ESL learners' intra-speaker variability in producing American English tense and lax vowels

Bruce L. Smith, Eric Johnson, and Rachel Hayes-Harb
The University of Utah

Nonnative (L2) English learners are often assumed to exhibit greater speech production variability than native (L1) speakers; however, support for this assumption is primarily limited to secondary observations rather than having been the specific focus of empirical investigations. The present study examined intra-speaker variability associated with L2 English learners' tense and lax vowel productions to determine whether they showed comparable or greater intra-speaker variability than native English speakers. First and second formants of three tense/lax vowel pairs were measured, and Coefficient of Variation was calculated for 10 native speakers of American English and 30 nonnative speakers. The L2 speakers' vowel formants were found to be native-like approximately half of the time. Whether their formants were native-like or not, however, they seldom showed greater intra-speaker variability than the L1 speakers.

Keywords: L2 vowel production, intra-speaker variability, speech acoustics

1. Introduction

A variety of factors have been found to influence second language (L2) learners' pronunciation, including age of arrival in the L2 country, the age formal instruction in the L2 began, amount of L2 exposure and use, and so forth (e.g., Best, 1995; Bongaerts, 1999; Cebrian, 2006; Flege, 1995; Moyer, 1999; Piske, MacKay, & Flege, 2001; Slevc & Miyake, 2006). Another important factor that can affect the acquisition of L2 pronunciation is that, to a greater or lesser extent, the first and second languages typically have differences in their vowel and consonant inventories. Even when the native language (L1) and the L2 share some of the "same" vowel or consonant phonemes, however, allophonic properties of those sounds may vary in subtle ways between the two languages (Bohn & Flege, 1992; Flege, 1995; Flege

https://doi.org/10.1075/jslp.15050.smi
Journal of Second Language Pronunciation 5:1 (2019), pp. 139–164. issn 2215-1931 | e-issn 2215-194x

& Hillenbrand, 1984; Flege, Schirru & MacKay, 2003) and between speakers (e.g., Kartushina & Fraunfelder, 2014). For example, Bradlow (1995) found that the vowels /i/, /e/, /u/, and /o/ had somewhat different first and second formant values between English and Spanish, both of which nominally contain these four vowels in their phonemic inventories. Strange, Weber, Levy, Shafiro, and Hisagi (2007) made similar observations when considering the acoustic properties of German, French, and American English vowels.

How similar L2 speakers' vowel and consonant productions are to those of native speakers has been the focus of a considerable amount of research (e.g., Cebrian, 2006; Flege, Munro, & MacKay, 1995; Flege et al., 2003; Schertz, Cho, Lotto, & Warner, 2015; Rallo Fabra & Romero, 2012; Smith, Hayes-Harb, Bruss, & Harker, 2009). However, whether or not L2 speakers may be more *variable* in producing L2 sounds than L1 speakers is an issue that has received much less attention. Indeed, variability has often been viewed as "statistical noise" rather than a potentially significant aspect of second language acquisition (see, e.g., Verspoor, Lowie & van Dijk, 2008).

Although native speakers of American English show a certain amount of acoustic variability in their speech (e.g., Smith, 2000), they typically make reasonably consistent and distinct contrasts between, for example, formant patterns of tense versus lax vowels, vowel duration preceding voiced versus voiceless obstruents, voice onset time of voiced versus voiceless stops, and so forth, especially in "clear speech" or other optimal speaking conditions (Bond & Moore, 1994; Bradlow & Bent, 2002; Ferguson, 2004). Why might we expect non-native speakers to exhibit more production variability than native speakers? According to Verspoor et al. (2008), variability may indicate an "unstable" system in the process of changing (2008, p. 215); we might thus expect variability to be a prominent feature of interlanguage systems. In the domain of speech production, learners' manifestations of individual phoneme targets in production may vary due to less well-developed motor control for those targets relative to native speakers. Another possibility is that the individual targets themselves may be less "stable" than those of native speakers due to the developing nature of L2 phonetic/phonological representations. Despite these possibilities, and their potential to inform our understanding of the nature and stability of learners' phonetic/phonological representations in the L2, very little research has examined L2 speech production variability from an acoustic perspective (c.f., Baese-Berk & Morrill, 2015; Hattori & Iverson, 2009; Lotto, Sato, & Diehl, 2004; Schertz et al., 2015).

In one of the few studies that has considered L2 speakers' acoustic variability, Wade, Jongman, and Sereno (2007) compared vowel productions by six native speakers of English and six native speakers of Spanish learning English as a second language (ESL). They found that the L1 English speakers' vowels (as a

group) tended to "cluster in regular, elliptical patterns" on measures of height and backness, whereas comparable vowels produced by the ESL speakers manifested "amorphous, heavily overlapping categories" (p. 130). They indicated that the standard deviation values of the group of L2 speakers averaged about one-third greater than those of the L1 group. It is not necessarily surprising that group (i.e., inter-speaker) variability among nonnative speakers would be somewhat greater than such measures for native speakers, since factors such as age of arrival in the L2 country, amount of formal study and experience with the L2, and so forth can make it difficult to assemble a homogeneous L2 group. For instance, Wade and colleagues' six native speakers of Spanish came from six different Central and South American Spanish-speaking countries, which likely contributed to at least some amount of the greater inter-speaker variability they showed relative to the L1 group. Wade et al. indicated that *intra-speaker* (i.e., within-speaker or token-to-token) variation is also an important factor to consider in L2 studies; however, they did not provide any data of this nature.

When considering the issue of speakers' acoustic variability, it is important to recognize the distinction between intra- and inter-speaker variability because they represent two crucially different circumstances pertaining to L2 speech production. Inter-speaker variability relates to how similar/disparate the performance of a *group* of L2 speakers is relative to the speech production patterns of a comparison group of native speakers. For example, are the L2 speakers' vowel formant values, voice onset time values, or vowel durations distributed in a manner comparable to those of a group of native speakers? That is, is the performance of the L2 group as *homogeneous* as that of the L1 group? By comparison, intra-speaker variability pertains to the speech of *individual* L2 speakers and how consistent (i.e., "stable") each person is in his/her own production of L2 speech sounds, as compared to the degree of variability shown by individual native speakers. Relatively few studies of L2 speech production have been directly concerned with either of these types of variability, but intra-speaker variability in particular has received very limited attention. Yet because intra-speaker variability pertains to the performance of individual speakers, it is a factor that (in addition to target similarity) is to some degree directly related to an L2 speaker's speech production abilities. By comparison, inter-speaker variability is largely indicative of the extent to which researchers are able to enter subjects with similar L2 backgrounds, experience, and abilities into their study.

The primary purpose of the present study was thus to determine whether L2 speakers would show greater or comparable *intra-speaker* variability relative to native speakers in attempting to produce American English vowels, in an effort to better understand the similarities and differences between native and non-native speech production. In this regard, four potential relationships (i.e.,

scenarios) seemed plausible between vowel production accuracy and variability, namely: (1) L2 speakers might be similar to L1 speakers in terms of both their vowel formants and the variability of those productions; (2) L2 speakers might produce vowel formants similar to those of L1 speakers, but be more variable in their productions; (3) L2 speakers might produce different vowel formants than L1 speakers, but be no more variable in their productions; (4) L2 speakers might produce different formants than L1 speakers and also be more variable in their productions. Stated otherwise, do L2 speakers who are more native-like in their vowel productions tend to be less variable than L2 speakers who are less native-like in their speech, or might L2 speakers have comparable variability to L1 speakers even if they do not have native-like productions?

The approach utilized for investigating these issues was to examine L2 speakers of American English who came from several different L1 backgrounds with vowel inventories that exhibit both similarities to and differences from one another, as well as similarities to and differences from American English. The primary goal of the study was to determine how *individual* L2 speakers would be distributed relative to the four scenarios outlined above with regard to American English tense and lax vowel formant patterns, and to assess their intra-speaker variability associated with producing those vowels. The American English contrast between tense and lax vowels was selected for analysis because of the difficulty it presents to ESL learners from various language backgrounds (Chang & Weng, 2013; Kondaurova & Francis, 2008). We were aware of no previous literature that considered these particular issues and, therefore, made no a priori assumptions about how the L2 speakers would be distributed among Scenarios 2, 3 and 4 in those instances when they did not produce native-like vowels and demonstrate comparable intra-speaker variability to the L1 speakers (i.e., as associated with Scenario 1).

