The Relationship between Second-Language Learners' Production and Perception of English Vowels: The Role of Native-Like Acoustic Correlates

Jae Yung Song and Fred Eckman

1. Introduction

The relationship between second-language (L2) learners' production and perception of target-language (TL) contrasts is one of the important questions in the investigation into the acquisition of L2 phonology. Whereas a few studies have found no connection in this area, others have claimed that there is a positive correlation between L2 learners' production and perception of TL contrasts. The conventional wisdom that has distilled from the inconclusive evidence is best expressed by Flege (1999), who argues that there is a correlation between L2 learners' production and perception, albeit a modest one. However, as pointed out by several researchers, the exact nature of the production-perception relation largely remains unclear (Flege, Bohn, & Jang, 1997; Levy & Law, 2010). The present paper is intended as a contribution to the current literature by reporting the results of a study showing that the way in which L2 learners implement the contrast is related to how they perceive the contrast. More specifically, our results indicate that participants who produce the contrast between the English vowels /i/ and /ɪ/ (as in *beat* vs. *bit*), on the one hand, and /ɛ/ and /æ/ (as in *bet* vs. *bat*) on the other, using the same acoustic correlates as do native speakers, also perceive these vowel contrasts better than those learners who distinguish the vowels using acoustic correlates other than those employed by the native speakers.

At the center of the discussion surrounding the connection between an L2 learner's ability to produce and perceive TL contrasts are several important phonetic models. These include the Speech Learning Model (SLM) (Flege, 1995), the Perception Assimilation Model (PAM) (Best, 1995) and PAM-L2 (Tyler, Best, Faber, & Levitt, 2014). The central claim of the SLM is that L2 learners' ability to acquire TL phonemes is a function of the learners' ability to perceive these phonemes. The model is predicated on the notion of "equivalence classification" in which an L2 learner sets up categories for TL phonemes on the basis of that learner's perception of the allophones of those phonemes. Thus, according to the SLM, many L2 production errors have a perceptual basis. A

* Jae Yung Song and Fred Eckman, University of Wisconsin-Milwaukee, songjy@uwm.edu, eckman@uwm.edu. We thank our research assistants for their help with collecting and coding the data.

 Proceedings of the 45th annual Boston University Conference on Language Development, ed. Danielle Dionne and Lee-Ann Vidal Covas, 693-706. Somerville, MA: Cascadilla Press.

fundamental assumption behind the claim is that the perception and production modalities are interconnected.

However, in practice, studies have come down on either side of the issue as to whether the L2 production of TL contrasts is correlated with the perception of those contrasts. Some studies have reported a positive connection between L2 production and perception. For example, Cheng and Zheng (2015) showed that Chinese speakers' production (or "accentedness" as rated by two native speakers of English) and perception of English consonants were significantly correlated in the syllable-initial position, but not syllable finally. Flege, MacKay and Meador (1999) showed that accuracy in L2 vowel production (intelligibility scores) was related to how accurately the vowels were discriminated by the L2 learners. Further evidence of the production-perception relationship comes from Flege (2003) who discussed the results of a study that showed a moderately positive correlation between Italian students' production (as measured by the goodness ratings) and discrimination scores of English vowel contrasts.

Other studies on this same topic found either weak or no correlation between L2 production and perception. For example, Kartushina and Frauenfelder (2014) found no correlation between individuals' production and perception, although the French vowel contrasts that the Spanish speakers perceived poorly (in terms of the identification of vowels) were generally also those that they produced poorly (as measured by the distance between the L1 and L2 vowels). As another example, Hattori (2009) found that Japanese speakers' perception and production of the English /r/-/l/ contrast were overall unrelated. Of the various measures of perception accuracies (identification, discrimination sensitivity, and perceptually best exemplars) and production accuracies (acoustic measurements, and intelligibility by native-English speakers) of /r/-/l/, the only connection that turned out to be significant was the correlation between participants' identification scores and their production intelligibility (However, see also Bradlow, Akahane-Yamada, Pisoni, and Tohkura (1999) who showed a positive link between perception and production of English /r/-/l/ in Japanese speakers). Furthermore, Levy and Law (2010) showed no link between perception and production accuracy (percent correct as judged by native speakers) for American English learners of French. Similarly, Peperkamp and Bouchon (2011) found that French-English bilinguals' perception did not correlate with production (as measured by a nativelikeness rating by native speakers). Moreover, Bohn and Flege (1997) showed that experienced German learners of English who were able to produce the English /ɛ/-/æ/ contrast in a native-like manner were not able to discriminate the contrast with the precision of native speakers. Therefore, the learners' production skills exceeded their perception skills. And finally, Strange (1995) gives a thorough historical review of several decades of phonetic research and found that segmental production and perception, even for experienced learners, were not correlated.

