Chiharu TSURUTANI and Shunichi ISHIHARA
Griffith University and Australian National University

Prosodic factors that affect the accentedness of L2 Japanese utterances: What matters most?

Abstract

This study aims to identify which of two crucial prosodic factors has greatest impact on native speakers' judgement of the accentedness of second language (L2) pronunciation. Prosodic features are already known to have more impact on the accentedness of L2 learners' pronunciation than segmental features do. In this study, timing and pitch are looked at as major prosodic factors that affect native speakers' accentedness judgement of L2 pronunciation. To examine the relative importance of timing and pitch, two types of speech samples—natural speech and prosody-modified speech—were used. In two experiments, native Japanese listeners assessed the accentedness of these stimuli and the results were compared. Both experiments obtained the same result: that timing is more important than pitch in improving the perceived naturalness of L2 Japanese speech.

1. Introduction

For many second language (L2) learners, the ultimate goal in language learning will be to attain native-like speech. However, in the multilingual society of the 21st century, accented speech due to first language (L1) interference is commonly heard and, intelligibility or naturalness has been more focused by educators than the identical copying of native speakers' speech (Jenkins 2000). It is useful to know which phonological factors help improve the quality of pronunciation in an economical way. It is known, for instance, that the acquisition of prosody has more impact on learners' intelligibility, as judged by native speakers, than does the acquisition of segmental features (Anderson-Hsieh, Johnson and Koehler 1992; Sato 1995). However, the significance of prosodic features that influence native speakers' judgements could differ depending on the language.

In general across languages, three major prosodic features—timing (duration of segments), pitch, and intensity—are coordinated by phonological rules to constitute the rhythm, or prosody of languages. In L2 speech, prosodic transfer effects in the form of inappropriate prosodic organisation of phonetic gestures are imposed on L2 targets through speech production habits formed in L1. Japanese spoken by English speaking learners has been selected as the target language for this study, because Japanese and English are contrastive in prosodic organisation. Japanese is traditionally classified as a pitch-accent language with moraic timing (Vance 1987), whereas English is classified as a stress-accent language with foot timing. Thus L1

English learners of Japanese have to learn a completely new system of accentuation and rhythmic organisation. Figure 1 outlines the accent (prosodic) typology proposed by Archibald (1998).

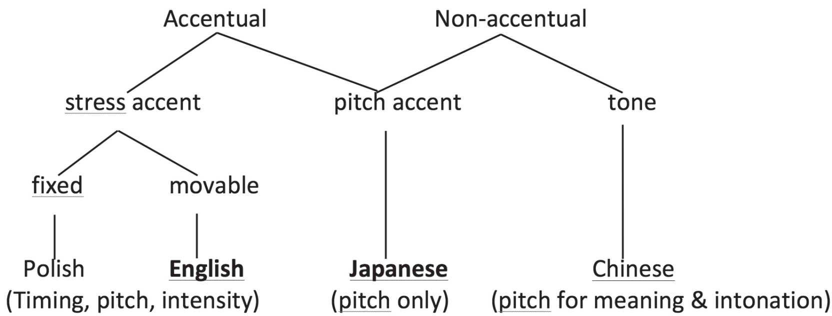

Figure 1: Accent typology (Archibald 1998, modified).

As can be seen in Figure 1, the two languages—English and Japanese—are distinct in their use of pitch. While stress-accent languages use duration, pitch, and loudness, pitch-accent languages use only pitch as an accent marker. With regard to their timing, Japanese and English are located at opposite ends of a continuum (see Figure 2) that includes, at one end, languages that try to equalise the duration of syllables (syllable-based), and, at the other end, languages that tend to equalise the duration between stresses (stress-based). In Japanese, a sub-syllabic unit of weight (mora) which takes a Consonant Vowel (CV) syllable as a basic form, plays an important role for durational contrast in both consonants and vowels. The durational differences are used phonemically to distinguish meanings, such as *kookoo* (high school) and *koko* (here) or kita (came) and kitta (cut). L2 English learners of Japanese have difficulty with this rigid durational contrast as vowel lengths may be partial or only allophonic in most varieties of English.

Figure 2: Rhythmic categories (Tsurutani 2008).

The prosodic differences of the languages described above imply that different languages require correctness in different prosodic features in order to achieve the level of intelligibility necessary for communication. On the other hand, there is a possibility that one prosodic feature might act as the dominant criterion for L2 pronunciation across languages. Tajima, Port and Dalby (1997) found the importance of timing in the improvement of English phrases spoken by Chinese-speaking learners of English. Cheng (2011), who investigated read-aloud English passages spoken by 126 L2 speakers with different L1 backgrounds, also reported that duration information was the best predictor of human prosody ratings.