A comment concerning our use of comparisons between ESL learners and native speakers of English is warranted. The goal of the present work is to attempt to characterize the intra-speaker variability of ESL learners (associated with how this variability relates to the phonetic "targets" they exhibit). In characterizing L2 production variability, we believe it is useful to consider the variability shown by native speakers for comparison, as doing so permits us to determine whether the variability exhibited by L2 learners is a unique feature of L2 production and thus requires attention in L2 speech models. However, despite our focus on comparisons between native and L2 speakers, we recognize the problematic nature of the so-called "native speaker model" in the context of language learning and teaching (see, e.g., Cook, 2015 for discussion).

2. Methods

2.1 Speakers

Forty individuals participated as speakers in the study. Ten (5 M, 5 F) were native speakers of American English (NE); their mean age was 24 years. The other 30 speakers consisted of three groups, each with 10 ESL learners. The native language backgrounds of these 30 ESL learners were Korean (5 M, 5 F), Mandarin (5 M, 5 F), and Spanish (5 M, 5 F). Speakers from these three language backgrounds were selected because their native language vowel systems have similarities to and differences from English, as well as similarities to and differences from one another. For example, all four languages' vowel inventories nominally contain the phonemes /i/, /e/, and /u/. However, the four languages differ in that the American English phonological system has lax vowel counterparts to these three tense vowels, whereas Korean has a long/short contrast among its tense vowels, but no tense/lax vowel contrast like English. Mandarin and Spanish have neither lax counterparts to their tense vowels, nor vowel length contrasts. Thus, in selecting native speakers of Korean, Mandarin, and Spanish, we are able to capture native languages with a range of phonological relationships to English, increasing the generalizability of our findings. For the same reason, each of the four sub-groups included both five male and five female speakers, so as not to limit the findings to only males or only females. The formant values were predictably different between the male and female speakers; however, this was not considered problematic since all four groups were balanced with five male and five female speakers. Furthermore, similar to including speakers from different language backgrounds, having both male and female speakers added to the diversity and thus representativeness of the subject pool. Finally, the diversity among the language backgrounds and speakers was not viewed as problematic given that the study focused on the production variability shown by *individual* speakers, as opposed to group variability.

Table 1. Mean age, age of arrival, length of residence, age began learning English, and length of formal English study for each of the three nonnative speaker groups; all values are in years with standard deviation in parentheses

Native language	Mean age at time of study	Age of Arrival (AoA) in US	Length of Residence (LoR)	Age began learning English	Length of formal English study
Korean (K1-K10)	21.9 (1.2)	17.9 (2.8)	3.8 (3.4)	10.5 (2.8)	6.9 (2.9)
Mandarin (M1-M10)	22.1 (2.3)	20.0 (2.9)	2.3 (1.3)	12.4 (4.3)	6.2 (4.3)
Spanish (S1-S10)	28.8 (12.1)	16.0 (10.1)	12.8 (6.8)	13.1 (4.9)	7.3 (2.6)

As can be seen in Table 1, the three L2 groups were similar in most respects, with the exception of the notably longer LoR for the native Spanish speakers than for the other two groups. However, as noted above, because the primary purpose of the study was not to compare the three different L2 groups with one another, but rather to examine the nonnative (NN) speakers as individuals (and as a larger group relative to the native speakers), the LoR difference by the native speakers of Spanish was not viewed as problematic for the interpretation of the findings.

2.2 Stimuli

All speakers were recorded in a sound-attenuated booth as they produced CVC words embedded in the carrier phrase, "I like to say ____ some of the time." A set of 58 sentences was repeated four times in a separate random order each time. The 58 sentences included 18 containing CVC sequences specifically associated with the American English tense/lax vowel contrast, as well as sentences associated with various other vowel and consonant patterns not related to this study and sentences containing different "filler" words at the beginning and end of each page. The 18 tense/lax target words were of the form CVd: *bead, bid, deed, did, heed, hid, bade, bed, fade, fed, shade, shed, cooed, could, shoed, should, who'd, hood*. The speakers produced each target word four times, resulting in a total of 72 target tokens for each speaker, or 12 tokens per vowel per speaker. Because the thirty L2 subjects came from different native language backgrounds and had at least somewhat varying amounts of experience with English, no attempt was made to control for the target words' familiarity or lexical frequency. Moreover, the primary analyses undertaken did not focus on individual words; rather multiple word pairs represented each tense/lax vowel contrast. Thus, in the event that a subject significantly mispronounced a particular target word on a routine basis (e.g., said "co-ed" rather than "cooed"), that token was omitted from acoustic analysis, and only the remaining words associated with a particular tense/lax contrast were analyzed. Such errors were very infrequent, occurring less than two percent of the time.

2.3 Measurements and analyses

Each speaker's 72 target words were measured acoustically utilizing Praat 5.3.23 (Boersma & Weenink, 2012), which simultaneously displayed waveforms and spectrograms of each utterance. The settings utilized included: Burg method with 50 Hz pre-emphasis, a frequency range from 0–6000 Hz, a Gaussian window shape, and a window length of .005 msec. For each subject, the first three formants of each of the tense/lax vowels /i-ɪ/, /e-ɛ/, and /u-ʊ/ were measured (at their temporal midpoints, so as to minimize possible co-articulatory effects

of surrounding consonants). Mean formant frequency was calculated across the various words that contained a particular vowel. Then mean formant frequency, standard deviation (SD), and Coefficient of Variation (CoV) were computed across the four repetitions of each of the tense and lax vowels. Coefficient of Variation was calculated for the four repetitions of the vowel in each target word using the formula: $sd/\bar{x}$ (standard deviation divided by the mean), which provided a relative value for the variability associated with each tense or lax vowel in each individual word produced by each of the 40 subjects. The analyses that were conducted focused on each talker's F1 and F2 values for the six vowels and the intra-speaker variability associated with those productions. This allowed us to evaluate issues regarding the four scenarios outlined in the Introduction pertaining to relationships between native versus nonnative vowel formant patterns and the associated intra-speaker variability.

Both intra-judge and inter-judge measurement reliability were assessed. Intra-judge reliability was determined by having the same investigator re-measure all of the productions by one randomly-selected NE speaker and one randomly-selected NN speaker. Inter-judge reliability was determined by having a second investigator measure the same NE speaker and the same NN speaker that had been previously measured by the first investigator. Intra-judge measurements for F1 and F2 (averaged across all vowels) differed, respectively, by 1 Hz and 6 Hz for the NE speaker and by 5 Hz and 6 Hz for the NN speaker (all a 1% difference or less). Inter-judge measurements for F1 and F2 differed by 17 Hz and 30 Hz for the NE speaker and by 3 Hz and 18 Hz for the NN speaker (all a 4% difference or less).

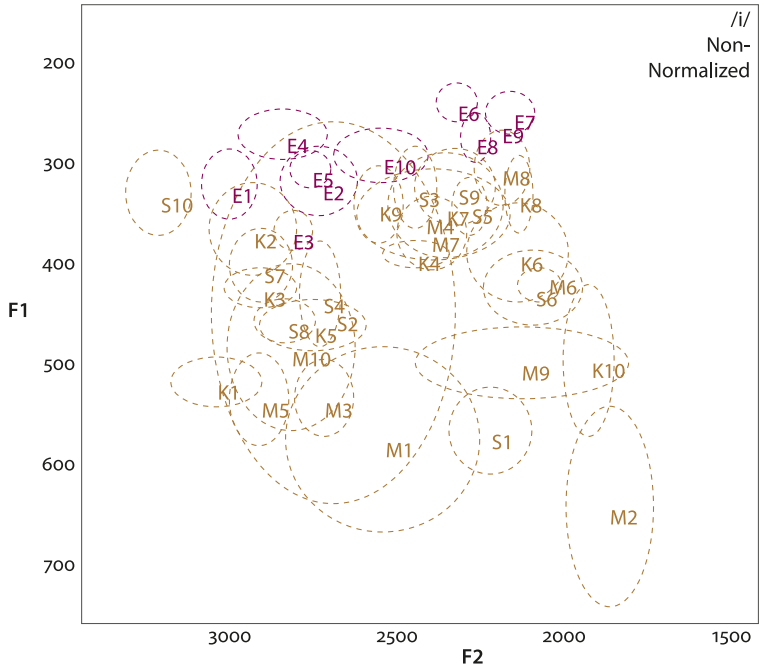
There exist a number of vowel normalization procedures, varying in their suitability for various data set characteristics (for a review, see Adank, Smits & van Hout, 2004). We performed a Bark Difference Metric normalization (Syrdal & Gopal, 1986) of the vowel formants produced by each of the subjects as a possible method of adjusting for the formant differences between males and females, as well as the differences among the vowel systems of the different language backgrounds of the nonnative speakers. Because comparisons using the original formant frequency measurements (in Hertz) and the Bark Difference normalized formant values resulted in similar statistical findings, we have presented the data in the Results section based on formant frequency values (i.e., in Hertz), since they are a more familiar unit of measure to most readers. In addition, they are a more direct measure of speech production per se, as opposed to Bark normalized vowels, which are associated with how the auditory system processes vowels that speakers produce.

