ACOUSTIC CHARACTERISTICS OF FOREIGN ACCENT IN L2 JAPANESE: A CROSS-SECTIONAL STUDY

Misaki Kato[1], Kaori Idemaru[1] and Kimiko Tsukada[2]

[1]University of Oregon, [2]Macquarie University
misaki@uoregon.edu, idemaru@uoregon.edu, kimiko.tsukada@gmail.com

ABSTRACT

This study reports an exploratory analysis of the acoustic characteristics of second language speech that affect perception of a foreign accent. Japanese speech samples were collected from native Japanese speakers and native English learners of Japanese across different instructional levels and learning backgrounds. Native Japanese raters rated the speech samples for foreign accentedness. While pitch accent, articulation rate, and vowel duration influenced perceived accentedness of the speech samples in general, the relative importance of these acoustic features varied across speaker groups. The current results shed light on issues related to development of second language speech, and the perceptual relevance of the development as perceived by lay listeners.

Keywords: foreign accent, second language speech, acoustic cues, Japanese.

1. INTRODUCTION

It has been well documented that adult second language (L2) learners retain a discernible foreign accent [16, 17, 20]. Understanding the nature of foreign accents is important because the perception of an accent can lead listeners to think that the speaker is not understandable [13] and/or less credible [12], even when the message is accurately conveyed.

While prior studies demonstrated that a broad range of speaker characteristics (e.g., the onset age of learning) possibly influence foreign accents [14, 18], it is less clear what it is that we call a 'foreign accent' in the first place. In particular, it is not well-understood what acoustic components of non-native speech give rise to the perception of a foreign accent.

Previous studies have demonstrated mixed results regarding the acoustic sources of perceived foreign accents. Some studies have reported a stronger influence of segmental features over prosodic features [21, 30], and others reported the opposite pattern [1, 9, 11, 27, 29]. It is possible that some of these divergent findings are due to the different methodologies used in the studies, and also due to speakers' different first language (L1) backgrounds, given that L2 speech learning is influenced by the acoustic and perceptual relationship between learners' L1 and their target L2 sound systems [2, 5].

However, it is also possible that acoustic sources of L2 learners' accents change in the course of learning the target language. While linguistic factors that affect comprehensibility ratings (i.e., how easy it is to understand the speech) differ for low-level vs. high-level learners [10], we know very little as to whether acoustic characteristics that affect perceived foreign accents also vary for learners of different proficiency levels. The current study investigates this question by examining acoustic correlates of foreign accents of L2 learners across various instructional levels and learning backgrounds. The majority of studies of foreign accents have examined English as the target L2 [14, 18]. By examining Japanese as the target language, the current study aims to broaden the scope of foreign accent research.

2. METHOD

2.1. Speakers and speech materials

Thirty one native English learners of Japanese, 10 heritage speakers of Japanese (see Table 1) and 10 native Japanese speakers provided Japanese speech samples. All were residing in the US at the time of testing. We recruited 10 beginning-level learners from the first-year Japanese language course (7 females; henceforth, 1Y learners) and 21 intermediate-level learners from the second- (n = 10; 5 females) and fourth-year course (n = 11; 9 females; henceforth, 2Y4Y learners). In addition to those classroom learners, we recruited 10 heritage speakers of Japanese (7 females; henceforth, HS). For these heritage speakers, the age of acquisition of Japanese was at birth or 1 year. Finally, we recruited 10 native Japanese speakers (henceforth, NS), who were all from Tokyo.

Stimuli were 6 short Japanese sentences (see Table 2), chosen from a beginning level Japanese course book and deemed appropriate for their length and the range of segments included in them. The aforementioned native English learners and heritage speakers of Japanese, and native Japanese speakers recorded the stimulus sentences in a sound-attenuated booth, as they completed a delayed repetition task [6, 26] delivered via E-Prime [19]. A female native

Japanese speaker provided the model for all sentences in the task. The order of the sentences was randomized for each speaker. The recordings were amplitude normalized to 75dB.

Table 1: Learner information in years with the mean and the range in parenthesis.

	1Y	2Y4Y	HS
Age	19 (18-19)	21.3 (19-26)	20 (18-27)
Age of acquisition	16.9 (15-18)	17 (14-21)	.2 (0-1)
Length of instruction	1.56 (0-4)	4.64 (1.5-9)	13.2 (2-27)

Table 2: Test sentences and translation (Long vowels are underlined).

