



The possible features of natural English pronunciation for Japanese learners and native speakers of English

Maito Ueno¹, Toshiya Magoku², Atsuko Nishiyama³

¹Graduate School, Kansai University, Japan, ²Graduate School, Wakayama University, Japan, ³Wakayama University, Japan

¹*uenomight@gmail.com*, ²*toahiya25@gmail.com*, ³*nishi@center.wakayama-u.ac.jp*

Abstract

This study examines whether so-called native-like English pronunciation by Japanese English learners (JELs) is perceived as “natural” by both JELs and native speakers of English (ENSs) and which features contributed to their perceptions. In order to compare the influence of phonetic features, especially that of intonation patterns, with that of vowel insertions, JELs read sample sentences in four patterns, that is, in a fall-rise intonation or in a fall intonation, with or without vowel insertions between consonants. JEL and ENS listeners were asked to rate the naturalness of the speeches based on a 10-point Likert scale. Their evaluation of the recordings without vowel epenthesis was significantly higher than those with a vowel epenthesis. While JELs and ENSs showed similar tendencies in their perceptual evaluation, JELs tended to rate recordings with vowel epentheses lower than ENSs did. These findings suggest that, although vowel epenthesis is a significant predictor of accent ratings, JELs may be too strict with it in judging the naturalness of their English pronunciation.

Keywords: naturalness, accentedness, pitch, intonation, vowel epenthesis

1. Introduction

There is a consensus that teaching intelligible English pronunciation to learners is much more important than getting rid of their foreign accents or awkwardness. However, it seems that many Japanese English learners (JELs) still wish to acquire so-called native-like English pronunciation [1, 2]. The question is whether they know what comprises native-like English pronunciation and whether their native-like pronunciation really sounds native to actual native speakers of English (ENSs).

According to Munro and Derwing [3], foreign accents are defined based on a listener’s perception of the difference between pronunciation and that of the L1 community. This means that if there is no perception of the difference, that is, if it sounds

unmarked or natural to ENSs, the pronunciation does not have a noticeable foreign accent.

Thus, this study requested that both JELs and ENSs rate the naturalness of English sounds recorded by Japanese learners and studied which features affect their perception of naturalness. This reveals that JELs and ENSs share similar judgments on the naturalness of English pronunciation and that segmental features, such as vowel epentheses have more influence on their perception of foreign accents than suprasegmental features, such as intonation patterns. This also demonstrates that JELs are slightly stricter with segmental features than ENSs.

2. Previous studies

Many studies have attempted to verify which phonological feature has the most influence on L2 English pronunciation. Some studies have argued that suprasegmental features, such as intonation and rhythm, have strong effects on L2 English pronunciation. For example, Cruz-Ferreira [4] states that intonation is the last “stronghold” of a foreign accent. Yabuuchi and Sato [5] suggest that the following three suprasegmental features are important in making English pronunciation sound more natural: i) a wider range of fundamental frequency (F0); ii) the proper location of pauses; and iii) a fast speech rate. Kang [6] reports the effects of lexical stress in addition to these three factors. Magoku, Ueno and Nishiyama [7] compare the effect of a wider F0 range on the perception of natural English pronunciation with that of intonation patterns, and show that a wider F0 range does not necessarily lead to the perception of more natural English pronunciation, although speeches with contextually appropriate intonation patterns are perceived as more natural.

Other studies have suggested the importance of segmental features. For example, Kashiwagi and Snyder [8] demonstrate that the pronunciation of some segmentals (/l/, /ɹ/, and /ð/) significantly impacts both JEL and ENS listeners’ perceptions of

accentedness. Trofimovich and Isaacs [9] show that the best predictors of accent ratings are syllable-level errors and non-native rhythmic patterns. Saito, Trofimovich, and Isaacs [10] also point out that, according to ENS listeners, accentedness is more strongly associated with segmental errors, including the insertion of extra vowels, than with intonation.

Some of these previous studies do not present clear definitions of naturalness or accentedness [5, 7]. Some of them mainly examine L2 speech by European-language speakers [4, 6, 9], while one of them focuses on how L2 speech sounds to ENSs [10].

The following studies focus on one segmental and one suprasegmental feature, namely, vowel epenthesis and contextually-odd intonation patterns, of L2 speech by JELs. Then, they examine which feature has a stronger influence on JEL and ENS listeners' ratings of the naturalness of English pronunciation and verify whether their ratings have agreed with each other.

