A COMPARATIVE STUDY ON PERCEPTION OF FOREIGN-ACCENTED JAPANESE BY L2 JAPANESE LISTENERS HAVING DIFFERENT L1 BACKGROUNDS: ENGLISH, CHINESE AND INDONESIAN

Shunichi Ishihara[a], *Yiran Fan*[a], *Dia Jalil*[a] *& Chiharu Tsurutani*[b]

[a]School of Culture, History and Language, Australian National University, Australia;
[b]School of Languages and Linguistics, Griffith University, Australia
shunichi.ishihara@anu.edu.au; u4399527@anu.edu.au;
u4297952@anu.edu.au; chiharu.tsurutani@griffith.edu.au

ABSTRACT

This study concerns *good pronunciation* from the L2 Japanese listeners' point of view. More precisely, focusing on prosodic variations, this study investigates whether L2 Japanese listeners from different L1 backgrounds perceive foreign-accented Japanese differently. The experimental results obtained from the two L2 Japanese groups who speak Mandarin Chinese and Bahasa Indonesian, respectively as L1 are presented in this study, and then are compared with the results of previous studies on native Japanese listeners and L2 Japanese listeners whose L1 is English.

Keywords: L2 speech, non-native listeners, Japanese, timing error, pitch error

1. INTRODUCTION

Native speakers can detect foreign accents in their own mother tongue even from a single word or phrase [3]. If this is indeed true, what contributes to native speakers' judgement on accented speech? Although there are some conflicting reports, the majority of previous studies across languages provide evidence to suggest that speech with prosodic errors is more likely to be perceived as foreign-accented speech than speech with segmental errors. This superiority of prosody to segment has been confirmed for Japanese as well [7]. Further relating to the superiority of prosody, some researchers [8, 9] hypothesise from their experimental results that the accuracy in timing is more important than the accuracy in pitch for a speech to be judged as having good pronunciation while others argue that the relative importance of pitch and timing may vary according to the linguistic nature of a given language [4].

Although there are a large number of studies focusing on native speakers' judgement of accented speech [1, 6], to date, studies on non-native listeners' perceptions of L2 speech are scarce or heavily dominated by studies with English as the target language. Consequently, we do not know; how the learners of i.e. Japanese perceive accented Japanese; whether the learners having different L1 backgrounds perceive accented Japanese differently; and how their perception changes as a function of their level of proficiency; and so on.

Thus, using non-native listeners of Japanese whose L1 is either Mandarin Chinese or Bahasa Indonesian, both of which are prosodically significantly different from Japanese (and English), the aim of this study is to investigate how differently/similarly the non-native listeners of Japanese having different L1 backgrounds perceive foreign-accented Japanese. The results will be compared with those of native Japanese listeners [8] and the L2 Japanese listeners whose L1 is English [9], which were previously reported. These two previous studies will be summarised in §3.1 and §3.2, respectively.

We will focus on timing and pitch errors in this study. Japanese is a so-called mora-timed pitch-accent language while Mandarin Chinese is a well-known tone language in which stress-like patterns can be observed [5]. Bahasa Indonesian seems to have no strict prosodic rules with conflicting results evident in a number of studies [2]. However, stress occurs in Indonesian, yet, stress differences in the same word will not result in a different word. Thus, as can be predicated from the unique prosodic nature of Japanese, Chinese and Indonesian listeners may rely on different cues for their judgement of accented Japanese speech.

The current study will not only contribute to a better understanding of the mechanism of L2 perception theoretically but will also provide pedagogically useful insight into learning and teaching Japanese.

2. EXPERIMENTS

2.1. Materials

The stimuli were extracted from recordings of English speakers who had studied Japanese for approximately 160 hours at university at the time of recording. The recordings were taken from a computer exercise that was developed as a self-assessment for pronunciation, in which students were required to utter various kinds of sentences. These recordings were compiled as a speech database. Utterances that contain timing errors (e.g. errors in long/short contrast in vowels and consonants) or pitch errors (error in pitch-accent) or both, with no obvious segmental errors, were chosen, together with utterances without any obvious errors, from this large speech database. The judgement of errors was made by the fourth author of this paper and two other native speakers of Japanese who had been teaching Japanese for many years. The four patterns given in Table 1 were considered.