1. Tanoshii desu yo.
   It is fun.
2. Kuruma ga kaitai desu.
   I want to buy a car.
3. Nihongo no jisho ga hoshii 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.

2.2. Raters and foreign accentedness rating task

Twenty three native Japanese speakers participated as raters in the accentedness rating task, with 13 raters residing in the US, and 10 raters residing in Japan at the time of testing. The rating pattern showed strong inter-rater reliability (ICC r = .995). Thus, all raters were included in the analyses.

The recordings were presented to the native Japanese raters in a foreign accentedness rating task conducted via E-Prime. Prior to the task, participants were told that they would hear productions by native and non-native speakers. Each trial began with an auditory presentation of a test sentence and a visual presentation of an analog scale [15]. Raters were then prompted to rate each sentence for a degree of foreign accent by sliding the bar in the middle of the scale. Raters could drag the bar anywhere between the leftmost (0: "like a native speaker") and the rightmost (100: "extremely strong foreign accent") points. The order of sentence and speaker was randomized for each rater. The rating task took approximately 30 minutes. The ratings were z-score normalized for each rater.

2.3. Acoustic measurements

Segmental features: The frequencies of the first and second formants (F1 and F2) were measured in short vowels [i], [e], [a], [ɯ], [o], and in long vowels [iː], and [eː], at the vowel mid-point, using Praat 5.2.18 [3]. The vowel formants were then Lobanov normalized to control for differences due to the size of vocal tract [25]. The duration of stop closure and voice onset time (VOT) were measured for [t] and [k], using the standard reference points [8]. Each closure and VOT duration was rate-normalized to the average CV mora duration of each sentence to control for talker-inherent speaking rate.

Prosodic features: As measures of global linguistic rhythm, VarcoΔV, V%, VarcoΔC, and nPVI were computed (see [9] for details of these measurements). The long vowel duration divided by short vowel duration (CVV/CV) was calculated as a normalized measure of long vowel for the sentences that included long vowels (i.e., sentences 1, 3, 5 in Table 2). As a tonal measure, the Japanese Tones and Break Indices (J-ToBi: [28]) was adopted to evaluate pitch-accent and intonation patterns of the speech samples. The target tones for the test sentences were identified using the native Japanese productions as the target. The learners' tones that did not match the target were counted as pitch-accent errors. Finally, as fluency measures, pause duration and pause frequency were measured for a silent period longer than 100 ms [26]. Articulation rate (the rate of speech without pauses) was computed for each sentence.

3. RESULTS

3.1. Relative importance of acoustic cues

We examined the relative contribution of the acoustic variables to perceived accentedness in an exploratory fashion, using Random forests [4, 24], implemented by the *party* R package [22] with 500 trees. For a discussion of these techniques in the context of linguistics and sociolinguistics, see [24]. To examine whether the importance of acoustic variables for perceived accentedness persisted over different items, sentences were added to the predictors; thus the analysis was item-based, with each item (sentence) contributing one data point. Variable importance was computed via conditional permutation scheme, a method that accounts better for collinearity [23].

Fig. 1 shows the relative importance of predictors with all speaker groups collapsed (1Y, 2Y4Y, HS, and NS). Only the top 10 predictors are shown. As the figure shows, 3 of the 4 most important variables were prosodic variables, with the pitch-accent error being the most relevant. Other important prosodic variables were CVV/CV (rhythm measure) and articulation rate (fluency measure). The most relevant segmental variable was F2 in [e]. These patterns are consistent with [9], which examined intermediate native English and Chinese learners of Japanese.

Figure 1: Relative variable importance of the top 10 variables based on a random forest analysis for all speakers' accentedness ratings.

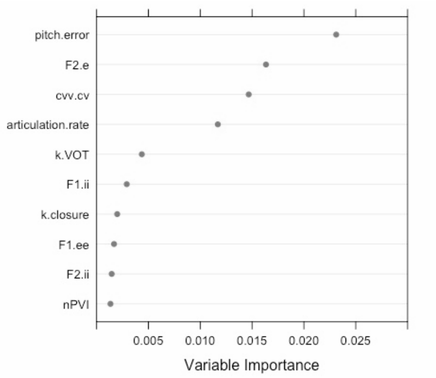


Figure 2: Relative variable importance of the top 10 variables based on a random forest analysis for each speaker group.