Because Japanese is a pitch-accent language, pitch and intonation, a feature of phrases or sentences expressed by the movement of pitch, are expected to have a major influence on the evaluation of Japanese utterances. It has been reported that, of the prosodic features, pitch is the most dominant cue to accent patterns at the word level (Beckman and Pierrehumbert 1986). In other words, pitch is the most significant feature to determine word accent for both English and Japanese. As already partly noted above, Sato (1995) reported the predominance of prosodic factors over segmental factors, and that pitch was most influential among prosodic factors in assessing L2 speech.

L2 prosody has been studied extensively in English but has not been fully explored in many other languages (Derwing and Rossiter 2003; Tajima et al. 1997). The extent of the prosodic divergence between English and Japanese should be able to shed light on the mechanism of L2 prosody and provide the opportunity to study the fundamental components of rhythmic properties. In this study, we attempt to identify the relative importance of the prosodic features of accentedness in Japanese, as judged by native listeners. Accentedness was chosen as a criterion to evaluate L2 speech as it is easier to assess for native listeners than intelligibility or comprehensibility, which can be influenced by the context and settings where the utterance occurs. It is also known that native listeners give a harsher score for accentedness than for intelligibility and comprehension (Derwing and Munro 1997); thus, we should be able to observe clear differences in the scores given by participants. The following types of stimulus were employed for this study:

1. L2 speech of English speaking learners of Japanese
2. Synthesised L2 speech containing artificially reproduced errors

Both types of stimulus have advantages and disadvantages as the targets of human evaluation. Natural speech contains genuine errors and reflects the actual linguistic behaviour of L2 learners, while the number and degree of errors cannot be controlled. In this study we used errors identified by native listeners from a large sample of speech data, which are genuine audible errors rather than relying on any differences measured by acoustic analysis. In order to control errors, we could use synthesised speech. However, it is not known if the errors created by synthesis actually occur in real-world settings. The researchers attempted to recreate genuine errors and their authenticity was confirmed by native listeners. The use of two types of stimulus will compensate for the shortcomings of each type and should be sufficient to address the aims of the research. We acknowledge that there are more rigorous ways to create synthesized speech by controlling the actual length and pitch of phonemes. However, the technique available at the time of experiment was morphing system to manipulate speech, which could be a limitation of this study.

2. Experiment 1

This experiment was designed to test how pitch and timing errors influence native speakers' judgements of utterances produced by learners. We used natural speech in this experiment and synthesized speech in Experiment 2.

2.1 Materials

The stimuli were extracted from recordings of Japanese speech by Australian English speakers who had been studying Japanese for 160 hours at university at the introductory level. The recordings were taken from a computer-based exercise that was developed for a self-assessment of pronunciation. Utterances that contained timing errors or pitch errors or both, with no obvious segmental errors, were chosen, together with prosodically correct utterances from a large sample of speech data (all together 1440 samples were collected). As the sample size was large and learners were in their second year of Japanese language study, it was possible to choose stimuli that did not have typical segmental errors, such as incorrect production of palatalized consonants or flap. The judgement of errors was made by the first author of this study and two other native speakers who have many years' experience in Japanese teaching. When two out of the three native speakers agreed, the judgement was accepted. The following four patterns were considered:

1. incorrect pitch, correct timing PiTc
2. correct pitch, incorrect timing PcTi
3. correct pitch, correct timing PcTc
4. incorrect pitch, incorrect timing PiTi

With regard to materials sections for testing, using the same sentence would be convenient for analysis but could strain the listeners' concentration. Thus, six sentences that contained the listed patterns were chosen and formed 24 stimuli. The four patterns, 1-4, of the same sentence were played consecutively to the participants, but the order of patterns was randomized. Since the stimuli were natural utterances, it was not possible to control the degree of incorrectness as systematically as in synthesised speech. However, efforts were made to select speech with similar speech rates and number of errors. Errors at both word and phrase levels were counted inclusively. Table 1 provides the list of the sentences (Vance 2008 was used as a reference for the IPA transcriptions).

2.2 Participants

The participants in Experiment 1 were 80 university students (22 females, 58 males) from the Kansai region of Japan. Their ages ranged from 18-20 years and they had not left Japan, apart from short overseas trips. Kansai dialect employs pitch accents differently when compared with standard (Tokyo) Japanese. However, young generations have been exposed to both Tokyo and Kansai dialects through mass media and school education, and are equipped with the standard perceptual criteria for common Japanese. Twenty of the 80 participants (chosen randomly) recorded the stimulus sentences with the instruction to read them as if they were presenting a model of pronunciation for learners. All participants used the standard accent pattern, showing that participants had a good understanding of the standard accent pattern and thus, their judgement of accentedness was regarded as that of the average native listener.

