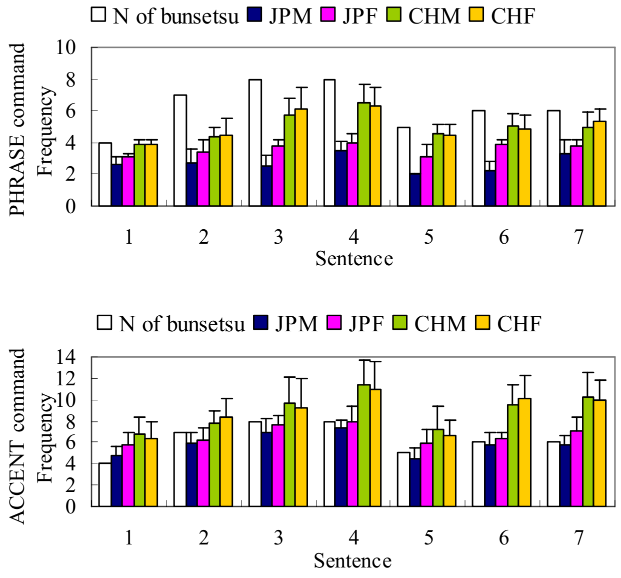

Figure 4: *Mean and SD (protruding T-bar) of the numbers of phrase (top) and accent (bottom) commands. Each leftmost bar (white) indicates the number of* bunsetsu *in the sentence.*

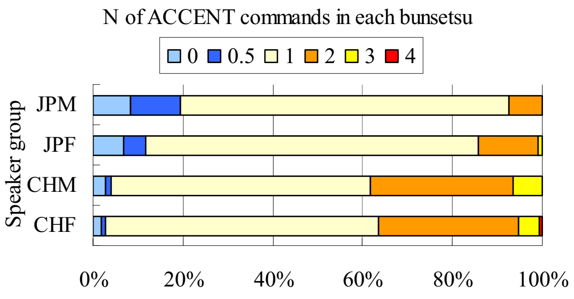

Figure 5: *Ratio of the number of accent commands within a bunsetsu. The number 0 indicates no accent command appeared within the bunsetsu. 0.5 indicates two bunsetsu share an accent command.*

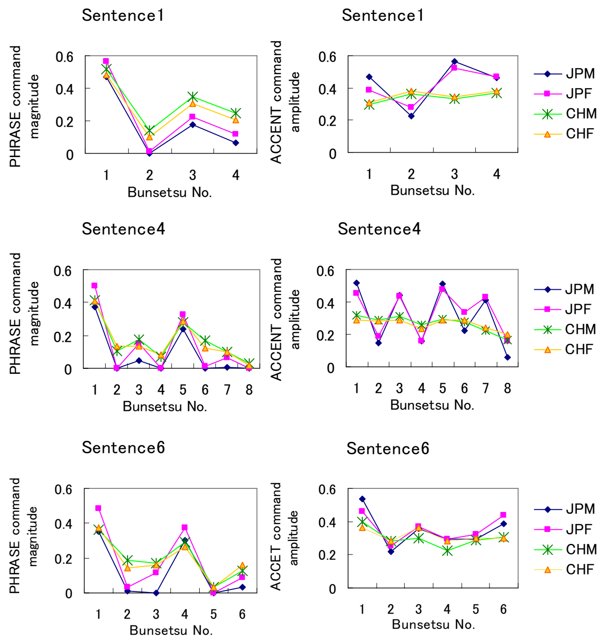

Figure 6: *Mean magnitudes of phrase commands (left) and accent command (right) in each* bunsetsu *for sentences 1 (top), 3 (center) and 6 (bottom). n=105 for each group. Note diversity across Japanese* bunsetsu *(seen as vertical zigzagging). By contrast, Chinese* bunsetsu *are uniform (seen as flatter line segments), resulting in lack of focus on meanings and cadence. Japanese males and females talk differently, whereas Chinese males and females talk the same.*

per-*bunsetsu* ratio) is relatively common in Japanese. Accent sandhi across *bunsetsu*, or a rise-fall followed by a flat-pitch or damped rise-fall are occasionally seen in Japanese but rarely in Chinese.

Hence it may be concluded that (a) Chinese tend to overproduce phrase/accent commands, (b) Japanese males and females talk differently, whereas (c) Chinese males and females talk the same.

4.3. Magnitudes of phrase/accent commands

Figure 6 shows mean magnitudes of phrase and accent commands. Japanese magnitudes tend to consist of a strong command followed by zero or more weak commands.

