Strong influence of prosody on the perception of foreign accent

Peipei Wei[1], Lucy Gubbins[2] and Kaori Idemaru[1]

[1]University of Oregon, [2]National Geographic
peipei@uoregon.edu, lucygubbins@gmail.com, idemaru@uoregon

ABSTRACT

This study investigates the acoustic characteristics of non-native speech which give rise to the perception of a foreign accent. Japanese speech samples were collected from L1 English and Mandarin speakers studying Japanese. The acoustic patterns of the L2 and native speech were related to the accentedness ratings of the speech provided by native Japanese listeners. Results indicate the important role that prosodic features play in affecting perception of foreign accent across different L1 backgrounds. They also indicate intriguing cross-linguistic differences: vowel production was related to English accent, whereas consonant production was related to Chinese accent in Japanese.

Keywords: foreign accent, second language speech, prosody

1. INTRODUCTION

Most second/foreign language (L2) learners who acquire a foreign language in adulthood retain a discernible foreign accent however fluent they might otherwise become, e.g., [9]. Understanding the nature of foreign accents is important because the perception of an accent leads some to decide that the speaker is not understandable [6] and/or less credible [5], even when the message is accurately conveyed.

A broad range of speaker characteristics has been examined, and many studies converge in demonstrating that the onset age of learning exerts a crucial influence on the development of an accent, e.g., [2], whereas studies are inconclusive about the influence of other factors, such as length of residence and formal instruction ([7] for review).

In addition to these speaker characteristics, it is also important to understand what acoustic components of non-native speech give rise to the perception of a foreign accent. Whereas the investigation of speaker characteristics deepens our understanding of the possible cause of a foreign accent, an acoustic investigation helps us better understand what it is that we call a 'foreign accent' in the first place.

A few studies that have examined the acoustic correlates of foreign accents, i.e., in American-accented Thai [14]; Korean-accented English [11]; Chinese-accented English [3], seem to indicate that different acoustic features affect perception of a foreign accent across different target languages, but suggest a potentially strong influence of prosodic features. These studies have renewed interest in the question of whether prosodic features contribute to the sense of an accent more than segmental features do, or vice versa.

An important consideration for research on foreign accent is the question of the generality or specificity of accent perception across different languages. What comprises a sense of accent may well be different from language to language, while also depending on the native language of the non-native speaker. It is also conceivable that, given the common auditory faculty, we all detect a foreign accent in similar ways cross-linguistically.

Foreign accent research examining both segmental and prosodic features in multiple language pairings is needed to address these issues. The present study attempts to contribute to this research by examining the effect of both segmental and prosodic features on perceived foreign accent in L2 production of Japanese by L1 English and Chinese learners.

2. METHODOLOGY

2.1. Participants

A total of 42 speakers participated in the production task, providing Japanese speech samples (Table 1). All were residing in the US at the time of testing. Among the 32 learners of Japanese, 17 (7 L1-English and 10 L1-Chinese) were enrolled in a second-year Japanese language course, and the remaining 15 were either enrolled in or had completed a fourth-year Japanese language course at the time of testing. Learners were recruited from these levels to include L2 samples at a range of abilities. All L1-English speakers reported English as their primary language. All L1-Chinese participants, including those from Taiwan, reported Mandarin as their primary language. Ten native Japanese speakers, all from Tokyo, also participated.

An additional 22 native Japanese participated as raters in the rating task. They were also residing in the US at the time of testing. Eleven were assigned to rate a set of speech samples including L1-English

learners' and native speakers' (Raters 1) and the other 11 were assigned to rate another set including L1-Chinese learners' and native speakers' (Raters 2). The rating pattern showed strong inter-rater reliability: ICC r = .987 for Raters 1, and r = .988 for Raters 2, and thus, all raters were included in the analyses.

Table 1: Participants.

