Acquisition of Japanese pitch accent by American learners

Irina A. Shport (University of Oregon)

Abstract

Native speakers of Tokyo Japanese distinguish accented versus unaccented pitch patterns in their production and perception. However, accentual patterns are typically not explicitly taught in JFL classes. This study investigated whether intermediate American learners of Tokyo Japanese had acquired accentual patterns merely from Japanese input without formal instruction focusing on accent. It was found that the learners failed to produce or perceive prescribed pitch patterns at a higher accuracy rate than chance level (56 percent correct in production, 46 percent correct in perception). They tended to produce all words as accented. It was shown that the segmental structure of words affected the accuracy of native and non-native subjects' performance in production and perception tests. Specifically, non-syllabic morae negatively affected the production accuracy of unaccented pitch patterns in both native-speaker and learner groups. In the perception task, however, this effect was found only in the learner group. These findings suggest that (a) the prosody of the accentual patterns is not acquired by learners implicitly, at the intermediate level at least, and (b) teaching of accentual patterns should emphasize how those patterns vary in words with non-syllabic morae, as well as in unaccented words. The presentation of accentual patterns as either rising or falling may be an oversimplification that might contribute to a stronger foreign accent. Additionally, it was found that, contrary to expectations, the performance of native speakers was not perfect in the test (81 percent correct in production, 59 percent correct in perception). Given the variability found among native speakers, objectives for Japanese instruction should not aim for perfect accuracy in the production and perception of accentual patterns.

1. Introduction

To teach or not to teach pitch accent? That is the question that teachers of Japanese as a second language ask themselves. In popular Japanese textbooks, used in schools and colleges in North America (for example, *Genki*, *Nakama*, *Yōkoso*), one can find brief explanations of the Japanese accentual system. On one or two pages, it will typically be stated that there are two types of accentual patterns in Tokyo Japanese, accented and unaccented. In accented patterns, the pitch is falling after an accented mora. In unaccented patterns, the pitch is rising to a plateau in the first mora of a word. Mora is introduced as a tone-bearing unit, and each mora in a word is associated with either a high or low tone.

Such textbook descriptions are based on a traditional Japanese phonological representation of pitch patterns as HL or LH, where H stands for a high tone and L for a low tone, and the pitch pattern of a word is represented as a sequence of high and low tones (Haraguchi 1999; Shibatani 1990; Vance 1987). Textbooks give examples of homonyms that are different only in their accentual patterns ("Accent Type" hereafter), as in [háʃi] 'chopsticks', [haʃí] 'bridge' and [haʃi] 'edge'. Those examples are illustrated by phonological sequences of tones (for example, HLL, LHL, LHH, in the case of the different *hashi* followed by the particle *ga*) and sometimes by graphs visualizing those phonological tone sequences. Information on how the pitch patterns are phonetically manifested is not offered in textbooks or teacher manuals.

Acoustically, pitch is a perceptual correlate of the fundamental frequency (F0) of a sound wave. The human ear can perceive pitch as relatively high or low, and in production, the distinction between accented and unaccented pitch patterns is made in terms of F0 contours. It has been shown in previous empirical research that the traditional phonological representation of pitch patterns where each mora is associated with high or low tone is not true, because F0 contours do not exhibit such sequences (Pierrehumbert and Beckman 1988). Therefore, by providing only a phonological description of the pitch patterns, textbooks set an unnatural goal for students.

To explore textbooks' practices further, even though a phonological explanation of the Tokyo Japanese accentual system is often given, most textbooks do not mark accents in their vocabulary sections or dialogues. As a result, students forget about the existence of pitch accent in Japanese, let alone practice accented and unaccented patterns. Introduction of the pitch-accent contrast thus depends solely on the goodwill of an instructor. According to my observations, instructors most often do not teach or emphasize accentual patterns, except for pointing them out sporadically in homonyms (for example, [ipːai] has a falling HLLL pattern in [ɯdon ípːai] 'one bowl of udon' and a rising LHHH pattern in [ɯdon ipːai] 'lots of udon'). There are numerous reasons for excluding accentual contrast from lessons. To name but a few here: incorrect production of Japanese accentual patterns does not impede comprehension of students' utterances in context; the number of minimal accentual pairs in Japanese is small; pitch contrast in Japanese is viewed as an easier prosodic feature than tone contrast in, for example, Mandarin; mastering of constructions and their functions is prioritized in the popular approach of communicative language teaching.

The question one might ask therefore is whether students acquire accentual patterns of Tokyo Japanese naturally, in the course of being exposed to the target language in the classroom or in other Japanese speaking contexts. In the process of the first language acquisition, young native speakers of Japanese are not taught accentual patterns explicitly. Nevertheless, adult speakers can perceive pitch accent and produce both accented and unaccented pitch patterns (see, for example, Fujisaki, Ohno and Tomita 1996; Kubozono 1987; Pierrehumbert and Beckman 1988). Therefore, the question arises as to whether accentual patterns do have to be taught explicitly in foreign language learning, or whether they can simply be acquired from input.

Ayusawa (2003) offered a comprehensive review of previous studies on the acquisition of the Japanese pitch accent that were conducted for over 30 years, from 1972 to 2003. Those studies differed in terms of the participants' backgrounds and methodology, and their results were not consistent. In production, accentual patterns had been analysed in free speech samples of learners and in elicited dialogues. Ayusawa (2003: 52) found that methodological issues did not allow us to generalize from the first type of production studies. As for elicitation studies, they revealed that learners failed to acquire correct accentual patterns of nouns in their production, regardless of their native language background, length and type of exposure to the Japanese language. In perception, a lot of studies were carried out using *Tōkyōgo akusento kikitori tesuto* [A Perception Test of Accent in Tokyo Japanese], in which participants were asked to mark the location of the accent for accented words and leave unaccented words unmarked. In those studies, the average accuracy of identification of the accent location by learners of Japanese varied from 41 percent to 87 percent. However, the interpretation of the results is difficult, because (a) the average accuracy of Japanese native speakers in this task was not close to 100 percent, and (b) participants who did not study Japanese (such as music majors) performed as well as learners of Japanese in this task (Ayusawa 2003). Therefore, one might wonder if this test is valid for accessing the acquisition of pitch accent. The question of the acquisition of Japanese pitch accent was addressed in this study again, using different methodology. Elicitation of nouns in a sentence frame was used for the production test, and a discrimination task was used for the perception tests. It was examined whether intermediate learners of Japanese who had not been consciously learning or practicing pitch-accent contrast were able to produce and perceive it.

