Syllable Structure Universals and Native Language Interference in Second Language Perception and Production: Positional Asymmetry and Perceptual Links to Accentedness

Bing Cheng[1,2] and Yang Zhang[2,3,4]*

[1] Department of English, School of Foreign Studies, Xi'an Jiaotong University, Xi'an, China, [2] Department of Speech-Language-Hearing Sciences, University of Minnesota, Minneapolis, MN, USA, [3] Center for Neurobehavioral Development, University of Minnesota, Minneapolis, MN, USA, [4] Center for Applied and Translational Sensory Science, University of Minnesota, Minneapolis, MN, USA

The present study investigated how syllable structure differences between the first Language (L1) and the second language (L2) affect L2 consonant perception and production at syllable-initial and syllable-final positions. The participants were Mandarin-speaking college students who studied English as a second language. Monosyllabic English words were used in the perception test. Production was recorded from each Chinese subject and rated for accentedness by two native speakers of English. Consistent with previous studies, significant positional asymmetry effects were found across speech sound categories in terms of voicing, place of articulation, and manner of articulation. Furthermore, significant correlations between perception and accentedness ratings were found at the syllable onset position but not for the coda. Many exceptions were also found, which could not be solely accounted for by differences in L1–L2 syllabic structures. The results show a strong effect of language experience at the syllable level, which joins force with acoustic, phonetic, and phonemic properties of individual consonants in influencing positional asymmetry in both domains of L2 segmental perception and production. The complexities and exceptions call for further systematic studies on the interactions between syllable structure universals and native language interference with refined theoretical models to specify the links between perception and production in second language acquisition.

Keywords: native language neural commitment, phonetic learning, speech perception, speech production, accentedness, syllable structure, allophonic variations

INTRODUCTION

Accentedness in speech refers the degree to which the pronunciation of an utterance sounds deviated from an expected version of its production (Munro et al., 2006). Foreign accent seems to be a natural and unavoidable consequence as adult learners of a second language (L2) are considered to be "phonologically deaf" to the new language (Listerri, 1995). It is well

acknowledged that some L2 sounds are much harder to learn than others. In particular, native-like perception and production for the consonants and vowels that either do not occur in the first language (L1) or are phonologically realized differently can be very difficult to achieve for adult L2 learners. A fundamental question in second language acquisition research is to explain the patterns of perceptual difficulties for different speech sound categories in various phonological contexts in connection with production issues in phonetic learning.

There are three general accounts that, respectively, emphasize the roles of the learner characteristics, input properties, and L1 influences. The *Critical Period* account emphasizes age dependency, which postulates genetically guided maturation of domain-specific language learning mechanisms as the reason for the declining abilities to learn L2 (Lenneberg, 1967; Johnson and Newport, 1989). The *Environmental Influence* account shifts the focus to assimilation of external factors by arguing that L2 proficiency variations reflect the different degrees of environmental influences as learners of various ages are inherently at different levels of cognitive, social, and cultural maturation (Snow, 1983; Jia and Aaronson, 2003). The *L1 Transfer/Interference* account highlights the role of structural compatibility, demonstrating that L2 perception/production is directly influenced by the L1 phonological system (Beckman, 1986; Munro and Derwing, 1995; Guion et al., 2000; Eckman et al., 2003; Iverson et al., 2003; Aoyama et al., 2004; Wang et al., 2004; Rogers and Dalby, 2005). Converging evidence suggests that the L1–L2 structural differences may interfere with speech learning at multiple levels, which include the acoustic and phonetic properties of speech sounds, allophonic variations, phonological rules, and syllable structure. The present exploratory study contributes data to the line of L1 transfer/interference research by investigating the role of syllable structure differences between L1 and L2 in L2 segmental perception and production. A corollary question is to investigate the strength of perception-production links as a function of syllabic position and sound category.

A number of theoretical models have been proposed for the L1 Transfer/Interference account at the segmental level. For instance, the Speech Learning Model (SLM; Flege, 1995, 2003, 2007) addresses how adult speakers acquire L2 consonants and vowels based on phonetic similarity to L1. The model predicts that production accuracy is limited by how accurately the L2 sounds are perceived (MacKay et al., 2001). Thus a major determinant of L2 accentedness is the underlying perceptual problem with the L2 phonology. According to the SLM, learners of all ages retain the capacity to align their production of L2 phonetic segments to long-term memory representations for vowels and consonants in the L2. Similarly, the Perceptual Assimilation Model (PAM; Best, 1995; Best and Tyler, 2007) proposes that L2 listeners classify sound contrasts via different routes of assimilation. L2 contrasts are classified as a Two Category (TC), Category Goodness (CG), or Single Category (SC) contrast, which depends on the similarities between the L2 and L1 sounds associated with different degrees of learning difficulties. The model attributes "malleability" to the perceptual systems of adult L2 learners, and predicts that speech

discriminability may improve as a function of L2 experience (Best and Strange, 1992).

While the SLM and PAM are based on behavioral studies alone, the Native-Language-Neural-Commitment (NLNC) model adds a neural-level specification and explanation from both developmental and cross-linguistic perspectives (Kuhl, 2004). There are four main theoretical claims in the NLNC model. (a) Early learning produces neural commitment to the abstract phonetic units and their statistical and combinatorial patterns in the native language in the process of establishing prototypical representations of the phonemic inventories and phonological structures, which can be predictive of language skills at word and sentence levels at later ages (Kuhl et al., 2005). For instance, specific brain regions in adulthood become more sensitive to native speech contrasts with more focal and efficient activation in comparison with non-native speakers (Zhang et al., 2005). (b) The effects of NLNC are self-reinforcing and bidirectional – it enhances efficient processing of compatible higher-order linguistic patterns, while at the same time hindering the detection of non-conforming patterns contained in foreign languages, as shown behaviorally (e.g., Iverson et al., 2003) and neurally at the pre-attentive level (e.g., Zhang et al., 2005). (c) Neural commitment is subject to continual shaping and reshaping by experience – Enriched exposure (including high stimulus variability and talker variability, exaggerated speech, and audiovisual training) not only provides enhanced stimulation to the infant brain (e.g., Zhang et al., 2011) but also can induce substantial plasticity in the adult brain for second language learning, producing hemispheric reallocation of resources for enhanced phonetic sensitivity and more efficient linguistic processing (Zhang et al., 2000, 2009). (d) Neural commitment involves the binding of perception and action systems to facilitate speech communication, and this process depends on social/affective learning early in life (Imada et al., 2006; Kuhl, 2007; Stevens and Zhang, 2014). These claims are consistent with the developmental framework that views language acquisition as an adaptive computational process to extract the abstract speech categories and higher-order linguistic structures. Similar views were expressed by other researchers (e.g., Werker and Tees, 2005). Unlike the SLM and PAM, the NLNC model does not specify how different degrees of acoustic, phonetic, or phonological conformity independently or jointly influence L2 learning in childhood or adulthood, which may result in varying outcomes in the cortical and subcortical brain circuits and networks dedicated to L1 and L2 phonological acquisition.

It is noteworthy that the three models for speech learning draw heavily on data from perceptual tests of consonants at the syllable-initial position alone. Apart from what these segmental learning models can explain and predict, suprasegmental differences between L1 and L2 strongly influence L2 phonetic perception and production. The adult L2 perception and production literature indicates the existence of a positional asymmetry that is attributable to differences in L1 and L2 syllable structures (Anderson, 1983; Flege, 1989; Stockman and Pluut, 1999; Broselow and Xu, 2004; Tsukada et al., 2004; Rogers and Dalby, 2005; Cebrian, 2007; Hayes-Harb et al.,

2008). For instance, Mandarin Chinese allows no syllable-final consonants other than two nasals (/n/ and /ŋ/; Duanmu, 2000). Correspondingly, Chinese ESL (English as a Second Language) learner's perception of stop contrasts was more accurate at the syllable-initial (onset) position than at the syllable-final (coda) position (Flege, 1989), and they also made more production errors in the coda than in the onset (Rogers and Dalby, 2005; Bent et al., 2007). Rogers and Dalby (2005) developed a minimal-pair word list targeting all the English phonemic contrasts that posed difficulty for Mandarin ESL learners and analyzed the intelligibility of Mandarin-accented English (MAE). The MAE segmental misperceptions for native English listeners occurred more frequently at the word-final position than at the word-initial position. The frequent word-final misperceptions included nasals, voiced obstruents, and consonantal clusters. These patterns indicate that Mandarin speakers have more difficulty in accurately producing consonants in the coda than in the onset. Similar positional asymmetry findings were previously reported in other studies, confirming the effect of L1 syllable structure and phonotactics on L2 learning (Eckman, 1981; Flege, 1995).

One straightforward explanation is that the syllable-final difficulties in L2 consonantal perception and production may reflect a general pattern of universal structural asymmetry in favor of the onset in terms of accessibility, stability, and learnability (Ohala, 1996; Carlisle, 2001; Nam et al., 2009). If the positional asymmetry is universally applicable and strong regardless of language experience, one would expect it to occur even in normal L1 learners. Moreover, if the positional asymmetry is a dominant force in L2 phonetic learning, it could potentially overrule the influences of acoustic, phonetic, and phonemic factors and be applicable to all classes of consonants. Given the constraint of syllable structure universals, it also follows that correlations between perception and production would be stronger for consonants in the CV context as opposed to VC. While these three conjectures provide the initial theoretical impetus for the current investigation, our interest is not limited to the verification or falsification of this language-universal structural preference account. Rather, we are intrigued by three basic observations in L2 phonetic learning: (a) positional asymmetry occurs in both domains of perception and production, (b) certain classes of sounds appear to be more affected than others in terms of positional asymmetry, and (c) the strength of correlations between perception and production appears to vary depending on sound category and syllabic position (Cheng and Zhang, 2009). Speech learning theories generally assume that accurate production of many L2 sounds tends to be challenging for adult learners due to their perceptual difficulties, but there has been a lack of studies to examine systematically the extent to which production errors or accentedness are perceptually based as a function of syllabic position and sound category. There is also no theoretical model that readily provides an exact account for the presence/absence (or rather the strength/weakness) of perception-production correlations in L2 phonetic learning.

The relationship between perception and production has been an outstanding issue in speech science (Massaro and Chen, 2008). In L2 acquisition, only moderate correlations have been reported between perception and production (Flege, 2003; Rauber et al., 2005; Cheng and Zhang, 2009). Some studies show either weak or no correlation between L2 segmental perception and production (e.g., Hattori and Iverson, 2010; Levy and Law, 2010; Kartushina and Frauenfelder, 2014). According to the SLM, perception precedes and guides production with perception accuracy imposing an upper limit on production accuracy. Although this contingency relation takes place in a position-sensitive allophonic level as opposed to the abstract phonemic level, it does not necessarily imply correlation between perception and production (Flege, 2003). Consistent with the argument, research findings show that while appropriate perceptual training methods produce sizable improvements in both perception and production, the amount of perceptual gains does not necessarily align with the amount of gains in production accuracy (Rochet, 1995; Bradlow et al., 1997, 1999; Best et al., 2001). The SLM specifically points out the two possible exceptional scenarios to the general rule that production is perceptually based. The first concerns the lack of sufficient input to allow L2 learning in either production or perception. The second applies to a small set of typologically rare L2 phonetic segments that are mastered relatively late even in normal native speakers of that language. If these two exceptional cases are removed, one would generally expect to see a significant correlation between perception and production in L2 learners. While the SLM spells out the exceptions and emphasizes the need to consider allophonic variations in different contexts, it stops short of providing predictions regarding the relative strength of position-dependent perception-production correlations in adult L2 learners for the different classes of speech sounds.