Participants	n	Mean age	Note
L1-English	16	21.3	All from US
L1-Chinese	16	22.3	12 from PRC, 4 from Taiwan
Japanese speakers	10	21.1	All from Tokyo
Japanese raters	22	21.8	17 from Tokyo, 5 from other areas

2.2. Production task

Six short Japanese sentences (Table 2) were used in a delayed repetition task [2], [11] to elicit speech samples. The first author (female) and a male Japanese speaker recorded 6 dialogues ("prompts"), in which one of the test sentences was a response to a question, for example:

(1) Question (male): Nihongo no kurasu wa doo desu ka? *(How is your Japanese class?)*
Response (female): Tanoshii desu yo. (*It is fun.)*

Each dialogue was recorded twice, except that the second time the response was not included, so as to prompt participants to produce the response.

The 6 recorded prompts were presented to participants auditorily using E-Prime (Psychology Software Tools, Inc.) and participants were recorded producing the test sentences in a sound booth using a flash digital recorder (Marantz PMD 670) and a microphone (SHURE Beta 87) at a sampling rate of 44 kHz and 16-bit quantization. The 6 prompts were presented in random order, with each presented three times consecutively. All measures reported here were taken from either the second or third repetition of the target sentences.

2.3. Rating task

The production of test sentences by L2 and native speakers was amplitude normalized to 75dB and presented to native Japanese raters to examine perceived accentedness. Each trial began with an auditory presentation of an utterance (test sentence) and the presentation of a visual analog scale [12] on a computer monitor. Raters then rated each utterance for degree of foreign accent by sliding the bar in the middle of the scale using a computer mouse. An accent score between 0 and 100 was registered depending on where the bar was moved to between the leftmost point = 0 ("no accent") and the rightmost point = 100 ("extremely heavily accented"). All raters completed the rating task in approximately 15 minutes. Accent scores were z-score normalized and averaged across 6 sentences for each speaker.

Table 2: Test sentences

	Test Sentences
1	Tanoshii desu yo. *It is fun.*
2	Kuruma ga kaitai desu. *I want to buy a car.*
3	Nihongo no jisho ga kaitai desu ne. *I want to buy a Japanese dictionary.*
4	Daigaku no tonari ni arimasu yo. *It's next to the university.*
5	Ashita wa eega o mitai desu ne. *I want to see a movie tomorrow.*
6	Rokuji ni okimasu. *I get up at six o'clock.*

2.4. Acoustic measurements and analysis

A number of acoustic measurements were made of features potentially characterizing native and non-native Japanese productions.

Segmental features: The frequencies of the first and second formants (F1 and F2) were measured in vowels [i], [e], [a], [u], [o], [i:], and [e:], at the mid-point of each vowel using Praat 5.2.18, and were Lobanov normalized [10]. The duration of stop closure and voice onset time (VOT) were measured using the standard reference points [4]. Each closure and VOT duration was rate-normalized to the average CV mora duration of each utterance.

Prosodic features: As measures of global linguistic rhythm VarcoΔV, V%, VarcoΔC, nPVI, e.g, [15], were computed based on durational measurements. J_ToBi, Japanese Tones and Break Indices, [13] was adopted to evaluate pitch accent and intonation patterns of the speech samples. The target tones for the test sentences were identified based on the native Japanese productions. The L2 learners' tones that did not match the target were counted as pitch accent errors. As measures of fluency, speaking rate (utterance duration including pauses divided by the mora count), pause duration in milliseconds, and pause frequency, were examined. All the acoustic measurements were averaged across 6 test sentences for each speaker.

Stepwise multiple regression analysis was conducted separately for the L1-English and L1-

Chinese data, with the foreign accent rating scores as the outcome variable and the acoustic measurements as the predictors, in order to explore the relationship between perceived foreign accent and acoustic characteristics of the speech samples. The native speaker data were included in both data sets.