More recent studies have continued the investigation into the relationship between L2 perception and production, though the number of languages

involved has been expanded, and some of the focus seems to have shifted. Thus, some recent studies have investigated the relationship between perception and production from the point of view of training studies. For example, Carlet and Kivisto de Souza (2018) found that their instructional treatment improved the learners' perception, but not their production. Sakai and Moorman (2018), on the other hand, found that perception training "can induce positive change in the productive mode as well" (p.187). Similarly, Bradlow et al. (1999) showed that successful perceptual learning led to improvement in production.

The only clear picture that emerges from our review of the literature is that the relationship between L2 production and perception still remains largely uncertain due to inconsistent results across numerous studies. As several authors have pointed out (Levy & Law, 2010), the inconclusive findings may be partly attributed to differences in methodologies and techniques used for assessing perception and production in different studies. In particular, we noted that a common way of measuring L2 learners' production is to have native-speakers of the TL listen to the pronunciations and determine how target-like they are, yielding some global measures such as ratings (e.g., accentedness, nativelikeness, goodness) or scores (e.g., percent correctly perceived). Only a few studies (e.g., Flege et al. 1997) have employed more fine-grained assessments of production, such as acoustic correlates of phonological contrasts. The present study is intended to fill this critical gap by analyzing L2 participants' implementation of the acoustic correlates of specific sound contrasts. Examining L2 production at the level of acoustics could provide more sensitive measures of production than solely relying on the global production scores provided by the native transcribers. These, in turn, could give us more insight into the production-perception relationship.

Given this context, the primary goal of the present study is to extend our understanding of the relationship between the production and perception of L2 contrasts by showing that L2 learners who implement the production of these distinctions using native-like acoustic correlates also score better on the perception of these contrasts. In particular, we examined the acquisition of the two English vowel contrasts, /i/-/ɪ/ and /ɛ/-/æ/, by native speakers of Korean, Portuguese, and Spanish. We chose to examine native speakers of these three languages because none of them has the vowel contrasts in question. That is, all three languages have /i/, but not /ɪ/. Although /ɛ/ is present in Korean and Portuguese (Spanish has /e/ instead), /æ/ is absent in all three languages.

In American English, /i/ is longer than /ɪ/, and /æ/ is longer than /ɛ/ (Flege et al., 1997). These vowels also differ in their spectral properties. Tense /i/, which occupies a more high-front peripheral position in the vowel space than lax /ɪ/, tends to have a lower F1 and a higher F2 compared to /ɪ/. The spectral properties of English /ɛ/ and /æ/ are known to be highly variable among individual speakers, resulting in a considerable degree of overlap in F1-F2 space (Hillenbrand et al., 1995). Because there is much overlap in formants, duration is the primary cue in the /ɛ/-/æ/ distinction for native speakers of English (Hillenbrand, Clark, & Houde, 2000). On the other hand, duration plays only a

small role in the recognition of /i/-/ɪ/, because the vowels are already sufficiently well separated on the basis of spectral properties (Hillenbrand et al., 2000).

In the present study, we conducted two analyses in order to examine the relationship between L2 learners' production and perception. In the first analysis, we used a more traditional, global measure of L2 production, which is the percentage of correct target-like productions as judged by native-speaker transcribers. This analysis was compared with a second one, where we employed a novel, more fine-grained acoustic measure of L2 production.

2. Method

2.1. Participants

We collected data from 62 participants, 29 of whom were native-speakers of American English, serving as controls, and 33 of whom were non-native speakers of English from three native-language backgrounds: Korean (n: 22, mean age: 24 years), Portuguese (n: 13, mean age: 22 years), and Spanish (n: 5, mean age: 31 years). All of our participants were either university students or staff members recruited through campus advertisements at University of Wisconsin-Milwaukee. All but three of the native controls reported that they had grown up in Wisconsin. Of the three, two grew up elsewhere in the Midwest (Illinois, Indiana) whereas only one grew up outside of the Midwest (Arizona). Therefore, the majority of our native controls used the Midwest dialect of American English. Like the native controls, all of our L2 participants were either university students or staff members recruited on campus, implying that the primary language they used for their everyday study or work was English. All of our L2 participants had more or less homogeneous experience with the TL; all of them started learning English as a school subject in their home country during their teens, then came to the U.S. around twenty years of age (thus having similar ages of arrival), and had their length of residence (LOR) between 1-12 months.