3. Study 1: Pilot Study

3.1. Method

3.1.1. Materials

One JEL (male), who had studied abroad and undergone phonetic training, read four variations of Dialogue 1 (see Figure 1), whose underlined part is in a fall-rise intonation or in a fall intonation with or without vowel insertions between consonants.

A: What can we have for \tea?
 B: Well we've 'got some \strawberries.
 A: So 'what's the \problem?
 B: We 'haven't got any \cream.

Figure 1: Dialogue 1 ([11], p. 28, underline ours).

After recording these variations, the authors checked the recordings using *Praat* [12] in order to ensure that the recordings clearly present the target intonation patterns and vowel insertions.

3.1.2. Participants

A total of 38 listeners, including 28 JELs and 10 ENSs, were asked to rate the recordings. The JELs are college students who do not major in English, while four out of 10 ENSs teach or have taught English at university in Japan.

3.1.3. Procedures

Participants were asked to rate the naturalness of each recording based on a 10-point Likert scale. They examined Dialogue 1 and listened to the recordings of the underlined sentence. Then, they chose one of the 10 rating criteria that ranged from 1 (*unnatural*) to 10 (*natural*).¹ They were also able to write optional comments on each recording or the entire survey.

3.2. Results

This study verified which recording was perceived as more natural by analyzing the ratings of naturalness based on a 10-point Likert scale.

Inter-rater reliability was computed using Cronbach's alpha. ENSs' alpha was .50, suggesting that the assessments of two out of the 10 ENSs appeared unreliable. When the two listeners' assessments were excluded, the value increased to .80. The JELs' alpha was .88. These two alphas suggested that the ratings of eight ENSs and 28 JELs were considered sufficiently reliable.

The descriptive results are summarized in Table 1 and 2.

Table 1: Descriptive results of ENSs' ratings of naturalness in different intonation patterns and vowel epentheses.

Intonation pattern	With a fall-rise intonation (Appropriate)		With a fall intonation (Not appropriate, but typical without a context)	
Vowel epenthesis (VE)	M	SD	M	SD
-VE	6.4	1.19	5.9	1.55
+VE	4.9	1.24	4.6	0.92

Table 2: Descriptive results of JELs' ratings of naturalness in different intonation patterns and vowel epentheses.

Intonation pattern	With a fall-rise intonation (Appropriate)		With a fall intonation (Not appropriate, but typical without a context)	
Vowel epenthesis (VE)	M	SD	M	SD
-VE	5.6	1.93	6.8	1.95
+VE	5.1	1.96	4.7	2.04

To examine whether these ratings significantly differed between the recordings, these descriptive results were submitted separately to the Kruskal-Wallis H tests.² The tests revealed a significant difference: ENSs, $H(3) = 9.14$, $p = .03$, Cramer's $V = .62$; JELs, $H(3) = 16.40$, $p = .00$, Cramer's $V = .44$.³

As post hoc analyses, pairwise comparisons were conducted with their alpha level set at $p < .05$. This showed that JELs' evaluation of both recordings without a vowel epenthesis between consonants was significantly higher for naturalness than those with a vowel epenthesis, and ENSs showed the same tendency. Among the recordings without vowel epentheses, ENSs rated the one with a contextually appropriate intonation higher, whereas JELs rated the one with an inappropriate but typical intonation higher. This difference may be due to the fact that JELs do not know which intonation is contextually appropriate.

4. Study 2: Follow-Up

4.1. Method

4.1.1. Materials

Two JELs (male graduate students) read Dialogue 1 and Dialogue 2 (see Figure 1 and 2). They read the underlined sentences in four ways: in a fall-rise intonation (FR) or in a fall intonation (F), and with or without vowel insertions between consonants (\pm VE). Then, the authors recorded them and collected 16 recordings ($2_{\text{speakers}} \times 2_{\text{sentences}} \times 4_{\text{patterns}}$) and additional 4 recordings as distractors.

A: How's your department's new faculty member?
 B: He is very nice and extremely helpful.
 A: What's the problem?
 B: Well, he might leave soon. He said he got a job offer from a computer company.

Figure 2: Dialogue 2.

The authors checked the recordings using *Praat*, as was done in Study 1.