If students are capable of distinguishing pitch-accent contrast in their production and perception, the next question to ask is whether there is an effect of the segmental structure of a word on their performance. A mora is a tone-bearing unit in Japanese, but not all morae have the same segmental structure, and this factor ("Mora Type" hereafter) may affect the production and perception of tones. A prototypical Japanese mora is represented by one orthographic symbol in Japanese *kana* writing, and it is syllabic. It has a (C)V structure, where C stands for a consonant, and V stands for a vowel. There are, however, three types of non-syllabic morae (Imada 1989; Kubozono 1999; Shibatani 1990): the first part of a geminate consonant /C:/, the second part of a long vowel /V:/, and a moraic nasal /N/. These morae are final in bi-moraic heavy syllables of Tokyo Japanese – CVC(:), CV: and CVN, respectively.

The segmental structure of a phrase-initial word appears to affect production of accentual patterns by native speakers. It has been observed in previous research that some Japanese unaccented phrases with phrase-initial heavy syllables do not start from a low tone, and thus violate the "initial lowering rule" proposed earlier for Tokyo Japanese (Hattori 1954; Wietzman 1970; Poser 1984 as cited in Pierrehumbert and Beckman 1988; Tanaka and Kubozono 1999). It was also reported that initially accented words had the same F0 pattern across the accent type, whereas unaccented words did not (Ishihara 2003; Pierrehumbert and Beckman 1988; Sugito 1982; Venditti 2005). Specifically, unaccented CVCVCV and CVC:V words have a rising F0 contour (referred to hereafter as Type 1), whereas unaccented CV:CV and CVNCV words have a flat (or slightly falling) F0 contour (referred to hereafter as Type 2) (Shport 2003, 2006). The F0 contours of these two groups of words were different in terms of the inter-moraic and intra-moraic F0 movements.

Similarly, in second language acquisition research, it has been reported that non-syllabic morae affect the production of accentual patterns (Andreev 2002; Hirata 1993; Minagawa-Kawai and Kiritani 1998; Tsurutani 1996) and identification of the pattern of unaccented words (Ayusawa 2003; Nagano-Madsen 1992). Andreev (2002) observed that, regardless of the student's intention to produce a word as accented or unaccented, the outcome was heavily influenced by the syllabic structure of the word. Ayusawa (2003) summarized that, in the perception test, learners often identified non-syllabic morae as accented, even when they were explicitly instructed that the pitch accent could not occur on such morae. To my knowledge, no study has been conducted to examine the effect of all four possible mora types in bi-moraic syllables (CV, CVC, CV: and

CVN) on the two types of accentual patterns (accented and unaccented) in production and perception. This study aimed to fill this gap by examining the interaction between Accent Type and Mora Type.

A final issue that this study addressed was that of a possible correlation between perception and production of accentual patterns. Ogawara (1997) reported a statistically significant correlation between the perceptual ability to identify accentual patterns in L2 learners' own speech and their production of those patterns. It was concluded that, if students could identify accentual patterns in their own speech, they could produce patterns accurately, and vice versa. However, there was no further experimental research to support this result. One of the reasons for the lack of experimental evidence might be the uncertainty about what the perception mechanisms of one's own speech are, and what would be a valid task through which to access self-perception. This study examined perception of pitch-accent contrast in others' speech and its correlation with production for particular types of tasks.

To summarize, this study investigated the production and perception of pitch patterns by American learners of Japanese as a Foreign Language (JFL learners hereafter) for words with varied segmental structure. The following three hypotheses were tested.

(1) Intermediate JFL learners who are not explicitly taught Tokyo Japanese accentual patterns produce and perceive them less accurately than native speakers of Tokyo Japanese.
(2) Non-syllabic morae negatively affect JFL learners' production and perception of accentual patterns.
(3) A correlation between the accuracy in the production and perception of patterns exists in JFL learners' data.

2. Participants and Materials

2.1. Participants

Two groups of speakers volunteered to participate in this study: (1) an experimental group of 16 American JFL learners at the intermediate level, and (2) a control group of 16 Japanese native speakers. There were eight male and eight female speakers in each group, and their ages varied from 17 to 40 years of age. All JFL learners were students at the University of Oregon enrolled in the third-year Japanese language course. Eleven students had studied Japanese in high school before. Twelve students had been to Japan with lengths of stay not exceeding 12 months. All the JFL learners reported that they had not had any formal instruction in Japanese accentuation prior to this study. Three students had learned a tone language (Mandarin) in a classroom setting, but for no more than one year.

Although the length of exposure of students to the Japanese language varied, they formed a homogeneous group in terms of their listening and speaking proficiency. To establish this, students' listening comprehension scores in the third-year course tests were analysed. Furthermore, seven native speakers of Tokyo Japanese were recruited to holistically rate the foreign accent of students' production on a scale from one to seven, where one corresponded to "foreign-accented" and seven corresponded to "native-like". The interrater reliability coefficient of native speakers' judgements was 0.85 (p<.05), indicating that the judges agreed with each other to a high degree in their ratings. Both listening-comprehension scores and foreign-accent scores were normally distributed, and their variation did not exceed two standard deviations. It can thus be concluded that the JFL learners had similar levels of Japanese proficiency and could be treated as a homogeneous group of speakers for the purposes of the study.

Japanese native speakers took the same production and perception tests as the JFL learners in order to ensure the reliability of the experiment. They reported in a questionnaire that they used Tokyo Japanese on a daily basis, and never used other dialects. All of the speakers were raised and lived in prefectures where Tokyo-type Japanese dominates, according to the *NHK Accent Dictionary of the Japanese Language* (1998): Tokyo, Kanagawa, Saitama, Chiba, Aichi, and Sapporo. Therefore, we can safely assume that the participants spoke the Tokyo-type variety of the Japanese language.

2.2. MATERIALS

48 real Japanese words were selected for the experiment: 24 words with the pitch-accent on the first mora and 24 unaccented words (Table 1). The accentual type of the words was verified with the help of the *NHK Accent Dictionary of the Japanese Language* (1998). In cases when a word had more than one acceptable accentual pattern, the pattern that Japanese native speakers used more frequently in the pilot study was chosen as the target one.