2.2. Stimuli

Participants performed both production and perception tasks. For the production task, there were 12 minimal or near-minimal word pairs for each of the two vowel contrasts: /i/-/ɪ/ and /ɛ/-/æ/. The perception task, unlike the production task, required all target words to be minimal pairs. Thus, for perception, we used some minimal pairs from the production task plus some additional minimal pairs for a total of 10 word pairs for each of the two vowel contrasts. All target words were monosyllabic CVC words, ending in a voiceless coda (e.g., *seek*). We used only high-frequency words from the Washington University's English Lexicon Project database (http://elexicon.wustl.edu) to ensure that all participants knew the meaning of the target words. In addition, we presented a list of the words to the non-native speakers before their participation. If there was any word that was unfamiliar to them, the

experimenter, who was a native speaker of English, explained the meaning of the word to the satisfaction of the participant.

The target words in both the production and perception tasks were presented twice in a random order. Ideally, the production task would have yielded 5952 words to analyze (4 vowels × 12 target words × 2 repetitions × 62 participants). However, the participants inadvertently skipped 23 tokens thereby failing to produce them. In addition, 169 tokens were omitted from the analysis for one of two reasons. Forty-four tokens were excluded for acoustic reasons (clipped signal, formants not clear, etc.), and 125 words were excluded because they were mispronounced, and the intended target was not clear. Therefore, the number of words analyzed for the production task was 7116, and the total number of words analyzed for the perception task was 4960 (4 vowels × 10 target words × 2 repetitions × 62 participants).

2.3. Procedures

For the production task, a target word containing one of the vowel phonemes in question was presented on a computer screen one at a time, and the participant was directed to read the word aloud. The productions were recorded at a sampling frequency of 44.1 kHz and 16-bit quantization via a Shure KSM137 unidirectional microphone. After the recording of the production data, the participants took part in the perception task by remaining seated in front of the computer screen and listening to a carrier sentence containing one of the target vowel phonemes, as in the example: *I wanted to say bet for him*. The participants indicated with a key-press which member of the minimal pair they heard (e.g., *bet*, *bat*). The two options were presented on the screen orthographically. All of the stimuli were produced by a male native speaker of American English who did not participate in the experiment.

2.4. Coding

The production data was phonetically transcribed and then analyzed acoustically. For phonetic transcription, we had phonetically-trained native-speaker transcribers listen to each word and transcribe the vowel. Then we calculated the percentage of target-like productions of a vowel for each participant. For example, if the transcriptions for a participant's productions of the vowel [ɛ] showed that 18 of the 24 tokens were transcribed as target-like, that participant's production score for [ɛ] was 75%. The acoustic analysis of vowel productions was carried out using three acoustic measures: the vowel duration (in milliseconds) and the first and second formant frequencies (F1 and F2 in Hz) measured from the mid-point of each vowel. In addition, for each participant, we calculated the percentage of correct perceptions for each target vowel. These scores ranged between 0% and 100%, with 0% indicating that none of the words containing the target vowel was perceived correctly by the participant, and 100% showing that all of them were.

3. Results

We conducted two analyses investigating the relationship between L2 learners' production and perception. In the first analysis, we used a more traditional, global measure of vowel production, which is the percentage of correct target-like productions as judged by native-speaker transcribers. In the second analysis, we employed a new, more fine-grained acoustic measure to gauge L2 learners' production. More specifically, the participants' vowel productions were described using the type of acoustic correlates (duration, F1, or F2) that they implemented to produce the target vowel contrasts. Therefore, the first analysis was done at the level of perceived phonemes, whereas the second analysis was performed in terms of individual acoustic correlates.

We begin by presenting the results of the first analysis. Table 1 shows the production (percentage of correct target-like productions) and perception scores for each of the four vowels for native-speaker controls and L2 learners. As can be seen in the table, all native-speaker controls had production scores between 99% and 100%, and perception scores between 94% and 100%. The L2 learners' production and perception scores for /i/ and /ɛ/, which are both acoustically close to the front vowels in their native language, were higher than those for /ɪ/ and /æ/, which do not occur in their native language. The production scores for /i/ and /ɛ/ were actually greater than the perception scores.

Table 1. Percent correct production and perception for the four target vowels. Percentages are rounded to whole numbers. The numbers in the parentheses are standard deviations.