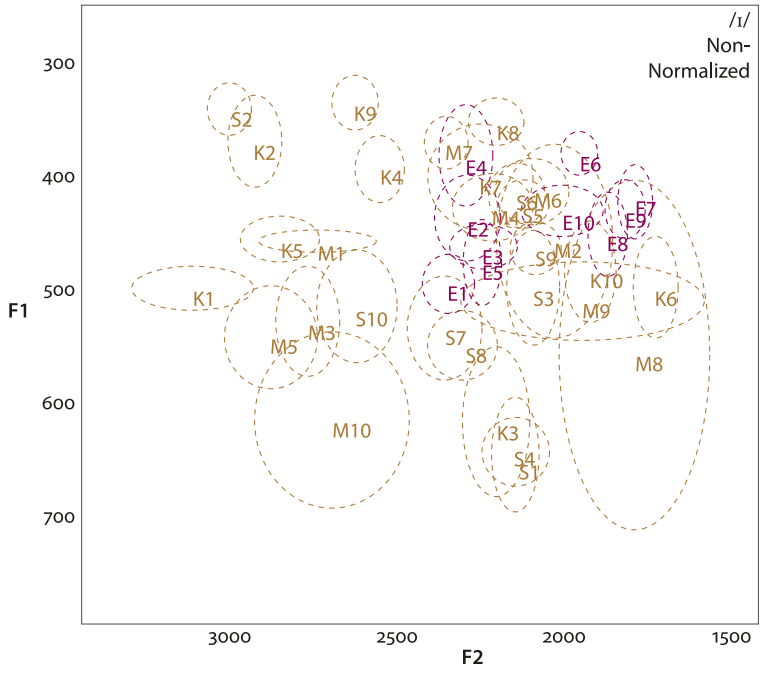
3. Results

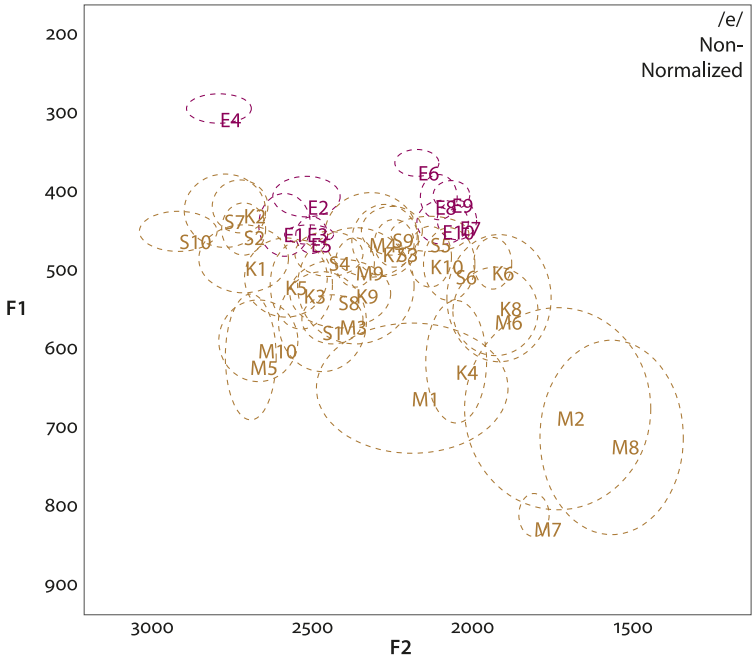
3.1 Vowel formants

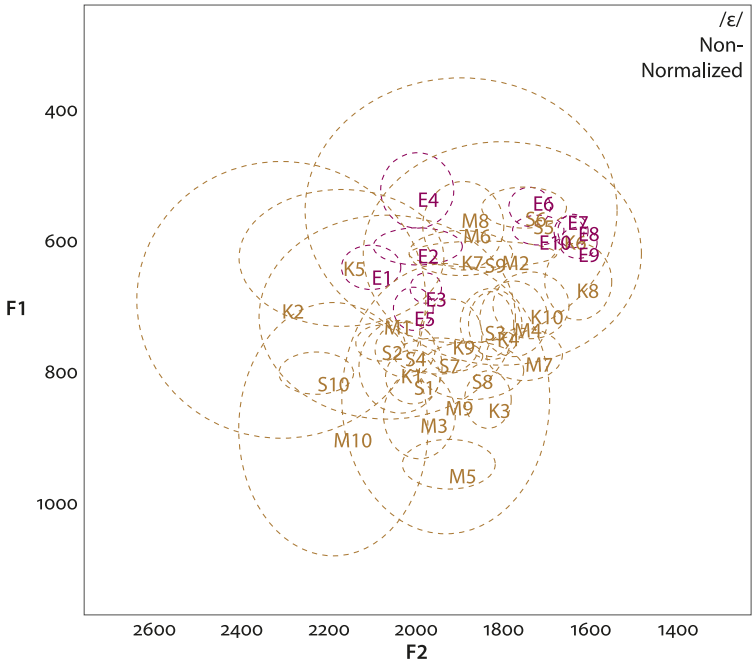
To provide an overview of the vowels produced by the L1 and L2 speakers, Figure 1 presents a series of six plots (one for each vowel) of the median F1 and F2 values for each of the forty speakers, with ellipses indicating one standard deviation around the mean.

In interpreting the plots presented in Figure 1, it is helpful to remember that F1 is inversely related to tongue height, where lower F1 values are associated with a higher tongue body, and that F2 is linearly related to tongue frontness/backness, where lower F2 values are associated with a tongue position that is further back. Note that in Figure 1, the (horizontal) F2 scale is ascending and the (vertical) F1 scale is descending. In this way, the top left corners are associated with higher and more front vowels, while the bottom right corners are associated with lower and more back vowels. Visual inspection of these vowel plots suggests that in general individual NE speakers tend to exhibit less intra-speaker variation than the NN speakers, evidenced by the smaller ellipses associated with NE vowels. In addition, the NE speakers as a group tend to have less inter-speaker variability, evidenced by the relatively closer proximity of their means to each other in the vowel space. We next turn to visual inspection of each individual vowel plot in turn. For the vowel /i/, several of the NN speakers as a group exhibited relatively higher F1 values (associated with lower tongue position), in particular the Mandarin speakers, while the F2 values for the non-native speakers are largely within the native speaker range. For the vowel /ɪ/, NN speakers differed from the native speakers on both the F1 and F2 dimensions, with several exhibiting higher F1 (lower tongue position) and/or higher F2 (more fronted tongue position). The vowel plot for /e/ indicates that the while the Spanish and the Korean speakers clustered nearer the NE speakers, several of the native Mandarin speakers exhibited higher F1 and lower F2 values. Of note in this figure are the relatively larger ellipses for M_1, M_2, and M_8 (but not M_7). For /ɛ/, native Mandarin speakers exhibit the highest F1 values and the largest ellipses; however, unlike for /e/, some of the largest ellipses are associated with speakers within the NE speaker acoustic space. For /u/, several Mandarin and Spanish speakers exhibited generally higher F1 values than the native speakers, and the Mandarin speakers had relatively lower F2 values than the other native language groups. The visual pattern for /ʊ/ is very similar to that of /u/.