All stimuli were three-mora words with a CV-X-CV structure that varied in the segmental type of the word-medial mora (X). In each accent type, CVCVCV, CVC:V, CV:CV and CVNCV structures were represented by six words each (see Table 1). Thus, the second mora was either a prototypical CV, or one of the non-syllabic morae.

The phonetic environment of the stimuli was controlled as strictly as possible with regard to the vowel quality and the voicing of consonants. Half of the stimuli had the structure Co-X-kV (words 1–3 in each cell of Table 1); the other half of the stimuli had the structure Ce-X-kV (words 4–6 in each cell of Table 1), where X is a varied word-medial mora. Thus, either [o] or [e] preceded the second mora and the voiceless velar stop [k] followed it, with one exception, [koŋgo] 'hereafter'. Word-initial consonants were voiceless obstruents /t/ or /k/ with two exceptions, [hoŋki] 'seriousness' and [hoːka] 'arson'.

Table 1: Stimuli words

	CVCVCV	CVCːV	CVːCV	CVNCV
Accented	1. kókokɯ 故国 'homeland' 2. tónoko との粉 'polishing powder' 3. kósaka 小阪 '(name)' 4. tétsɯko てつ子 '(female name)' 5. téɾɯko てる子 '(female name)' 6. tébako 手箱 'box'	kókːo 国庫 'treasury' kókːa 国家 'state' tókːa 特価 'bargain price' kékːɯ 結句 'poem's last line' kékːi 決起 'rouse to action' tékːi 適期 'good timing'	tóːka 等価 'equivalence' kóːka 校歌 'school song' kóːko 公庫 'municipal treasury' téːki 定期 'regular' téːko 艇庫 'boathouse' kéːko 稽古 'practice'	kóŋki 今期 'this term' kóŋka 婚家 'husband's family' kóŋkɯ 困苦 'suffering' téŋka 天下 'whole country' ténko̥ 典故 'authentic precedent' kéŋka 県下 'prefecture'
Unaccented	1. kokʲakɯ 顧客 'patron' 2. kohakɯ 琥珀 'amber' 3. komakɯ 鼓膜 'tympanum' 4. tenɯki 手抜き 'negligence' 5. temaki 手巻き 'hand-rolled' 6. terako 寺子 'pupil'	tokːa 徳化 'moral influence' tokːi 突起 'projection' kokːi 国旗 'national flag' tekːa 鉄火 'gunfire' tekːi 鉄器 'ironware' kekːa 結果 'result'	hoːka 放火 'arson' toːka 透過 'penetration' koːka 降下 'landing' keːka 経過 'development' keːki 景気 'market' keːkɯ 警句 'epigram'	koŋki 根気 'patience' koŋgo 今後 'hereafter' hoŋki 本気 'seriousness' teŋka 点火 'ignition' teŋkɯ 転句 'revolving phrase' keŋka 喧嘩 'quarrel'

Note. C = consonant; V = vowel; N = moraic nasal. Words are given in the IPA transcription.

In order to control for the familiarity of the stimuli, most stimuli used in the study were words rarely used in everyday conversation. Both groups of participants were thus asked to produce words they did not necessarily know or use frequently. It was therefore assumed that similar mechanisms were employed by both groups in the production of accentual patterns of the target words on the basis of accent-type graphs (see the procedure below). In both production and perception tasks, the same 48 stimuli were used.

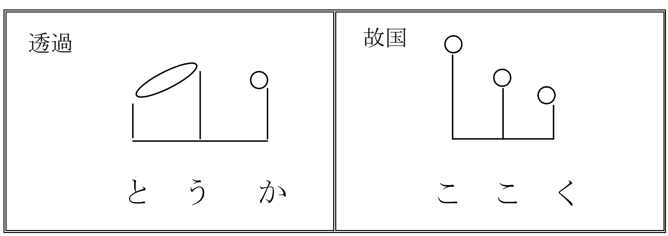
3. Procedure and Analysis

3.1. Production Test

In the production experiment, the target words were randomized and presented to the speakers on stimuli cards, with one word per card (Figure 1). The representation of accentual patterns of words was adopted from Matsuzaki *et al*. (as cited in Hirata 1999: 102 and Kawano 2004). Each small circle represented one mora (Figures 1a and 1b), and ellipses represented two-mora syllables where the syllable coda was a non-syllabic mora (Figure 1a). The relative pitch of each syllable was shown by its height relative to other syllables. Two pitch patterns were represented on graphs: a rising, or unaccented, pattern (Figure 1a) and a falling, or accented, pattern (Figure 1b).

Fig 1: Samples of the stimuli cards

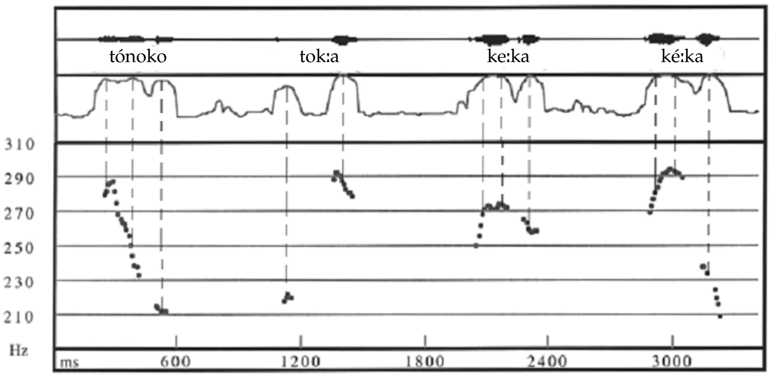
a) b)



Word readings in the Japanese orthography (*hiragana*) were printed under the graphs. Appropriate Chinese characters (*kanji*) appeared at the upper-left hand corner of a card in order to establish the desired degree of familiarity of a word for native speakers of Japanese. For example, the word [toːka] has the meaning 'the tenth day' when written with the characters 十日 and the meaning 'penetration' when written with the characters 透過 . Undoubtedly, 'the tenth day' is much more frequently used than 'penetration'. To establish an association with the less frequently used meaning 'penetration', the characters 透過 were printed on the card.

All speakers read brief explanations regarding the Japanese accentuation system, the graphic representation of pitch patterns and the production test task. The explanations were similar to what one can find in Japanese textbooks (see the introduction section of this paper). Plenty of examples (written and recorded) were provided along with the task instructions. After a short practice with the stimuli cards, the speakers were instructed to read the words embedded in the carrier sentence *Jíʃo ni ____ ga arimaseɴ* 辞書に________がありません 'There is no ______ in the dictionary' at their normal speaking rate. They were allowed to self-correct themselves.