The main purpose of the present study was to investigate positional asymmetry and the influences of Mandarin Chinese syllable structure on the relationship between perception and production of American English consonants. In one previous report, Mandarin speakers' perceptual gains on a French /b-p/ contrast during perceptual training resulted in improved production (Rochet, 1995). Importantly, the training effects were found to generalize to other stop contrasts in the syllable-initial position but not in the syllable-final position, suggesting that there might be a stronger correlation between perception and production in the syllable-initial position. There are at least two related issues here in terms of L1 transfer/interference. First, as previous studies have shown and argued, production error patterns at the onset and coda positions reflect interference of the native language syllable structure. Second, poorer L2 perception may lead to more accented speech as well as production errors, and the strength of the perception-production relationship could be further influenced by L1 syllable structure. By contrast, an alternative possible explanation is that the positional asymmetry as well as the strength of perception-production links may reflect universal structural preference of CV over VC. In this perspective, native language interference could be a subordinate factor. Although the positional asymmetry effect is attributable to structural differences between L1 (Mandarin Chinese) and L2 (American English) at the syllable level, it remains to be tested whether the effect applies across the board to all classes

of L2 consonants that are not permissible in the L1 syllable-final position. It is also equally important to examine whether the positional asymmetry can be found in both native and non-native speakers and whether it applies to the L2 consonants that are allowed in the L1 syllable-final position, both of which would be consistent with the universal structural asymmetry account.

A comparison of Mandarin Chinese (also known as Putonghua in mainland China) and American English phonemes and syllable structures is necessary to help understand the syllable structure differences and the consonant systems of the two languages. There are approximately 400 syllables in Mandarin without tones and 1100 syllables with the tones added (Chao, 1968; Duanmu, 2000). In contrast, English has a much larger inventory of permissible syllables, which exceeds 80,000. At the individual sound level, Mandarin has three nasal sounds as in English. Unlike the English voicing distinction, the bilabial, alveolar, and velar stops in Mandarin are distinguished by the presence or absence of aspiration. In other words, all the Mandarin Chinese stops are voiceless sounds. Similarly, the fricatives and affricates in Chinese also rely on the aspiration distinction instead of voicing. English has more fricatives (9 in total) than Mandarin (5 in total). Except for the glottal /h/, English fricatives carry the voicing distinction for the place of articulations (POAs) of labiodental, interdental, alveolar, and palatoalveolar. There are no Mandarin equivalents for the English interdental fricatives, /θ/ and /ð/. In contrast, Mandarin has a richer set of affricates, namely, the alveolar /ts/ and / tsʰ/, the retroflex /tʂ/ and /tʂʰ/, and the palatoalveolar /tɕ/ and /tɕʰ/, each pair carrying the aspirated-unaspirated distinction. In English, there is only one pair of palatoalveolar affricates with the voicing distinction (/tʃ/ and /dʒ/). In the approximant category, Mandarin has liquids and glides like English. But the Mandarin 'r' sound is often transcribed as /ʐ/ (a voiced retroflex fricative). The two languages also differ in phonotactic constraints in terms of combinatory and distributional restrictions. While consonants can form clusters in American English words, no consonant clusters are allowed in Mandarin. The inventory of coda structures in English is considerably richer than Chinese. For the current study, we take into account 18 consonants that could appear in both initial and final position of the syllable in English (see **Table 1** for sample words). English also allows all consonants to appear in either syllable-initial or syllable-final position except /h/, / j/ and /w/. In Mandarin Chinese, only two nasal sounds, /n/ and /ŋ/, are permissible in the syllable-final position. In both languages the velar nasal sound /ŋ/ cannot occur in the word-initial position.

In the present study, an overall assessment of consonantal perception and production skills in adult ESL learners in China was attempted by examining effects for all the major individual classes of English consonants in terms of voicing, POA, and manner of articulation (MOA) in the CVC (consonant-vowel-consonant) syllabic context. Testing all the major classes of English consonants was partly motivated by the pedagogical need to identify general patterns and issues in perception and production among Chinese college students for whom English is part of their core curriculum. For

TABLE 1 | Sample words used in the study.

Target Phoneme	Examples	
	Syllable-initial	Syllable-initial
m, n, ŋ*	Meat, neat	Seem, seen, sing
b, p, d, t, g, k	Bit, pit, dean, teen, gum, come	Rib, rip, need, neat, dug, duck
f, v	Fine, vine	Leaf, leave
θ, ð	Thin, then	Bath, bathe
s, z	Sip, zip	Loose, lose
ʃ	Sheet	Leash
tʃ, dʒ	Chin, gin	Rich, ridge
l, ɹ*	Lead, read	Heal, here

*/ ŋ / is not allowed in the onset. / ɹ / in the syllable-final position is rhotacized into the r-colored vowel in American English. The data for / ŋ / and / ɹ / were not included in the statistical analysis of positional asymmetry.

comparison purpose, perception data were also collected from adult native speakers of English using the same set of speech stimuli to verify whether native English speakers would show a similar positional asymmetry to the Chinese subjects. The voicing subcategories in this study refer to the two-way distinction of voiced and voiceless consonants. The POA subcategories refer to the six-way distinction of bilabial, labiodental, dental (interdental), alveolar, palatal-alveolar, and velar consonants. The MOA subcategories refer to the five-way distinction of nasal, plosive, fricative, affricate, and approximant (including both liquids and glides) consonants.

There were two specific hypotheses for our study motivated by the three models on L2 phonetic learning and the universal structural asymmetry theory. First, adult Chinese ESL learners would experience more difficulties in perceiving and producing English consonants at the syllable-final position relative to the syllable-initial position across all the voicing, POA, and MOA subcategories due to the limitations of Mandarin Chinese syllabic and phonotactic structures. We did not expect the native speakers of English would have the same perceptual issue although the universal structural asymmetry theory would predict otherwise. Second, the position-dependent difficulties would be reflected in stronger correlations between L2 perception and production at the syllable onset than at the coda.

Considering the syllable structure of Mandarin Chinese and the complexity of the causes for L2 phonetic learning difficulties due to the relative importance of acoustic, phonetic, phonemic properties of the individual speech sounds in each subcategory, we expected to see category-dependent variations in the positional asymmetry phenomenon as well as exceptions to our two general hypotheses at the individual sound level. In particular, as the Chinese language allows /n/-/ŋ/ contrast at the coda position, our expectation, in line with the three models of L2 phonetic learning, was that the Chinese ESL learners would not experience much difficulty in producing or perceiving this syllable-final nasal contrast in English. In other words, the nasal sounds might not show the positional asymmetry. As the POA has more subcategories tied to individual

sounds (e.g., the labiodental /f/ and /v/, interdental /θ/ and /ð/, palatal /ʃ/, palatoalveolar /tʃ/ and /dʒ/; see **Table 1** for the list of target phonemes) than voicing and the MOA, we might see a constellation of exceptions in the POA subcategories. In accordance with the SLM and PAM, English sounds such as the voiced stops, voiced fricatives/affricates, interdental fricatives, and palatoalveolar affricates that do not exist in Chinese would show different degrees of perception and production difficulties with some of these sounds demonstrating equal difficulties or reduced positional disparity between the syllable onset and coda.

MATERIALS AND METHODS

Participants

The reported study was conducted with approval from the institutional review boards for human subject protection at the two home institutions. Thirty-nine Chinese college students participated in this study. One Chinese female student did not complete all the tests in the study and was dropped in statistical analysis. The Chinese participants were between 20 and 21 years old (29 females and 10 males). All of them were native speakers of Mandarin Chinese. When selecting participants, we excluded students who spoke Mandarin Chinese (also referred as "Putonghua") with a noticeable accent. Our survey of language background indicated that all our Chinese participants grew up speaking standard Mandarin Chinese from the first year of age. While eight (six females and two males) of the 38 participants were raised in regions where the local dialects (Wu and Min) were very different from Putonghua (Wang, 1991), they all learned the standard Mandarin early on and predominantly spoke Putonghua since early childhood. For cross-language comparison, a group of 12 native American English speakers participated in the perception test subjects (age: 19–22; 7 females and 5 males). All the participants were volunteers under informed consent and recruited after screening for speech, hearing, and language background.

To reduce possible confounds on the perception-production correlations due to insufficient input/learning as specified in the SLM, we were rather restrictive in recruiting the Chinese subjects with respect to their L2 background. The Chinese subjects had received at least 6 years of formal English education in junior and senior high school, in which reading and writing proficiency in English were emphasized rather than conversational skills. None had lived in an English-speaking community before joining the study. All the Chinese participants had passed the National Matriculation English Test in China for college admission, and all except one had passed the national standardized English proficiency test, Test for English Majors, TEM-4 (Zou, 2003; Cheng, 2008), before joining the study. Their average score for TEM-4 was 71.3 out of 100 points with a standard deviation of 7.7, and the average listening comprehension score was 24.8 out of 30 points with a standard deviation of 3.2. The one participant who failed the TEM-4 was also included in our study because this person had a 70% passing grade (21 out of 30) for the listening

comprehension section. According to the TEM-4 syllabus, a passing listening grade for TEM-4 indicated that the examinee had a vocabulary knowledge of approximately 6000, and was able to understand speeches or conversations by native English speakers about daily and social life, news broadcasts of BBC and VOA at normal speed, and listening passages comparable to the mini talks in TOEFL (Test of English as a Foreign Language).

Perception Test

The perception data were collected from both subject groups with a revised Speech Assessment and Training software program (SAT) based on versions used in previous publications (Zhang et al., 2000; Cheng and Zhang, 2009, 2013; Zhang and Cheng, 2011). In the current implementation, isolated word stimuli were used, containing all consonants with Standard American English pronunciation (see **Table 1** for examples and the target phonemes). To control stimulus familiarity, all word candidates were selected from the published vocabulary list required for College English Test (CET Band 4) in China (Zheng and Cheng, 2008). The word stimuli were spoken by four adult native speakers of American English (two males and two females in the age range of 20 ∼ 32), and digitally recorded with Presonus Firebox sound card and Sennheiser ME65 microphone in a sound treated booth (Acoustic Systems, ETS-Lindgren). None of these four native English speakers participated in the perception test. Before the digital recording, the native English speakers were asked to familiarize themselves with all the words on the word list. They were then instructed to read each word three times in isolation with approximately 1 s silence in between. One clear pronunciation from each speaker was then chosen for each word. For the purpose of testing positional asymmetry, words with consonant clusters were not included to avoid complications due to phonotactic constraints for permissible consonant clusters in the onset and coda positions. The sound levels for all the chosen words were normalized to have equal RMS (root mean square) average intensity using the SONY Sound Forge 9 program. Allophonic variations in different vocalic contexts were included for each English consonant.