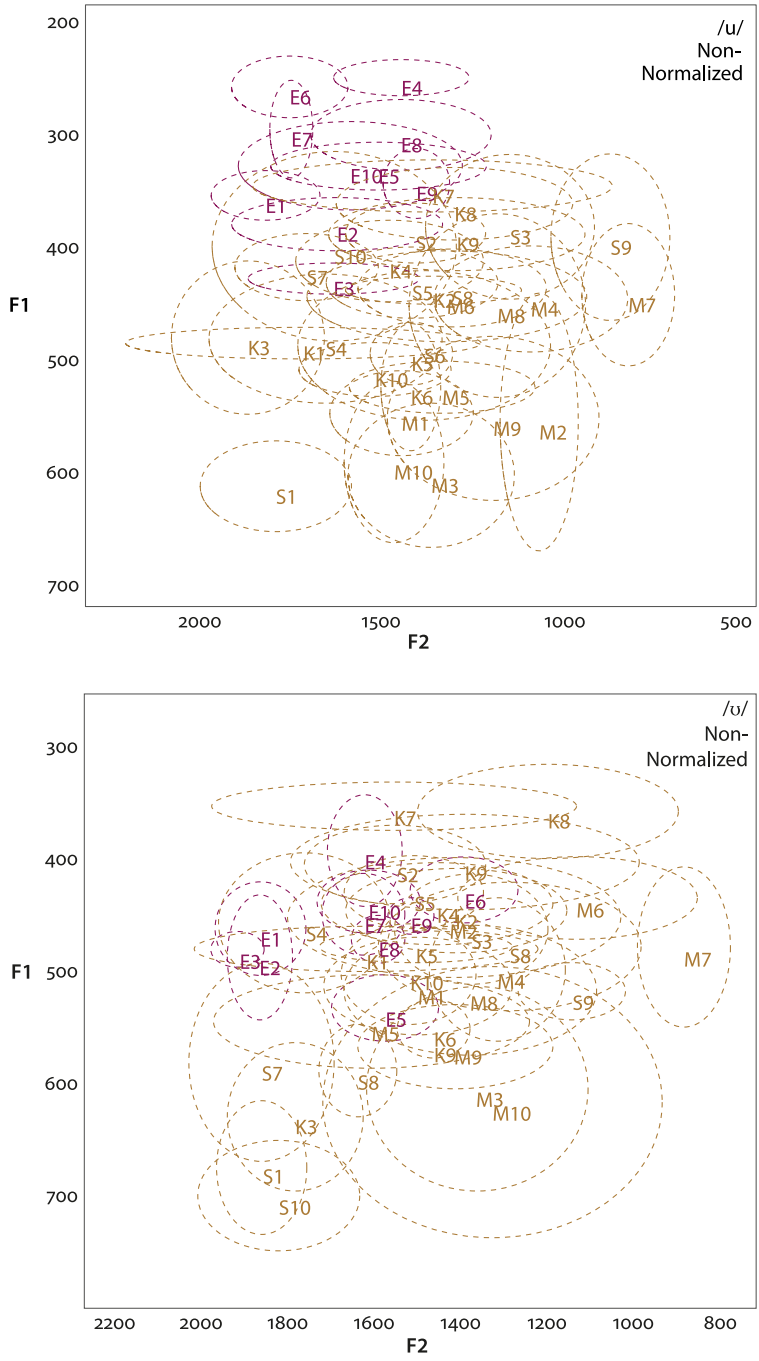


Figure 1. Vowel plots; F1/F2 values by speaker; ellipses indicate one standard deviation. These plots were created using NORM (Thomas & Kendall, 2007)

Figure 2 presents box and whisker F1 plots for the three tense-lax vowel pairs and the four native language groups; Figure 3 presents the same for F2. Because of the possibility that formant values might be distributed differently for the NE versus the NN speakers, non-parametric statistics were employed for making group comparisons. For F1 of /i/, a Kruskal-Wallis ANOVA determined that a significant difference existed across the groups ($H = 21.74$; $p < .0001$). Dunn's multiple comparisons indicated that the median F1 of /i/ was significantly different for the NE speakers compared to each of the three NN groups ($p < .05$ in all cases). However, there were no statistically significant differences among any of the NN groups for this comparison, which suggests that LoR or other differences among the NN groups had little, if any, impact for this particular measure. There was not a significant difference in the average formant values among the four groups of speakers for F1 of /ɪ/ (Kruskal-Wallis $H = 5.48$; $p > .05$). There were no significant F2 differences among the groups for either /i/ or /ɪ/ ($p > .05$ in both cases).

A Kruskal-Wallis ANOVA indicated that a statistically significant difference existed among groups for F1 of /u/ ($H = 21.82$; $p < .0001$). Dunn's multiple comparisons indicated that the median F1 of the tense vowel /u/ was significantly different for the NE speakers compared to each of the NN groups. (Once again, no significant differences were observed between any of the L2 sub-groups for F1 of /u/.) As with the lax vowel /ɪ/, there was also not a significant difference in the formant values among the NE and NN groups for F1 of the lax vowel /ʊ/ (Kruskal-Wallis $H = 6.492$; $p > .05$). There were also no significant F2 differences across groups for either /u/ or /ʊ/ ($p > .05$ in both instances).

As with the tense vowels /i/ and /u/, a significant difference was also found for F1 of /e/ (Kruskal-Wallis $H = 24.47$; $p < .0001$). Dunn's multiple comparisons indicated that the median F1 of /e/ was significantly different for the NE group versus the NK and the NM groups, but not between the NE and NS groups. (Although this finding could relate to the greater LoR by the native Spanish speakers, it should be noted that /e/ was the only one of the six vowels that showed such an effect for the Spanish-speaking sub-group. Therefore, such an interpretation seems unwarranted.) Similar to the other two lax vowels, F1 of /ɛ/ did not show a significant difference among the four groups (Kruskal-Wallis $H = 6.089$; $p > .05$). As with /i – ɪ/ and /u – ʊ/, there were also no significant differences across groups for F2 (bottom half of Figure 4) for either /e/ or /ɛ/ ($p > .05$ in both instances).

Summarizing across these three tense/lax vowel pairs, some important general patterns were observed. Namely, significant differences were found between the NE versus the NN groups only for F1 of /i/, /u/, and /e/, but not for the lax vowel counterparts of those vowels. In addition, no significant differences were observed between the NE and the NN groups for F2 for any of the six vowels. These findings regarding formant similarities and differences for the various tense and lax vowels

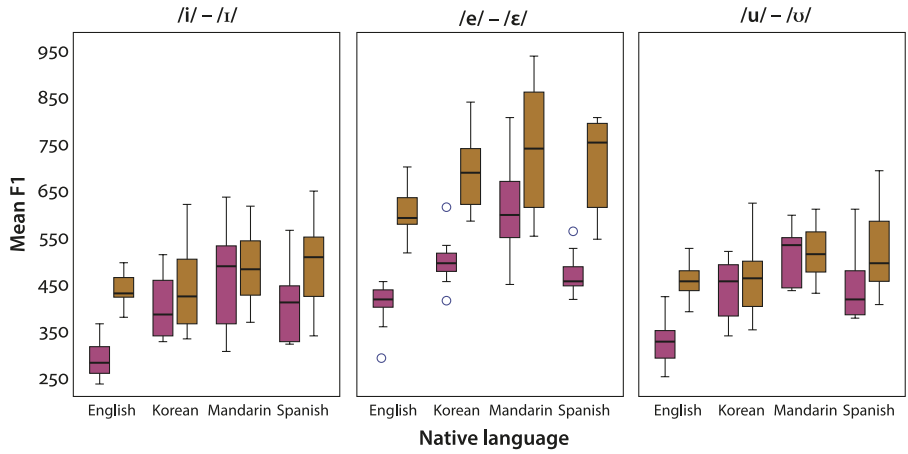


Figure 2. F1 box and whisker plots by native language and vowel. Brown bars represent the lax vowels and purple bars the tense vowels

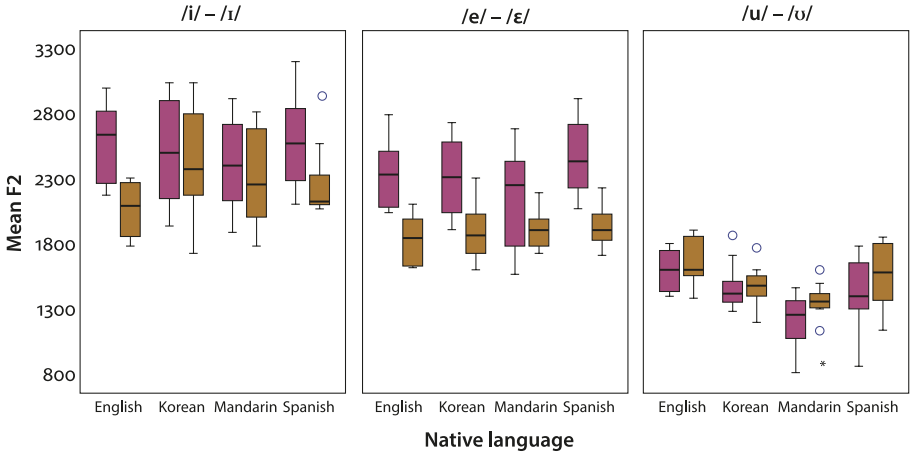


Figure 3. F2 box and whisker plots by native language and vowel. Brown bars represent the lax vowels and purple bars the tense vowels

produced by the L1 and L2 speakers provide data that is relevant for evaluating the four scenarios outlined in the Introduction, but additional information concerning formant variability is also needed before addressing those issues.