Fig. 2: Measurements of F0 in the data



Speakers' production of the target words (768 tokens) were tape-recorded in a quiet room, digitized with the 22050 Hz sampling rate, and analysed using the speech analysis software MacQuirer, SCICON R&D, INC. The fundamental frequency (F0) was recorded for each mora of the target words by using pitch-contour displays. In CV syllables, as in [tónoko] (Figure 2), the F0 values were measured in the steadiest parts of the vowels, according to the intensity display. In CVː and CVN syllables (as in [keːka] and [kéːka]), the F0 of the first mora was measured in the first quarter of the syllable, and the F0 of the second mora was measured in the last quarter of the syllable according to the waveform displays. No F0 measurements were taken for the second mora of CVCː syllables (as in [tokːa], Figure 2), because geminate obstruents manifest in phonetic silence. Therefore, only two F0 values were recorded for words with geminates, whereas three F0 values were recorded for all other words.

To make the devoicing of word-final vowels [i] and [ɯ] less likely, the target words were followed by the voiced obstruent [g] in the sentence frame *Jíʃo ni ____ ga arimaseɴ*. This notwithstanding, some speakers still devoiced [i] and [ɯ] so that their F0 could not be measured. Tokens with only one F0 measurement (for example, words with geminate consonants where the last vowel was devoiced, as in [toːki̥]) constituted 0.7 percent of the data and were not included in the analysis.

After the F0 values were recorded, the difference in F0 between the first and the third morae ($\Delta\ F0_{\mu1\text{-}\mu3}$) was calculated for each token (see examples in Table 2). This measure estimated an overall pitch pattern of each word. A positive value of $\Delta\ F0_{\mu1\text{-}\mu3}$ indicated that the pitch was falling across the word, whereas a negative value indicated that it was rising across the word. In the latter case, the word's pitch pattern was undoubtedly unaccented. However, in the former case, it was necessary to assess whether the pitch fall was sharp enough to consider the word as accented.

Tab. 2: Accuracy assessment of the pitch patterns produced by a female speaker

Tokens	µ1 (Hz)	µ3 (Hz)	$\Delta\ F0_{\mu1\text{-}\mu3}$ (Hz)	Produced	Expected	Score
tonoko	270	250	20	unaccented	accented	0
kéŋka	281	233	48	accented	accented	1
tokːa	221	289	–77	unaccented	unaccented	1
kéːka	280	212	68	accented	unaccented	0
99 %-lower bound $\Delta\ F0_{\mu1\text{-}\mu3}$			36			

In order to make such judgements, the lower value of the 99 percent confidence interval of the positive $\Delta\ F0_{\mu1\text{-}\mu3}$ was calculated for each individual speaker. This measure allows us to say with a high degree of confidence how steep the F0 fall across a word should be for each speaker to consider its pitch pattern as being accented. Consider as an example one female speaker's data in Table 2. The lower-bound $\Delta\ F0_{\mu1\text{-}\mu3}$ value of the 99 percent-confidence interval, which served as a cutoff value for accented and unaccented words in the data of this speaker, was 36 Hz. If the $\Delta\ F0_{\mu1\text{-}\mu3}$ of a particular token was higher than 36 Hz (e. g. [kéŋka] and [kéːka] in Table 2) the token was considered with 99 percent certainty to be accented. Otherwise, it was considered to be unaccented (for example, [tonoko] and [tokːa]). Lastly, the pitch patterns of words produced by the speakers were compared with expected pitch patterns, and judged as being accurate (1) or inaccurate (0).

The average cutoff $\Delta\ F0_{\mu1\text{-}\mu3}$ value between accented and unaccented words was 63 Hz for female and 44 Hz for male native speakers, and 56 Hz for female and 41 Hz for male JFL learners. It may be roughly estimated that the accent F0 fall was around 60 Hz for females and around 40 Hz for males, regardless of their L1.

3.2. Perception Test

It has been established in previous perceptual studies that a change in fundamental frequency is directly related to a change in pitch (Ladefoged 2001: 166). In this study, it was assumed that a listener's ability to perceive F0 contours of words was directly related to his or her ability to perceive the pitch patterns of those words as accented or unaccented.

In the perception test, the subjects listened to 48 word sets. Each set consisted of three isolated words: one target word (see Table 1) and two distracters having the CV-X-CV structure (Figure 3). In the test worksheets, each word was written in Chinese characters (*kanji*) and Japanese orthography (*hiragana*). The pitch pattern of the target word differed from those of the distracters. For example, in Figure 3, the target word [kohakɯ] 琥珀 'amber' is unaccented, whereas the other two words have accented pitch patterns. The subjects were asked to listen to the stimuli set and blacken out the word which they perceived as having a different pitch pattern from the other two words (AXY-discrimination task). Problem sets were presented to the listeners only once, with two-second intervals between words, and with six-second intervals between sets. Subjects' answers were judged as being correct (1) or incorrect (0) in each set.

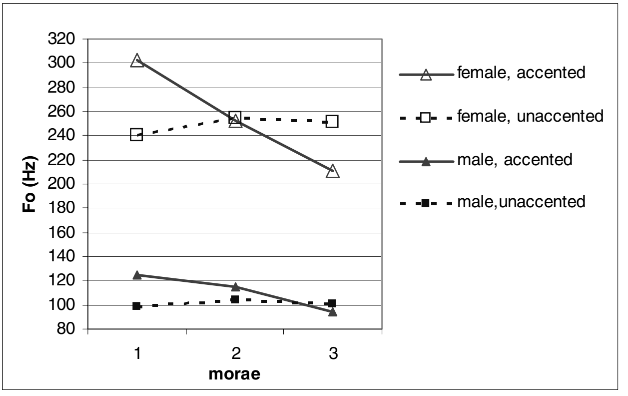
Fig. 3: Sample of the AXY discrimination task.

S	月下	琥珀	短歌
	D げっか	D こはく	D たんか

S = the same pitch pattern in all three words; D = a different pitch pattern from other two words.

Words produced by one female and one male native speaker of Tokyo Japanese were used to compile the perception test materials. 48 stimuli presented in Table 1 and 96 distracters having the same CV-X-CV structure were recorded. To establish the validity of the words as the test stimuli, the F0 of each mora in the target words was measured and the F0 contour of each word was analysed. Averaged F0 values of words are plotted in Figure 4. Pitch patterns in female and male production exhibited the expected phonetic cues for distinguishing accented and unaccented words, that is, falling vs. rising or flat patterns. Each individual production had an expected pitch pattern as well. The perception test stimuli thus had valid pitch patterns to be used in the AXY discrimination task.