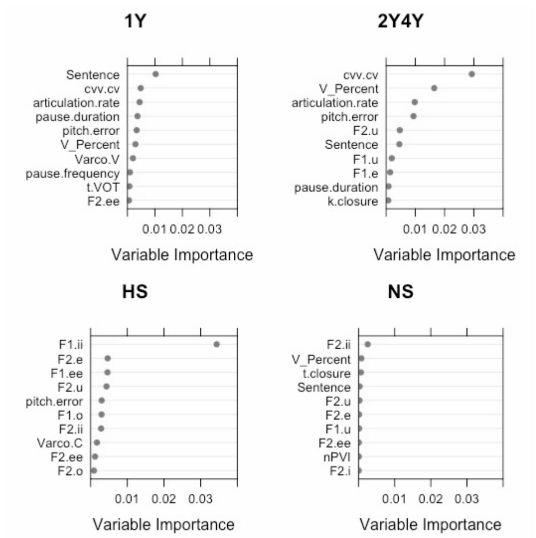


Fig. 2 reports the relative importance of predictors separately for each speaker group. The figure shows that the important variables were similar for 1Y and 2Y4Y. In particular, they shared many of the top 6 acoustic variables: articulation rate, CVV/CV (long vowel duration/short vowel duration), V% (proportion of vowel duration in a sentence), and pitch-accent error, all of which were prosodic variables. In contrast, segmental variables (e.g., F1 in [iː]) were more important than prosodic variables for HS. We note that sentence was the most important variable for 1Y, but it was ranked lower for other groups (e.g., the 6th for 2Y4Y). This seems to suggest that the sentence-by-sentence variance in speech materials affected 1Y learners' accent the most, and the influence of the sentences themselves were less salient for more advanced learners. Also, it is noteworthy that variable importances were overall much lower for 1Y and NS compared to 2Y4Y and HS, suggesting that raters based their evaluations of accentedness for 1Y and NS less straightforwardly on the acoustic variables that we examined here.

3.2. The effects of learners' backgrounds and acoustic variables on perceived accentedness

As the next step, we examined the effects of the important acoustic variables (selected based on the exploratory analyses above) on perceived accentedness across speaker groups. Five acoustic variables were selected based on both overall analysis and group-by-group analysis. Pitch-accent error (error count per utterance), F2 in [e], CVV/CV (duration ratio of CVV to CV, where VV is a long vowel), and articulation rate (utterance duration in milliseconds per mora) were selected because they were the 4 most important acoustic variables emerged from the overall analysis (Fig 1). Additionally, we selected V% (percentage of the total duration of vowels in utterance), which was another highly ranked prosodic variable for 1Y and 2Y4Y (Fig. 2). Since the maximum pitch-accent error varied across sentences, we analysed by-speaker averages of the acoustic variables.

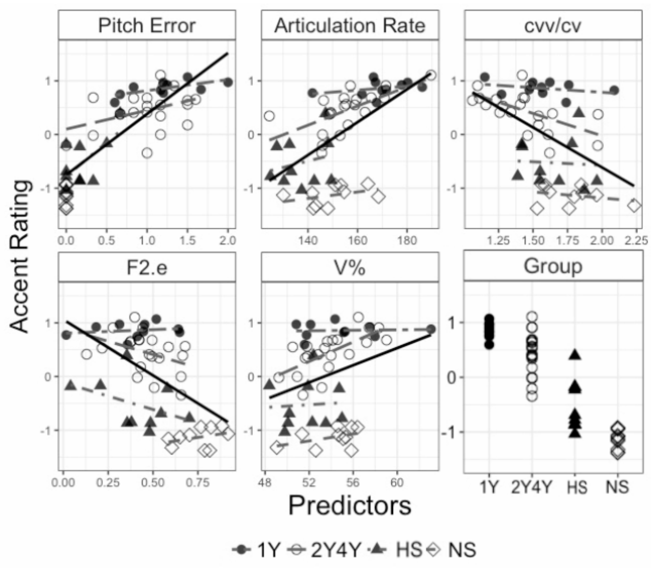
Fig. 3 illustrates the correlations between these predictors (5 acoustic variables and speaker groups) and accent ratings, with each data point representing a speaker. It is clear that the accent ratings differed by speaker groups (right bottom panel): 1Y learners were perceived to be the most accented, followed by 2Y4Y, HS, and NS. Further, the effects of some acoustic variables on accent ratings seem to differ across groups, indicated by slopes of the dotted lines. For example, the effect of F2 in [e] seems greater for 2Y4Y and HS compared to other groups.

