ABSTRACT

It has been claimed that a correlation does not exist between how accurately experienced late learners produce and perceive phonetic segments in a second language (L2). According to one theory, learners of an L2 are no longer able to align segmental production and perception after the passing of a critical period. This contribution reviews studies that have examined L2 production and perception. All of the studies yielded significant, albeit modest, correlations. Possible explanations for why stronger correlations have not been observed are presented.

1. INTRODUCTION

The later a second language is learned, the more strongly foreign accented its pronunciation is apt to be, both globally (e.g., [3]) and in terms of the production of particular vowels and consonants (e.g., [6]). Two accounts have been offered for age effects on L2 production accuracy. According to a critical period hypothesis ([12]), the ability to learn speech diminishes because of lost neural plasticity. According to the Speech Learning Model [4], on the other hand, speech learning mechanisms remain intact across the life span, although they may cease to yield completely native-like production because of age-related changes in how the L1 and L2 phonetic systems interact.

A core tenet of the Speech Learning Model (SLM) is that production errors often result from native vs. non-native differences in segmental perception. By hypothesis, perceptual differences may arise from an age-related decrease in the likelihood that new categories will be established for vowels and consonants that are found in the L2 but not the L1 and/or from changes in how L1 and L2 categories mutually influence one another. The decreased likelihood of category formation is, itself, attributed to changes in the state of development of L1 phonetic categories at the time L2 learning begins.) According to the SLM, the accuracy with which L2 segments are perceived limits how accurately they will typically be produced, although not all aspects of perceptual learning may be incorporated in production. (That is, production and perception may not be brought into perfect alignment, as in L1 speech acquisition.) Thus, the SLM predicts that modest correlations will exist between L2 segmental production and perception for highly experienced speakers of an L2.

Strange [10], on the other hand, observed that although the segmental production errors of inexperienced L2 learners can often be predicted from their perceptual errors, perceptual difficulties may persist even after segmental production has been mastered. Strange concluded (p. 79) that L2 segmental production and perception by experienced speakers of an L2 may be "uncorrelated". According to Bever [1], L2 speech is typically foreign-accented because the "psychogrammar" used by children to align segmental production and perception during L1 acquisition becomes unavailable once the L1 sound system has been established. However, a review of recent studies supports the prediction of the Speech Learning Model ([5]) that a modest correlation between production and perception exists, even for highly experienced speakers of an L2.

2. EMPIRICAL FINDINGS

2.1. Consonants

Flege [3] examined the production and perception of vowel duration cues to the word-final English /t/-/d/ distinction. Both late Chinese/English bilinguals (n = 30) and early Chinese/English bilinguals (n = 9) made vowels significantly longer before /d/ than /t/ (p < 0.01). However, the duration differences produced by the late bilinguals were smaller than those produced by the early bilinguals or by native speakers of English (p < 0.01). Perception was assessed using the method of adjustment. In separate blocks, subjects selected the best instances of bat and bead from the 17 members of a natural-edited beat-to-bead continuum, and the best instances of bat and bad from a bat-to-bad continuum. The late Chinese/English bilinguals chose stimuli with significantly longer vowels as the best instances of /d/-final than of /t/-final English words. However, their perceptual differences were smaller than those of the early bilinguals or native English speakers (p < 0.01). The size of vowel duration differences observed for the Chinese subjects in production and perception showed a modest positive correlation (r = 0.54, p < 0.01).

Another study ([7], [11]) examined voice onset time (VOT) in the production and perception of word-initial English stops by 40 native speakers of Spanish who arrived in the United States as young adults. As expected, many of the native Spanish subjects produced English /p/ with shorter VOT values than the native English control subjects. The subjects' perception of /p/ was assessed by having them rate for goodness, as instances of English /p/, the members of /bat/-to-/bad/ VOT continuum. As VOT increased beyond values typical for English /b/, the native English subjects gave increasingly high ratings to members of the VOT continuum. Then, as VOT increased beyond values typical for English /p/, the native English subjects began to give lower ratings.