Fig. 4: The mean F0 values of the external perception test stimuli



In the test, half of the stimuli sets were produced by the female speaker, and the other half were produced by the male speaker. The order of presentation of the sets was counterbalanced. A perception test followed a production test with at least a two-week interval between them.

4. RESULTS

Factorial analysis of variance (ANOVA) is a statistical procedure that was used to examine the possible effects and interactions of factors controlled in this study by comparing mean accuracy scores for all word types. The production and perception scores of the subjects were submitted to two separate ANOVAs with three factors: (1) Native Language, (2) Accent Type, and (3) Mora Type. The Native Language factor had two levels: native speakers and JFL learners. The Accent Type factor also had two levels: initially accented and unaccented words. The Mora Type factor had four levels corresponding to four types of words with different segmental structure: CVCVCV, CVC:V, CV:CV, and CVNCV words.

A significant effect of Native Language was found for both production [$F(1,1520)=118.68$, $p<.0001$] and perception [$F(1,1520)=30.94$, $p<.0001$] scores. Consequently, since native speakers and JFL learners produced and perceived pitch patterns with significantly different accuracies, their data were analysed separately in the following Accent Type x Mora Type ANOVAs and pairwise comparisons (Tukey HSD tests) comparing mean F0 values in all word types. An alpha level of 0.05 was used for all the statistical tests, indicating that the probability that a significant result was yielded by chance was only 5 percent.

4.1. Production Test Scores

As was expected, native speakers produced pitch patterns more accurately than JFL learners, 81 percent vs. 56 percent respectively. Because the scores of the students were slightly above 50 percent, their performance, on average, was at chance level. In other words, their accurate productions of pitch patterns could have been the result of chance rather than proficiency. The performance of the native speakers of Tokyo Japanese was clearly above that of chance, but, surprisingly, they did not have close-to-perfect scores.

Fig. 5: The mean percentage of correct answers for each word type in the production test

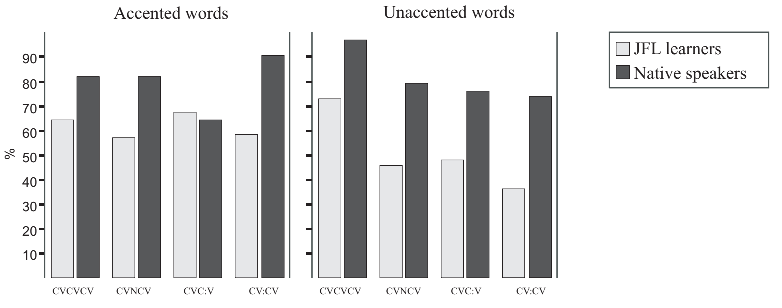


Figure 5 shows the percentage of words produced with correct pitch patterns with regard to native language, accent type and mora type. Overall, native speakers produced approximately the same percentage of accented and unaccented words correctly: 80 percent and 82 percent, respectively. The graphs show that the scores of native speakers seem to be affected more by mora type rather than by accent type. For accented words, 82 percent of CVCVCV words, 82 percent of CVNCV words, 65 percent of CVC:V words, and 91 percent of CV:CV words were produced with correct pitch patterns. For unaccented words, 97 percent of CVCVCV words, 79 percent of CVNCV words, 76 percent of CVC:V words, and 74 percent of CV:CV words were produced with correct pitch patterns. The accuracy of pitch patterns in accented words with geminate obstruents and unaccented words with non-syllabic morae yielded the lowest score, below 80 percent.

An Accent Type x Mora Type ANOVA on production scores of the native speakers revealed a significant effect of mora type [$F(3,760)=8.16$, $p<.05$] and its interaction with accent type [$F(3,760)=6.70$, $p<.05$]. The effect of accent type, on the other hand, failed to yield a significant level.

Pairwise comparisons of mean accuracy scores for all word types within the accentual patterns revealed that the score for accented words with geminates was significantly smaller than the scores for other accented word types. The score for unaccented CVCVCV words was significantly higher than scores for unaccented words with non-syllabic morae.

As for JFL learners' production scores, Figure 5 suggests that unaccented words with non-syllabic morae were produced less accurately than accented words with non-syllabic morae. The percentage of correct scores in unaccented versus accented words was 46 percent vs. 57 percent in CVNCV words, 48 percent vs. 68 percent in CVC:V words, and 37 percent vs. 58 percent of CV:CV words. However, unaccented words of the CVCVCV structure were produced with better accuracy than accented words of the same structure, 73 percent and 65 percent, respectively.

An Accent Type x Mora Type ANOVA on the production scores of the JFL learners yielded significant effects of accent type and mora type, [$F(1,760)=10.24$, $p<.05$] and [$F(3,760)=7.05$, $p<.05$], respectively. The interaction of factors was also significant [$F(3,760)=3.87$, $p<.05$], and this was explored in subsequent pairwise comparisons of mean accuracy scores for all word types within the accentual patterns. Among all the word types, only unaccented CVCVCV words were significantly different from unaccented words with non-syllabic morae. Specifically, the difference was 31 percent between unaccented CVCVCV and CVNCV words, 32 percent between unaccented CVCVCV and CVC:V words, and 45 percent between unaccented CVCVCV and CV:CV words, $p<.05$.

In summary, native speakers produced pitch patterns with higher accuracy than JFL learners, except for accented words with geminates. Both groups of speakers produced pitch patterns of unaccented CVCVCV words more accurately than patterns of unaccented words with non-syllabic morae. This result supported previous research findings that non-syllabic morae negatively affected correct production of pitch patterns of unaccented words (Aoki 1990; Oguma 2000). However, this did not hold true for accented words. Therefore, hypothesis 2, that non-syllabic morae negatively affect the production of pitch patterns by JFL learners, was only partially supported by the empirical evidence of the present study.