	/i/ production	/i/ perception	/ɪ/ production	/ɪ/ perception
English	100 (0)	99 (2)	100 (0)	100 (1)
L2 participants	71 (29)	58 (19)	51 (31)	57 (16)
Korean	66 (31)	57 (15)	53 (30)	60 (19)
Portuguese	76 (24)	58 (21)	49 (28)	49 (10)
Spanish	74 (37)	63 (27)	51 (46)	67 (15)
	/ɛ/ production	/ɛ/ perception	/æ/ production	/æ/ perception
English	100 (1)	94 (7)	99 (3)	95 (9)
L2 participants	84 (21)	57 (15)	45 (30)	54 (22)
Korean	82 (23)	59 (14)	39 (27)	59 (14)
Portuguese	82 (20)	53 (17)	36 (24)	52 (17)
Spanish	98 (2)	61 (16)	88 (13)	74 (32)

Because our data were not normally distributed, with many participants' production scores being at 100% (especially for /i/ and /ɛ/), we ran Kendall's tau-b correlations to determine any relationship between productions and perception. The results showed that there was no significant correlation between

L2 participants' production and perception for all four vowels (see Figure 1). The results suggested that L2 participants who were transcribed as producing target-like vowels did not necessarily perceive the vowels more accurately.

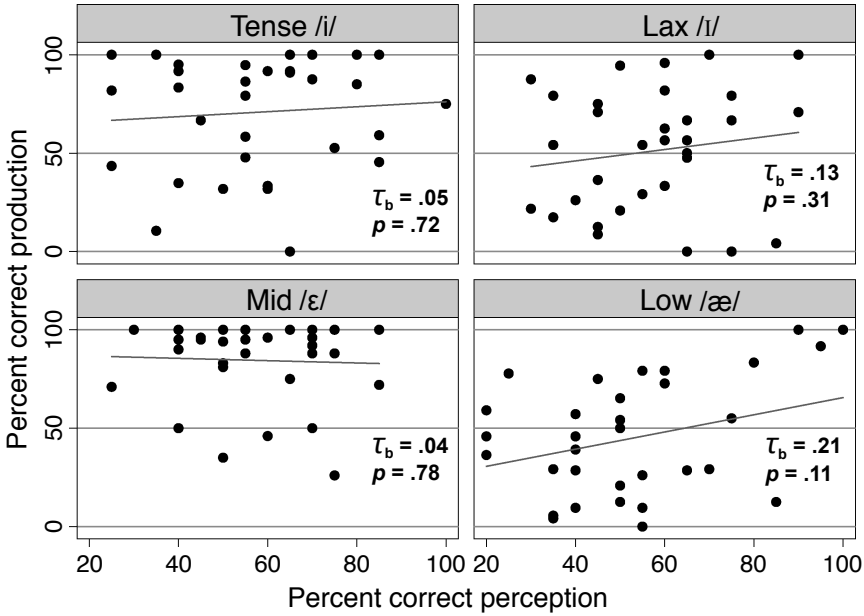

Figure 1. Correlation between production and perception scores for L2 participants. τ_b indicates Kendall's tau-b coefficient.

In the second analysis, we examined whether the L2 participants who produced distinctions between the vowels using native-like acoustic correlates also have better perception scores. We define native-like and non-native-like acoustic correlates based on which of the three acoustic measures in question (duration, F1, and F2) were used by native speaker controls. To examine which of the three acoustic correlates each participant employed to distinguish between /i/ and /ɪ/ and between /ɛ/ and /æ/, we performed separate linear mixed-effects regression analyses for the individual participants. Because the goal of this analysis was to identify the individuals who use the same types of acoustic correlates, the analysis was necessarily conducted on individual participants, rather than on the basis of a group. Then once the individuals were divided into three groups (native-like, non-native-like, and none of the acoustic correlates) on the basis of the use of acoustic distinctions, we performed a subsequent analysis to determine whether there were any differences between the groups.

For the analysis on individuals, each mixed-effects regression model included one fixed-effect factor, the vowel (/i/ vs. /ɪ/ or /ɛ/ vs. /æ/), and one random-effect factor, the word. The dependent measure was one of the three

acoustic correlates: duration, F1, and F2. Analyses were carried out using the R statistical computing software, and in particular the *nlme* package.