4.1.2. Participants

Participants are 45 listeners (33 JELs and 12 ENSs). All JELs are university students aged 18 to 22 years. Their English proficiency level is roughly equivalent to the CEFR (the Common European Framework of Reference for Languages) A1 to B2 levels, based on their EIKEN grades⁴ or TOEIC L & R scores.⁵ One ENS is from the United States and lives in Japan, while all other ENSs live in the United States.

4.1.3. Procedures

Following Study 1, the listeners rated the naturalness of the recordings based on a 10-point Likert scale after they had listened to the recordings.

4.2. Results

Regarding the inter-rater reliability of the rating scores, ENSs' alpha was .95 and JELs' was .85. These two alphas suggest that their ratings were considered highly reliable.

Table 3 and 4 show the descriptive results, namely, ENSs' and JELs' mean scores and the standard deviation of their ratings of the recordings of Dialogue 1 and 2 (D1 and D2) read by two speakers in a fall-rise intonation or in a fall intonation with or without vowel insertions.

Table 3: Descriptive results of ENSs' and JELs' ratings of naturalness of Dialogue 1 in different intonation patterns and with/without vowel epentheses.

Speakers and Listeners	Speaker 1		Speaker 2	
	ENSs' rating (M)	JELs' rating (M)	ENSs' rating (M)	JELs' rating (M)
Recordings				
-VE, FR	7.1 (SD=1.82)	7.2 (SD=2.04)	7.2 (SD=1.58)	4.1 (SD=1.98)
-VE, F	7.5 (SD=1.87)	8.1 (SD=1.45)	8.1 (SD=1.44)	8.0 (SD=1.34)
+VE, FR	4.2 (SD1.62)	2.8 (SD=1.43)	5.3 (SD=1.64)	2.1 (SD=1.19)
+VE, F	5.8 (SD=1.61)	4.0 (SD=1.65)	4.4 (SD=1.60)	2.9 (SD=1.85)

Table 4: Descriptive results of ENSs' and JELs' ratings of naturalness of Dialogue 2 in different intonation patterns and with/without vowel epentheses.

Speakers and Listeners	Speaker 1		Speaker 2	
	ENSs' rating (M)	JELs' rating (M)	ENSs' rating (M)	JELs' rating (M)
Recordings				
-VE, FR	6.7 (SD=1.64)	6.5 (SD=2.34)	6.7 (SD=1.94)	6.9 (SD=2.05)
-VE, F	7.8 (SD=1.72)	8.3 (SD=1.59)	7.5 (SD=1.55)	7.8 (SD=1.80)
+VE, FR	4.2 (SD=1.17)	2.8 (SD=1.23)	4.6 (SD=1.64)	2.2 (SD=1.08)
+VE, F	4.9 (SD=1.14)	2.6 (SD=1.21)	4.8 (SD=2.07)	2.1 (SD=1.07)

The mean scores of these tables are summarized in Figures 3 and 4.

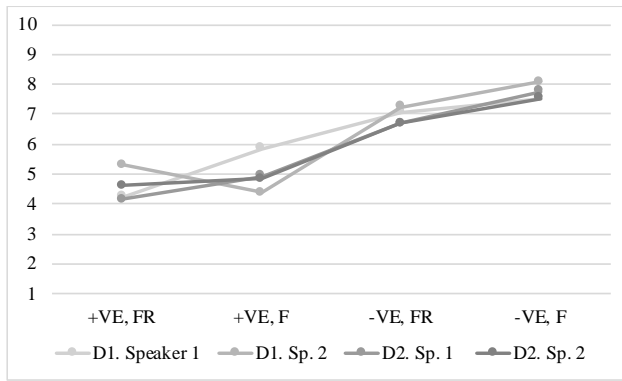


Figure 3: ENSs' means of ratings in naturalness in different intonation patterns and vowel epentheses.

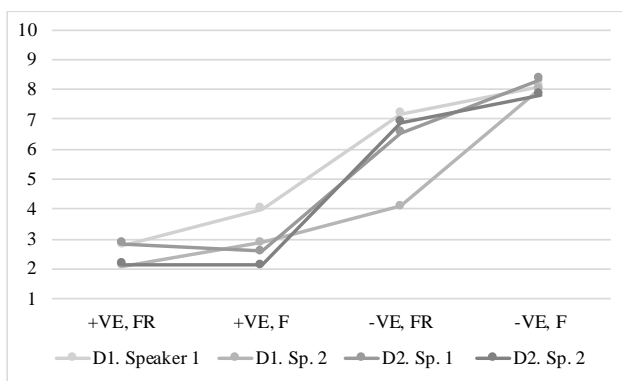


Figure 4: JELs' means of ratings in naturalness in different intonation patterns and vowel epentheses.