Before the perception test, the subjects were familiarized with the visual phonetic symbols along with alphabetic letter spelling shown on a computer monitor with target sounds marked in each word. After ensuring that the subjects were confident about recognizing all the symbols for the consonants and verifying their symbol recognition accuracy, each phoneme in the perception test was presented 10 times using different words, respectively, in the syllable onset and coda positions. Listeners were explicitly instructed to pay attention to possible target consonant candidates as prompted on computer screen and make choices by clicking the proper symbol. The listeners were asked to rely on their immediate impression of the consonant in the target onset or coda position irrespective of word candidacy or meaning. The word stimuli were presented at 70 dB SPL in random order. The subjects were tested individually in a sound booth. The selected target responses were automatically registered in a Microsoft ACCESS database in the SAT program

and sorted into syllable onset and coda positions for each subject.

Production Data Collection and Accentedness Rating

Production data were digitally recorded from each Chinese participant, featuring isolated words that contained the contexts for allophonic variations of the target English consonants. The order of perception and production tasks was counterbalanced among the subjects. Each word was presented on computer screen, the subjects were asked to pronounce it in isolation three times with approximately 1 s pause in between. The subjects were instructed to pronounce the words clearly each time. The spoken words were then normalized in RMS intensity and assessed by two native speakers of English, who had formal training in phonetics. The SONY Sound Forge 9 program and a Sennheiser HD280 Pro binaural headphone were used on a workstation installed with the Presonus Firebox sound card for the rating task.

While production accuracy judgment only permits a binary choice, the accentedness rating measure is designed to allow more gradient reflection of the degrees of deviation from the target pattern of speech production as expected of a native speaker. For this reason, the accentedness rating measure was adopted in the present study. The ratings for the target consonants were in the range of 1 (wrong consonant or very strong foreign accent) and 5 (correctly produced English consonant, no foreign accent) following a foreign accent rating scale similar to Munro and Derwing (1995). The raters were instructed to focus on the highlighted target consonant for each word written in a column in a Microsoft Excel spreadsheet. The raters used continuous scores with decimal points to allow finer distinctions in rating. The raters were instructed to adjust the sound volume so that they could hear the words clearly. In cases of uncertainty, they were allowed to listen one more time to help determine the score. The rating scores for each target consonant in each position were based on the overall assessment of the three consecutive pronunciations of each word. The scores were individually entered and saved for further analysis.

Data Analysis

The raw scores for each subject were sorted, grouped, and averaged for syllable-initial and syllable-final positions based on the subcategories of voicing, POA, and MOA. The data for the glides (/w/ and /j/) and the glottal fricative /h/ were not included as they did not occur in the syllable-final position in English. Statistical analyses were performed using Systat12. For the perception and production scores, repeated-measures ANOVA tests were performed separately to examine the main effects of syllabic position (onset vs. coda), consonant classification (two classes for voicing, six classes for POA, and five classes for MOA), and possible interactions. As the raters used continuous rating scores, inter-rater and intra-rater reliability for the production data was assessed using Pearson's correlation analysis. *Post hoc* pair-wise *t*-tests were performed for the individual consonant subcategories to examine the effects of syllable structure on perception and production. To examine the overall positional

effect on the perception-production links, correlational analysis was first performed on the overall pooled perception and production scores separately for syllable-initial and syllable-final positions. Correlational tests between perception and production as a function of syllabic position were also run for each class of speech sounds. To examine possible dialectal influences, we additionally performed statistical analyses with dialect as a between subjects factor using a multivariate analysis of variance (MANOVA) model. Given the heterogeneity of the dialectal background of the eight subjects who were from regions where the local dialects are very different from Putonghua, we also performed further analysis by removing these eight subjects from our data set to verify the statistical results.

RESULTS

Speech Perception Data

The pooled perception data and average scores sorted in terms of voicing, POA and MOA categories are shown in the left panels of **Figures 1** and **2** for the Chinese subjects. As expected, the native speakers of English showed ceiling-level perceptual performance for all the consonant categories tested (accuracy range: 95.4–99.2%; standard deviation range: 0.02–0.06). Repeated measures ANOVA results and paired *t*-tests for the grouped consonant categories showed no statistically significant positional asymmetry effects for the monolingual English speakers. As the focus of the study was on L2 perception and production in terms of positional asymmetry, the reported statistics in the remainder of the text would be on the Chinese speakers only. As expected, the Chinese subjects performed significantly better for consonants in syllable-initial position than for those in syllable-final position. There were also exceptions.

In the two voicing subcategories (voiced vs. voiceless), significant effects were found for the main factor of syllabic position [onset versus coda; $F(1,37) = 49.2$, $p < 0.00001$, $\eta^2 = 0.571$] and an interaction between position and voicing [$F(1,37) = 5.05$, $p < 0.05$, $\eta^2 = 0.120$]. There was no main effect of voicing [voiced versus voiceless; $F(1,37) = 0.82$, $p = 0.4$, $\eta^2 = 0.021$]. Further tests of simple main effects showed that significant position effects (onset vs. coda differences) for both the voiced [$F(1,37) = 6.6$, $p < 0.05$, $\eta^2 = 0.151$] and voiceless [$F(1,37) = 30.2$, $p < 0.00001$, $\eta^2 = 0.449$] consonants. A *post hoc* two-tailed paired *t*-test was run to compare the difference scores (onset minus coda) for voiced and voiceless consonants. The result indicated that the voiceless sounds in English had larger onset-coda perceptual difference than the voiced sounds for the Chinese subjects [$t = 2.248$, $p < 0.05$, $df = 37$, Cohen's $d = 0.589$].

In the six POA subcategories (bilabial, labiodental, interdental, alveolar, palatal-alveolar, and velar), significant effects were observed for the main factors of position [onset versus coda; $F(1,37) = 43.5$, $p < 0.00001$, $\eta^2 = 0.541$], place [$F(5,185) = 42.0$, $p < 0.00001$, $\eta^2 = 0.532$; Greenhouse-Geisser corrected]. There was also an interaction between position and place [$F(5,185) = 7.1$, $p < 0.001$, $\eta^2 = 0.159$; Greenhouse-Geisser corrected]. Simple main effects tests showed significant
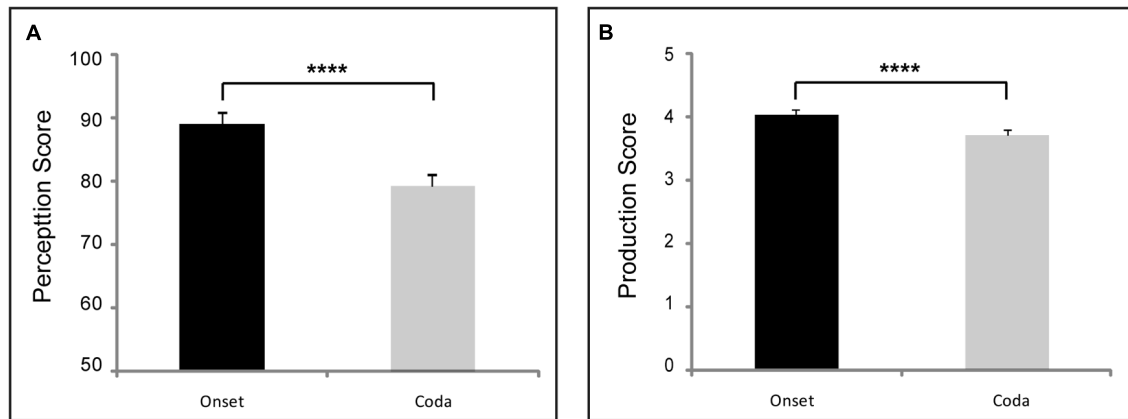
FIGURE 1 | Positional (syllable onset vs. offset) asymmetry in pooled perception (A) and production (B) data of all 38 Chinese subjects. [**** stands for $p < 0.0001$; error bars for a within-subject design are used (Masson and Loftus, 2003)].

differences among the POA subcategories at both the onset [$F(5,185) = 33.9$, $p < 0.00001$, $\eta^2 = 0.478$] and coda [$F(5,185) = 20.0$, $p < 0.00001$, $\eta^2 = 0.351$]. Significant position effects were observed for the bilabial [$F(1,37) = 16.5$, $p < 0.001$, $\eta^2 = 0.308$], labiodental [$F(1,37) = 12.4$, $p < 0.01$, $\eta^2 = 0.251$], alveolar [$F(1,37) = 134.3$, $p < 0.00001$, $\eta^2 = 0.784$], and velar [$F(1,37) = 4.2$, $p < 0.05$, $\eta^2 = 0.101$] sounds. *Post hoc* paired t-tests (two-tailed; Bonferroni-corrected) were performed for each POA category, which confirmed significant syllabic position effects for the bilabial, labiodental, alveolar and velar sounds: bilabial [$t = 4.06$, $p < 0.001$, $df = 37$, Cohen's $d = 0.968$], labiodental [$t = 3.518$, $p < 0.01$, $df = 37$, Cohen's $d = 0.815$], alveolar [$t = 11.59$, $p < 0.00001$, $df = 37$, Cohen's $d = 2.38$], and velar [$t = 2.041$, $p < 0.05$, $df = 37$, Cohen's $d = 0.489$]. The interdental and palatoalveolar sounds, which were not part of the Mandarin Chinese phonemic inventory, were the exceptions: interdental [$t = 0.116$, $p = 0.91$, $df = 37$, Cohen's $d = 0.024$], and palatoalveolar [$t = 1.886$, $p = 0.07$, $df = 37$, Cohen's $d = 0.369$].

In the five MOA subcategories (nasal, plosive, fricative, affricate, and approximant), the significant effects were as follows: position [onset vs. coda; $F(1,37) = 194.1$, $p < 0.000001$, $\eta^2 = 0.84$], manner [$F(4,148) = 35.4$, $p < 0.0001$, $\eta^2 = 0.489$; Greenhouse-Geisser corrected], interaction of position and manner [$F(4,148) = 37.8$, $p < 0.0001$, $\eta^2 = 0.505$; Greenhouse-Geisser corrected]. Simple main effects tests showed significant differences among the MOA subcategories at both the onset [$F(4,148) = 13.0$, $p < 0.00001$, $\eta^2 = 0.260$] and coda [$F(4,148) = 48.1$, $p < 0.00001$, $\eta^2 = 0.565$]. In addition, significant position effects were observed for all the MOA subcategories, namely, the nasals [$F(1,37) = 39.3$, $p < 0.00001$, $\eta^2 = 0.515$], stops [$F(1,37) = 50.2$, $p < 0.00001$, $\eta^2 = 0.576$], fricatives [$F(1,37) = 8.7$, $p < 0.01$, $\eta^2 = 0.191$], affricates [$F(1,37) = 42.3$, $p < 0.00001$, $\eta^2 = 0.539$], and approximants [$F(1,37) = 150.3$, $p < 0.00001$, $\eta^2 = 0.803$]. *Post hoc* paired t-tests (two-tailed; Bonferroni-corrected) confirmed significant syllabic position effects for each MOA subcategory: the nasals [$t = 6.266$, $p < 0.00001$, $df = 37$, Cohen's $d = 1.356$], stops

[$t = 7.085$, $p < 0.00001$, $df = 37$, Cohen's $d = 1.55$], fricatives [$t = 2.953$, $p < 0.01$, $df = 37$, Cohen's $d = 0.579$], affricates [$t = 3.212$, $p < 0.001$, $df = 37$, Cohen's $d = 0.916$], and approximants [$t = 12.26$, $p < 0.00001$, $df = 37$, Cohen's $d = 2.398$].