Table 1: Four stimulus types.

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

The stimuli were taken from the six sentences given in Table 2. Note that the sentences given in Table 2 contain many words in which long vowels/consonants appear. Needless to say, these long vowels/consonants need to be articulated with enough duration in contrast to corresponding short vowels/consonants. For each sentence, an utterance which fits each of the four patterns given in Table 1 was extracted from the database, and compiled as 24 stimuli (= 6 sentences x 4 patterns). It is important to point out here that since all stimuli are natural utterances, they are not perfectly controlled in a mutually comparable manner, yet every effort was made to select utterances which are as comparable as possible in terms of characteristics such as speech rate and number of errors.

The main point of using these stimuli is to see whether different error types (PiTc vs. PcTi) are weighted differently or equally by listeners.

2.2. Participants

27 Mandarin Chinese and 23 Bahasa Indonesian native speakers who fall under the category of having an intermediate to advanced level of Japanese proficiency participated in the perception tests. Those participants who were categorised as intermediate are the ones who passed level 2 of the Japanese Language Proficiency Test (JLPT) or were judged by one of the authors of this paper as having an equivalent level of proficiency. *Mutatis mutandis*, the same criteria apply to the advanced group with level 1 of the JLPT.

Table 2: Stimuli sentences.

1.	*shachoo-no kekkonshiki-ni okyakusan-ga sennin kita.* 社長の結婚式にお客さんが千人来た。 1000 people attended the president's wedding reception.
2.	*tsugi-no jugyoo-no suugaku-wa chotto muzukashii-desu.* 次の授業の数学はちょっと難しいです。 Mathematics in the next class is a bit hard.
3.	*watashi-no kookoo-de isshoni shashin-o tori-mashoo.* 私の高校で一緒に写真をとりましょう。 Let's take a photo together at my high school.
4.	*otooto-no okusan-wa ryokoo-ni ikuno-ga suki-desu-yo.* 弟の奥さんは旅行に行くのが好きですよ。 My younger brother's wife likes travelling.
5.	*tanjoobi-ni tomodachi-kara kireena hana-o moratta.* 誕生日に友達から綺麗な花をもらった。 I received beautiful flowers from my friend on my birthday.
6.	*shuumatsu-kara futarino hito-to shigoto-o suru yotee-desu.* 週末から二人の人と仕事をする予定です。 From the weekend I'm planning to work with 2 people.

2.2.1. Mandarin Chinese

The 27 Chinese participants consist of 13 advanced and 14 intermediate participants. Many of them are university students studying in an Australian or a Japanese university, but there are some participants, in particular, in the advanced group, who are engaged in a profession in Japan after completing tertiary education.

2.2.2. Bahasa Indonesian

The 23 Indonesian participants, who are all university students either in Australia or Japan, comprise of 14 advanced and 9 intermediate learners.

2.3. Procedures

The 24 stimuli explained in §2.1 were played to the two groups of listeners (Chinese and Indonesian) for their judgements on pronunciation goodness. Each stimulus was played twice with a 4-second interval. Before the task, 3 practice sentences were played for the listeners to become accustomed to the task and the proficiency level of the stimuli. The participants were asked to rate the naturalness of utterances on a Likert scale ranging from 1 (not at all native-like) to 7 (native-like). The listening task took about 15 minutes, including the time for instructions.

3. RESULTS AND DISCUSSIONS

In this section, the experimental results of the current study are described and discussed by referring to the results of previous studies. The main findings of previous studies on native Japanese [8] and on English listeners [9] are summarised first in §3.1 and §3.2, respectively before the results of Chinese and Indonesian are given in §3.3.

3.1. Native Japanese listeners

80 Japanese native listeners participated in [8] using exactly the same stimuli and experiential procedure as the current study. The bar plots given in Figure 1-a show the mean values of the scores pooled separately for the four different stimulus types for the Japanese native listeners.