As shown in Tables 3 and 4 and Figures 3 and 4, first, both ENSs and JELs rated speech samples without vowel epentheses as higher in naturalness than those with vowel epentheses. Second, both JELs and ENSs evaluated speech in fall-rise intonation as lower in naturalness than in fall intonation. Third, JELs rated speech samples with vowel epenthesis as lower than ENSs did.

The results of the Kruskal-Wallis H tests for the rating scores by ENSs show significant differences in each data set (D1-Sp.1, $H(3) = 24.54$, $p = .00$, Cramer's $V = .83$; D1-Sp.2, $H(3) = 16.39$, $p = .00$, Cramer's $V = .68$; D2-Sp.1, $H(3) = 18.63$, $p = .00$, Cramer's $V = .72$; D2-Sp.2, $H(3) = 24.01$, $p = .00$, Cramer's $V = .82$) and JELs (D1-Sp.1, $H(3) = 85.73$, $p = .00$, Cramer's $V = .93$; D1-Sp.2, $H(3) = 92.83$, $p = .00$, Cramer's $V = .97$; D2-Sp.1, $H(3) = 70.98$, $p = .00$, Cramer's $V = .85$; D2-Sp.2, $H(3) = 94.08$, $p = .00$, Cramer's $V = .98$). Particularly, pairwise comparisons revealed that the influence of vowel epentheses on ENSs' and JELs' perception of naturalness was significantly greater than that of intonation patterns (alpha level set at $p < .05$).

5. Discussion

In both studies, ENSs and JELs rated recordings without vowel epenthesis as higher in naturalness than those with vowel epentheses. This suggests that vowel epenthesis can have a more negative influence on naturalness perception than contextually inappropriate intonation. This concurs with the findings on ENSs' ratings by Saito et al. [10], that is, accentedness is more strongly associated with the insertion of extra vowels than with intonation. This may be due to the fact that vowel epenthesis can lead to segmental and suprasegmental errors, such as rhythmic changes.

ENSs in Study 2, as well as JELs, evaluated the recordings in a fall intonation as higher than those in a fall-rise intonation, while ENSs in Study 1 evaluated recordings in a fall intonation as lower. This difference may come from the different backgrounds of the ENS raters. Four out of eight reliable ENSs in Study 1 are or were English teachers in Japan, while three have advanced degree in Linguistics. They may have evaluated the recordings based on their linguistic and pedagogical backgrounds. On the other hand, all ENSs in Study 2, live in the United States (except for one) and seem to have less linguistic or pedagogical backgrounds. They may have regarded the contextually-odd intonation pattern as a possible intonation pattern in a naturally-occurring language setting, considering it as just one element among all other kinds of clues, including pauses, facial expressions, or gestures, to convey the speaker's intention.

Finally, JELs in Study 2 rated the recordings with vowel epentheses as lower in naturalness than ENSs did. It seems that JELs may be slightly too strict with vowel epenthesis, which is one of the most typical features in Japanized English resulting from the Japanese moraic syllable structure. As Zhang and Elder [13] have discussed, non-native speakers of English may obey the norms of standard English even more strictly than native speakers. Since ENSs' evaluation is not as strict as JELs', current English education in Japan may focus more on teaching "correct" pronunciation than is necessary, even for those who want to acquire native-like English pronunciation.

Thus, it may be effective for JELs to learn to pronounce segmental elements correctly in order to improve their English pronunciation, although they may do not have to work that hard to get rid of a vowel between consonants to make their English pronunciation sound natural.

6. Conclusions

This study examined whether JELs share similar judgment on natural English pronunciation with ENSs and what feature affects their perceptions. ENSs and JELs rated four types of recorded L2 speech (a fall-rise intonation or a fall intonation with/without vowel insertions between consonants) in terms of naturalness. It was discovered that JELs and ENSs had similar tendencies in their perceptual evaluation of English speech; that is, vowel epenthesis have a more negative influence on the evaluation of English speech in naturalness than contextually inappropriate intonation patterns. In addition, JELs may have tougher standards in their evaluation of English pronunciation than ENSs, as shown in Study 2.