3. RESULTS

3.1. L1-English learners

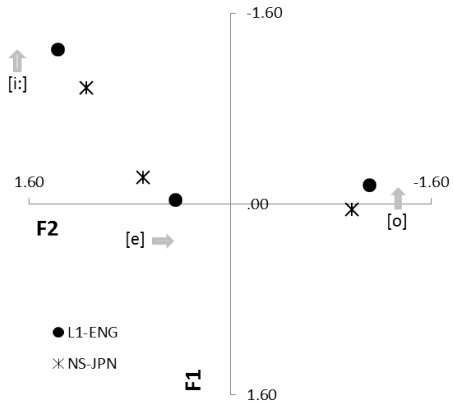
The best-fit model predicting foreign accent rating of L1-English learners' (and native Japanese speakers') speech samples showed that pitch error score, F1 in [o], F2 in [e], F1 in [i:], VarcoΔV and speaking rate influenced foreign accent rating ($R = .967$, $p < .0001$), explaining 93.6 % of the accent rating pattern. Among the predictors, pitch error score was the strongest predictor (.410), followed by F1 in [i:] (-.325), F2 in [e] (-.245), VarcoΔV (-.233), F1 in [o] (-.189) and speaking rate (.155) (values indicating Beta, $p < .05$ for all predictors). The variance inflation factors (VIFs) of the predictor variables were considerably smaller than 10, indicating the independent contribution of each factor. The means of the factors retained in the model are reported in Table 4.

It is noteworthy that pitch error score, a prosodic feature, emerged as the strongest predictor in the model. The mean F1 and F2 values of the three vowels, [o], [e], [i:], retained in the model, are plotted in Figure 1 separately for L1-English (circles) and Japanese speakers (asterisks). The arrows indicate the direction of the greater degree of perceived accent along the F1 and F2 dimensions. The results indicate that speakers were rated more accented when vowels [i:] and [o] were produced in a higher tongue position, and [e] was produced in a more centralized position. In addition to pitch error score, two more prosodic features were identified as significant predictors. VarcoΔV is a rate-normalized measure of variance in vowel durations. The results here indicate that smaller variance in vowel durations was predictive of greater foreign accent. In fact, the mean VarcoΔV in Japanese speakers' speech was 54.35, whereas that of L1-English learners' was 44.87 (Table 4). A vowel can assume various statuses vis-à-vis mora structure in Japanese: it can be part of a mora (in the case of a CV), an entire mora (in the case of a stand-alone V), or it can comprise two moras (in the case of a long vowel V:). As mora is related to durational regularities [8], vowels implemented in these various mora statuses must vary in their durations. It is possible that L1-English learners' Japanese did not include native-like variation in vowel duration that should emerge due to the variant statuses of a vowel in a mora. The results regarding the other prosodic predictor, speaking rate, indicated that slower speaking rate contributed to a stronger foreign accent in English-accented Japanese.

Table 3: Mean and (SD) of the significant predictors

	L1-English learners	Japanese speakers
Pitch error	1.02 (.45)	00 (00)
F1 in [o]	-.16 (.19)	.05 (.23)
F2 in [e]	.43 (.18)	.69 (.27)
F1 in [i:]	-1.29 (.23)	-.97 (.36)
VarcoΔV	44.87 (6.06)	54.35 (5.56)
Speaking rate	154.12 (15.11)	147.62 (10.83)

Figure 1: F1 and F2 means for the vowel predictors



3.2. L1-Chinese learners

The best-fit model predicting foreign accent rating on L1-Chinese learners' (and native Japanese speakers') speech samples showed that pitch error score, closure duration of voiceless stops, speaking rate and VarcoΔV affected the degree of perceived foreign accent ($R = .964$, $p < .0001$). This model explained 92.9 % of the accent rating and indicated that pitch error score was the most predictive factor (.505), followed by speaking rate (.265), closure of voiceless stops (-.236), and VarcoΔV (-.205) (values = Beta, $p < .05$ for all predictors). VIFs of all the predictors were considerably smaller than 10.

As with English-accented Japanese, pitch error score was the strongest predictor of perceived foreign accent in L1-Chinese learners' Japanese. Also consistent with the L1-English data, speaking rate and VarcoΔV contributed to the perception of foreign accent in Chinese-accented Japanese, with slighter variation in vowel durations and slower speech being more accented.

In contrast to the English-accented data, one consonant factor, but none of the vowel factors, was retained in the model. Shorter closures in voiceless stops were related to a stronger accent rating. The mean of this factor for L1-Chinese learners is quite short relative to that for Japanese speakers (Table 4). While Japanese speakers' stop closure for voiced stops was on average .203 (SD = .031), the L1-Chinese value hovers almost at the mid-point between Japanese voiceless and voiced closure durations.