The most common pattern that native controls exhibited for the /i/ vs. /ɪ/ contrast was to make distinctions in all three acoustic measures: duration, F1, and F2, with 26 out of 29 (90%) showing this pattern (see Table 2). Also, three control participants (10%) made distinctions in both F1 and F2, but not in duration. Therefore, all of the native speakers produced systematic distinctions between /i/ and /ɪ/ in both F1 and F2. These two patterns were identified as native-like correlates for the /i/ vs. /ɪ/ contrast, as all of the native controls showed one of these patterns. When native controls made significant acoustic distinctions between /i/ and /ɪ/, the differences were not only reliable, but also in the expected direction. That is, the duration was longer, F1 was smaller, and F2 was larger for /i/ than /ɪ/. Therefore, in order to be designated as producing native-like acoustic correlates, L2 participants were required to produce a significant difference using the same acoustic correlates as native controls, and also the produced difference had to be in the same direction as native controls.

Table 2. Number of participants producing native-like, non-native-like, and none of the acoustic correlates for the /i/ vs. /ɪ/ contrast

Grouping	*Acoustic correlates*	*English*	*Korean*	*Portuguese*	*Spanish*
Native-like	Duration, F1, F2	26	2	2	1
	F1, F2	3	0	1	1
Non-native-like	Duration, F1	0	1	1	0
	Duration, F2	0	0	0	1
	Duration	0	6	3	1
	F1	0	0	1	0
	F2	0	0	0	1
None		0	6	5	0
	Total	29	15	13	5

There were seven L2 participants (21.2%) who also produced native-like correlates (Korean: 2, Portuguese: 3, Spanish: 2, as shown in Table 2). Initially, there were eight L2 participants in this group. However, one of the participants who produced distinctions in both in F1 and F2 produced F1 in the opposite direction to the native controls (significantly larger F1 for /i/ than /ɪ/). Therefore, the participant was determined as producing a native-like distinction only in F2, which was not a native-like pattern. The other seven L2 participants not only produced native-like acoustic correlates, but they were also in the same direction as the native controls. The 15 other L2 participants (45.5%), in contrast, exhibited five types of acoustic patterns that were different from those of the native controls (see Table 2). In the first two patterns, the L2 participants made a difference between the two vowels both in duration and in either F1 or F2. In the other three patterns, the L2 participants made a distinction only in duration, or only in F1, or only in F2. Thus, unlike our native controls, none of the 15 L2 participants produced a distinction in both F1 and F2. Finally, there were 11 L2

participants (33.3%) who made no distinction in any of the three measures, and therefore these participants were designated as producing none of the correlates.

The analysis of the /ɛ/ and /æ/ contrast was more complex in that it was difficult to determine the acoustic patterns of the native controls. The three most common such patterns involved making a distinction in duration along with differences in one or both F1 and F2 (see Table 3). In addition, two of the native controls made a reliable difference only in duration, and in neither of the formants. Finally, one native speaker made a distinction only in F1 and F2. We defined these five patterns as native-like.

Table 3. Number of participants producing native-like, non-native-like, and none of the acoustic correlates for the /ɛ/ vs. /æ/ contrast

Grouping	*Acoustic correlates*	*English*	*Korean*	*Portuguese*	*Spanish*
Native-like	Duration, F1	11	0	0	1
	Duration, F1, F2	8	0	0	0
	Duration, F2	7	0	0	0
	Duration	2	0	1	0
	F1, F2	1	0	0	4
Non-native-like	F1	0	3	1	0
None		0	12	11	0
	Total	29	15	13	5

With respect to the direction of each acoustic correlate, duration was longer and F1 was larger for /æ/ than /ɛ/ for all native controls. However, F2 showed individual variation. Of the 16 native controls who produced significant differences in F2, 11 had a larger F2 for /æ/ than for /ɛ/, whereas the other 5 had a smaller F2 for /æ/ than for /ɛ/. Thus, consistent with the literature, there was considerable variability in F2 for the two vowels. Because the difference in F2 between /æ/ and /ɛ/ moved in both directions for native speakers, we considered both as native-like.

Using these patterns of the native controls, we designated six L2 participants (18%) as producing native-like correlates. Four L2 participants (12%) exhibited two types of acoustic patterns that were different from those of the native controls in that they produced a distinction only in F1. These participants were designated as producing non-native-like correlates. The 23 of 33 L2 participants (70%) who did not make distinctions in any of the three measures were designated as producing none of the correlates. As for the direction of the differences in duration and F1, all of the L2 participants showed the same direction as the native controls. On the other hand, all 4 of the L2 participants who produced a significant distinction in F2 had larger values for /ɛ/ and /æ/. This was different from the native controls who showed distinctions in both directions**.** After dividing our L2 participants into three groups (native-like, non-native-like, and none of the correlates), the next step was to determine

whether the three groups of L2 participants differed in their perception skills. To this end, we conducted two one-way between-group analyses of covariance (ANCOVA) tests, one for each vowel contrast.