Based on the statistically confirmed relationship of PcTc > PiTc > PcTi > PiTi, [8] argues the superiority of timing to pitch for the judgement of L2 Japanese speech by Japanese native listeners. [8] also discusses that the inferiority of pitch can be accounted for in light of the variations of pitch patterns observed across Japanese dialects and generations. That is, since there are various patterns observed even across native speakers of Japanese depending on their dialects and generations, Japanese native speakers are more tolerant of pitch errors than timing errors.

3.2. English listeners

The results of the English speakers (15 beginners and 11 advanced students) [9] are plotted in Figure 1-b. Again, the stimuli and experiential procedure are identical to those of the current study. The main finding of their study is that like Japanese native listeners, non-native listeners speaking English as L1 also put more weight on accuracy in timing than in pitch for the assessment of L2 accented Japanese, regardless of their levels of proficiency.

An interesting observation which was made is that in contrast to the native listeners, L1 English speaking listeners assessed the PcTc type very harshly (refer to the arrow of Figure 1-b), which should get the highest average score.

3.3. Chinese and Indonesian

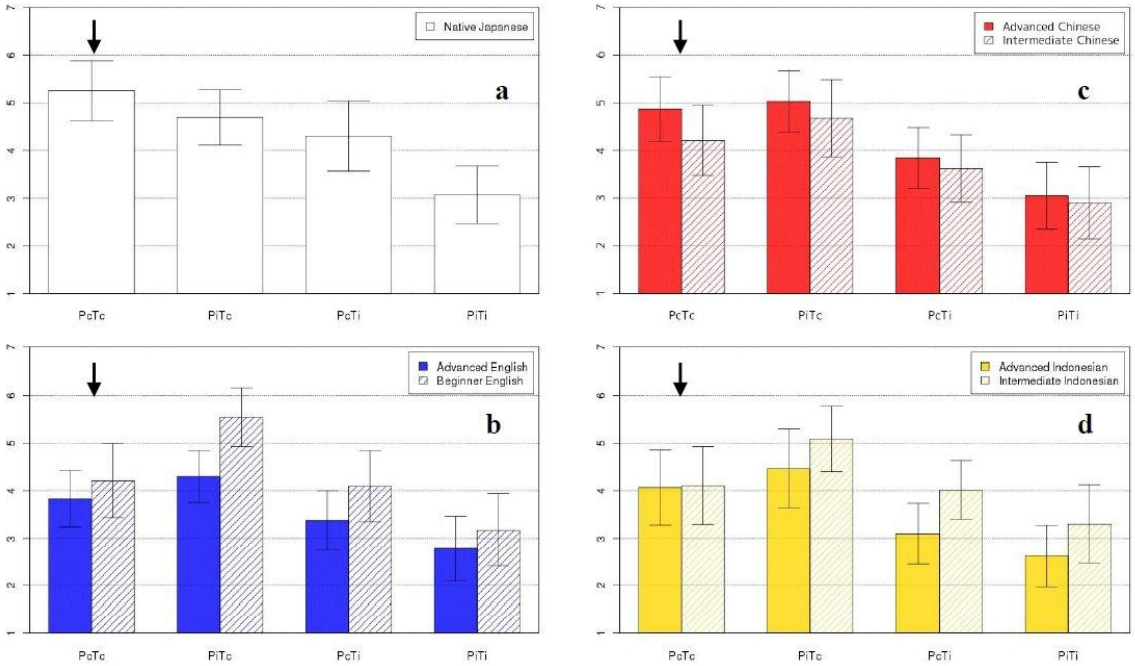
The results of the Chinese and Indonesian listeners (intermediate and advanced) are given in Figure 1-c and Figure 1-d, respectively. The presentational procedure is the same as Figure 1-a. For the PiTc, PcTi and PiTi types, both levels of the two language groups statistically significantly ($p<0.001$) show the relationship of PiTc > PcTi > PiTi (a one-way repeated ANOVA with the stimulus types as a factor, followed by Tukey HDS tests with a confidence level of 95%). That is, regardless of different L1 languages and levels of proficiency, more weight is put on accuracy in timing than in pitch when they assess L2 accented speech. This result echoes the results of the previous studies that both native Japanese and L2 Japanese listeners speaking English as L1 are more sensitive to timing information than pitch information. The current result gives further insight showing that this sensitivity towards timing information is evident regardless of whether or not one's L1 is a stress-accent language (English), a tonal language (Chinese) or a language in which pitch and timing do not lexically play significant roles (Indonesian). This may mean that timing is generally more salient, accessible and, hence, easier to acquire than pitch in Japanese.