Table 1: Stimulus sentences for Experiment 1

	No. of mora	Pitch pattern
1) Shachoono kekkonshikini okyakusanga sen-ninkita. ɕatɕoːno kekkonçi̥kiɲi okʲakɯ̥saŋŋa seɲɲiŋki̥ta (1000 people attended the president’s wedding reception.)	24	LHHH LHHLLLL LHHHHH HLLL HL
2) Tsuginojugyoono suugakuwa chotto muzu-kashiidesu. tsɯgino ǰɯgʲoːno sɯːgakɯwa tɕotto mɯzɯkaɕiːdesɯ̥ (Mathematics in the next class is a bit hard.)	24	LHH HLLL LHHHH HLL LHHHL LL
3) Watashino kookoode isshoni shashino tori-mashoo. wataɕino koːkoːde iɕɕoɲi ɕaɕin.o toɾimaɕoː (Let’s take a photo together at my high school.)	23	LHHH HHHHH LHHH LHHH LHHHL
4) Otootono okusanwa ryokooni ikunoga sukidesuyo. otoːtono okɯsaŋwa ɾʲokoːɲi ikɯnoŋa sɯ̥kidesɯjo (My younger brother’s wife likes cooking.)	24	LHHHH HLLLL LHHH LHLL LHLLL
5) Tanjoobini tomodachikara kireenahanao moratta. tanǰoːbini tomodatɕi̥kaɾa kiɾeːna hanao moɾatta (I received beautiful flowers from my friend on my birthday)	23	LHHHLL LHHHHH HLLL LHL LHHH
6) Shuumatsukara futarinohitoto shigotoo suruyoteedesu. ɕɯːmatsɯ̥kaɾa ɸɯ̥taɾinoçi̥toto ɕigoto.o sɯɾɯjoteːdesɯ̥ (From the weekend I’m planning to work with 2 new people.)	24	LHHHHH LHHL LHL LHHH LHHHH HL

2.3 Procedure

The L2 learners’ utterances containing the four error patterns (24 in total) were played to the native listeners for accentedness judgement. The four versions of the same sentence were played consecutively for ease of comparison. Each stimulus was played twice with a four-second interval, and an inter-trial interval of eight seconds. Before the task, three practice sentences were played for the listeners to become accustomed to the task and the proficiency level of the L2 speakers. The participants were asked to judge the accentedness of utterances using a Likert scale with potential responses ranging from 1 (not at all native-like) to 7 (native-like).

2.4. Results

The average and standard deviation of scores obtained from the judgements were calculated for the 24 stimuli. First, the four patterns were compared using the average scores for 480 sentences (6 sentences for 80 participants) in order to determine whether the four error patterns were distinguished by the listener participants. The mean score for each stimulus type is plotted in Figure 3 together with one standard deviation around the mean.

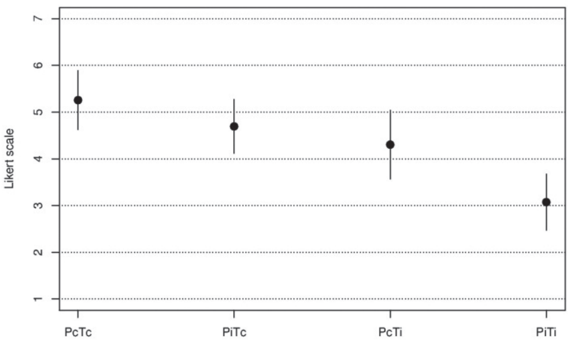

Figure 3: Average scores in Experiment 1 (7=Native like, 1 = Non-native like).

The average scores received were highest in sentences with PcTc type, followed by PiTc and PcTi, and lowest in sentences with PiTi errors. This means that the participants, on average, judged utterances with timing errors as worse than those with pitch errors. From the results shown in Figure 3, it is apparent overall that the judgement of speech containing Types PcTc and PiTi is different, but there is likely overlap in the ranking of speech with pitch and timing errors. The relative ranking of errors by individual participants was then examined to further understand whether individuals more specifically distinguished between all four types. We performed both a one-way ANOVA test, and the Tukey test ($p < 0.01$) to see if the ranking presented in Figure 3 is also statistically significant.