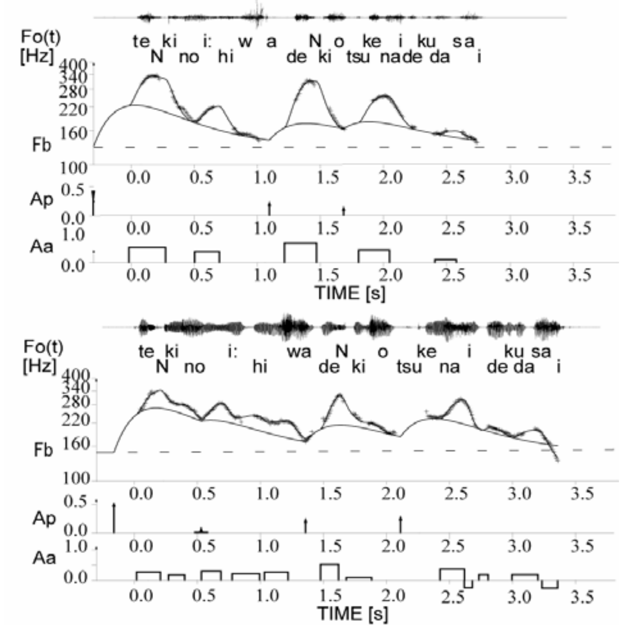

Figure 7: *Analyses of F_0 contours of a Japanese female (above) and a Chinese female (below). This Japanese declarative sentence consists of 4 bunsetsu: "teNkino (weather) / iihiwa (good day) / deNkio (lights) / tsuke (turn on) nai (don't) dekudasai (please)". ("Please don't turn on the lights on days with good weather"). Note negative accent commands for Chinese speakers. None are theoretically necessary, nor were found among Japanese speakers. Also note the regularity in Chinese accent commands*

Chinese magnitudes tend to be the same size, resulting in unexpressive speech.

Figure7 shows F_0 contours of a Japanese female and a Chinese female. Of particular interest are negative accent commands for Chinese speakers. A total of 95 negative accent commands were found in the 210 utterances said by Chinese speakers. None are theoretically necessary, nor were found among Japanese speakers. Furthermore, native speakers perceive the profusion of and regularity within and across *bunsetsu* in Chinese accent commands as phrasal decomposition.

5. Analysis and discussion

We analyzed Japanese spoken by native speakers of Mandarin Chinese. Compared to native speakers of Japanese, F_0 analysis using a command-response additive model showed that Chinese speakers have: (a) higher baseline F_0, (b) more phrase commands, (c) more accent commands, and (d) utterance-final negative commands.

The cause and effect of higher baseline F_0 is unknown, although we could speculate on cultural preferences and L1 marking (i.e., identifying the speaker's L1 based on L1-specific features).

Phrase and accent commands probably increased due to a combination of poor prosodic planning, and tonal and syllable-timed interference. Poor planning is evident in frequent pausing [8]. Pauses might be avoided through syntactic training.

Tonal interference appears as rapid pitch changes within syllables. Pre-pausal pitch plunges preceding phrase-initial pitch rises cannot be adequately explained as phrase control responses that reset the phrase component [2]. Negative accent commands were required in 45 percent of Chinese utterances (95 cases in 210 utterances). Negative accent commands likely correspond to Chinese negative tone commands. We are designing methods to train Chinese talkers to resist intrasyllabic pitch changes.

Learning to adjust pitch across multiple syllables (instead of changing pitch rapidly within syllables) should be the top priority for Chinese speakers. Supplanting lexical tones with phrasal prosody is the key. Accomplishing this should prove particularly effective in avoiding phrase-final pitch plunges, because Japanese syntax places information-laden modals at sentence-final position. Prosody on these function words can drastically alter the sentence's paralinguistic meaning (e.g., the negative accent commands on *tsukenaidekudasai* in Figure 7). Utterance-final negative accent commands are potentially hazardous because abrupt utterance termination might be misconstrued as emphatic, aggressive, accusatory, angry, or rude. Negative accent commands within *bunsetsu* or at non-pre-pausal locations are not misinterpreted, however -- they merely tag the talker as non-native.

The greater number of commands in females may support the common notion that females talk more emphatically than males. Males start utterances with phrase command magnitudes smaller than females, necessitating compensatory larger accent command magnitudes. Sounding like a man or a woman means understanding gender-specific prosody. Text-to-speech systems may benefit by using more commands for female speech, for instance.

We observed that Chinese talkers indiscriminately apply strikingly similar pitch contours to all phrases. Adding contour variety and shift pitch gradually are the first step towards native-sounding speech.

6. References

[1] Hirano, H. and Kawai, G. "Pitch patterns of intonational phrases and intonational phrase groups in native and non-native speech" Proc. INTERSPEECH 2005, Lisbon, Portugal, 761-764, 2005.

[2] Fujisaki, H. and Hirose, K. "Analysis of voice fundamental frequency contours for declarative sentences of Japanese" J. Acous. Soc. Japan (E), 5(4), 233-242, 1984.

[3] Hirano, H., Gu, W., and Hirose, K. "Model-based Analysis of F_0 Contours of Japanese Sentences Uttered by Chinese Speakers" Proc. The 7th Phonetic Conference of China and International Forum on Phonetic Frontiers, Beijing, China, 2006.

[4] Boersma. P. and Weenink, D. "Praat speech analysis software, version 4.5.17" http://www.praat.org/ (accessed 2007-03-27)

[5] "Julius/Julian automatic speech recognition toolkit, version 3.1" http://julius.sourceforge.jp/ (accessed 2007-03-27)

[6] Narusawa, S., Minematsu, N., Hirose, K., and Fujiaski, H. "A method for automatic extraction of model parameters from fundamental frequency contours of speech" Proc. ICASSP, Orlando, FL, 1:509-512, 2002.

[7] Hirose, K., Fujisaki, H., and Seto, S. "A scheme for pitch extraction of speech using autocorrelation function with frame length proportional to the time lag" Proc. IEEE ICASSP, 1:149-152, 1992.

[8] Hirano, H., Kawai, G., Hirose, K., and Minematsu, N., "Unfilled pauses in Japanese sentences read aloud by non-native learners", Proc. INTERSPEECH 2006-ICSLP, Pittsburgh, Pennsylvania, USA, 725-728, 2006