Speech Production Data

The accentedness ratings for the Chinese subjects in terms of voicing, POA and MOA subcategories are summarized in the right panels of **Figures 1** and **2**. The inter-rater correlation was significant [Pearson's $r = 0.73$, $p < 0.001$]. Intra-rater correlations were further performed by half-splitting the individual ratings for the same consonants in each subject, and the results showed a high level of consistency for both raters [Pearson's $r = 0.86$ for rater 1, Pearson's $r = 0.89$ for rater 2; $p < 0.0001$ for both raters]. The scoring data indicated that the two native English speakers with training in phonetics were highly consistent in their accentedness judgment. The two raters' scores were thus averaged for each sound tested for each subject. Overall, the subjects' consonantal production was rated significantly better (i.e., less accent) for syllable-initial position than syllable-final position. There were also some exceptions in individual subcategories as reported below.

In the two voicing subcategories (voiced vs. voiceless), the significant effects were as follows: syllabic position [onset versus coda; $F(1,37) = 37.5$, $p < 0.0001$, $\eta^2 = 0.764$], voicing [voiced versus voiceless; $F(1,37) = 119.5$, $p < 0.00001$, $\eta^2 = 0.503$], interaction of position and voicing [$F(1,37) = 70.2$, $p < 0.00001$, $\eta^2 = 0.655$]. In tests of simple main effects, a significant position effect (onset vs. coda differences) was observed for the voiced consonants [$F(1,37) = 61.8$, $p < 0.00001$, $\eta^2 = 0.626$], and significant differences between the voiced and voiceless sounds were found at the coda [$F(1,37) = 125.4$, $p < 0.00001$, $\eta^2 = 0.228$]. *Post hoc* paired t-tests showed a significant position effect for accentedness in voiced consonants [$t = 7.863$, $p < 0.00001$, $df = 37$, Cohen's $d = 1.819$] but not in the voiceless consonants [$t = -1.44$, $p = 0.136$, $df = 37$, Cohen's $d = -0.188$].
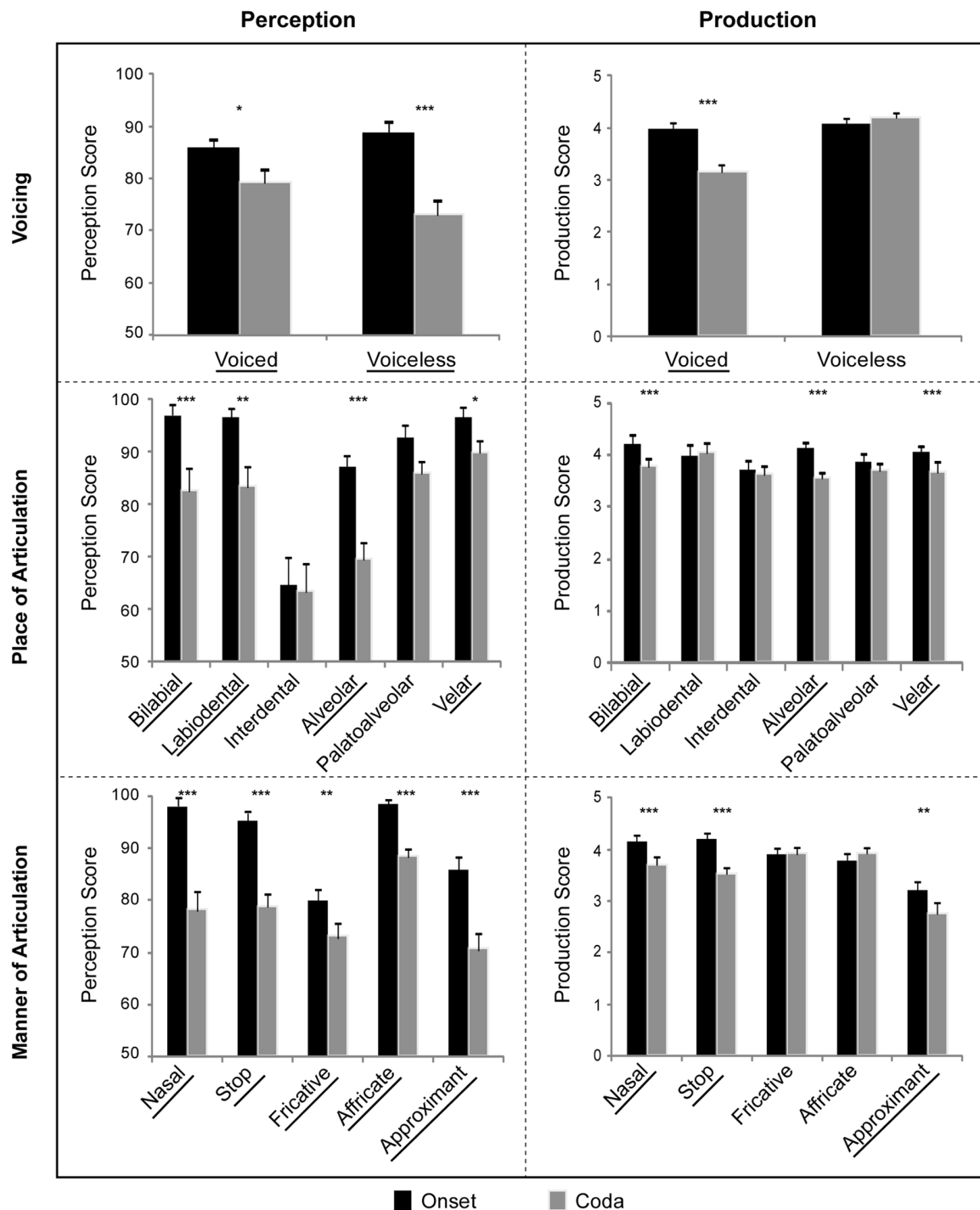
FIGURE 2 | Perception and production data of all 38 Chinese subjects separated by subcategories of voicing, place of articulation, and manner of articulation. Each consonant subcategory that showed significant positional asymmetry effect carries the underline mark. (***$p < 0.001$, **$p < 0.01$, *$p < 0.05$.)

In the six POA subcategories (bilabial, labiodental, interdental, alveolar, palatal-alveolar, and velar), significant effects in accentedness were found for the main factors of position [onset versus coda; $F(1,37) = 34.9$, $p < 0.00001$, $\eta^2 = 0.485$] and place [$F(5,185) = 10.2$, $p < 0.0001$, $\eta^2 = 0.217$; Greenhouse-Geisser corrected]. There was also an interaction between position and place [$F(5,185) = 10.4$, $p < 0.0001$, $\eta^2 = 0.22$; Greenhouse-Geisser corrected]. Simple main effects tests showed significant differences among the POA subcategories at both the onset [$F(5,185) = 6.4$, $p < 0.001$, $\eta^2 = 0.148$] and coda [$F(5,185) = 12.6$,

$p < 0.0001$, $\eta^2 = 0.253$]. Significant position effects were observed for the bilabial [$F(1,37) = 19.8$, $p < 0.0001$, $\eta^2 = 0.348$], interdental [$F(1,37) = 8.8$, $p < 0.01$, $\eta^2 = 0.191$], alveolar [$F(1,37) = 47.8$, $p < 0.00001$, $\eta^2 = 0.563$], and velar [$F(1,37) = 45.7$, $p < 0.00001$, $\eta^2 = 0.553$] sounds. *Post hoc* paired *t*-tests were performed for each POA subcategory, which confirmed positional asymmetry in the same four POA subcategories: bilabial [$t = 4.447$, $p < 0.0001$, $df = 37$, Cohen's $d = 0.848$], interdental [$t = 2.959$, $p < 0.01$, $df = 37$, Cohen's $d = 0.463$], alveolar [$t = 6.911$, $p < 0.00001$, $df = 37$, Cohen's $d = 1.416$], and velar [$t = 6.761$, $p < 0.00001$, $df = 37$, Cohen's $d = 1.593$]. The two exceptions were labiodental [$t = -0.98$, $p = 0.3$, $df = 37$, Cohen's $d = -0.221$], and palatoalveolar [$t = 0.513$, $p = 0.6$, $df = 37$, Cohen's $d = 0.099$].

In the five MOA subcategories (nasal, stop, fricative, affricate, and approximant), the significant effects in accentedness were as follows: position [onset vs. coda; $F(1,37) = 64.9$, $p < 0.000001$, $\eta^2 = 0.637$], manner [$F(4,148) = 50.1$, $p < 0.00001$, $\eta^2 = 0.575$; Greenhouse-Geisser corrected], interaction of position and manner [$F(4,148) = 5.9$, $p < 0.01$, $\eta^2 = 0.138$; Greenhouse-Geisser corrected]. Simple main effects tests showed significant differences among the MOA subcategories at both the onset [$F(4,148) = 39.1$, $p < 0.00001$, $\eta^2 = 0.514$] and coda [$F(4,148) = 12.6$, $p < 0.0001$, $\eta^2 = 0.431$]. In addition, significant position effects were observed for the nasals [$F(1,37) = 21.8$, $p < 0.0001$, $\eta^2 = 0.371$], stops [$F(1,37) = 80.5$, $p < 0.00001$, $\eta^2 = 0.685$], and approximants [$F(1,37) = 9.4$, $p < 0.01$, $\eta^2 = 0.202$]. *Post hoc* paired *t*-tests confirmed significant syllabic position effects for the nasals [$t = 4.671$, $p < 0.0001$, $df = 37$, Cohen's $d = 0.986$], stops [$t = 8.97$, $p < 0.00001$, $df = 37$, Cohen's $d = 1.984$], and approximants [$t = 3.06$, $p < 0.01$, $df = 37$, Cohen's $d = 0.591$]. There were two exceptions; namely, fricatives [$t = 1.305$, $p = 0.2$, $df = 37$, Cohen's $d = 0.235$], and affricates [$t = 1.374$, $p = 0.1$, $df = 37$, Cohen's $d = 0.241$].

Correlations between Perception and Production

Pearson's correlation results for the grand mean scores of the Chinese subjects showed significant positive links between perception and production in the syllable-initial position [Pearson's $r = 0.4$; $p < 0.05$] (**Figure 3A**). In contrast, no significant correlational effect was observed between perception and production for the syllable-final position [Pearson's $r = 0.12$; $p = 0.5$] (**Figure 3B**).

Detailed correlational results for each of the voicing, POA, and MOA subcategories were summarized in **Table 2**. For all the voicing and MOA subcategories, significant positive correlations were observed only in the onset but not in the coda. It is noteworthy that contrary to our prediction, the nasal sound category also showed positional asymmetry in perception and production as well as in the correlational analysis. Exceptions were also found in the POA subcategories. In particular, the labiodental and interdental sounds did not show significant correlations in the onset; and contrary to our hypothesis, the palatoalveolar sounds showed significant correlation in the coda.