This analysis showed significant differences not only between the most extreme assessments, but also between all other combinations, supporting the initial view of the error hierarchy presented in Figure 3. In order to see the influence of error types, the stimuli were listed by their average score in a descending order by 80 participants together with the number of errors (see Table 2). The errors were grouped into three types: (a) word timing (durational errors within the word); (b) word pitch (pitch errors within the word); and (c) phrase-final (phrase-final lengthening and/or phrase-final pitch error).

Table 2: Average scores for each stimulus

	ID No.	Error pattern	Number of errors			
			Word pitch	Word Timing	Phrase-final (pitch errors unless specified)	score
1	10	PcTc			1	6.39
2	17	PcTc				5.86
3	19	PcTi			3 (lengthening)	5.51
4	6	PcTc				5.37
5	1	PiTc	2		1	5.26
6	3	PcTc			1	5.24
7	18	PiTc	3			5.2
8	7	PcTi			1 (lengthening)	5.07
9	24	PiTc	4		1	4.99
10	4	PcTi		1	1 (lengthening)	4.97
11	14	PiTc		1	4	4.7
12	15	PcTc				4.55
13	16	PcTi		2		4.25
14	8	PiTc	2			4.12
15	22	PcTc			1	4.12
16	9	PiTc	2		1 (lengthening.)	3.89
17	11	PiTi	1	2		3.8
18	5	PiTi	3	4	1 (lengthening)	3.67
19	20	PiTi	3	1		3.26
20	12	PcTi	1	2		3.24
21	13	PiTi	3	4		2.86
22	21	PcTi		3		2.77
23	23	PiTi	2	3	2	2.66
24	2	PiTi		3	1	2.17

NB: Errors at both word and phrase levels were counted inclusively. The judgement of errors was made by the auditory impression of the native speakers. A distinction in counting was made between more evident and very subtle errors. The latter, particularly at phrase final position, were counted but were not registered as 'Incorrect' in the word pitch and word timing columns.

Overall, stimuli at the bottom of the table had more word-internal timing errors. The top nine utterances did not contain any word-internal timing errors, but had some phrase-final lengthening or pitch raising in some cases (e.g. items 3 and 7). Phrase-final lengthening and pitch raising are seen across the ranking and do not seem to affect the native listeners' judgement. As can be seen from Table 2, while a high incidence of timing errors was associated with lower scores, high incidences of pitch errors or of phrase-level errors correlated with both higher and lower scores.

The judgements were not affected by whether the timing error involved lengthening rather than shortening, or vowels rather than consonants. This result also indicates that phrase-final lengthening was irrelevant to accentedness judgement; it is suggested that this is because phrase-final lengthening is a common phenomenon in speech. There were no particular types of pitch error that affected native listeners' judgement, either. Phrase-final pitch raising errors occurred in several utterances, from the highest scoring to the lowest scoring and showed no impact on the accentedness score. Thus, the results show that timing errors affect native listeners' judgement most significantly.

3. Experiment 2

The same perception task was conducted using synthesised L2 stimuli with a different set of Japanese native listeners. In the following sections, the method of creating stimuli and the procedure for the perception task used to assess the accentedness of speech will be discussed.

3.1. Methods

STRAIGHT[1], a speech manipulation and morphing system, was employed to separately manipulate different prosodic variables while maintaining the quality of speech samples. Five stimulus sentences which contain prosodic difficulties (vowel and consonant length, flat pitch pattern) for L2 English learners of Japanese were used. The stimulus sentences had to be kept short (13-14 mora long) due to technical issues, and, thus, a different set from Experiment 1 was used. It becomes harder for STRAIGHT to maintain the quality of synthesised speech as the speech becomes longer, although the STRAIGHT can produce synthesised speech of high quality, and has been used in many psychological and cognitive science studies (Kawahara and Morise 2009; Uchida 2009).

3.2. Participants

A total of 45 Japanese native speakers (11 males, 34 females), many of whom were university students, participated in the task for a small payment in Brisbane and Canberra, Australia. The participants were all familiar with standard Japanese through school education, although their places of birth spanned several regions across Japan.

3.3. Materials

Segmental errors are not examined in this study; however, phonemes that could induce prosodic errors (e.g. palatalised consonants) were included in the original sentences. Of ten trial L2 sentences mimicked by the first author, five sentences which two native Japanese listeners could not identify easily, were chosen as the original L2 materials for synthesis. The stimuli, which are based on these five sentences, are given in Table 3. Each sentence is transcribed in IPA together with an English translation. The number of mora and the pitch pattern of each sentence are also included in Table 3 below.