We then tested the effects of speaker groups, the 5 acoustic variables, and the interactions between the speaker groups and acoustic variables, on accent ratings. All the acoustic variables were mean centered. The speaker group was Helmert-coded to compare between 1Y vs. 2Y4Y (beginning learners vs. intermediate learners), between the means of 1Y and 2Y4Y vs. HS (classroom learners vs. heritage speakers), and between the means of 1Y, 2Y4Y, and HS vs. NS (learners + heritage speakers vs. native speakers). These contrast-coded speaker group terms were entered as the only predictors in the base model, and they were significant predictors of accent ratings [$F(3, 47) = 85.03$, $p < .001$, $R^2 = .83$]. All the group terms were significant predictors ($p < .01$).

The 5 acoustic variables were then added to the base model using a step-wise method, first entering the least important variable (Fig. 1), and then adding the next least important variable, and so forth. All variables significantly improved the model fit except for F2 in [e]: V% [$F(1, 46) = 6.93$, $p < .05$, $p\eta 2 = .13$], articulation rate [$F(1, 45) = 7.03$, $p < .05$, $p\eta 2 = .16$], CVV/CV [$F(1, 44) = 4.28$, $p < .05$, $p\eta 2 = .09$], F2 in [e] [$F(1, 43) = 3.87$, $p = .056$, $p\eta 2 = .08$], pitch-accent error [$F(1, 42) = 7.42$, $p < .01$, $p\eta 2 = .15$]. The variance inflation factors of these predictor variables were all less than 4, indicating the independent contribution of each predictor. The interactions between the acoustic variables and the speaker group terms were then added to the model. The significant effect was found in the interaction between pitch-accent error and the 1Y2Y4Y vs. HS group term [$F(1, 41) = 8.17$, $p < .01$, $p\eta 2 = .17$].

These results demonstrated that group factors explained a large amount of variability in the accent ratings ($R^2 = .83$), suggesting that native Japanese listeners were sensitive to the differences in speakers' characteristics (i.e., instructional levels and learning backgrounds). Further, while 4 of the 5 acoustic variables tested here (pitch-accent error, articulation rate, CVV/CV, and V%) significantly influenced the accent ratings in general, the effect of pitch-accent error was greater for HL compared to 1Y2Y.

Figure 3: Correlation between predictors and accentedness ratings. Lines are best-fitting linear regression lines (dotted lines for different speaker groups; solid lines for the whole data).



4. DISCUSSION AND CONCLUSION

This study explored acoustic sources of foreign accents in L2 Japanese produced by native English learners of different instructional levels and learning environments. The results demonstrated that error in the tonal pattern (pitch-accent and intonation) most strongly affected the degree of perceived foreign accents when learner groups were pooled together. Additionally, vowel duration measure (CVV/CV) and articulation rate were also important, suggesting the strong influence of prosodic features on perceived accents of L2 Japanese in general.

In addition to these general patterns, we also observed that the relative importance of these acoustic variables differed across groups. There were clear differences between classroom learners (first-, second-, fourth-year learners) and heritage speakers; prosodic variables influenced accents of classroom learners more than segmental variables, whereas the pattern was the opposite for heritage speakers (3.1). While the influence of pitch-accent error was greater for heritage speakers' speech than classroom learners' speech (3.2), pitch-accent error mattered less than vowel formant measures for heritage speakers' speech (3.1).

It is possible that learning Japanese prosodic features are inherently more challenging than learning segmentals for native English learners as [9] argued. The L1 English and L2 Japanese differ markedly in terms of prosodic systems, e.g., [7], while L1 segmentals map fairly straightforwardly onto L2 segments. This explains the general patterns we observed for the pooled data as well as the results of classroom learners (3.1); however, it does not explain the pattern we observed with heritage speakers. Why segmental features influence perceived accents more than prosodic features for heritage speakers' speech is an open question.

It is also noteworthy that the effects of acoustic variables on perceived accents were generally different for first-year learners from other groups. The weight of acoustic variables were overall low explaining accent ratings of first-year learners (3.1). Even the 4 most important acoustic variables (pitch-accent error, articulation rate, CVV/CV, and F2 in [e]) did not seem to affect accent ratings, as the linear regression lines were rather flat (Fig. 3). These results suggest that the acoustic variables examined in this study did not characterize the variation in perceived accents in the beginning learners' speech; yet their speech was clearly more accented than intermediate (second- and fourth-year) learners' speech as indicated by the rating scores. Further research is needed to examine what other factors affect perceived accents of beginning learners.