Statistical Results Concerning Dialectal Influences

Multivariate analysis of variance analyses showed that none of the main effects or interaction effects concerning the dialect factor reached the level of statistical significance ($p < 0.05$) in perception or production (see Tables S1 and S2 in Supplementary Material for details). In the production data for MOA, the main effect of dialect was at the very edge of significance [$F(1,36) = 4.104$, $p = 0.050$]. In addition, in the production data for POA, there was one interaction effect (syllabic position * dialect) approaching the level of significance [$F(1,36) = 3.717$, $p = 0.062$].

Further analysis by removing the eight subjects who grew up in the Wu and Min dialect regions essentially replicated the results reported above (see Supplementary Tables S3–S5). The perception data did not appear to be much affected by removing the eight subjects (Supplementary Table S3). In the production data (Supplementary Table S4), however, there was a significant position effect for the interdentals [$F(1,29) = 4.242$, $p < 0.05$], which was not observed in the statistical results with all the 38 subjects (**Figure 2**). In the correlational analysis (Supplementary Table S5), the MOA affricate subcategory failed to reach the level of significance [$r = 0.357$, $p = 0.053$, $df = 29$] at the syllable onset, but it was a very close miss.

DISCUSSION

Positional Asymmetry in L2 Perception

The perception data in terms of voicing, POA, and MOA showed a strong effect of position on perceptual performance at the segmental level in the ESL learners but not in the native English speakers. Overall, the Mandarin Chinese speakers performed significantly better in consonantal perception of English phonemes in the syllable onset position as compared with the coda position for all the subcategories in terms of voicing and MOAs. There were several exceptions in terms of POA subcategories in the perception data. These results are consistent with the positional asymmetry effect reported earlier on stop consonants (Flege, 1989), showing that syllable structure can have a significant impact on phonetic perception (Eckman, 1991; Carlisle, 2001). The ceiling-level performance observed for the control group reflected the relative ease of the perception task for native English speakers, which was aided by the use of a limited number of common words in the CVC context and clearly articulated exemplars as stimuli.

The structural difference at the phonemic and syllabic levels between L1 (Chinese) and L2 (English) jointly showed a direct impact on the perceptual accuracy of individual L2 consonants. Interestingly, we observed a lack of balance among the subcategories of each consonantal classification. The main effects and significant interaction of position and voicing indicate that the Chinese ESL learners' lack of integration of acoustic cues in the vocalic context for perceiving English stop sounds not only depends on syllabic position but also differs between the voiced and voiceless sounds. For instance, syllabic
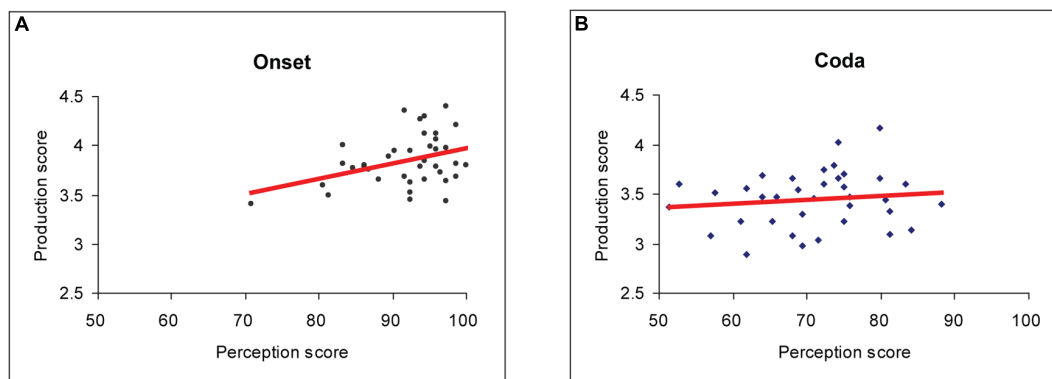
FIGURE 3 | Pearson's correlation analysis results for pooled perception and production data from all 38 Chinese subjects. (A) Significant correlation was found at the syllable-initial position (Pearson's $r = 0.37$; $p < 0.05$; Bonferroni-corrected). **(B)** No significant correlation was found at the syllable-final position.

TABLE 2 | Pearson's correlation analysis results for the perception and production scores in the onset and coda positions from all 38 Chinese subjects, respectively, for the voicing, place of Articulation (POA), and manner of articulation (MOA) subcategories.

Perception-Production Correlation		Onset	Coda
Voicing	Voiced	$r = 0.325, p = 0.046*$	$r = 0.274, p = 0.096$
	Voiceless	$r = 0.439, p = 0.006**$	$r = 0.072, p = 0.668$
POA	Bilabial	$r = 0.322, p = 0.048*$	$r = 0.283, p = 0.085$
	Labiodental	$r = 0.202, p = 0.076$	$r = 0.117, p = 0.483$
	Interdental	$r = 0.243, p = 0.142$	$r = -0.029, p = 0.861$
	Alveolar	$r = 0.492, p = 0.002**$	$r = -0.093, p = 0.577$
	Palatoalveolar	$r = 0.366, p = 0.024*$	$r = 0.322, p = 0.049*$
	Velar	$r = 0.359, p = 0.027*$	$r = 0.309, p = 0.059$
MOA	Nasal	$r = 0.482, p = 0.002**$	$r = -0.093, p = 0.577$
	Stop	$r = 0.346, p = 0.033*$	$r = 0.205, p = 0.216$
	Fricative	$r = 0.362, p = 0.026*$	$r = 0.068, p = 0.685$
	Affricate	$r = 0.328, p = 0.045*$	$r = 0.283, p = 0.086$
	Approximant	$r = 0.406, p = 0.012*$	$r = 0.179, p = 0.284$

*Significant correlations are bold-underlined. (**p < 0.01; *p < 0.05.)*

position showed a stronger effect on perceptual accuracy of voiceless consonants than voiced consonants. This result is not unexpected, suggesting that both syllable structure universals and language-specific factors are at play here (Carlisle, 2001). As the Chinese stops, fricatives, and affricates are voiceless except for /z̩/ (a voiced retroflex fricative), the difference in the voiceless vs. voiced sounds could be due to disparity in transferring L1 experience with the voicing distinction to L2 learning. Due to the highly restrictive syllable structure of Mandarin, Mandarin speakers would have more perception and production experience with voiceless consonants in the onset position than voiced consonants, which could partly explain why their perception for the voiceless sounds showed larger onset-coda difference than the voiced sounds. Previous studies showed that Chinese ESL learners had difficulty in the use of vowel lengthening for word-final voiced stops in both perception and production (Flege et al., 1992; Bent et al., 2008).

The POA perception data suggest that L2 phonetic learning difficulties exist at multiple levels, among which structural

differences at the syllable level may become a subordinate factor for certain sounds. In particular, the interdental and palatoalveolar sounds did not show a significant effect of syllabic position. As the palatoalveolar sounds happen to be fricatives that are relatively long in duration, the positional asymmetry effect could be partly offset by the presence of long-duration acoustic cues from the preceding vowels that may aid the perception of syllable-final fricative/affricate sounds (Dorman et al., 1980; Jongman et al., 2000; Broersma, 2010). It was previously demonstrated that unlike American English listeners, Mandarin-speaking subjects perceptually weighted the frication noise more heavily than formant transitions and were particularly sensitive to fricative contrasts in the VC context which was phonotactically illegal in their native language (Chiu and Babel, 2010). In agreement with the PAM (Best, 1995), the L1–L2 phonemic inventory differences may also play a major role here. The interdental and palatoalveolar consonants in English do not occur in Mandarin Chinese. Moreover, the interdental fricatives, which were known to be the weakest in terms of acoustic energy

in all English consonants, showed the lowest perceptual accuracy. These factors could jointly explain why the fricatives showed the lowest perceptual accuracy among the subcategories of manner of articulation. Further analysis of misperceptions indicated that the interdental sounds were frequently identified as alveolar sounds. This is not surprising as the phonetic learning models, SLM, PAM, and NLNC, would all predict perceptual difficulty with the interdental fricatives for Chinese ESL learners. But the existing models are largely based on primary data of the CV structure, which cannot fully predict or explain the interaction effects as reflected in the presence or absence of the positional asymmetry for particular classes of sounds in L2 perception. For instance, although the Chinese language allows syllable-final nasal contrasts, the Chinese subjects' perception of English nasals showed a significant positional asymmetry, which seems to support the universal structural preference account (See further discussions below).

Positional Asymmetry in the Production Data

Consistent with the syllable structure universals theory (Carlisle, 2001), the Mandarin Chinese speakers showed significantly better scores (i.e., lower accentedness) for the production of English consonants in the syllable-initial position than in the syllable-final position. The positional asymmetry echoes previous intelligibility data of Mandarin speakers' production of English (Rogers and Dalby, 2005; Bent et al., 2007), indicating that Chinese speakers are more likely to produce not only speech errors but also stronger degrees of accentedness in the syllable-final position than in the syllable-initial position.

The significant interaction effects between position and consonant categories indicated that there were differences in the production of various consonantal classes. We observed some discrepancies between perception and production. First, the positional asymmetry effect in terms of voicing was found only for the voiced consonants in production. The overall production of voiceless English consonants by Mandarin Chinese speakers was rated to be more native-like than the voiced ones, reflecting the fact the most of the Chinese consonants are voiceless sounds. Rogers and Dalby (2005) reported a similar production error pattern for word-final voiced obstruents. Second, there was no positional asymmetry effect in the production of labiodentals, interdentals and palatoalveolar sounds. These sounds except for the labiodental /f/ do not exist in Chinese and are typologically rare and even hard for the normal L1 learners, which would fall into the exceptional cases as described in the SLM. Third, the MOA subcategories showed lack of the positional effect for fricatives and affricates. Here syllable structure universals may become subordinate in comparison with the native language interferences and acoustic/articulatory efforts needed for producing these sounds: first, Mandarin Chinese has a richer set of palatoalveolar fricatives and affricates than English, and second, the English interdental and palatoaveolar fricatives and affricates are not in the Chinese phonemic inventory (Ladefoged and Wu, 1984; Duanmu, 2000). As in the case of fricative perception, the amount of duration and effort needed for

articulating these fricatives may also play an important role (See Keating et al., 1999 for a discussion about the lack of positional asymmetry in fricative articulation).