3.2 Vowel formant variability

Figure 4 displays box and whisker plots of F1 Coefficient of Variation (CoV) for the three tense/lax vowel pairs. Regarding intra-speaker variability in F1 shown by the NN versus the NE speakers, Kruskall-Wallis H values for /i, ɪ, e, ɛ, u, ʊ/ were 3.982, 4.873, 5.303, 4.923, 3.958, and 6.810, respectively; none of these H values was significant ($p > .05$ in all cases). Thus, although significant differences were observed between the NE and the NN groups for F1 of each of the three tense vowels, there were no significant F1 CoV differences between the NE and the NN groups for any of the tense or lax vowels.

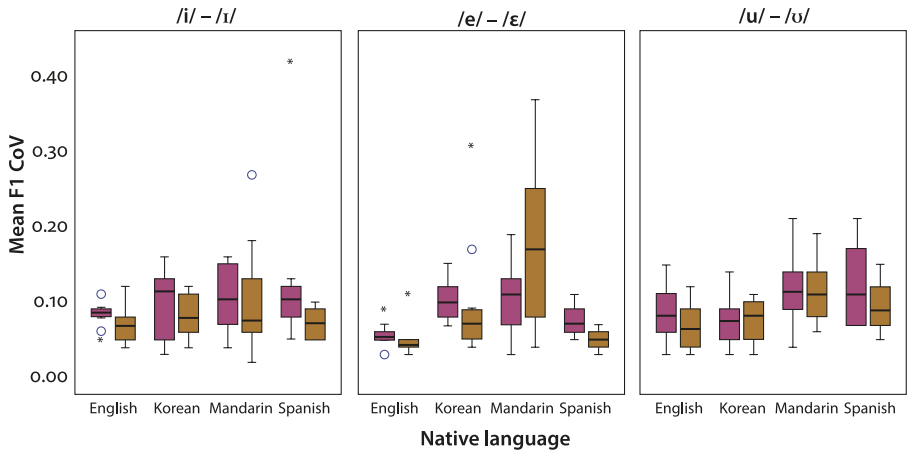


Figure 4. F1 CoV box and whisker plots by native language and vowel. Brown bars represent the lax vowels and purple bars the tense vowels

Figure 5 presents box and whisker F2 CoV plots for the three tense/lax vowel pairs. Significant differences were found among the groups for both /i/ and /ɪ/ (Kruskal-Wallis $H = 11.45$ and 16.28, respectively; $p < .01$ in both cases). Dunn's multiple comparisons determined that F2 CoV was significantly less for the NE versus the NM speakers for both /i/ ($p < .05$) and /ɪ/ ($p < .01$), and for the NK versus NM speakers for F2 CoV of /ɪ/ ($p < .01$). There were no other significant between-group differences for F2 CoV for either /i/ or /ɪ/. In addition, Kruskall-Wallis comparisons indicated that there were no significant differences across groups for F2 CoV for any of the other four vowels (i.e., /e, ɛ, u, ʊ/; $H = 1.511$, 2.968, 3.086, 4.527, respectively; $p > .05$ in all cases).

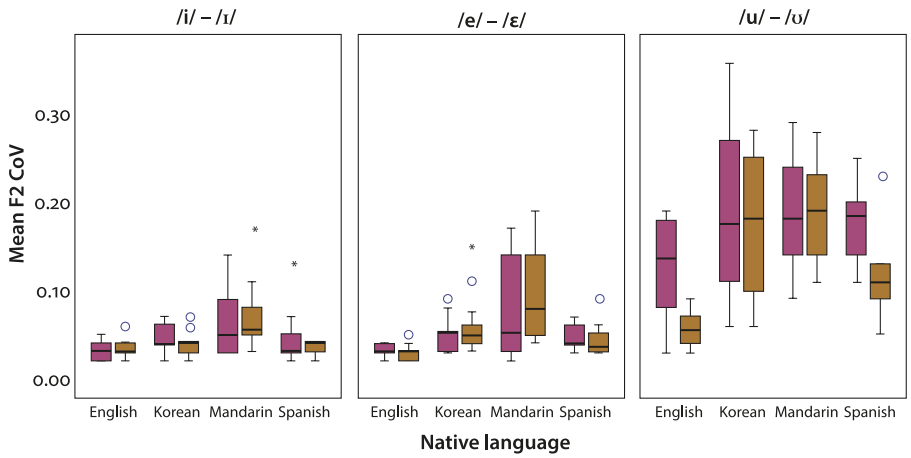


Figure 5. Mean F2 CoV box and whisker plots by native language and vowel. Brown bars represent the lax vowels and purple bars the tense vowels

3.3 Analyses of individual speakers' vowel formants and variability

The results discussed thus far have focused on vowel formant and intra-speaker variability among groups of speakers; here we attempt to characterize each of the thirty nonnative speakers in terms of their individual vowel formant and production variability profiles. In particular, we are interested in the assignment of individual NN speakers to the four L2 production scenarios outlined in the Introduction: (1) L2 speakers might be similar to L1 speakers in terms of both their vowel formants and the variability of those productions; (2) L2 speakers might produce vowel formants similar to those of L1 speakers, but be more variable in their productions; (3) L2 speakers might produce different vowel formants than L1 speakers, but be no more variable in their productions; (4) L2 speakers might produce different formants than L1 speakers and also be more variable in their productions

There were not enough tokens of each vowel produced by each individual speaker to conduct inferential analyses comparing their productions to those of the NE speakers. In an effort to nonetheless determine individual NN speakers' production performance relative to the four scenarios, each NN speaker's F1 and F2 vowel formant and CoV values were compared to the range of values associated with the NE group. Whether or not an individual NN speaker was specified as showing similar vowel formants and/or intra-speaker variability was based on whether their average F1 and/or CoV F1 values were greater than or less than two standard deviations different from the NE speakers' average formant and CoV values.

The focus in this analysis is on F1 and F1 CoV values because only F1 was associated with both native-like and nonnative-like formant values across the six vowels, whereas F2 values did not differ significantly between the NN and the NE speakers. As there were 30 nonnative speakers, each row in Table 2 totals to 30 and shows the distribution of speakers across the four L2 production scenarios in terms of vowel formant values and intra-speaker variability relative to the NE speakers. As can be seen in the "Total" row in Table 2, for all six vowels combined, Scenario 1 was most common (92/180 = 51%), i.e., the NN speakers' F1 and CoV values were both within +/− 2 standard deviations of the NE speakers' average F1 and CoV values. Among the remaining three scenarios, in which the F1 and/or CoV values differed by more than +/− two standard deviations from the NE speakers' formant or CoV values, Scenario 3 (28%) was two to three times more common than either of the other two scenarios, occurring more often as Scenario 2 (12%) and Scenario 4 (9%) combined. A Chi-square goodness of fit analysis again indicated that the observed distribution was significantly different from an equal occurrence of each of the four scenarios ($\chi^2 = 80.09$, $df = 3$, $p < .0001$). It is also notable that Scenarios 1 and 3 combined were associated with 79% of the NN learners' productions, indicating that whether their F1 values were native-like or not, the NN speakers most often (i.e., ~80% of the time) did not manifest greater intra-speaker variability than did the NE speakers.