These findings provide a step forward in understanding what makes an L2 speaker sound foreign accented, and whether it differs depending on speaker characteristics. We suggest that different acoustic features may need to be targeted in pronunciation instructions for learners with different backgrounds.

5. REFERENCES

[1] Anderson-Hsieh, J., Johnson, R., Koehler, K. 1992. The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Lang. Learn., 42*(4), 529–555.

[2] Best, C. T. 1995. A direct realist view of cross-language speech perception. In: W. Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research*. Baltimore, MD: York Press. 171-204.

[3] Boersma, P., Weenink, D. 2015. Praat: Doing phonetics by computer (Version 5.2.18) [Computer software].

[4] Breiman, L. 2001. Random forests. *Mach. Learn. 45*, 5-32.

[5] Flege, J. E. 1995. Second language speech learning: Theory, findings, and problems. In: W. Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research*. Baltimore, MD: York Press. 233-277.

[6] 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.

[7] Hyman, L. M. 2009. How (not) to do phonological typology: the case of pitch-accent. *Lang. Sci. 31*(2), 213–238.

[8] Idemaru, K., Guion, S. G. 2008. Acoustic covariants of length contrast in Japanese stops. *J. Int. Phon. Assoc. 38*(2), 167-186.

[9] Idemaru, K., Wei, P., Gubbins, L. 2018. Acoustic sources of accent in second language Japanese speech. *Lang. Speech*, 1-25.

[10] Isaacs, T., Trofimovich, P. 2012. Deconstructing comprehensibility. *Stud. Second Lang. Acquis. 34*(3), 475–505.

[11] Kang, O. 2010. Relative salience of suprasegmental features on judgments of L2 comprehensibility and accentedness. *System, 38*(2), 301–315.

[12] 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.

[13] 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.

[14] Munro, M. J., Derwing, T. M. 2011. The foundations of accent and intelligibility in pronunciation research. *Lang. Teach. 44*(3), 316–327.

[15] Munson, B., Schellinger, S. K., Carlson, K. U. 2012. Measuring speech-sound learning using visual analog scaling. *Perspectives on Language Learning and Education*, *19*(1), 19-30.

[16] Oyama, S. 1976. A sensitive period for the acquisition of a nonnative phonological system. *J. Psycholing. Research. 5*(3), 261–283.

[17] Patkowski, M. 1994. The critical age hypothesis and interlanguage phonology. In: M. Javas (ed.), *First and second language phonology*. San Diego, CA: Singular Publishing Group. 205–221.

[18] 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.

[19] Schneider, W., Eschman, A., Zuccolotto, A. 2002. *E-Prime: User's guide*. Psychology Software Incorporated.

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

[21] Sereno, J., Lammers, L., Jongman, A. 2016. The relative contribution of segments and intonation to the perception of foreign-accented speech. *Applied Psycholinguistics, 37*(2), 303–322.

[22] Strobl, C., Boulesteix, A.-L., Zeileis, A., Hothorn, T., 2007. Bias in random forest variable importance measures: illustrations, sources and a solution. *BMC Bioinform. 8*(25). http://www.biomedcentral.com/1471-2105/8/25

[23] Strobl, C., Malley, J., Tutz, G. 2009. An introduction to recursive partitioning: Rationale, application and characteristics of classification and regression trees, bagging and random forests. *Psych. Methods. 14*, 323–348.

[24] Tagliamonte, S. A., Baayen, R. H., 2012. Models, forests, and trees of York English: was/were variation as a case study for statistical practice. *Lang. Var. Change. 24*(2), 135--178.

[25] Thomas, E. R., Kendall, T. 2015. NORM: The vowel normalization and plotting suite. *Online Resource: http://lingtools.uoregon.edu/norm/norm1.*

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

[27] Trofimovich, P., Isaacs, T. 2012. Disentangling accent from comprehensibility. *Bilingualism: Lang. Cog. 15*(4), 905–916.

[28] Venditti, J. J. 2005. The J_ToBI model of Japanese intonation. In: S.-A. Jun (ed.), *Prosodic typology: The phonology of intonation and phrasing*. New York, NY: Oxford University Press. 172–200.

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

[30] Winters, S., O'Brien, M. G. 2013. Perceived accentedness and intelligibility: The relative contributions of F0 and duration. *Speech Comm. 55*(3), 486–507.