One theoretical challenge with the current data is to explain why L1 syllabic structure does not appear to influence L2 production as much as it does in the perceptual domain. Part of the answer may lie in the different scales of measurement we adopted for our study. The perceptual data used a percentage scale of 0–100 whereas the production data were in a much reduced range of 1–5 with no absolute zero involved. Another factor to consider is the relative strength of links between perception and production. The correlation analysis results indicate that there was hardly any significant correlation between perception and production at the syllable-final position. In the present design, the Chinese subjects were instructed to produce the words in isolation, which could have led to increased conscious efforts to articulate each sound in a word regardless of the position. As discussed earlier, phonemic inventory differences and allophonic variations likely play an important role here, too. For example, the English interdental fricatives were the most difficult for the subjects as Mandarin Chinese does not use the interdental POA. The interdentals were so difficult for the Mandarin ESL learners that syllabic position did not seem to matter - their performance for the interdentals was equally poor in the onset and coda for both perception and production. There are also other factors to be taken into account. For instance, successful L2 perceptual training and the perception-induced improvement in production are shown to be dependent on the variability of the input that require multiple talkers and multiple phonetic contexts (Pisoni and Lively, 1995; Zhang and Cheng, 2011). Corpus analysis shows that syllabic constituents in onset and coda positions can be highly predictable as nearly 75% of coda consonants in English are alveolars (Greenberg, 2005). There could be insufficient input/learning for some of the non-alveolar sounds at the syllable-final position, which would then fall into the first exceptional scenario as specified by the SLM. As we could not retrospectively collect data and analyze the input or prior perception/production training that our participants (English majors in a Chinese college) received, our statements about the input quality and quantity remain purely speculative here. Future phonetic training studies can be designed to control and characterize L2 input in terms of combinatory and distributional probabilities in determining the amount of exposure to syllable-initial and syllable-final consonants in different phonetic contexts, which may help explain some of the disparities in learning outcomes that we observed between the consonantal classes.

Correlations between Perception and Production and Exceptional Cases

Despite the existence of large variances, overall perception and production scores were positively correlated in the syllable-initial position but not in the syllable-final position, which appear to conform to the universal syllable structure pattern in favor of CV over VC. Our data are consistent with previous studies that show a lack of direct correspondence between L2 speech perception

and production at the syllable-final position. For instance, Flege (1989) reported that the adult Chinese subjects could identify the final stops in words like 'beat' and 'bead,' but they pronounced these sounds poorly. In our study, the rating scores of Chinese ESL learners' production on /t/ and /d/ in the initial position of a word were also better than those in the final position.

As predicted, we saw a constellation of exceptions in the individual POA subcategories. Specifically, significant positive correlations between perception and production were found for palatoalveolar sounds in both onset and coda positions. No significant correlation was found for labiodentals and interdentals in either onset or coda. These exceptions cannot be solely explained with either syllable structure universals or the L1 syllable structure interference account. Given the range of scores, the existence of these exceptions (**Figure 2** and **Table 2**; also see supplementary data) is not due to a lack of variance in either perception or production measures. One technical issue concerning the investigation of the perception-production relationship lies in the lack of commensurability in methodology in terms of experimental protocols and measures.

Dissociation of speech production from perception is well acknowledged in the brain research literature. In studies with brain-lesioned patients, speech perception, and production can be selectively impaired without apparent damage on the other (Praamstra et al., 1991; Dronkers, 1996), which applies to not only to speech but vocal skills in general (including singing; see Hutchins and Moreno, 2013 for a discussion on the Dual Linked Representation model.). Brain imaging studies indicate that white matter morphometric measures and diffusion tensor imaging measures in the left and right auditory cortex as well as other language-related regions (such as the left inferior parietal cortices, the left and right insula) show a dissociation between perception and production in learning foreign speech sounds (Golestani and Pallier, 2007; Golestani et al., 2007).

Consistent with the brain research findings, behavioral data also indicate that L2 perceptual learning is not necessarily paralleled with equivalent gains in the production domain or vice versa. While perception-training-induced improvement in production can be sustained in the long term, it does not necessarily show alignment of proportional gains in the two domains (Bradlow et al., 1997). A number of studies have reported that L2 production may surpass perception (Goto, 1971; Sheldon and Strange, 1982; Flege and Eefting, 1987; Underbakke, 1993; Bohn and Flege, 1996; Beach et al., 2001; Kluge et al., 2007; de Jong et al., 2009a,b; Hattori and Iverson, 2009), which is not in line with predictions of existing theories or models including SLM. The speech learning/acquisition trajectory is most likely non-linear (Zhang and Wang, 2007). Perceptual skills may improve faster than productive skills or vice versa, which varies in individuals that differ in L2 proficiency, age of acquisition, and length of exposure or residence. For instance, it is possible that production might precede perception as a result of production training with explicit explanations of the target articulatory gestures and movements and rigorous exercises with feedback. In this regard, it can be technically difficult to assess measurements and the effects of learning in the two domains

in a balanced way using one time measure from a relatively homogenous subject sample.

Positional Asymmetry in the Special Case of Nasal Sounds

Although the nasal contrast, /n/ and /ŋ/, is permitted in the syllable coda in Mandarin Chinese, there were strong positional effects in both perception and production for the /n/ sound. This pattern is contradictory to our prediction. Moreover, consistent with previous findings (Rogers and Dalby, 2005), we found that the coda /n/ and /ŋ/ sounds in English were a difficult contrast for Chinese ESL learners. The current data showed that the English /n/ and /ŋ/ sounds in coda were not consistently assimilated into the Chinese /n- ŋ/ contrast in either perception or production. Similar results were previously reported in adult Japanese ESL learners, who also showed considerable difficulty in distinguishing /ŋ/ from /n/ syllable-finally despite the fact that the Japanese language allows nasal coda (Aoyama, 2003).

Acoustic analysis of Chinese words and English loanwords suggests that Chinese coda nasals could be treated as part of the syllable nucleus rather than a consonantal coda (Chen, 2000; Hsieh et al., 2009; see Duanmu, 2000 for a different theoretical view). In this perspective of syllabification, the positional asymmetry effects in perception, production, and the perception-production correlation for the nasal sounds are consistent with the overall patterns across the consonants.

Positional Asymmetry: A Universal Phonological Constraint?

The perception and production data in this study indicate that in addition to acoustic, phonetic, and phonemic similarities/differences in the set of speech sounds between L1 and L2, there is a strong influence of the L1 syllabic structure in L2 phonetic learning. Could the observed positional asymmetry simply be a reflection of language-universal preference of the CV structure (Tarone, 1978; Singh, 1985; Beckman, 1998; MacNeilage, 1998; Lin, 2003), which is independent of possible interference from the L1 syllable structure? For instance, Golestani et al. (2007) found that when English and Japanese listeners were asked to identify POA for the English consonants, both subject groups performed equally well in the CV context and both had substantially more errors in the VC context. Other studies have reported similar results in adults (Redford and Diehl, 1999; Kochetov, 2004) as well as in infants (Jusczyk et al., 1999; Zamuner, 2006). In production models, consonants in the CV structure are considered to have an advantage over those in VC in terms of accessibility, stability, and learnability (Nam et al., 2009). It is generally agreed that consonants in the syllable onset tend to be produced with greater articulatory effort and precision and longer duration than those in coda (Greenberg, 2005). According to the Positional Faithfulness Theory (Beckman, 1998), the onset-coda differences may reflect a subset of phonological asymmetries and typological optimization within and across languages due to universal positional faithfulness constraints that are driven by perceptual or psycholinguistic prominence/salience in favor of the onset

position being the privileged and perceptually strong position. Furthermore, auditory neurophysiology data from animals and humans also support the dominance of onset response over the offset response (Greenberg, 2005).

However, the lack of the positional asymmetry in the perception data from the native speakers of English did not support a straight language-universal interpretation. Phonological theories arguing for the language-universal preference of CV would have difficulty in explaining the exceptional cases as discussed above in both perception and production domains without considering the acoustic, perceptual, articulatory, and co-articulatory details of each sound category in the syllable onset and coda positions and how the interphonology of L1 and L2 accommodates the similarities and differences at the syllable level. We acknowledge that our selection of CVC word stimuli that were clearly articulated in both syllable-initial and syllable-final positions without using a carrier phrase could have increased the predictability of target sounds and eliminated some of the onset-coda articulatory and acoustic differences. For instance, voiced English fricatives in natural connected speech tend to be devoiced in coda positions, but it was not the case in our exemplar stimuli produced in isolation. Developmental data demonstrate that when the onset and coda consonants are clearly articulated, there may be no inherent asymmetry in perceptibility or discriminability for CV and VC syllables for an infant listener (Jusczyk, 1977; Nazzi and Bertoncini, 2009; Swingley, 2009). When careful acoustic manipulation is implemented and experiential factors are controlled, psychoacoustic, and neurophysiological studies suggest that offset coding and the detection of consonants in coda position can be as effective and precise as onset coding (e.g., Pols and Schouten, 1978; Qin et al., 2007; Baltzell and Billings, 2014).

Native Language Interference: Dialectal Influences

As native speakers of Chinese may represent geographically and linguistically diverse regions with distinct local dialects, it is important to control for dialectal influences in speech research studies. We selected our subject sample by excluding those who spoke Putonghua with an accent. But our sample included eight subjects who are early bilingual speakers of both Putonghua and local dialects that are very different from Putonghua. Statistical tests that specifically examined and removed the dialectal (or rather bidialectal) influences provided verification of the perceptual test results (see **Figure 2** and Table S3 in Supplementary Material). In the production data, there were some minor differences involving the dialect factor that approached the level of statistical significance, which also slightly affected the perception-production correlational test results when those eight subjects were removed in the analysis (see **Table 2** vs. Table S5 in Supplementary Material). As our subject sample did not include those who acquired Mandarin Chinese late and spoke Mandarin Chinese with a noticeable local accent, further studies can probe how dialectal influences (in particular, early vs. late Chinese bidialectalism) factor into the effects of

native language interferences in perception and production of the second language.

Limitations and Alternative Explanations

One methodological limitation of the current study is the use of natural speech stimuli in a CVC context instead of non-sense syllables in the CV and VC formats. Speech research has well established that natural speech contains many redundant cues to segment identity. Both internal cues (heard during the sound) and external cues (heard on the neighboring sounds/syllables) contribute important perceptual information, and language experience may fundamentally change perceptual weighting of such cues (e.g., Cooper et al., 1952; Polka and Strange, 1985; Keating et al., 1999; Zhang et al., 2009). An alternative possible explanation is that the positional effects reflect asymmetry in learning allophonic variations for the CV and VC contexts. However, in the present design, the consonant clusters were not included, and the vocalic influences on segmental identities were not systematically controlled and examined. A number of studies have shown that native speakers of English rely on different acoustic cues and the integration of adjacent consonant and vowel for perceiving the same consonant at syllable-initial, intervocalic, and syllable-final positions (Hillenbrand et al., 1984; Gordon, 1989; Samuel, 1989; Whalen, 1989; Jongman et al., 1992; Pickett et al., 1995; Decoene, 1997; Sinnott et al., 1998; Kochetov, 2004; Nittrouer, 2004). Presumably, the L2 learners would also need to acquire the same set of detailed phonetic knowledge for all the allophonic variations (including consonant cluster contexts) in order to reach native-like performance in perception and production.

A second limitation of the current study is that the results were obtained from English major students in one of the top universities with a highly selective admission rate in China. As perception and production may proceed at different time courses in L2 acquisition, it would be necessary to apply well-defined subject selection criteria for a proper investigation. While efforts were made to include a normally distributed sample size in terms of L2 ability using the standardized TEM-4 scores, the Mandarin-speaking ESL population would be expected to exhibit a much wider range of English proficiency. Further work would be needed to examine whether similar patterns of positional asymmetry in perception and production at the syllable level can be obtained from Mandarin-speaking subjects with lower and higher English proficiency levels as assessed by CET and TEM tests (Zou, 2003; Cheng, 2008).