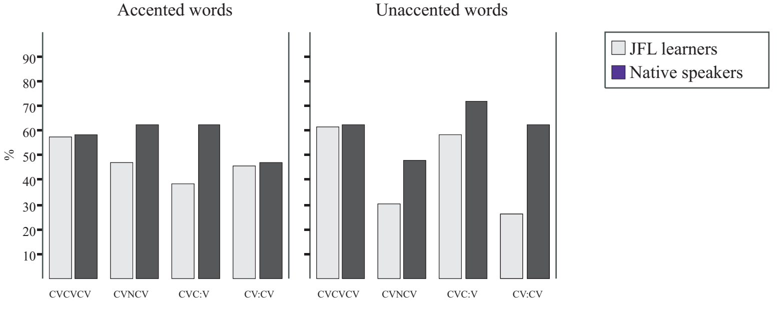
The performance of native speakers was not perfect, as it was not close to 100 percent correctness, except for a few word groups. Several explanations of this result come to mind. One is that the stimuli in this study were not frequently used words, and the low familiarity of the stimuli influenced the scores of native speakers. Another explanation might be that the real-life pitch accent norms differ from those given in the *NHK Accent Dictionary of the Japanese Language* (1998), or that accent norms became ambiguous among speakers of the Tokyo dialect, especially in the case of accent-contrastive pairs like [ték:i] 適期 'good timing' and [tek:i] 鉄器 'ironware'. Yet another possibility is that individual variation in native speakers overcomes pitch accent norms. Additionally, the production task in the experiment was not ecological, because native speakers do of course not have pitch pattern graphs in front of their eyes when they produce pitch patterns in natural speech. Thus, such factors amounting to the "unnaturalness" of the task might have influenced the native speakers' performance.

4.2. Perception Test Scores

Native speakers performed significantly better than JFL learners in the perception test: 59.4 percent and 45.6 percent of correct responses, respectively, $p<.05$. In contrast to the production task, both groups of the speakers performed at close to a chance level (50 percent) in this task.

Figure 6 shows the percentage of pitch patterns correctly discriminated by native speakers and JFL learners. For accented patterns, native speakers correctly discriminated 58 percent of CVCVCV words, 63 percent of CVNCV words, 63 percent of CVC:V words and 47 percent of CV:CV words. For unaccented patterns, 63 percent of CVCVCV words, 48 percent of CVNCV words, 72 percent of CVC:V words and 63 percent of CV:CV words were discriminated correctly. There was no evidence that accented words were discriminated better than unaccented words. The lowest scores were observed in accented words with long vowels and unaccented words with moraic nasals.

Fig. 6: Percentages of correct answers in the perception test



The overall scores of the native speakers were higher than the scores of JFL learners, except for CVCVCV words and accented words with long vowels, where the scores were approximately the same. As Figure 6 depicts, scores for accented words with long vowels were similar because native speakers performed worse for accented words with long vowels than for other accented words, not because JFL learners performed better for this group of words.

An Accent Type x Mora Type ANOVA on the perception scores of the native speakers yielded similar results to the analysis of variance on their production scores, in that the effect of Mora Type and its interaction with Accent Type were significant, [$F(3,760)=2.73$, $p<.05$] and [$F(3,760)=3.43$, $p<.05$], respectively. But the effect of Accent Type failed to yield a significant level. The interaction was examined in subsequent pairwise comparisons of mean accuracy scores for all word types within the accentual patterns. Only one significant difference in the scores was found, namely, unaccented words with geminates were discriminated significantly better (by 24 percent) than unaccented words with moraic nasals, $p<.05$.

As for the JFL learners' scores for accented words, 57 percent of CVCVCV words, 47 percent of CVNCV words, 39 percent of CVC:V words and 46 percent of CV:CV were discriminated correctly. In the case of unaccented words, the numbers for correct discrimination stand at 61 percent of CVCVCV words, 30 percent of CVNCV words, 58 percent of CVC:V words and 26 percent of CV:CV words. Figure 6 suggests that non-syllabic morae negatively affected the perception of pitch patterns by JFL learners, whereas the segmental structure of words had no effect on the perception scores of the native speakers.

Similarly to native speakers' results, in JFL learners' perception scores, the effect of Mora Type was significant [$F(3,760)=9.28$, $p<.05$], as was its interaction with Accent Type [$F(3,760)=7.07$, $p<.05$]. Accent Type, on the other hand, turned out to be non-significant. Pairwise comparisons of word types revealed that:

(1) accented CVCVCV words were discriminated 19 percent better than accented words with geminates, $p<.05$;
(2) unaccented CVCVCV words were discriminated 31 percent better than unaccented words with moraic nasals, and 35 percent better than words with long vowels, $p<.05$;
(3) unaccented CVC:V words were discriminated 28 percent better than unaccented words with moraic nasals, and 32 percent better than unaccented words with long vowels, $p<.05$.

These results thus support hypothesis 2 of this study, that non-syllabic morae negatively affect perception of pitch patterns by JFL learners.

4.3. Correlation between Perception and Production

In order to examine a possible correlation between production and perception, the Pearson correlation coefficients between scores of both tests were calculated for native speakers, JFL learners, and each individual speaker. The correlation was significant in the native speakers' data (r =.07, p<.05), but not in the students' data. Correlation coefficients in some individual speakers' data were also significant, but they were very small. They ranged from r=.65 (the biggest correlation between production and perception scores of a female native speaker) to r=-.27 (correlation between production and perception scores of a male native speaker). Small correlation coefficients revealed that there was no meaningful relationship between the correctness of answers in the production and perception tests, although some of them reached statistical significance.

There are several possible explanations for the absence of a meaningful correlation between production and perception scores. First, production and perception of pitch accent may be two different processes not related to each other. People who produce Japanese pitch patterns correctly do not necessarily have an ability to distinguish them in their perception, and vice versa. Secondly, it might be the case that the nature of the production and perception tasks was different, and not ecological: in real speech, speakers do not produce pitch patterns according to graphs, and do not have to discriminate between pitch patterns, except for a limited number of homonyms.

5. Discussion

This study revealed that the task performance accuracy was affected by the segmental structure of words more than by the accentual type of words. Regardless of the native language, pitch patterns were discriminated poorly in both accented and unaccented words. This contradicts the results obtained in the studies of Nagano-Madsen (2002) and Nishinuma *et al*. (1996), who found that unaccented words were identified significantly better than accented words. Although a discrimination task was used in this study and an identification task was used by Nagano-Madsen and Nishinuma *et al*., the discrepancy of results cannot be solely explained by methodological differences. For example, Funatsu and Inouchi (1997a, 1997b) used the identification task as well in their perception test, but they did not find a significant effect of the accentual type on the accuracy of the pattern perception, thus supporting the results of the present experiment.