Table 3: Stimulus sentences for Experiment 2 (Vance (2008) was used as a reference for the IPA transcriptions)

Sentences	No. of mora	Pitch pattern (assigned on mora)
Senen kuretara kaemasuyo *sen.eŋ kɯɾetaɾa kaemasɯjo (If you give me 1000 yen, I can buy it.)	12	HHHH LHHL LHHLH
Tsumaranai hono kattekita ʦɯmaɾanai hoN.o kattekḭta (I bought a boring book.)	13	LHHLL HLL LHHHL
Watashino iega miemasuka wataɕino ieŋa miemasɯ̥ka (Can you see my house?)	12	LHHH LHL LHHLH?
Gaikokuni ryokoo shimashoo gaikokɯɲi ɾʲoko: ɕimaɕo: (Let’s travel overseas.)	12	LHHHH LHH LHHL
Daigakuno sotsugyoo shashin desu daigakɯno soʦɯgʲo: ɕaɕin desɯ̥ (This is my university’s graduation photo.)	14	LHHHH LHHH HLL LL

**Full stop was placed on mora boundary to distinguish [en] and [nen].*

Synthesising perceptually L2-like stimuli by morphing requires a near-bilingual speaker who can utter the speech materials with perfect model pronunciation, and can also mimic an absolute beginner’s pronunciation, containing all required errors. This requirement is due to the nature of STRAIGHT; morphing two different voices as one will increase the unnaturalness of the resulting synthesised speech. The first author performed the speech task, attempting to include in the non-native speech the errors presented in Table 4. Due to the short length of the sentences, phrase-final lengthening and pitch raising could not be included in this experiment.

Table 4: Errors included in the non-native speech (highlighted in bold for timing errors and yellow for pitch pattern errors; Harrington, Cox and Evans (1997) was used as a reference for the IPA transcriptions)

Timing errors	*Incorrect phonemes which affected timing	No. of timing errors	Pitch pattern errors	No. of pitch errors
Senen kuRetaRa kaemasyo. sɛnɛn kʰʉːɹɛtʰɐːɹə kʰeːmɐsjoː	2R	3	HHH **HL**LL **HHLL**H	2
TsumaranAI hono katekita. ʦmɐɹənɑe hɔnɔ kʰɐtɛkɪtɐ	AI	2	**LLLHL** HL **HLLL**	2
Wataashino IEga mIImasuka. wətɐːʃinə ɪegɐ miːmɐːskɐ	IE, II	1	**LHLLHLLHHLL**H	3
gAIkokkuni Rookoo shi-imashIO. gɑekɔknɪ ɹoːkoː ʃiːməʃəʉ	AI, R IO	3	LL**HLLHHLLHLLL**	3
dEIgakkuno sotsuGIO shaashin desu. deɪgɐkkʊno soʦʉgɪo sʲɐːʃin dɛs	EI, GIO	2	**LLHLLHHLL**HHLL LL	2

The number and degree of errors in the non-native speaker model required careful consideration. In the current study, the types of errors could be controlled, because the first author, who knew what kind of errors needed to be included in the stimuli, produced the erroneous utterance. However, making errors deliberately is not an easy task. In addition, making errors on every single word was both impossible and unrealistic, and yet we needed a reasonable number of both pitch and timing errors. Efforts were made to change a flat pitch pattern to a HL pitch pattern, which is a common error observed in L2 English learners' production (Tsurutani 2008), and to make durational errors on the length contrast between long and short vowels and consonants (see Table 4). Durational errors at segmental level and at phrase/sentence levels could not be separated since both would be judged as temporal deviations by native listeners. Incorrect flap sounds, labial fricatives, and English-sounding vowels were also added to assist natural production of the non-native speech model.

3.3.1. Constructing stimuli

First, the original speech samples were separately decomposed into three independent acoustic parameters of segment, pitch, and duration. With the segmental parameters being constant, only the pitch and duration features of the native Japanese samples were morphed with the corresponding pitch and duration features of the non-native samples with different percentages (0%, 50%, and 100%) as shown in Figure 4. This was done using STRAIGHT.