We focus first on the /i/ vs. /ɪ/ contrast. The independent variable in the ANCOVA test was the participants' production skills with the three levels being designated as native-like, non-native-like, and none of the correlates, and the dependent variable being the participants' perception scores averaged across /i/ and /ɪ/. Also, LOR was included as a covariate to control for any potential effect of this variable. Results indicated that the three groups of L2 participants had significantly different perception scores, $F(2, 29) = 11.87$, $p < 0.001$, after controlling for the effect of LOR. Bonferroni-corrected pairwise comparisons revealed significant differences in perception scores between the native-like correlates group ($M = 74$, $SE = 4$) and the non-native-like correlates group ($M = 53$, $SE = 3$) ($p < 0.001$), the native-like correlates and none-of-the-correlates groups ($M = 53$, $SE = 3$) ($p < 0.01$), but not between the non-native-like correlates and none-of-the-correlates groups ($p = 1.00$).

As for the /ɛ/-/æ/ contrast, results showed that the three groups of L2 participants had significantly different perception scores, $F(2, 29) = 5.70$, $p < 0.01$. Bonferroni post hoc tests showed that the L2 participants who made distinctions between /ɛ/ and /æ/ using native-like acoustic correlates on average had significantly better perception scores of /ɛ/ and /æ/ ($M = 66$, $SE = 4$) compared to the L2 participants who made distinctions using non-native-like acoustic cues ($M = 51$, $SE = 4$) ($p < 0.05$). The participants who used native-like acoustic correlates also had significantly better perception scores compared to those who failed to make distinctions using any of the three acoustic correlates (none-of-the-correlates groups) ($M = 53$, $SE = 2$) ($p < 0.05$). However, there was no significant difference in perception scores between those who used non-native-like acoustic correlates and those who used none of the three acoustic distinctions ($p = 1.00$).

4. Discussion

The primary goal of this study was to broaden our understanding of the relationship between L2 learners' production and perception of TL vowel contrasts through two analyses. In the first analysis, we used a more traditional measure of L2 learners' production, namely, the percentage of correct target-like production as judged by native-speaker transcribers. This was compared with the second analysis, where we proposed a novel measure to gauge L2 learners' production, namely, implementation of native-like acoustic correlates in making the vowel contrasts. The first analysis found no correlation between the L2 participants' production and perception. That is, the L2 participants whose productions were transcribed as target-like did not necessarily perceive the vowels more accurately. On the other hand, the second analysis showed that those participants who implemented the vowel contrasts using the same acoustic correlates as native speakers also had higher scores on the perception task for

the vowels than did those participants using non-native-like acoustic correlates, and those who failed to make any distinctions using the three acoustic measures. At the same time, no significant difference was found in perception scores between those participants who used non-native-like acoustic correlates and those who implemented none of the acoustic measures. This suggests the possibility that the perception skills of the L2 participants making vowel distinctions using non-native-like acoustic correlates are not necessarily better than those of the L2 participants who do not make any distinctions in the three acoustic measures we employed.

The findings from the present study contribute to the L2 phonological acquisition literature in several ways. First of all, the second analysis provides strong supporting evidence for a link between L2 learners' production and perception skills. Due to the mixed results in prior research, the relationship between the two domains has remained largely unclear. Our study suggests that using a well motivated measure of L2 learners' production can help uncover the relationship, as our fine-grained analysis of L2 learners' production based on individual acoustic correlates of phonological contrasts suggests that the relationship is significant.

Our study was primarily motivated by the need for a better measure of L2 production in studying the production-perception relationship. Therefore, this study necessarily considers this connection from the point of view of the learners' production. The main findings also suggest that the L2 participants' perception scores are a function of their production. To the best of our knowledge, this study is one of the very few to consider the production-perception connection from the point of view of the learners' production. However, based on numerous studies (Flege, 2003), along with the well-known phonetic models including the SLM (Flege, 1995), which suggest that accurate perception needs to be in place before production can be mastered, we argue that the relationship is bi-directional. If this is indeed defensible, it in turn raises the possibility that the two modalities share common phonetic representations.