Table 4. Mean and (SD) of the significant predictors

	L1-Chinese learners	Japanese speakers
Pitch error	.85 (.34)	00 (00)
Voiceless closure	.32 (.05)	.43 (.06)
Speaking rate	172.66 (15.81)	147.62 (10.83)
Varco V	46.72 (5.78)	54.35 (5.56)

4. DISCUSSION AND CONCLUSION

Three features have emerged as common acoustic sources of foreign accent across English-accented and Chinese-accented Japanese: error in pitch accent and intonation, insufficient variation in vowel durations, and slow speech, all of which are prosodic features. In particular, error in pitch accent and intonation exerted the strongest influence across both L1 accent types.

It is noteworthy that prosodic features were identified as affecting perceived foreign accent. Some previous research, e.g. [11], took steps to filter speech samples to reduce segmental information, out of a concern that the influence of segmental features could overwhelm prosodic features in processing speech for foreign accent. Our results here have shown that relevant prosodic features may exert a strong influence even in the presence of notable deviations in segmental features.

There were also interesting differences between English-accented and Chinese-accented Japanese with regards to the effect of segmental features. Whereas deviation in vowel production affected perception of accent in L1-English learners' Japanese, deviation in stop production had the same effect in L1-Chinese learners' Japanese.

The current study of foreign accent in L2 Japanese has shown that there is a possible generality and specificity of accent perception across different languages. Important future questions are whether the strongest influence of pitch accent and intonation, or prosodic features in general, is common among learners of other L1 backgrounds learning Japanese, and whether it is common for other target L2 languages as well. The current results and previous work together suggest this may be a possibility.

5. REFERENCES

[1] Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., Tohkura, Y. I. 1997. Training Japanese listeners to identify English/r/and/l: IV. Some effects of perceptual learning on speech production. *J. Acoust. Soc. Am*. *101*(4), 2299-2310.

[2] Flege, J. E., Munro, M. J., MacKay, I. R. 1995. Factors affecting strength of perceived foreign accent in a second language. *J. Acoust. Soc. Am*. *97*(5), 3125-3134.

[3] Hardman, J. B. 2010. *The intelligibility of Chinese-accented English to international and American students at a US university*. Dissertation, The Ohio State University.

[4] Idemaru, K. 2008. Acoustic covariants of length contrast in Japanese stops. *J. Int. Phon. Assoc. 38*(02), 167-186.

[5] Lev-Ari, S., Keysar, B. 2010. Why don't we believe non-native speakers? The influence of accent on credibility. *J. Exp. Soc. Psychol. 46*(6), 1093-1096.

[6] Munro, M. J., Derwing, T. M. 1995. Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Lang. Learn. 45*(1), 73-97.

[7] Piske, T., MacKay, I. R., Flege, J. E. 2001. Factors affecting degree of foreign accent in an L2: A review. *J. Phon. 29*(2), 191-215.

[8] Port, R. F., Dalby, J., & O'Dell, M. (1987). Evidence for mora timing in Japanese. The Journal of the Acoustical Society of America, 81(5), 1574-1585.

[9] Scovel, T. 1969. Foreign accents, language acquisition and cerebral dominance. *Lang. Learn. 19*(3-4), 245-253.

[10] Thomas, E. R., Kendall, T. 2007. NORM: The vowel normalization and plotting suite. *Online Resource: http://ncslaap.lib.ncsu.edu/tools/norm*.

[11] Trofimovich, P., Baker, W. 2006. Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Stud. Second Lang. Acquis. 28*(01), 1-30.

[12] Urberg-Carlson, K., Munson, B., Kaiser, E. 2009. Gradient measures of children's speech production: Visual analog scale and equal appearing interval scale measures of fricative goodness. *J. Acoust. Soc. Am. 125*, 2529.

[13] Venditti, J. J. 2005. The J_ToBI model of Japanese intonation. *Prosodic typology: The phonology of intonation and phrasing*, 172-200.

[14] Wayland, R. 1997. Non-native production of Thai: Acoustic measurements and accentedness ratings. *Appl. Linguist. 18*(3), 345-373.

[15] White, L., Mattys, S. L. 2007. Calibrating rhythm: First language and second language studies. *J. Phon. 35*(4), 501-522.