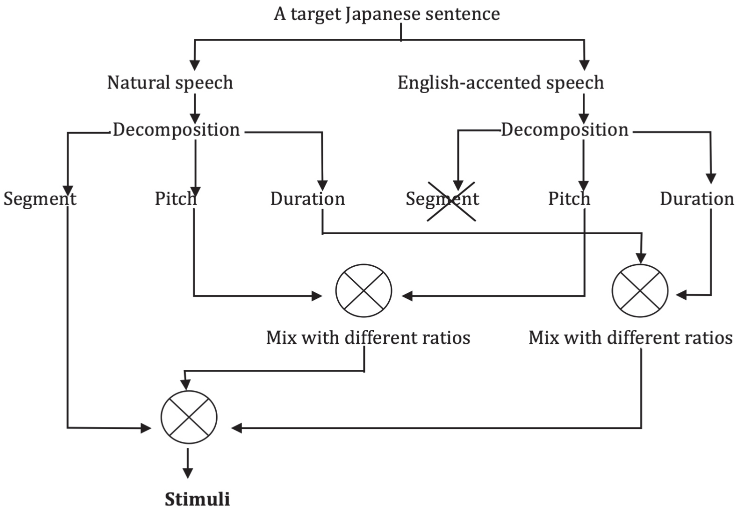

Figure 4: Morphing process

All possible combinations of the morphing features with respect to their percentages are listed in Table 5. For example, stimulus p000t000 means the pitch and timing of the original native Japanese samples were not morphed with any of the non-native speech component, i.e. p000t000 is identical to the original native speaker Japanese sample. Stimulus p050t050 is a morphed sample with 50% pitch and 50% timing of the non-native sample.

Table 5: The possible permutations of the morphing parameters

Native			Non-native		Stimuli
segment	pitch	timing	pitch	timing	type
100%	100%	100%	0%	0%	p000t000
100%	100%	50%	0%	50%	p000t050
100%	100%	0%	0%	100%	p000t100
100%	0%	100%	100%	0%	p100t000
100%	50%	100%	50%	0%	p050t000
100%	50%	50%	50%	50%	p050t050
100%	0%	0%	100%	100%	p100t100

3.3.2 Experimental procedure

Participants were asked to rate the accentedness of utterances on a Likert scale (1 to 7). A listening task was conducted on-line, using the following instructions in Japanese:

> You will listen to short Japanese sentences recorded by one female speaker. Some of them are natural and some of them are foreign accented. Evaluate the naturalness of the utterance using a Likert scale with potential responses ranging from 1 (non-native like) to 7 (native like). There are no right and wrong answers. Don't listen to the same speech sample more than twice. Just follow your intuition as a native listener.

The order of 35 sentences was automatically randomised in order to avoid an ordering effect.

3.4. Results

The scores were pooled together separately according to the stimulus types. The mean score for each stimulus type is plotted in Figure 5 together with one standard deviation around the mean. For Figure 5, the ranking of scores, p000t000 > p100t000 > p000t100 > p100t100 was statistically confirmed by the Tukey HSD test ($p < 0.01$). That is, the stimulus that had the correct prosodic features (p000t000) received the highest score while the stimulus that had 100% incorrect prosodic features (p100t100) had the lowest score, as expected. The important point is that there is a statistically significant difference between the ratings of the p100t000 and p000t100 stimuli (p100t000 > p000t100). This result indicates that the native Japanese speakers are sensitive to both incorrect pitch and timing but nevertheless put more weight on accuracy in timing than in pitch when judging the naturalness of speech, which confirms the findings of Experiment 1.

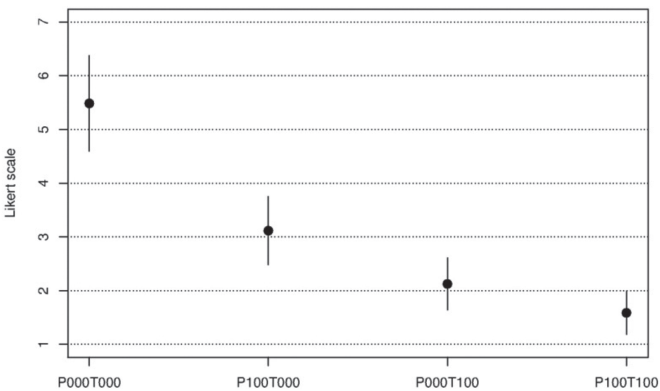

Figure 5: The average scores in Experiment 2 (7=Native like, 1 = Non-native like)

4. Discussion

Based on the assumption that prosodic factors are the most crucial features in improving L2 pronunciation, two experiments were carried out, using natural L2 speech and synthesised L2 speech, both of which contained pitch and timing errors. The results from both experiments suggest that timing is the most important feature for L2 learners of Japanese to work on to improve their pronunciation, based on the judgements of native Japanese listeners. The study was conducted targeting English speakers whose Japanese utterances had noticeable timing errors as well as pitch errors, and used declarative sentences, which uniformly have declination towards the end of a sentence. Under this setting of stimulus sentences (no interrogatives), pitch was not a crucial factor, and timing was more important than pitch pattern for accentedness judgement.