Another contribution of our study is that it provides a better characterization of native-like and non-native-like vowel production. Numerous studies on the pronunciation of English vowels have shown considerable variability in the acoustic measurements of vowels within and between native-speaker participants (Hillenbrand et al., 1995). This issue becomes even more important when we consider that vowel productions are highly affected by dialects. One major issue that has been consistently overlooked in the L2 literature is the fact that there are various ways of implementing vowel contrasts for native speakers. Much of the research has focused on differences between native-speakers and non-native speakers as a group. Since the native controls in the present study exhibited a few different types of acoustic patterns in producing the vowel contrasts, future studies should take this into account when characterizing native-like pronunciations.

We used three acoustic measures (duration, F1, F2) to define native-like and non-native-like productions of the target vowels. However, there could be other

acoustic properties that also vary reliably with the with the way native speakers produce these vowels. For example, Ladefoged and Johnson (2015) suggested the F2-F1 as an alternative measure of vowel backness instead of using F2 alone. If this is the case, and if we employed other acoustic measures, some of the L2 learners who were designated as producing non-native-correlates in this study could in fact be re-categorized as producing native-like acoustic distinctions. That being said, the three acoustic measures we chose have been widely used to characterize vowel production in English, and we believe that the data from our study could serve as a basis for future studies. In making such choices caution obviously needs to be exercised because the selection of acoustic measures could affect the characterization of native-like vowel production.

Given the present results, the question arises as to what factors led to L2 participants' implementation of native-like acoustic correlates. Within this context, several possible answers suggest themselves, ranging from the LOR to the learners' L1. Numerous studies have focused on the role of LOR on L2 production and perception (Flege et al. 1997). This research suggests that exposure to the TL could alter the mechanisms underlying speech production and perception, and that both age of onset of acquisition and the amount of experience a learner has with the TL could influence how native-like the learner's performance is. In light of this, it is possible that the correlations using a global measure of production in the present study were not significant due to the fact that the LOR range of our participants was relatively small and uniform (only between 1-12 months). If our study had included L2 participants with longer LOR, that could have introduced more variability into the production and perception scores, leading to significant correlations.

We have preliminary results from a follow-up study that could provide insight into this possibility. In our follow-up study, we performed the same two statistical analyses on 42 L2 participants, with 33 original participants with shorter LOR and 9 new participants with longer LOR. The LOR of the nine additional L2 participants (one Korean-speaking and eight Spanish-speaking) ranged from 5-15 years. On average, their productions were more often transcribed as target-like and they had higher perception scores. The first analysis showed that the correlations between production and perception of the four target vowels became significant when the data included the L2 participants with longer LOR. The second analysis revealed that the L2 participants who made distinctions between the target vowel phonemes using native-like acoustic correlates also had significantly better perception scores for the vowels compared to the group of participants who made distinctions using non-native-like acoustic correlates. Therefore, our fine-grained analysis of L2 learners' production was able to reveal the production-perception relationship more consistently, regardless of the LOR of the participants. In contrast, the correlations using a traditional, global measure of vowel production appeared to be meaningfully affected by participants' LOR.

Our decision to include three typologically diverse languages was motivated by an effort to show that the production-perception relationship was not limited to L2 learners of a specific language background. To this end, we had similar numbers of speakers of an Asian language (15 Korean-speaking participants) and from two European languages (13 Portuguese-speaking and 5 Spanish-speaking participants). Although we were not able to conduct any systematic analysis of the effect of L1, we observed some patterns that are in line other studies (Escudero & Boersma, 2004; Kim, Clayards, & Goad, 2018). For example, all of our Korean-speaking participants used temporal cues to distinguish /i/ and /ɪ/. In fact, Korean- and Spanish-speaking participants on average produced longer duration for /i/ compared to the native controls, although the average duration of their lax vowel was more or less similar to that of native speakers. Another observation from our data is that the Spanish-speaking participants on average produced larger spectral differences between English /ɛ/ and /æ/ compared to the native controls, which is consistent with Flege et al. (1997).

As the focus of our research was on attesting the potential of our novel measure of production in studying the L2 production-perception link, we did not have any specific hypotheses as to the possible role of participant characteristics, such as their L1 and their proficiency, in explaining their production and perception patterns, and this remains as a limitation of the present study. Therefore, in the future it would be of interest to expand this research to manipulate these factors more systematically to see if similar results would be still be found. The novel approach presented in this study is a step towards developing a more complete understanding of the L2 production and perception relationship, and it opens up new possibilities for developing more sophisticated measures of L2 production.