A third limitation is that the reported results were pooled across consonants for the major classes of voicing, POA, and manner of articulation. While this allows us to see the general patterns of the major consonantal classes, refined studies at the individual sound level are needed to allow a detailed examination of the acoustic, phonetic, and phonological factors at play in the syllable-initial, intervocalic, and syllable-final positions. We currently do not have an exact model for explaining all the exceptions we reported. Existing theories and models including SLM, PAM, NLNC, the Automatic Selective Perception model (Strange, 2011), and the recently revised Second Language Linguistic Perception model (L2LP; Van Leussen and Escudero,

2015) tend to focus on the perceptual side. While it makes sense to argue following the seminal work by Polivanov and Trubetzkoi that foreign accent originates from "the use of language-specific perceptual strategies that are entrenched in the learner" (Escudero, 2007), it is hard to make precise predictions regarding the dissociation patterns between perception and production for the current project of investigation that covers all consonants in the VOT, POA, and MOA classes with respect to the positional asymmetry phenomenon. As such, our study remains rather descriptive and exploratory, which would require more refined research on the individual cases of exceptions to have a better understanding the interactions among the acoustic, phonetic, phonological, and semantic factors in the learning process in both domains of perception and production. In this regard, studying positional asymmetry and the relationship between perception and production in second language acquisition may also benefit from studies of the same nature in the domain of first language acquisition to have a better understanding of the physiological, articulatory, and perceptual factors as well as language-universal and language-specific properties in phonetic learning (e.g., Lin and Demuth, 2015).

Implications for ESL Education and Future Studies

In practice, ESL teachers are advised that articulatory training would be ineffective unless the problem of perception is overcome (Ur, 1996). This pedagogical practice is supported by evidence from many studies that accurate perception is a prerequisite for good production (Rochet, 1995). However, the current results suggest that only moderate links exist between perception and production and that the perception-production correlation is highly position-dependent at the syllable level. It is important that ESL programs pay special attention to the syllable structure differences between L1 and L2 and the patterns of allophonic variations in both languages to enhance L2 phonetic learning in both perception and production domains. From a pedagogical perspective, the unbalanced relationship between perception and production requires further specification for the role of talker variability in influencing the learning of different classes of speech sounds including both consonants and vowels in different syllabic contexts.

The segmental phonetic learning models such as SLM, PAM, and NLNC all posit that the perceived relationships between L1 and L2 categories play an important role in how accurately L2 segments are perceived or produced. But these models have not fully taken into account production measures such as accentedness rating and the role of syllable structure that interacts with acoustic, phonetic, and phonemic similarities between L1 and L2 speech sounds. The current data supplement these models by highlighting the role of syllable structure in learning L2 consonants. Further research efforts, including speech training, are needed to extend to the study of more complex syllable structure involving consonant clusters and the contribution of vowel contexts (e.g., Mok, 2010) and refine the multi-factor models as well as the underlying neural mechanisms that promote or limit neural plasticity in L2

acquisition (Greenberg, 2005; Zhang and Wang, 2007; Flege and MacKay, 2011). In addition to phonemic inventory, acoustic, and allophonic variations, suprasegmental phonotactic and phonological factors (including syllable-structure universals, co-articulation, assimilation, dissimilation, stress, and prosody), and the nature of linguistic input as shown in the recent development of corpus phonology (Durand et al., 2014) would all influence adult L2 learners' perception and production. It is important to combine the segmental and suprasegmental approaches to understand how the matchup of phonetic/phonological units (including syllable structure) and allophonic variations between L1 and L2 shapes the perception and production in the course of L2 acquisition. The nature of L2 exposure experience, age of acquisition, language input, and attention would have to be incorporated in a successful model to account for and improve the perceptual and productive performance of L2 learners (Best and Tyler, 2007; Flege, 2007; Guion and Pederson, 2007; Aoyama et al., 2008; Flege and MacKay, 2011). As some Chinese dialects do not have some of the Mandarin phonemic distinctions (e.g., /n/ vs. /ŋ/ and /n/ vs. /l/; Wang, 1991), it would also be interesting and important to study how dialectal background influences the positional asymmetry effect and the strength of perception-production links.

SUMMARY

The present study investigated the role of native language syllable structure in second-language phonetic perception and production. The overall scores show a strong positional asymmetry effect (i.e., better performance in syllable-initial position than in syllable-final position) on L2 consonantal perception and production across the categories of voicing, place of articulation, and manner of articulation. Moreover, a significant positive correlation between L2 perception and production was found in the syllable-initial position but not in the syllable-final position. There are also exceptions to the overall positional asymmetry pattern, suggesting that syllable structure is not the sole determinant. The results highlight the complexity of L2 speech learning and the necessity to consider syllable structure universals and native language interference by including syllable structure and allophonic variations in a multi-factor model for speech learning, which have important implications for future behavioral and brain research.

FUNDING

ACKNOWLEDGMENT

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fpsyg.2015.01801

REFERENCES

Anderson, J. I. (1983). The difficulties of English syllable structure for Chinese ESL learners. *Lang. Learn. Commun.* 2, 53–61.

Aoyama, K. (2003). Perception of syllable-initial and syllable-final nasals in English by Korean and Japanese speakers. *Second Lang. Res.* 19, 251–265. doi: 10.1191/0267658303sr222oa

Aoyama, K., Flege, J. E., Guion, S. G. S. G., Akahane-Yamada, R., and Yamada, T. (2004). Perceived phonetic dissimilarity and L2 speech learning: the case of Japanese /r/ and English /l/ and /r. *J. Phonet.* 32, 233–250. doi: 10.1016/S0095-4470(03)00036-6

Aoyama, K., Guion, S. G., Flege, J. E., Yamada, T., and Akahane-Yamada, R. (2008). The first years in an L2-speaking environment: a comparison of Japanese children and adults learning American English. *Int. Rev. Appl. Linguist. Lang. Teach.* 46, 61–90. doi: 10.1515/IRAL.2008.003

Baltzell, L. S., and Billings, C. J. (2014). Sensitivity of offset and onset cortical auditory evoked potentials to signals in noise. *Clin. Neurophysiol.* 125, 370–380. doi: 10.1016/j.clinph.2013.08.003

Beach, F. E., Burnham, D., and Kitamura, C. (2001). Bilingualism and the relationship between perception and production: greek/English bilinguals and Thai bilabial stops. *Int. J. Biling.* 5, 221–235. doi: 10.1177/1367006910050020501

Beckman, J. (1998). *Positional Faithfulness,* Doctoral dissertation, University of Massachusetts, Amherst, MA.

Beckman, M. E. (1986). *Stress and Non-Stress Accent.* Dordrecht: Foris.

Bent, T., Bradlow, A. R., and Smith, B. L. (2007). "Segmental errors in different word positions and their effects on intelligibility of non-native speech: all's well that begins well," in *Language Experience in Second Language Speech Learning: In honor of James Emil Flege.* eds M. J. Munro and O.-S. Bohn (Amsterdam: John Benjamins), 331–347.

Bent, T., Bradlow, A. R., and Smith, B. L. (2008). Production and perception of temporal patterns in native and non-native speech. *Phonetica* 65, 131–147. doi: 10.1159/000144077

Best, C. T. (1995). "A direct realist view of cross-language speech perception," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, ed. W. Strange (Baltimore, MD: York Press), 171–204.

Best, C. T., McRoberts, G. W., and Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *J. Acoust. Soc. Am.* 109, 775–794. doi: 10.1121/1.1332378

Best, C. T., and Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *J. Phonet.* 20, 305–330.

Best, C. T., and Tyler, M. D. (2007). "Nonnative and second-language speech perception: commonalities and complementarities," in *Language Experience in Second Language Speech Learning: In honor of James Emil Flege*, eds M. J. Munro and O.-S. Bohn (Philadelphia: John Benjamins), 13–34.

Bohn, O.-S., and Flege, J. E. (1996). "Perception and production of a new vowel category by adult second language learners," in *Second-Language Speech. Structure and Process*, eds A. James and J. H. Leather (New York, NY: Mouton de Gruyter), 53–73.

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., and Tohkura, Y. (1999). Training Japanese listeners to identify English /r/ and /l/: long-term retention of learning in perception and production. *Percept. Psychophys.* 61, 977–985. doi: 10.3758/BF03206911

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., and Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *J. Acoust. Soc. Am.* 101, 2299–2310. doi: 10.1121/1.418276

Broersma, M. (2010). Perception of final fricative voicing: native and nonnative listeners' use of vowel duration. *J. Acoust. Soc. Am.* 127, 1636–1644. doi: 10.1121/1.3292996

Broselow, E., and Xu, Z. (2004). Differential difficulty in the acquisition of second language phonology. *Int. J. English Stud.* 4, 135–163.

Carlisle, R. S. (2001). Syllable structure universals and second language acquisition. *Int. J. English Stud.* 1, 1–19.

Cebrian, J. (2007). Interaction of L2 phonotactics and L1 syllable structure in L2 vowel Production. *New Sounds 2007: Proceedings of the Fifth International Symposium on the Acquisition of Second Language Speech* (Florianópolis: Federal University of Santa Catarina), 107–113.

Chao, Y.-R. (1968). *A Grammar of Spoken Chinese.* Berkeley, CA: University of California Press.

Chen, M. Y. (2000). Acoustic analysis of simple vowels preceding a nasal in Standard Chinese. *J. Phonet.* 28, 43–67. doi: 10.1006/jpho.2000.0106

Cheng, B., and Zhang, Y. (2009). The relationship between speech perception and production in second language learners. *J. Acoust. Soc. Am.* 134:2754. doi: 10.1121/1.4784618

Cheng, B., and Zhang, Y. (2013). Neural plasticity in phonetic training of the /i-I/ contrast for adult Chinese speakers. *J. Acoust. Soc. Am.* 125:4245. doi: 10.1121/1.4831610

Cheng, L. (2008). The key to success: english language testing in China. *Lang. Test.* 25, 15–37. doi: 10.1177/0265532207083743

Chiu, C. and Babel, M. (2010). "Effects of syllable positions on Taiwanes Mandarin sibilant perception," *Proceedings of the International Symposium on Chinese Spoken Language Processing, ISCSLP 2010* (Tainan: IEEE), 60–64.

Cooper, F. S., Delattre, P. C., Liberman, A. M., Borst, J. M., and Gerstman, L. J. (1952). Some experiments on the perception of synthetic speech sounds. *J. Acoust. Soc. Am.* 24, 597–606. doi: 10.1121/1.1906940

Decoene, S. (1997). The representational basis of syllable categories. *Percept. Psychophys.* 59, 877–884. doi: 10.3758/BF03205505

de Jong, K., Hao, Y.-C., and Park, H. (2009a). Evidence for featural units in the acquisition of speech production skills: linguistic structure in foreign accent. *J. Phonet.* 37, 357–373. doi: 10.1016/j.wocn.2009.06.001

de Jong, K., Silbert, N., and Park, H. (2009b). Segmental generalization in second language segment identification. *Lang. Learn.* 59, 1–31. doi: 10.1111/j.1467-9922.2009.00499.x

Dorman, M. F., Raphael, L. J., and Isenberg, D. (1980). Acoustic cues for a fricative-affricate contrast in word-final position. *J. Phonet.* 8, 397–405.