Pitch and timing errors can occur both at word and phrase levels. Due to the length constraint of stimulus sentences in Experiment 2, only Experiment 1 used sentences long enough to observe errors at phrase level. Results from Experiment 1 revealed that phrase-final pitch raising or lengthening was not regarded as a serious error. Final-syllable lengthening functions as a boundary marker and appears to be a putative universal across languages, but the lengthening ratio varies depending on the language, e.g. it is higher in English and lower in Spanish (Cruttenden 1986). Presumably, syllable-timed languages regulate timing more strictly than stress-timed languages. However, in Experiment 1, only timing errors in words lowered the performance score.

Pitch raising at phrase-final position is regarded as boundary tone, and a rising tone typically appears in English to signal non-finality. On the other hand, it is not common in Japanese sentences, as illustrated using the ToBI annotation in Figure 6, which shows that pitch is lowered at the end of each phrase.

sankaku-no	*yane-no*	*mannaka-ni*	*okimasu*
the triangular	roof-GEN	in the middle	put it

L% = Fall at the end of accentual phrase

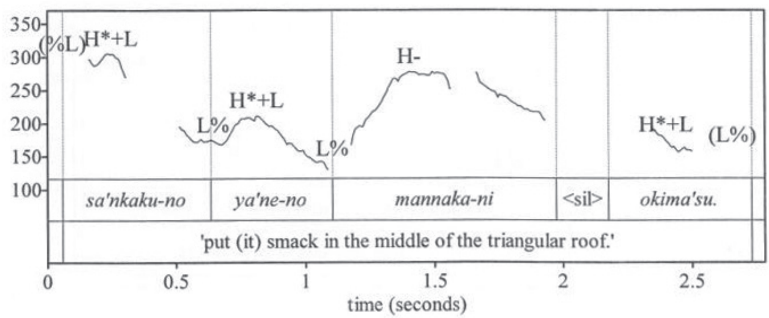

Figure 6: Pitch contour of a Japanese declarative sentence (Venditti, Maekawa and Beckman 2008)

At the same time, pitch raising and final-syllable lengthening are known to be common speech behaviours in Japanese young people (Inoue 1993), and the phenomenon has spread since it was first observed in the 1990s (Tsurutani 2010). This could be a possible reason that pitch errors at phrase level were tolerated.

Japanese has mora-based timing, and the duration of phonemes is more rigid than in stress-based languages. In addition, the pitch pattern of Japanese words can differ depending on the region and is not the most critical determinant for identifying a word. This phonological organisation of Japanese partly explains the fact that timing is more important than pitch to L1 listeners. At word level, pitch is the most significant feature to determine the word accent both for Japanese and for English (Beckman and Pierrehumbert 1986). However, the influence of prosodic features on overall performance needs to be looked at above word level. Previous studies (Cheng 2011; Maier et al. 2009; Tajima et al. 1997) found that timing is the most important factor for assessment of L2 speech of English and German. The result of this study suggests that the significance of correct timing in L2 speech can be observed widely across languages including L2 Japanese. It is considered that native listeners are more sensitive to timing than pitch in general (Tsurutani 2008). In fact, non-native speakers of Japanese are also more sensitive to timing than to pitch (Ishihara, Fan, Jalil and Tsurutani 2011). Our results indicate that people can easily and accurately tell whether a sound in their native tongue is short or long, but are less affected by differences in pitch. The significance of timing in accentedness judgement needs to be tested in other languages as well.

The result of this study has implications for how language learners can efficiently improve the naturalness of their L2 speech. In addition, the finding that timing is the most crucial factor for accentedness brings further interpretation when applied to language classrooms—at least with regard to the accurate learning of Japanese.

Timing is a core element of rhythm. Learners need to master all phonemes and basic grammar of the target language to build a foundation for the correct rhythm of the target language, which is heavily influenced by timing. It is quite often the case that the improvement of one phonetic feature reflects the progress of the learner’s overall language proficiency. The correctness of rhythm, particularly timing, has great potential as a criterion for assessing L2 speakers’ proficiency level.

5. Conclusion

The importance of moraic timing for accentedness has been recognised widely in Japanese phonology, whereas the importance of pitch has not been empirically tested. Due to the fact that Japanese is a pitch accent language, the significance of pitch in assessing pronunciation could be reasonably anticipated. On the other hand, pitch accent patterns have regional variation in Japanese, and pitch raising at the phrase-final position is becoming common. It was hypothesised that Japanese native listeners might be more lenient about pitch deviation than previous studies—where it was found that pitch was more significant than timing—suggest. The results of this study found that timing is the most crucial among the suprasegmental factors to improve L2 pronunciation of Japanese using both natural speech stimuli and synthesised stimuli. Pitch errors do affect the performance score, but not as significantly as do timing errors. Errors at phrase-final positions, pitch raising, and final-syllable lengthening did not affect native listeners' judgement and were treated as natural speech behaviour.