The VOT value of the stimulus in both continua that received the highest ratings, called the perceptually "preferred" VOT value, was determined for each subject. The native English subjects who produced /p/ with relatively long VOT values tended to perceptually prefer stimuli having relatively long VOT values, leading to a significant production-perception correlation.
(r = 0.54, p < 0.01). The native Spanish subjects were assigned to a "proficient" or a "non-proficient" group based on their overall degree of perceived foreign accent in English sentences. A significant production-perception correlation was observed for the proficient subjects (r = 0.49, p < 0.01), but not for the non-proficient subjects (r = -0.004, p > 0.10). This finding suggests that, contrary to the hypothesis of Bever [1], nonnative adults' L2 segmental production and perception begin to align as they gain proficiency in the L2.

2.2. Vowels

Flege et al. [5] examined the production and perception of English vowels by 20 native speakers each of German, Spanish, Korean and Mandarin who arrived in the United States as adults and had lived there for an average of 4 years (range: 0.2 to 23 years). The subjects' production of /i/ or /æ/ was assessed acoustically and also auditorily (through native English speakers' forced-choice identifications). Native English speakers produced large spectral differences, but only small temporal differences between /i/-/æ/ and /æ/-/æ/. Many non-native subjects failed to produce spectral differences between one or both pairs, instead producing exaggerated temporal differences. The subjects also identified vowels in synthetic beat-to-bit (/i/-/æ/) and bet-to-bit (/æ/-/æ/) continua in which spectral quality varied in 11 steps and vowel duration varied in three steps. The native English subjects relied primarily on spectral variation in the synthetic stimuli. On the other hand, many non-native subjects relied on temporal cues, some to the exclusion of spectral cues. This finding suggested that certain non-native subjects failed to differentiate /i/-/æ/ and/or /æ/-/æ/ on the basis of spectral differences.

The relation between the non-native subjects' production and perception of /i/-/æ/ was assessed by computing the percentage change in /i/ responses that occurred as F1 frequency increased in the synthetic stimuli, and the magnitude of spectral differences between /i/-/æ/ that were produced. The larger were the spectral differences that the non-native subjects produced, the greater was the effect of the spectral manipulation in the synthetic vowel stimuli on their identifications of vowels as /i/ or /æ/ (r = 0.53, p < 0.01). A similar production-perception correlation was observed when the same kind of analysis was applied to /æ/-/æ/ (r = 0.52, p < 0.01).

Inspection of scatter plots revealed that the non-native subjects who showed a large spectral effect in perception produced spectral differences that varied greatly in magnitude. Most, but not all, of the subjects who showed little spectral effect in perception produced little or no spectral difference between /i/ and /æ/. Each subject was classified as an "accurate" perceivers of /i/-/æ/ (a decrease in /i/ judgments > 80% as F1 values increased; n = 18), as a "moderately accurate" perceivers (shifts of 10% to 79%, n = 22), or as a "poor" perceivers (shifts < 10%, n = 40). The accurate perceivers were found to have produced larger spectral differences between /i/-/æ/ than the moderately accurate perceivers who, in turn, produced significantly larger spectral contrasts than the inaccurate perceivers (p < 0.05). Similarly, the 30 accurate perceivers of /æ/-/æ/ and the 28 moderately accurate perceivers were found to have produced larger spectral differences between /æ/-/æ/ than did the inaccurate perceivers of /æ/-/æ/ (p < 0.05).

A study by Flege et al. [6] examined highly experienced L2 learners. The subjects were 72 Italians who had lived in Canada for an average of 35 years. The subjects' accuracy in producing English vowels was evaluated auditorily by native English-speaking listeners. (Intelligibility scores were obtained for 10 vowels through forced-choice identifications; and goodness ratings were obtained by presenting 11 vowels in separate blocks.) The perception of English vowels was assessed using a categorial discrimination test. The later the native Italian subjects had arrived in Canada, the less accurately they produced certain English vowels, and the lower were their discrimination scores. There was a significant correlation between the subjects' discrimination of English vowels and the intelligibility of the vowels they produced (r = 0.62, p < 0.01). There was also a correlation between the discrimination scores and the goodness ratings accorded vowels produced by the native Italian subjects (r = 0.59, p < 0.01).