Perception scores were affected by the mora type for both groups of speakers, in particular with regard to pitch patterns of unaccented words. In CVCVCV words JFL learners discriminated pitch patterns as well as native speakers did, but non-syllabic morae negatively affected their performance. The question thus arises as to why the segmental structure of words affects the perception of JFL learners' pitch patterns. It has been shown that unaccented CVCVCV and CVC:V words have different F0 contours from unaccented CV:CV and CVNCV words (Shport 2003, 2006). One explanation could thus be that, if learners straightforwardly associate the unaccented pattern with the rising pitch contour, they will have difficulty in discriminating or identifying the pattern in unaccented words with long vowels and moraic nasals that have relatively flat F0 contours. On the other hand, they will identify patterns of CVCVCV words and words with geminates better, because they have a rising F0 contour. Indeed, JFL learners tended to better discriminate the accentual type of CVCVCV words than that of words with non-syllabic morae, whereas native speakers did not show such a tendency.

This finding is important from the pedagogical point of view, because even when the pitch accent is marked in the textbooks consistently (as in, for example, *Japanese for Everyone* and *Japanese: The Spoken Language*), it is marked very simplistically, assuming that accented pitch patterns are falling and unaccented pitch patterns are rising, regardless of the segmental structure of words. In other words, congruity of the accent-type pitch pattern is assumed, which is not the case for unaccented words in Tokyo Japanese.

How pitch patterns should be taught is a complex issue, because current phonological models do not necessarily lead to phonetically correct accentuation. Phonetic reality of the pitch pattern depends both on the presence of an accent and on the segmental structure of a word and its syntactic position. Hasegawa (1995) gave another example of textbook explanation of accentual patterns which might lead to unnatural pronunciation. When a lexically accented vowel is devoiced as in [ʃí̥ki] "four seasons", it does not, contrary to a textbook explanation, have a higher pitch than that of the flanking vowels. Also, when an unaccented vowel is devoiced, as in [háʃi̥ ga] "chopsticks-Nominative", it cannot be lower in pitch than the flanking vowels. Yet another phonetic manifestation of pitch realization is F0 "peak delay" in accented heavy syllables (Ishihara 2003; Sugito 1982). Hasegawa (1995) concluded that simplified phonological presentation of the pitch-accent contrast as rising vs. falling pitch patterns was insufficient for its successful acquisition. Without phonetic details, a textbook explanation can confuse students and lead to incorrect accentual pattern production. Hasegawa therefore suggested that learners should focus on imitation of native speakers' prosody, rather than trying to reproduce it from textbook pages. This recommendation can be easily incorporated into the classroom following the audiolingual approach, that is, classrooms where "native-speaker-like pronunciation is sought" (Richards and Rogers 2001: 156). However, in the framework of communicative language teaching, where the goal is simply comprehensible pronunciation, Hasegawa's recommendation is not applicable.

To summarize, the results of this study reveal that American learners of Japanese encounter the following problems in acquisition of pitch accent:

(1) learners tend to produce all words as being accented, with falling F0 movement, and
(2) learners tend to produce and perceive pitch patterns of words with non-syllabic morae (moraic nasals, geminates, and long vowels) incorrectly.

It was also found that there was no correlation between perception and production in both L1 and L2. In other words, better production does not imply better perception of pitch patterns and vice versa. Production and perception of suprasegments of speech appear to develop independently from each other. This result calls for further investigation into this issue. The further examination of acquisition of the Japanese pitch accent is motivated not only by the lexical contrastive function of the accent, but also by the role it plays in the prosodic system of Japanese. It was shown in previous research that prosodic contours are affected by the manifestation of the pitch accent, for example, the place and duration of the pitch rise in questions depends on the accentual type and the segmental structure of the question-final word (see Ayusawa's research on intonation 2003). Thus, to achieve a native-like intonation, the acquisition of pitch accent is necessary.

The methodology used in the production and perception tests constitutes a limitation of this study. It remains unknown at present which methods are more valid and reliable for the assessment of production and perception of suprasegments such as pitch accent. To investigate why the test scores of native speakers were far from perfect, an examination of whether the task formats influenced the scores might be beneficial. One could, for example, compare native speakers' performance in a reading task or discourse, where pitch patterns are not graphically represented, and their performance in a card task, where pitch patterns are graphically represented. Another possibility could be an accent identification test with a subsequent comparison of the results with those of an AXY-discrimination test. One could furthermore try using more familiar tokens instead of the unfamiliar ones used in this study. Using different methods might lead to better scores of learners in tests.

6 Conclusions

This study found that intermediate learners of Japanese who were not taught accentual patterns of the target language failed to produce or perceive them at a higher accuracy rate than the chance level. Overall, they tended to produce all words as being accented, with falling F0 movement. It seems that learners of Japanese cannot acquire Japanese pitch accent from input only, without explicit instructions and practice, after only two and a half years of studying the language at least. However, if accentual patterns are taught, they cannot be simplified as rising and falling pitch contours only. It was shown that learners are less accurate in the production of unaccented patterns of words with all non-syllabic morae, and in the perception of unaccented patterns of words with long vowels and moraic nasals. This suggests that teaching of the accentual patterns should also emphasize how those patterns vary in words with different segmental structures. Finally, this study found no relationship between production and perception accuracy, which suggests that practicing discrimination of pitch patterns might not help to improve students' L2 prosody.

References

Andreev, Anton (2002): *A Study on Acquisition and Interlanguage Features of Second Language Word-Accent*. Unpublished Abstracts of the Oxford Workshop on Japanese Linguistics.

Aoki, Haruo (1990): *Amerikajin ni yoru nihongo no hatsuon* [Japanese Pronunciation of American Learners]. In: Miyoko Sugito (ed.): *Kōza – nihongo to nihongo kyōiku* [A Course in Japanese Language and Japanese Language Education]. Vol III. Tokyo: Meiji Shoin, pp. 219–233.

Ayusawa, Takako (2003): Acquisition of Japanese Accent and Intonation by Foreign Learners. In: *Journal of the Phonetic Society of Japan* 7, 2, pp. 47–58.

Banno, Eri (1999): *Genki: An Integrated Course in Elementary Japanese*. Tokyo: The Japan Times.

Fujisaki, Hiroya, Ohno Sumio and Tomita Osamu (1996): *On the Levels of Accentuation in Spoken Japanese.* Paper Presented at the International Conference on Spoken Language Processing 96, Philadelphia.