Language instructors wish for their learners' mastery of the target language, and are inclined to point out all learner errors. This teaching approach tends to be translated to, and reflected in, the production of assessment tools. However, not all prosodic features have to be brought to the level of a textbook model. The threshold of native listeners' judgement requires mainly the accuracy of timing in Japanese. Their criteria, which count timing errors more heavily than pitch errors, should be taken into consideration for pronunciation teaching or the development of computer-based assessment tools. Other languages with a different prosodic organization could prioritize a different prosodic component. Future study of prosodic judgement for L2 speech from other languages will create an interesting comparison.

Note

1. For more information, see http://www.wakayama-u.ac.jp/~kawahara/PSSws/.

References

Anderson-Hsieh, J., R. Johnson and K. Koehler. 1992. The relationship between native speakers' judgements of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning* 42 (4), 529-555.

Archibald, J. 1998. *Second Language Phonology*. Amsterdam/Philadelphia: Benjamins.

Beckman, M. and J. Pierrehumbert. 1986. Intonational structure in Japanese and English. *Phonology Yearbook* 3, 155-300.

Cheng, J. 2011. Automatic assessment of prosody in high-stakes English tests. *Proceedings of Interspeech 2011*, 1589-1592. Florence: ISCA.

Cruttenden, A. 1986. *Intonation*. Cambridge: Cambridge University Press.

Derwing, T.M. and S. Munro. 1997. Accent, intelligibility and comprehensibility. *Studies in Second Language Acquisition* 19, 1-16.

Derwing, T.M. and M.J. Rossiter. 2003. What do ESL students say about their accents? *The Canadian Modern Language Review* 59 (4), 547-566.

Harrington, J., F. Cox and Z. Evans. 1997. An acoustic phonetic study of broad, general, and cultivated Australian English vowels. *Australian Journal of Linguistics* 17, 155-184

Inoue, F. 1993. The significance of new dialects. *Dialectologia et Geolinguistica* 1, 3-27.

Ishihara, S., Y. Fan, D. Jalil and C. Tsurutani. 2011. A comparative study on perception of foreign-accented Japanese by L2 Japanese listeners having different L1 backgrounds: English, Chinese and Indonesian. *Proceedings of the 17th International Congress of Phonetic Sciences*, 950-953. Hong Kong.

Jenkins, J. 2000. *The Phonology of English as an International Language*. Oxford: Oxford University Press.

Kawahara, H. and M. Morise. 2009. TANDEM-STRAIGHT and voice morphing: an application to emotional speech and singing research. *Journal of Phonetic Society of Japan* 13 (1), 29-39.

Maier, A. et al. 2009. A language-independent feature set for the automatic evaluation of prosody *Proceedings of Interspeech 2009*, 600-603. Brighton: ISCA.

Sato, T. 1995. Comparison of the influence of segmental information and prosody on the assessment of Japanese pronunciation. *Sekai no Nihongo Kyoiku* 5, 139-154.

Tajima, K., R. Port and J. Dalby. 1997. Effects of temporal correction on intelligibility of foreign-accented English. *Journal of Phonetics* 25, 1-24.

Tsurutani, C. 2008. *Pronunciation and Rhythm of Japanese as a Second Language*. Hiroshima: Keisuisha.

Tsurutani, C. 2010. Foreign accent matters most when timing is wrong. *Proceedings of Interspeech 2010*, 1854-1857. Makuhari: ISCA.

Pauk, B. 2007. Student participation in developing outcomes: a survey of students majoring in European Languages. In *Student Engagement:* Proceedings of the 16th Annual Teaching Learning Forum, 30-31 January 2007. Perth: The University of Western Australia.

Uchida, T. 2009. Proposing the PROSPECT model: relationship between the prosodic features of speech sound and the personality impressions. *Journal of Phonetic Society of Japan* 13 (1), 17-28.

Vance, T.J. 1987. *An Introduction to Japanese Phonology*. Albany, NY: State University of New York Press.

Vance, T.J. 2008. *The Sounds of Japanese*. Cambridge: Cambridge University Press.

Venditti, J., K. Maekawa and M. Beckman. 2008. Prominence marking in the Japanese intonation system. In S. Miyagawa and M. Saito (eds), *Handbook of Japanese Linguistics,* 458-514. Oxford: Oxford University Press.