References

Best, Catherine. 1995. A direct realist view of cross-language speech perception. In Winifred Strange (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, 171-204. Timonium, MD: York Press.

Bohn, Ocke-Schwen & James Flege. 1997. Perception and production of a new vowel category by second-language learners. In Allen James & Jonathan Leather (eds.), *Second-language Speech: Structure and Process*, 53-74. Berline: Walter de Gruyter.

Bradlow, Ann, Reiko Akahane-Yamada, David Pisoni & Yohichi Tohkura. 1999. Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics 61*(5), 977-985.

Carlet, Angélica & Hanna Kivistö de Souza. 2018. Improving L2 Pronunciation Inside and Outside the Classroom: Perception, Production and Autonomous Learning of L2 Vowels. Ilha do Desterro *71*(3), 99-123.

Cheng, Bing, & Yang Zhang. 2015. Syllable structure universals and native language interference in second language perception and production: Positional asymmetry and perceptual links to accentedness, *Frontiers in Psychology 6*, 1801.

Escudero, Paola, & Paul Boersma. 2004. Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition*, 551-585.

Flege, James. 1995. Second language speech learning: Theory, findings and problems. In Winifred Strange (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, 233-277. Baltimore: York Press.

Flege, James. 1999. Age of learning and second-language speech. In David Birdsong (ed.), *New Perspectives on the Critical Period Hypothesis for Second-Language Learning*, 101-132. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Flege, James. 2003. Assessing constraints on second-language segmental production and perception. In Niels Olaf Schiller & Antje Meyer (eds.), *Phonetics and Phonology in Language Comprehension and Production, Differences and Similarities*, 319-355. Berlin: Mouton de Gruyter.

Flege, James, Ocke-Schwen Bohn & Sunyoung Jang. 1997. The effect of experience on nonnative subjects' production and perception of English vowels. *Journal of Phonetics 25*, 437-470.

Flege, James, Ian MacKay & Diane Meador. 1999. Native Italian speakers' perception and production of English vowels. *Journal of the Acoustical Society of America*, *106*(5), 2973-2987.

Hattori, Kota. 2009. Perception and production of English /r/-/l/ by adult Japanese speakers. Ph.D. thesis, University College London, London, UK.

Hillenbrand, James, Laura Getty, Michael Clark & Kimberlee Wheeler. 1995. Acoustic characteristics of American English vowels. *Journal of the Acoustical society of America 97*, 3099-3111.

Hillenbrand, James, Michael Clark & Robert Houde. 2000. Some effects of duration on vowel recognition. *Journal of the Acoustical Society of America 108*(6), 3013-3022.

Kartushina, Natalia, & Ulrich Frauenfelder. 2014. On the effects of L2 perception and of individual differences in L1 production on L2 pronunciation. *Frontiers in Psychology 5*, 1246.

Kim, Donghyun, Meghan Clayards & Heather Goad. 2018. A longitudinal study of individual differences in the acquisition of new vowel contrasts. *Journal of Phonetics 67*, 1-20.

Ladefoged, Peter & Keith Johnson. 2015. *A Course in Phonetics* (7th edition). Stamford, CT: Cengage Learning.

Levy, Erika & Franzo Law. 2010. Production of French vowels by American-English learners of French: Language experience, consonantal context, and the perception-production relationship. *Journal of the Acoustical Society of America 128*(3), 1290-1305.

Peperkamp, Sharon & Camillia Bouchon. 2011. The relation between perception and production in L2 phonological processing. In *Proceedings of the International Speech Communication Association*, Florence, Italy, 161-164.

Sakai, Mari & Colleen Moorman. 2018. Can perception training improve the production of second language phonemes? A meta-analytic review of 25 years of perception training research. *Applied Psycholinguistics 39*(1), 187-224.

Strange, Winifred. 1995. Phonetics of second-language acquisition: Past, present, future. In *Proceedings of the ICPhS conference*, Stockholm, Sweden, 77-83.

Tyler, Michael, Catherine Best, Alice Faber & Andrea Levitt. 2014. Perceptual assimilation and discrimination of non-native vowel contrasts. *Phonetica 71*(1), 4-21.

Proceedings of the 45th annual Boston University Conference on Language Development

edited by Danielle Dionne
and Lee-Ann Vidal Covas

Cascadilla Press Somerville, MA 2021

ISSN 1080-692X
ISBN 978-1-57473-067-8 (2 volume set, paperback)

Ordering information

To order a copy of the proceedings or to place a standing order, contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, sales@cascadilla.com, www.cascadilla.com