When the effect of the subjects' age of arrival (AOA) in Canada was partialled out, the correlation between discrimination and vowel intelligibility remained significant (r = 0.32, p < 0.01), but not the correlation between the discrimination scores and the goodness ratings. In a multiple regression analysis, AOA accounted for 48% of the variance in the vowel intelligibility scores at Step 1. The discrimination scores accounted for 5% of variance at Step 2, and the vowel discrimination scores accounted for 3% of variance at Step 3.

Ceiling effects may have led to an under-estimation of the strength of correlation between the native Italian subjects' production and perception of English vowels. Most vowels spoken by the majority of subjects were identified as intended. However, 31 subjects'/æ/s were often misidentified. These subjects were used to test the SLM's prediction that while there may be individuals who perceive a particular L2 vowel well and yet produce it poorly, the reverse (i.e., good production but inaccurate perception) will not occur. If this prediction is correct, one would expect the 31 "poor producers" of /æ/ to have greater difficulty discriminating English /æ/ from Italian /a/ (which appears to be the Italian vowel that is perceptually closest to English /æ/) than subjects who were relatively good producers of /æ/. The poor producers of /æ/ were indeed found to discriminate English /æ/ from Italian /a/ less accurately than the good producers of /æ/ (p < 0.01). However, the two groups' discrimination of /æ/ from Canadian English /æ/, and /æ/ from English /æ/, did not differ.
significant \( (p < 0.01) \).

### 2.3. Relations at higher levels

There is evidence [2] that individuals who speak their L2 with a foreign accent may nonetheless be able to accurately gauge degree of foreign accent in L2 sentences. Groups of Chinese adults from Taiwan who had lived in the US for averages of 1.5 and 5.3 years were found to speak English with equally strong foreign accents. The same subjects were asked to rate English sentences spoken by native speakers of Chinese for degree of foreign accent. The ratings obtained from inexperienced Chinese subjects were correlated with the ratings obtained from native English-speaking listeners \( (r = 0.88) \), but the correlation obtained for the more experienced Chinese subjects \( (r = 0.95) \) was even more strongly correlated \( (\chi^2(1) = 7.8, p < 0.01) \). It appears from this finding that late learners of an L2 may be able to perceptually detect certain divergences from the phonetic norms of their L2 that are evident in their own speech production. If so, then this finding supports the view that perception at times may "precede" production.

Meador et al. [9], examined the repetition of semantically unpredictable English sentences presented in noise to 72 native Italian subjects who were highly experienced in English. The more words the native Italian subjects were able to repeat, the less foreign-accented was their pronunciation of English sentences \( (r = 0.69, p < 0.01) \), and the better able they were to identify English word-initial and word-final consonants presented in noise \( (r = 0.59 \) and \( 0.49, \) respectively; \( p < 0.10) \). This finding links word recognition to lower-level, segmental perception as well as to overall success in pronouncing an L2.

### 3. SUMMARY

Contrary to the hypotheses of [1] and [10], the research reviewed here indicates that L2 segmental production and perception are correlated. This appears to be true even for late learners who are highly experienced speakers of the L2. The correlations that are typically observed, however, have been modest, averaging about \( r = 0.50 \). It is uncertain why the observed correlations have not been larger. In this last section, several possible explanations for why the correlations have not been higher will be presented.

One possibility is that not all subjects who have learned to perceive an L2 sound accurately will update segmental production to conform to their new, or modified, long-term memory representation for the L2 sound. There does not appear to be a complete breakdown in the ability to align production and perception, as hypothesized by Bever ([11]). However the possibility exists that there is a partial breakdown. Alternatively, certain L2 learners may not bring production into conformity with perception for psychosocial reasons. For example, they may not want to sound completely native-like in the L2 for practical reasons, or because of L1 language "loyalty".