However, this result may be deserved judging from the nature of Japanese orthography. The short-long contrast of Japanese vowels/consonants is orthographically expressed whereas pitch contrast is not. That is, as soon as learners are taught how to read and write Japanese, which usually happens at a fairly early stage, they are constantly made aware of timing differences, but not so much with pitch-accent.

An interesting observation which can be made across English (Figure 1-b), Chinese (Figure 1-c) and Indonesian (Figure 1-d) is that in contrast to the native group (Figure 1-a), the non-native groups, regardless of their levels of proficiency, assessed the PcTc type very harshly, which should get the highest average score (refer to the arrows in Figure 1).

It is very difficult to find a sensible reason for this harshness from possible perceptual differences between the non-native and native groups. An explanation we can moot is the actual stimuli we used for the PcTc type. All PcTc type stimuli were read by male learners, and thus their f0 ranges (their average range = 75.2 Hz) were narrower than the average f0 range of the other utterances (= 120.0 Hz). In fact, they admittedly sound very monotonous. If this paralinguistic characteristic of the PcTc stimuli resulted in the difference between the non-native and native groups, it can be said that learners' perception is subject to the influence arising from paralinguistic differences of speech, causing misjudgement of good pronunciation.

Figure 1: Bar plots showing the mean values of the scores pooled separately for the four different stimulus types (PcTc, PiTc, PcTi, PiTi) for native Japanese (a), English (b), Chinese (c) and Indonesian (d) listeners. Colour-filled bars are for the advanced groups and bars with oblique lines are for non-advanced group. One standard deviation is also given around the mean. The arrows indicate the average scores of the PcTc type.



4. CONCLUSIONS

In this study, we examined the perception of the non-native listeners of Japanese whose L1 is either Mandarin Chinese or Bahasa Indonesian. We observed that, despite the different levels of proficiency and L1 languages, utterances with timing errors were deemed to be judged less natural than those with pitch errors. This result conforms to those of previous studies on native Japanese [8] and English-speaking L2 Japanese learners [9]. A possible explanation was discussed for this result, but this may mean that timing information is generally more salient, accessible and, hence, easier to acquire than pitch information in Japanese.

5. ACKNOWLEDGEMENTS

The authors thank two reviewers for their comments.

6. REFERENCES

[1] Bond, Z.S., Stockmal, V., Markus, D. 2003. Sentence durations and accentedness judgments. *J. Acoust. Soc. Am.* 113, 2330-2334.

[2] Cohn, A.C. 1989. Stress in Indonesian and bracketing paradox. *Natural Language and Linguistic Theory* 7, 167-216.

[3] Flege, J.E. 1984. The detection of French accent by American listeners. *J. Acoust. Soc. Am.* 76, 692-707.

[4] Holm, S. 2008. *Intonational and Durational Contributions to the Perception of Foreign-accented Norwegian: An Experimental Phonetic Investigation*. Ph.D. dissertation, Norwegian University of Science and Technology.

[5] Kochanski, G., Shih, C., Jing, H. 2003. Quantitative measurement of prosodic strength in Mandarin. *Speech Communication* 41(4), 625-645.

[6] Munro, M.J. 1995. Nonsegmental factors in foreign accent. *Studies in Second Language Acquisition* 17(1), 17-34.

[7] Sato, T. 1995. A comparison of phonemes and prosody in the evaluation of spoken Japanese. *Japanese Language Education around the Globe* 5, 139-154.

[8] Tsurutani, C. 2010. Foreign accent matters when timing is wrong. *Proc. Interspeech* Tokyo, 1854-1857.

[9] Tsurutani, C., Tsukada, K., Ishihara, S. 2010. Comparison of native and non-native perception of L2 Japanese speech varying in prosodic characteristics. *Proc. Australasian International Conference on Speech Science and Technology* Melbourne, 122-125.