Funatsu, Seiya and Mayako Inouchi (1997a): Perception of Japanese Accent by Native Russian Speakers. In: *Spoken Japanese Language Education: Looking to the 21st Century.* Tokyo: The National Language Research Institute, pp. 37–60.

—— (1997b): Sutoresu akusento gengo o bogo to suru washa no pitchi akusento no chikaku [Perception of the Pitch Accent by Native Speakers of a Stress-Accent Language]. In: *Dai 11 kai nihon onsei gakkai zenkoku taikai yokōshū* [Proceedings of the 11th National Conference of the Phonetic Society of Japan], pp. 65–70.

Haraguchi, Shosuke (1999): Accent. In: Natsuko Tsujimura (ed.): *The Handbook of Japanese Linguistics.* Malden: Blackwell Publishers, pp. 1–30.

Hasegawa, Yoko (1995): Against Marking Accent Locations in Japanese Textbooks. In: *Japanese Language Education Around the Globe* 5, pp. 95–103.

Hirata, Yukari (1993): *Perception of Japanese Geminated Stops at Word and Sentence Levels by Japanese and Learners of Japanese*. Paper Presented at the Midwest Conference on Asian Affairs.

—— (1999): *Effects of Sentence- vs. Word-Level Perceptual Training on the Acquisition of Japanese Durational Contrasts by English Native Speakers*. Paper Presented at the International Congress of Phonetic Sciences 99, San Francisco.

Imada, Shigeko (1989)*: Kyōshi-yō nihongo kyōiku handobukku: hatsuon* [The Teachers' Handbook of Japanese as a Foreign Language: Pronunciation]. Tokyo: Bonjinsha.

Ishihara, Takeshi (2003): *A Phonological Effect on Tonal Alignment in Tokyo Japanese.* Paper Presented at the International Congress of Phonetic Sciences 2003, Barcelona.

Jorden, Eleanor H. and Noda Mari (1987): *Japanese: The Spoken Language*. New Haven: Yale University Press.

Kawano, Tomonobu (2004): *Ichinichi 10-pun no hatsuon renshū* [Daily 10-Minute Practice for Pronunciation]. Tokyo: Kuroshio.

Kubozono, Haruo (1987): *The Organization of Japanese Prosody.* Unpublished Ph. D. Dissertation, University of Edinburgh.

—— (1999): Mora and Syllable. In: Natsuko Tsujimura (ed.): *The Handbook of Japanese Linguistics*. Malden: Blackwell Publishers, pp. 31–61.

Ladefoged, Peter (2001): *A Course in Phonetics*. USA: Harcourt College Publishers.

Makino, Sei'ichi, Yukiko Abe-Hatasa and Kazumi Hatasa (1998): *Nakama: Japanese Communication, Culture, Context.* Boston: Houghton Mifflin.

Minagawa-Kawai, Yasuyo and Shigeru Kiritani (1998): *Non-Native Productions of Japanese Single Stops that are too Long for One Mora Unit*. Paper Presented at the International Conference on Spoken Language Processing 98, Sydney.

Nagano-Madsen, Yasuko (1992): *Mora and Prosodic Coordination: A Phonetic Study of Japanese, Eskimo and Yoruba*. Lund: Lund University Press.

Nagara, Susumu (1990): *Japanese for Everyone*. Tokyo: Gakken.

NHK Hōsō Bunka Kenkyūjo (ed.) (1998): *NHK nihongo hatsuon akusento jiten* [NHK Accent Dictionary of the Japanese Language]. Tokyo: Nihon Hōsō Shuppan Kyōkai.

Nishinuma, Yukihiro, Masako Arai and Takako Ayusawa (1996): *Perception of Tonal Accent by Americans Learning Japanese*. Paper Presented at the International Conference on Spoken Language Processing 96, Philadelphia.

Ogawara, Yoshirō (1997): Hatsuon kyōsei bamen ni okeru gakushūsha no hatsuon to kikitori no kankei ni tsuite [The Relationship between Production and Perception of Learners in Situations of Pronunciation Correction]. In: *Nihongo Kyōiku* [Journal of Japanese Language Teaching] 92, pp. 83–94.

Pierrehumbert, Janet and Mary E. Beckman (1988): *Japanese Tone Structure*. Cambridge: MIT Press.

Oguma, Rie (2000): Nihongo gakushūsha no chō'on no sansei ni kan suru shūtoku kenkyū [Research on the Acquisition of the Production of Japanese Long Vowels]. In: *Nihongo kyōiku* [Journal of Japanese Language Teaching] 109, pp. 110–117.

Richards, Jack C. and Theodore S. Rodgers (2001): *Approaches and Methods in Language Teaching*. Cambridge: Cambridge University Press.

Shibatani, Masayoshi (1990): *The Languages of Japan*. Cambridge: Cambridge University Press.

Shport, Irina A. (2003): *Perception and Production of Japanese Pitch Patterns with Non-Syllabic Morae*. Unpublished MA Thesis, University of Oregon.

––– (2006): The Effect of Segmental Structure on the Production of Japanese Pitch Accent (Abstract). In: *Journal of the Acoustical Society of America* 120, 5, 2, p. 3088.

Sugito, Miyoko (1982): *Nihongo akusento no kenkyū* [Studies in Japanese Accent]. Tokyo: Sanseido.

Tanaka, Shin'ichi and Kubozono Haruo (1999): *Nihongo no hatsuon kyōshitsu*: *riron to renshū* [Introduction to Japanese Pronunciation: Theory and Practice]. Tokyo: Kuroshio.

Tohsaku, Yasuhiko (ed.) (2004): *Yōkoso*! [Welcome!]. Boston: McGraw-Hill College.

Tsurutani, C. (1996): A Study of the Correlation between Durational Mistakes and Pitch Pattern in English Speakers' Production of Japanese. In: International Symposium on Language and Linguistics (eds.): *Pan-Asiatic Linguistics: Proceedings of the Fourth International*

Symposium on Languages and Linguistics. Vol. IV. Bangkok: Mahidol University at Salaya, pp. 1478–1498.

Vance, Timothy J. (1987): *An Introduction to Japanese Phonology*. Albany: State University of New York Press.

Venditti, Jennifer J. (2005): The J_ToBI Model of Japanese Intonation. In: Sun-Ah Jun (ed.): *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press, pp. 172–200.