Although not a focus of the study, it is interesting to note that the L2 speakers tended to be more native-like in their productions of lax versus tense vowels, as shown in Table 2. For tense vowels, they most commonly (43%) showed the Scenario 3 pattern of being more than two standard deviations different from the NE speakers' F1 values (but not more variable). For lax vowels, the NN speakers most often (68%) showed the Scenario 1 category (no formant or CoV differences), which was more than twice the frequency of occurrence of the other three scenarios combined. Scenario 1 for the lax vowels (68%) was also twice as frequent as Scenario 1 for the tense vowels (34%).

In an effort to better understand individual speakers' profiles with respect to the four scenarios, we have selected four individual nonnative speakers representing relatively prototypical examples of each scenario for further inspection, using the same similarity criterion detailed above. Speaker S9, a 41-year-old male whose LoR was 18 years and who began learning English at 18 years old, produced mean F1 values and had mean F1 CoV values meeting the similarity criterion for each of the six vowels (Scenario 1). Speaker M5, a 22-year-old male, had a LoR of 1 year and began learning English at age 17. He exhibited F1 CoV not meeting the similarity criterion for five of the six vowels (/i, ɪ, e, ɛ, u/), but differed from the NE speakers in only three of the vowels (/i, e, u/); while not an excellent exemplar of Scenario 2, this speaker represents the most prototypical exemplar in the data set.

Table 2. F1 similarities/differences and within-subject variability (CoV) findings across 30 ESL speakers and the scenarios they reflect

	Scenario 1 Similar formants Similar variability	Scenario 2 Similar formants More variable	Scenario 3 Different formants Similar variability	Scenario 4 Different formants More variable
/i/	5	5	15	5
/ɪ/	16	6	3	5
/e/	14	1	14	1
/ɛ/	24	6	0	0
/u/	12	3	10	5
/ʊ/	21	1	8	0
Total	**92 (51%)**	**22 (12%)**	**50 (28%)**	**16 (9%)**

On the other hand, S5, a 19-year-old whose AoA was 4 years old, did not meet the similarity criterion for F1 values for any of the six vowels, though she nonetheless exhibited similar F1 CoV values for all vowels (Scenario 3). M10, a 20-year-old female, with a LoR of 2 years who began learning English at age 11, did not meet the criterion for similar F1 for any of the vowels, and met the F1 CoV similarity criterion for only two vowels (/e,u/). In summary, among the 30 nonnative speakers in the present data set, prototypical speakers were in evidence for only Scenarios 1 and 3, while Scenarios 2 and 4 were best represented by speakers only partially meeting the four scenarios' prototypes. As demonstrated next, this finding regarding the characteristics of individual speakers mirrors that captured by the distribution of all nonnative speakers among the four scenarios.

Table 3 integrates and summarizes the various findings described thus far regarding the relationships between the formant frequency patterns and the intra-speaker variability (CoV) associated with the tense/lax vowel productions by the NE versus the NN speakers. It can be seen in the first row of Table 3, for example, that for the vowel /i/, there was a significant difference between the NE versus the NN speakers for F1 (i.e., column 1 = "yes"), but not a significant difference for F1 CoV of /i/ (column 2 = "no"). This pattern thus fits with what was described in the Introduction as Scenario 3 (as noted in column 3); that is, there was a significant difference between the NE and the NN groups for their /i/ F1 values, but not for their associated (F1 CoV) variability measures. This same basic relationship was also observed for the two other tense vowels (/e/ (row 3) and /u/ (row 5)), which thus also both fit with the Scenario 3 pattern. That is, the 30 nonnative speakers produced F1 values that differed significantly from those of the NE speakers for each of the three tense vowels. However, the intra-speaker F1 CoV variability of the NN speakers for the tense vowels did not differ significantly from the intra-speaker variability shown by the NE speakers.

Table 3. Summary of F1 and F2 similarities and differences for six vowels and intra-speaker variability (CoV) across the four groups of subjects and the scenarios they reflect

Vowel	F1: Statistical difference?	F1 – CoV: Statistical difference?	Scenario	F2: Statistical difference?	F2 – CoV: Statistical difference?	Scenario
/i/	yes	no	3	no	yes	2
/ ɪ /	no	no	1	no	yes	2
/e /	yes	no	3	no	no	1
/ɛ /	no	no	1	no	no	1
/u /	yes	no	3	no	no	1
/ʊ /	no	no	1	no	no	1

In contrast, it can also be seen in Table 3 that the F1 values for each of the three lax vowels were not significantly different when comparing the NE and the NN speakers. There was also not a CoV difference for F1 of any of the lax vowels for the NE versus NN speakers (columns 1 and 2 in rows 2, 4 and 6 all = "no"). This thus fits with the pattern described as Scenario 1 (native-like formant values and variability). When examining the findings for F2, it can be seen in the right half of Table 3 that while both /i/ and /ɪ/ did not show a NE versus NN difference for their formant values (column 4), there were significant differences in F2 CoV for both of these vowels. These two high, front vowels thus both match the pattern described by Scenario 2, i.e., no difference in the formant values, but significant CoV differences. The other four vowels (/e, ɛ, u, ʊ/) also did not show an F2 difference (column 4), nor did they show an F2 CoV difference (column 5) when comparing the NE versus the NN speakers; this again reflects a Scenario 1 pattern. Therefore, across both F1 and F2 for the six tense and lax vowels and their associated intra-speaker variability, Scenario 1 was the most common pattern observed (7 times = 58%), followed by Scenario 3 (3 times = 25%), and then Scenario 2 (2 times = 17%). There were no examples of the Scenario 4 pattern, in which both the formant values and CoV differ across groups. A Chi-square goodness of fit analysis indicated that the observed distribution was significantly different from an equal occurrence of each of the four scenarios ($\chi^2 = 8.667$, $df = 3$, $p < .05$).

4. Discussion

In the past, intra-speaker variability has often been overlooked or viewed as a relatively unimportant artifact associated with the process of speech production.

However, more recently, researchers have begun to recognize that it is an important factor to consider, as it can provide insights into an individual's speech production performance and abilities (e.g., Romeo, Hazan, & Pettinato, 2013). Warner and Tucker (2011, p. 1606) indicated, for instance, that variability is "one of the hallmarks of human speech." Similarly, Larsen-Freeman (2006, p. 593) noted that "There is a great deal of variation…in learners' performances and clear instability over time. Variation and fluctuation are important characteristics of dynamic systems…and should not be dismissed as measurement error." In addition, Bosch and Ramon-Casas (2011, p. 517) reinforced the importance of examining nonnative speaker variability, stating that, "A detailed analysis exploring the range of variability in the realization of specific vowels by groups of bilingual speakers should be undertaken, as possible differences between monolinguals and bilinguals … could be found."

The present study focused on two factors related to native and nonnative speaker variability, that is, the extent to which nonnative speakers' tense and lax vowel formants were similar to or different from those of native speakers of English and whether intra-speaker acoustic variability shown by nonnative speakers was comparable to or greater than that of the native speakers. Four scenarios were proposed as plausible relationships that could exist when considering the tense and lax vowels that nonnative learners attempt to produce and the extent of intra-speaker variability they might show in conjunction with those productions. Specifically, we suggested that: (1) L2 speakers might be similar to L1 speakers in terms of both their vowel formants and the variability of those productions; (2) L2 speakers might produce vowel formants similar to those of L1 speakers, but be more variable in their productions; (3) L2 speakers might produce different vowel formants than L1 speakers, but be no more variable in their productions; (4) L2 speakers might produce different formants than L1 speakers and also be more variable in their productions.