Still another possibility is that there are age limits on learning new forms of articulation. According to the Speech Learning Model ([5]), L2 production accuracy is "limited" by the accuracy with which L2 segments are perceived. However, the SLM does not claim that all instances of non-native segmental production have a perceptual origin. It may be the case, especially for articulatorily complex sounds that are not commonly found in human languages, that true age-related limits exists for articulatory learning. To show this, however, it would be necessary to demonstrate that highly experienced learners produce an L2 sound inaccurately in the absence of a perceptual divergence from native speakers of the target L2. To my knowledge, such evidence has not yet been presented in the literature.

Methodological factors may also contribute to the modest strength of the production-perception correlations that have been observed in the literature reviewed here. One factor to consider is what is being measured in L2 research. As is well known, segmental phonetic contrasts are typically based on multiple dimensions that enter into "trading" relations. However, most previous research in the area of L2 speech acquisition has focused on a single dimension, such as VOT. As a result, these studies may have overlooked some other dimension(s) that changed concurrently in both production and perception. If so, then the correlations between production and perception may have been underestimated. One recent study (McAllister, Flege and Piske, this volume) examined the production and perception of duration differences between pairs of Swedish mid vowels that differed in phonological length. The subjects were native speakers of Spanish and English who had lived in Sweden for an average of 21 years. The correlation observed in this study, \( r = 0.70 \), was higher than in any previous study, perhaps because the production and perception variables were more nearly commensurable than those examined in previous studies.

Another methodological factor to consider is speech clarity. As is well known, the production of phonetic segments changes across variations in degree of clarity (or carefulness), as well as speaking rate. Researchers often attempt to characterize their speech samples along these dimensions. For example, words that have been read from a list may be referred to as having been produced in a "slow, careful" manner. Regardless of whether such characterizations are correct, it is important to note that the same is not usually done with respect to perceptual data that are collected. Thus, there is no guarantee that the perception and production tests are matched in terms of rate and clarity. The "hyperspace" effect described by Johnson et al. [13] illustrates the nature of this potential problem. The subjects in that study used the method of adjustment to choose the best instances of a vowel category. The vowels they chose tended to be more peripheral in the vowel space than were the subjects' own productions of the vowels. If a similar phenomenon exists for L2 learners, it would lead to an underestimation of the strength of correlation between their production and perception of L2 phonetic segments.
Finally, it may be the case that only modest production-perception correlations have been observed in the L2 literature because L2 research is not assessing the most meaningful perceptual variable. The Speech Learning Model [4] posits that category formation exerts a powerful influence on the accuracy of L2 segmental production. This implies a discontinuity in production accuracy in some instances. More specifically, late bilinguals who have established a category for a sound that is found in the L2 but not in the L1 would be expected to produce the L2 sound more accurately than those who have not established a category. If so, it may be useful in future research to examine the relation between production and perception using discrete tests of category formation instead of, or in addition to, continuous tests of subjects' use of specific perceptual "cues" such as VOT.

This approach can be exemplified by considering a recent study [8] that examined native Spanish speakers’ production and perception of English word-initial voiceless stops. Category formation was indexed by whether, in a goodness rating test, subjects’ preferred VOT values for /p/ shifted according to (simulated) speaking rate. Native Spanish subjects who showed the shift, and thereby resembled native speakers of English, might be credited with having established a category for the long-lag /p/ of English. The subjects who could be credited with having formed an English /p/ category were found to produce stops with English-like long-lag VOT values. On the other hand, the native Spanish subjects who were not credited with category formation were found to produce English stops with Spanish-like short-lag VOT values.

In summary, a useful strategy for future research may be to accompany a continuous evaluation of L2 production accuracy with discrete tests of L2 segmental perception that assess category formation. Such a research strategy may yield greater insight into the relation between segmental perception and production than has been afforded by studies examining single, continuous dimensions in both domains.

ACKNOWLEDGMENTS
This research was supported by a grant from NIDCD (DC00257). The author thanks Susan Guion and Thorsten Piske for their comments on a previous version. Thanks are also extended to the many colleagues who contributed to the work reviewed here, including Ocke-Schwen Bohn, Ian R.A. MacKay, Bob McAllister, Diane Meador, Murray Munro, and Anna Schmidt.

REFERENCES