The findings in this study should be interpreted within the context of several limitations. Further studies are needed to establish the construct validity of the measures. This study calls for future research to replicate and extend the current research framework.

7. Acknowledgements

We would like to thank the ISAPh 2018 organizing committee members and the international scientific committee members for their invaluable assistance and reviews.

8. References

- [1] Derwing, T. M. 2003. What do ESL students say about their accents? *Canadian Modern Language Review* 59, 547–567.
- [2] Scales, J., Wennerstrom, A., Richard, D., Wu, S. H. 2006. Language learners' perceptions of accent. *TESOL Quarterly* 40, 715–738.
- [3] Munro, M. J., Derwing, T. M. 1995. Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning* 45, 73–97.
- [4] Cruz-Ferreira, M. 1989. A test for non-native comprehension of intonation in English. *IRAL-International Review of Applied Linguistics in Language Teaching* 27, 23–40.
- [5] Yabuuchi, S., Sato, H. 2001. Prosodic characteristics of Japanese EFL learners' oral reading: Comparison between good and poor readers. *Language Education & Technology*, 38, 99–112.
- [6] Kang, O. 2010. Relative salience of suprasegmental features on judgments of L2 comprehensibility and accentedness. *System* 38, 301–315.
- [7] Magoku, T., Ueno, M., & Nishiyama, A. 2018. Eigo hatsuon no shizensa to pitch haba: Nihon no eigogakushusha to eigobogowasha no ninshikiteki hyoka [The relationship between naturalness of English pronunciation and pitch ranges: The perceptual evaluation by Japanese learners and native speakers of English]. *Gakugei* 64, 189–193.
- [8] Kashiwagi, A., Snyder, M. 2014. Intelligibility of Japanese college freshmen as listened to by native and nonnative listeners. *Japan Association of College English Teachers Journal* 58, 39–56.
- [9] Trofimovich, P., Isaacs, T. 2012. Disentangling accent from comprehensibility. *Bilingualism: Language and Cognition* 15, 905–916.
- [10] Saito, K., Trofimovich, P., Isaacs, T. 2016. Second language speech production: Investigating linguistic correlates of comprehensibility and accentedness for learners at different ability levels. *Applied Psycholinguistics* 37, 217–240.
- [11] Wells, J. C. 2006. *English Intonation: An Introduction*. Cambridge: Cambridge University Press.
- [12] Boersma, P., Weenink, D. 2016. Praat: Doing phonetics by computer Version 6.0.20 <http://www.praat.org/>
- [13] Zhang, B., Elder, C. 2010. Judgments of oral proficiency by non-native and native English-speaking teacher ratings: Competing or complementary constructs? *Language Testing* 28, 31–50.
- [14] TOEIC® SCORE VS CEFR <http://yourenglishtest.com/article/index/art/54>

¹ Although the notion of “naturalness” as a scale to rate L2 speech may still seem somewhat vague, it would avoid the ambiguity of “accentedness” or “foreign-accentedness” among listeners of different L1 communities regardless of their linguistic backgrounds. It was also assumed that JELs would rate the naturalness of L2 speech based on what they think of as L1 speech.

² Although a parameterized test should have been used for the rating scores by JELs, this study chose the Kruskal-Wallis H test, which is a rank-based non-parametric test, in order to use the same test used for ENSs' evaluation; ENS participants were extremely limited.

³ The values of test statistics for JELs' evaluation were higher than those for the ENSs' one because the number of JEL listeners who participated in this study was higher than the number of ENSs.

⁴ *EIKEN, Jitsuyo Eigo Gino Kentei* (Test in Practical English Proficiency), is one of the most popular English-proficiency tests in Japan. It has seven levels of tests: 1, Pre-1, 2, Pre-2, 3, 4 and 5, each corresponding to C1, B2, B1, A1 in CEFR. JELs in Study 2 were at levels 3 to Pre-1.

⁵ TOEIC is an acronym for the Test of English for International Communication. In Study 2, JELs' scores ranged from 450 to 515 on the TOEIC Listening & Reading Test (TOEIC L & R). This can be considered CEFR A2, according to the official TOEIC site [14].