Although examples of each of these four proposed scenarios were observed in several different analyses of the tense and lax vowel productions, the frequency of occurrence of the different scenarios varied substantially. As summarized in Table 2, for example, Scenario 1 (native-like vowel formants and variability) occurred most commonly, approximately 50% of the time for the 30 nonnative speakers. However, when their vowel productions did not fit the Scenario 1 pattern, they most commonly fit the Scenario 3 pattern of not producing native-like formants, but still having native-like intra-speaker variability. By comparison, Scenario 2 (native-like formants but greater intra-speaker variability) and Scenario 4 (not native-like vowel formants and also greater intra-speaker variability) each occurred less than half as frequently as Scenario 3. In general, the individual nonnative speakers were seldom more variable than the NE speakers (only about

20% of the time), whether they had native-like (Scenario 2) or nonnative-like (Scenario 4) vowel formants. The more frequent occurrence of the Scenario 3 pattern relative to Scenario 2 in particular makes it quite clear that not producing native-like vowel formants is not necessarily associated with greater intra-speaker variability by individual L2 speakers. Therefore, with regard to the issue mentioned in the Introduction as to whether L2 speakers who are more native-like in their vowel productions are also less variable than L2 speakers who are less native-like in their speech, the answer appears to be that this is not routinely the case. Having found for at least these tense and lax vowel pairs that a majority of the L2 speakers manifested intra-speaker variability comparable to that of NE speakers seems to support the idea that when nonnative speakers do not produce native-like vowels, they may nonetheless have relatively well-established "phoneme targets," even if those targets are not native-like (Cebrian, 2006; Flege, MacKay, & Meador, 1999; Jia, Strange, Wu, Collado, & Guan, 2006). Indeed, the relatively "stable" L2 productions observed here may reflect fossilization of these targets, especially among the groups of NN speakers who participated in the present study, whose mean length of formal English study ranged from 6.2 to 7.3 years and whose mean LoR ranged from 2.3 to 12.8 years. However, whether production targets are acoustic, articulatory, auditory, and/or gestural remains a topic of interest and uncertainty (Goldinger & Azuma, 2003; A. Smith, 2006).

Although not included as a potential option among the four scenarios that were originally proposed, another hypothesis could be that L2 speakers might be less variable than L1 speakers (whether their vowel productions are similar to those of L1 speakers or not). For instance, in a study of word duration and lexical properties of speech, Baker et al. (2011) reported that the Chinese and Korean L2 English learners they studied had less relative durational variance than native English speakers. They indicated that this occurred at least in part because their L2 speakers tended to reduce function words to a lesser extent than native speakers did. Although the findings of the current investigation may seem to differ from those of Baker and colleagues, it is difficult to make direct comparisons between the two studies since Baker and colleagues did not measure absolute word duration, nor did they examine individual segments from the passages their subjects read. In the present study, out of 180 sets of observations (30 NE speakers x six vowels), there were only two instances where NE speakers exhibited F1 CoV values that were less than two standard deviations below the NE speakers' mean (speakers K9 and M2, both for the vowel /i/). Therefore, the extent to which L2 speakers might sometimes be less variable than L1 speakers may relate to the nature of the speech task involved and/or the specific types of measurements employed; this is an issue that deserves additional attention.

For the individual speakers who did exhibit higher levels of intra-speaker variability, it is worthwhile asking whether there are real-world consequences of this greater variability. As yet, in keeping with the general lack of studies of L2 intra-speaker variability, very little is known about its effects on communication. A notable exception is a study by Witteman, Weber and McQueen (2014), who found that native Dutch listeners adapted more quickly to the accent of a native German L2 speaker of Dutch when the speaker produced the Dutch vowel /œy/ "consistently" than when the same speaker produced the vowel variably. However, Witteman et al. (2014) note that although the inconsistent speech delayed adaptation to the speaker's accent, listeners were able to adapt, leading them to conclude that the inconsistencies in the more variable speech "seem to create no major problems" (p. 519). Thus, based on the very little empirical evidence available so far, it appears that while increased variability may have measurable consequences for communication, these consequences may not rise to the level of significantly impeding communication.

Now we turn to discussion of the nature and sources of variability that occur in speech production. Many different biomechanical, neurophysiological, linguistic, and other factors that may contribute to both native and nonnative speaker variability have been suggested in the literature. Although the underlying sources related to inter- and intra-speaker variability may be similar and overlap to some extent and cannot always be unambiguously differentiated from one another, factors such as extent of reliance on neuromotor feedback (Tingley & Allen, 1975), motor system instability (Seddoh et al., 1996), degree of "economy" in energy expenditure, and degree of flexibility associated with utilizing different motor patterns (Seddoh et al., 1996; Sharkey & Folkins, 1985) may relate more to intra-speaker variability, i.e., how consistent a given L2 speaker is in achieving a particular production pattern relative to the "stability" shown by native speakers. By comparison, factors such as age of arrival in the L2 country, age that formal instruction in the L2 began, daily amount of L1 versus L2 exposure and use, and others (e.g., Cebrian, 2006; Moyer, 1999; Piske, MacKay, & Flege, 2001; Slevc & Miyake, 2006), as well as factors such as speech style and rate (e.g., Baese-Berk & Morrill, 2015; Ferguson, 2004; Warner & Tucker, 2011), and individual learning strategies (e.g., Koenig, Lucero, & Perlman, 2008) may exert a greater influence on inter-speaker variability, i.e., how similar the performance of a group of L2 speakers is compared to a group of L1 English speakers.

In summary, the present study examined tense and lax vowel productions and the associated intra-speaker variability of thirty proficient nonnative speakers from three language backgrounds. Examples of each of the four possible scenarios that were initially proposed were observed in acoustic analyses of their vowel productions, but not to the same extent. Overall these nonnative speakers most com-

monly showed native-like vowel formants and intra-speaker variability in their productions (Scenario 1), followed by productions that were not native-like in terms of their vowel formants (primarily F1), but that showed comparable intra-speaker variability as the native English speakers (Scenario 3). Approximately 80% of the individual nonnative speakers' productions showed the same degree of intra-speaker variability as the native English speakers, whether or not they were associated with comparable (i.e., native-like) vowel formant values. Furthermore, Scenario 3 was found to occur more often than Scenarios 2 and 4 combined, suggesting that L2 speakers whose productions are not native-like do not necessarily manifest greater intra-speaker variability, in possible contrast to a finding reported by Kartushina & Frauenfelder (2014) of a strong correlation between "compactness" and vowel production accuracy for three of four L2 vowels studied (though the two studies' findings may in fact not be comparable due to differences in the relationship between the L1-L2 vowel inventories between the studies). In conclusion, inter- and intra-speaker variability are both important aspects of speech production to consider when examining L2 speech. Intra-speaker variability, in particular, represents a distinctive aspect of L2 speech production that frequently does not appear be greater for the nonnative speakers in the present study, even in instances when they do not produce native-like vowels.

Acknowledgements

We would like to thank Hannah Allen, Micah Foster, and Catherine Showalter for their help in recording subjects and performing acoustic measurements.

References

Adank, P., Smits, R., & van Hout, R. (2004). A comparison of vowel normalization procedures for language variation research. *Journal of the Acoustical Society of America*, 116, 3099–3107. https://doi.org/10.1121/1.1795335

Baese-Berk, M., & Morrill, T. H. (2015). Speaking rate consistency in native and non-native speakers of English. *Journal of the Acoustical Society of America*, 138, 223–228. https://doi.org/10.1121/1.4929622

Baker, R. E., Baese-Berk, M., Bonnasse-Gahot, L., Kim, M., Van Engen, K. J., & Bradlow, A. R. (2011). Word durations in non-native English. *Journal of Phonetics*, 39, 1–17. https://doi.org/10.1016/j.wocn.2010.10.006

Best, C. T. (1995). A direct realist perspective on cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 167–200). Timonium, MD: York Press.

Boersma, P., & Weenink, D. (2012). Praat: doing phonetics by computer [Computer program, version 5.3.23], retrieved from http://www.praat.org/

Bohn, O.-S., & Flege, J. E. (1992). The production of new and similar vowels by adult German Learners of English. *Studies in Second Language Acquisition*, 14, 131–158. https://doi.org/10.1017/S0272263100010792

Bond, Z. S., & Moore, T. J. (1994). A note on the acoustic-phonetic characteristics of inadvertently clear speech. *Speech Communication*, 14, 325–337. https://doi.org/10.1016/0167-6393(94)90026-4

Bongaerts, T. (1999). Ultimate attainment in L2 pronunciation: The case of very advanced late L2 learners. In D. Birdsong (Ed.), *Second language acquisition and the critical period hypothesis* (pp. 133–158). Mahwah, NJ: Lawrence Erlbaum Associates.

Bosch, L., & Ramon-Casas, M. (2011). Variability in vowel production by bilingual speakers: Can input properties hinder the early stabilization of contrastive categories? *Journal of Phonetics*, 39, 514–526. https://doi.org/10.1016/j.wocn.2011.02.001

Bradlow, A. R. (1995). A comparative acoustic study of English and Spanish vowels. *Journal of the Acoustical Society of America*, 97, 1916–1924. https://doi.org/10.1121/1.412064

Bradlow, A. R., & Bent, T. (2002). The clear speech effect for non-native listeners. *Journal of the Acoustical Society of America*, 112, 272–284. https://doi.org/10.1121/1.1487837

Cebrian, J. (2006). Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics*, 34, 372–387. https://doi.org/10.1016/j.wocn.2005.08.003

Chang, D., & Weng, C. (2013). Late ESL learners' difficulties of distinction between lax and tense vowels. In J. Levis and K. LeVelle (Eds.). *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference* (pp. 129–139). Ames, IA: Iowa State University.

Cook, V. (2015). Where is the native speaker now? *TESOL Quarterly*, 50, 186–189. https://doi.org/10.1002/tesq.286

Ferguson, S. H. (2004). Talker differences in clear and conversational speech: Vowel intelligibility for normal-hearing listeners. *Journal of the Acoustical Society of America*, 116, 2365–2373. https://doi.org/10.1121/1.1788730

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* (pp. 233–277). Timonium, MD: York Press.

Flege, J. E., & Hillenbrand, J. (1984). Limits on accuracy in foreign language speech production. *Journal of the Acoustical Society of America*, 76, 708–721. https://doi.org/10.1121/1.391257

Flege, J. E., MacKay, I. R. A., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *Journal of the Acoustical Society of America*, 106, 2973–2987. https://doi.org/10.1121/1.428116

Flege, J. E., Munro, M. J., & MacKay, I. R. A. (1995). Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America*, 97, 3125–3134. https://doi.org/10.1121/1.413041

Flege, J. E., Schirru, C., & MacKay, I. R. A. (2003). Interaction between the native and second language phonetic subsystems. *Speech Communication*, 40, 467–491. https://doi.org/10.1016/S0167-6393(02)00128-0

Goldinger, S. D., & Azuma, T. (2003). Puzzle-solving science: The quixotic quest for units in speech perception. *Journal of Phonetics*, 31, 305–320. https://doi.org/10.1016/S0095-4470(03)00030-5

Hattori, K., & Iverson, P. (2009). English /r/ – /l/ category assimilation by Japanese adults: Individual differences and the link to identification accuracy. *Journal of the Acoustical Society of America*, 125, 469–479. https://doi.org/10.1121/1.3021295

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*, 97, 3099–3111. https://doi.org/10.1121/1.411872

Jia, G., Strange, W., Wu, Y., Collado, J., & Guan, Q. (2006). Perception and production of English vowels by Mandarin speakers: age-related differences vary with amount of L2 exposure. *Journal of the Acoustical Society of America*, 119, 1118–1130. https://doi.org/10.1121/1.2151806

Kartushina, N., & Frauenfelder, U. H. (2014). On the effects of L2 perception and of individual differences in L1 production on L2 pronunciation. *Frontiers in Psychology*, 5, 1246. https://doi.org/10.3389/fpsyg.2014.01246

Koenig, L. L., Lucero, J. C., & Perlman, E. (2008). Speech production variability in fricatives of children and adults: Results of functional data analysis. *Journal of the Acoustical Society of America*, 124, 3158–3170. https://doi.org/10.1121/1.2981639

Kondaurova, M. V., & Francis, A. L. (2008). The relationship between native allophonic experience with vowel duration and perception of the English tense/lax vowel contrast by Spanish and Russian listeners. *Journal of the Acoustical Society of America*, 124, 3959–3971. https://doi.org/10.1121/1.2999341

Larsen-Freeman, D. (2006). The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics*, 27, 590–619. https://doi.org/10.1093/applin/aml029

Lotto, A., Sato, M., & Diehl, R. L. (2004). Mapping the task for the second language learner: The case of the Japanese acquisition of /r/ and /l/. In J. Slifka, S. Manuel, and M. Matthies (Eds.), *From sound to sense: 50+ years of discoveries in speech communication* (pp. 181–186). Cambridge MA: MIT Press.

Moyer, A. (1999). Ultimate attainment in L2 phonology. *Studies in Second Language Acquisition*, 21, 81–108. https://doi.org/10.1017/S0272263199001035

Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, 29, 191–215. https://doi.org/10.1006/jpho.2001.0134

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, 24, 175–184. https://doi.org/10.1121/1.1906875

Rallo Fabra, L., & Romero, J. (2012). Native Catalan learners' perception and production of English vowels. *Journal of Phonetics*, 40, 491–508. https://doi.org/10.1016/j.wocn.2012.01.001

Romeo, R., Hazan, V., & Pettinato, M. (2013). Developmental and gender-related trends of intra-talker variability in consonant production. *Journal of the Acoustical Society of America*, 134, 3781–3792. https://doi.org/10.1121/1.4824160

Schertz, J., Cho, T., Lotto, A., & Warner, N. (2015). Individual differences in phonetic cue use in production and perception of a nonnative sound contrast. *Journal of Phonetics*, 52, 183–204. https://doi.org/10.1016/j.wocn.2015.07.003

Seddoh, S. A. K., Robin, D. A., Hageman, C., Sim, H.-S., Moon, J. B., & Folkins, J. W. (1996). Temporal control in apraxia of speech: an acoustic investigation of token-to-token variability. *Clinical Aphasiology*, 24, 65–81.

Sharkey, S. G., & Folkins, J. W. (1985). Variability of lip and jaw movements in children and adults: Implications for the development of speech motor control. *Journal of Speech and Hearing Research*, 28, 8–15. https://doi.org/10.1044/jshr.2801.08

Slevc, L. R., & Miyake, A. (2006). Individual differences in second-language proficiency: Does musical ability matter? *Psychological Science*, 17, 675–681. https://doi.org/10.1111/j.1467-9280.2006.01765.x

Smith, A. (2006). Speech motor development: Integrating muscles, movements, and linguistic units. *Journal of Communication Disorders*, 39, 331–349. https://doi.org/10.1016/j.jcomdis.2006.06.017

Smith, B. L. (2000). Variations in temporal patterns of speech production among speakers of English. *Journal of the Acoustical Society of America*, 108, 2438–2442. https://doi.org/10.1121/1.1290514

Smith, B. L., Hayes-Harb, R., Bruss, M., & Harker, A. (2009). Production and perception of voicing and devoicing in similar German and English word pairs by native speakers of German. *Journal of Phonetics*, 37, 257–275. https://doi.org/10.1016/j.wocn.2009.03.001

Strange, W., Weber, A., Levy, E. S., Shafiro, V., & Hisagi, M. (2007). Acoustic variability within and across German, French, and American English vowels: Phonetic context effects. *Journal of the Acoustical Society of America*, 122, 3728–3737. https://doi.org/10.1121/1.2749716

Syrdal, A. K. and Gopal, H. S. (1986). A perceptual model of vowel recognition based on the auditory representation of American English vowels. *Journal of the Acoustical Society of America*, 79, 1086–1100. https://doi.org/10.1121/1.393381

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

Tingley, B. M., & Allen, G. D. (1975). Development of speech timing control in children. *Child Development*, 46, 186–194. https://doi.org/10.2307/1128847

Verspoor, M., Lowie, W. & Van Dijk, M. (2008). Variability in second language development from a dynamic systems perspective. *The Modern Language Journal*, 92, 214–231. https://doi.org/10.1111/j.1540-4781.2008.00715.x

Wade, T., Jongman, A., & Sereno, J. (2007). Effects of acoustic variability in the perceptual learning of nonnative-accented speech sounds. *Phonetica*, 64, 122–144. https://doi.org/10.1159/000107913

Warner, N., & Tucker, B. V. (2011). Phonetic variability of stops and flaps in spontaneous and careful speech. *Journal of the Acoustical Society of America*, 130, 1606–1617. https://doi.org/10.1121/1.3621306

Witteman, M. J., Weber, A., & McQueen, J. M. (2014). Tolerance for inconsistency in foreign-accented speech. *Psychonomic Bulletin & Review*, 21(2), 512–519. https://doi.org/10.3758/s13423-013-0519-8

Address for correspondence

Bruce L. Smith
752 E. Verona Meadows Ct.
Murray, UT 84107
USA
blsmith1948@gmail.com

Co-author information

Eric Johnson
johnson.7259@osu.edu

Rachel Hayes-Harb
The University of Utah
r.hayes-harb@utah.edu