Dronkers, N. (1996). A new brain region for coordinating speech articulation. *Nature* 384, 159–161. doi: 10.1038/384159a0

Duanmu, S. (2000). *The Phonology of Standard Chinese*. Oxford: Oxford University Press.

Durand, J., Gut, U., and Kristoffersen, G. (2014). *The Oxford Handbook of Corpus Phonology*. Oxford: Oxford University Press.

Eckman, F. (1981). On the naturalness of interlanguage phonological rules. *Lang. Learn.* 31, 195–216. doi: 10.1111/j.1467-1770.1981.tb01379.x

Eckman, F. R. (1991). The structural conformity hypothesis and the acquisition of consonant clusters in the interlanguage of ESL learners. *Stud. Second Lang. Acquisit.* 13, 23–41. doi: 10.1017/S0272263100009700

Eckman, F. R., Elreyes, A., and Iverson, G. K. (2003). Some principles of second language phonology. *Second Lang. Res.* 19, 169–208. doi: 10.1191/0267658303sr2190a

Escudero, P. (2007). "Second-language phonology: the role of perception," in *Phonology in Context*, ed. M. C. Pennington (London: Macmillan).

Flege, J. (1995). "Second language speech learning: theory, findings, and problems," in *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues*, ed. W. Strange (Timonium, MD: York Press), 229–273.

Flege, J. (2003). "Assessing constraints on second-language segmental production and perception," in *Phonetics and Phonology in Language Comprehension and Production, Differences and Similarities*, eds A. Meyer and N. Schiller (Berlin: Mouton de Gruyter), 319–355.

Flege, J. (2007). "Language contact in bilingualism: phonetic system interactions," in *Laboratory Phonology*, Vol. 9, eds J. Cole and J. Hualde (Berlin: Mouton de Gruyter), 353–382.

Flege, J. E. (1989). Chinese subjects' perception of the word-final English /t/-/d/ contrast: performance before and after training. *J. Acoust. Soc. Am.* 86, 1684–1697. doi: 10.1121/1.398599

Flege, J. E., and Eefting, W. (1987). Cross-language switching in stop consonant perception and production by Dutch speakers of English. *Speech Commun.* 6, 185–202. doi: 10.1016/0167-6393(87)90025-2

Flege, J., and MacKay, I. (2011). "What accounts for "age" effects on overall degree of foreign accent?," in *Achievements and Perspectives in the Acquisition of Second Language Speech: New Sounds 2010*, Vol. 2, eds M. Wrembel, M. Kul, and K. Dziubalska-Koaczyk (Bern: Peter Lang), 65–82.

Flege, J. E., Munro, M. J., and Skelton, L. (1992). Production of the word-final English /t/-/d/ contrast by native speakers of English, Mandarine, and Spanish. *J. Acoust. Soc. Am.* 92, 128–143. doi: 10.1121/1.404278

Fujimura, O., Macchi, M. J., and Streeter, L. A. (1978). Perception of stop consonants with conflicting transitional cues: a cross-linguistic study. *Lang. Speech* 21, 337–346. doi: 10.1177/002383097802100408

Golestani, N., Molko, N., Dehaene, S., Le Bihan, D., and Pallier, C. (2007). Brain structure predicts the learning of foreign speech sounds. *Cereb. Cortex* 17, 575–582. doi: 10.1093/cercor/bhk001

Golestani, N., and Pallier, C. (2007). Anatomical correlates of foreign speech sound production. *Cereb. Cortex* 17, 929–934. doi: 10.1093/cercor/bhl003

Gordon, P. C. (1989). Context effects in recognizing syllable-final /z/ and /s/ in different phrasal positions. *J. Acoust. Soc. Am.* 86, 1698–1707. doi: 10.1121/1.398600

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds "L" and "R". *Neuropsychologia* 9, 317–323. doi: 10.1016/0028-3932(71)90027-3

Greenberg, S. (2005). "A multi-tier framework for understanding spoken language," in *Listening to Speech: An Auditory Perspective*, eds S. Greenberg and Ainsworth (Hillsdale, NJ: Erlburm).

Guion, S. G., Flege, J. E., Liu, S. H., and Yeni-Komshian, G. H. (2000). Age of learning effects on the duration of sentences produced in a second language. *Appl. Psycholinguist.* 21, 205–228. doi: 10.1017/S0142716400002034

Guion, S. G., and Pederson, E. (2007). "Investigating the role of attention in phonetic learning," in *Language Experience in Second Language Speech Learning*, eds O.-S. Bohn and M. Munro (Amsterdam: John Benjamins), 57–77.

Hattori, K., and Iverson, P. (2009). English /r/-/l/ category assimilation by Japanese adults: Individual differences and the link to identification accuracy. *J. Acoust. Soc. Am.* 125, 469–479. doi: 10.1121/1.3021295

Hattori, K., and Iverson, P. (2010). "Examination of the relationship between L2 perception and production: an investigation of English /r/-/l/ perception and production by adult Japanese speakers," in *Interspeech Workshop on Second Language Studies: Acquisition, Learning, Education and Technology*, ed. M. Nacano (Tokyo: Waseda University).

Hayes-Harb, R., Smith, B. L., Bent, T., and Bradlow, A. R. (2008). The inter language speech intelligibility benefit for native speakers of Mandarin: production and perception of English word-final voicing contrasts. *J. Phonet.* 36, 664–679. doi: 10.1016/j.wocn.2008.04.002

Hillenbrand, J., Ingrisano, D. R., Smith, B. L., and Flege, J. E. (1984). Perception of the voiced-voiceless contrast in syllable-final stops. *J. Acoust. Soc. Am.* 76, 18–26. doi: 10.1121/1.391094

Hsieh, F.-F., Kenstowicz, M., and Mou, X. (2009). "Mandarin adaptations of coda nasals in English loanwords," in *Loan Phonology: Issues and Controversies*, eds A. Calabrese and L. Wetzels (Amsterdam: John Benjamins).

Hutchins, S., and Moreno, S. (2013). The Linked Dual Representation model of vocal perception and production. *Front. Psychol.* 4:825. doi: 10.3389/fpsyg.2013.00825

Imada, T., Zhang, Y., Cheour, M., Taulu, S., Ahonen, A., and Kuhl, P. K. (2006). Infant speech perception activates Broca's area: a developmental magnetoencephalography study. *Neuroreport* 17, 957–962. doi: 10.1097/01.wnr.0000223387.51704.89

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., et al. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition* 87, B47–B57. doi: 10.1016/S0010-0277(02)00198-1

Jia, G., and Aaronson, D. (2003). A longitudinal study of Chinese children and adolescents learning English in the US. *Appl. Psycholinguist.* 24, 131–161. doi: 10.1017/S0142716403000079

Johnson, J., and Newport, E. (1989). Critical period effects in second language learning: the influence of maturation state on the acquisition of English as a second language. *Cogn. Psychol.* 21, 60–99. doi: 10.1016/0010-0285(89)90003-0

Jongman, A., Sereno, J. A., Raaijmakers, M., and Lahiri, A. (1992). The phonological representation of [voice] in speech perception. *Lang. Speech* 35(Pt 1–2), 137–152.

Jongman, A., Wayland, R., and Wong, S. (2000). Acoustic characteristics of English fricatives. *J. Acoust. Soc. Am.* 108, 1252–1263. doi: 10.1121/1.1288413

Jusczyk, P. (1977). Perception of syllable-final stop consonants by 2-month-old infants. *Percept. Psychophys.* 21, 450–454. doi: 10.3758/BF03199467

Jusczyk, P. W., Goodman, M. B., and Baumann, A. (1999). Nine-month-old's attention to sound similarities in syllables. *J. Mem. Lang.* 40, 62–82. doi:10.1006/jmla.1998.2605

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

Keating, P., Wright, R., and Zhang, J. (1999). Word-level asymmetries in consonant articulation. *UCLA Work. Papers Phonet.* 97, 157–173. doi: 10.1037/a0012481

Kluge, D. C., Rauber, A. S., Reis, M. S., and Bion, R. A. H. (2007). The relationship between the perception and production of English nasal codas by Brazilian learners of English. *Paper Presented at the Proceedings of Interspeech 2007*, Antwerp.

Kochetov, A. (2004). Perception of place and secondary articulation contrasts in different syllable positions: language-particular and language-independent asymmetries. *Lang. Speech* 47, 351–382. doi: 10.1177/00238309040470040201

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nat. Rev. Neurosci.* 5, 831–843. doi: 10.1038/nrn1533

Kuhl, P. K. (2007). Is speech learning 'gated' by the social brain? *Dev. Sci.* 10, 110–120. doi: 10.1111/j.1467-7687.2007.00572.x

Kuhl, P. K., Conboy, B. T., Padden, D., Nelson, T., and Pruitt, J. (2005). Early speech perception and later language development: implications for the 'critical period'. *Lang. Learn. Dev.* 1, 237–264. doi: 10.1097/MAO.0b013e3181f1cc6a

Ladefoged, P., and Wu, Z. (1984). Places of articulation: an investigation of Pekingese fricatives and affricates. *J. Phonet.* 12, 267–278.

Lenneberg, E. (1967). *The Biological Foundations of Language*. New York, NY: Wiley.

Levy, E. S., and Law, F. F. I. I. (2010). Production of French vowels by American-English learners of French: language experience, consonantal context, and the perception-production relationship. *J. Acoust. Soc. Am.* 128, 1290–1305. doi: 10.1121/1.3466879

Lin, S., and Demuth, K. (2015). Children's acquisition of English onset and coda /l/: articulatory evidence. *J. Speech Lang. Hear Res.* 58, 13–27. doi: 10.1044/2014_JSLHR-S-14-0041

Lin, Y. H. (2003). Interphonology variability: sociolinguistic factors affecting L2 simplification strategies. *Appl. Linguist.* 24, 439–464. doi: 10.1093/applin/24.4.439

Listerri, J. (1995). "Relationships between speech production and speech perception in a second language," in *Proceedings of the XIIIth Internationsl Congress of Phonetic Sciences*, Vol. 4, eds K. Elenius and P. Branderudi (Stockholm), 92–99.

MacKay, I. R., Flege, J. E., Piske, T., and Schirru, C. (2001). Category restructuring during second-language speech acquisition. *J. Acoust. Soc. Am.* 110, 516–528. doi: 10.1121/1.1377287

MacNeilage, P. F. (1998). The frame/content theory of evolution of speech production. *Behav. Brain Sci.* 21, 499–511. doi: 10.1017/S0140525X98001265

Massaro, D. W., and Chen, T. H. (2008). The motor theory of speech perception revisited. *Psychon. Bull. Rev.* 15, 453–457. doi: 10.3758/PBR.15.2.453

Masson, M. E. J., and Loftus, G. R. (2003). Using confidence intervals for graphically based data interpretation. *Can. J. Exp. Psychol.* 57, 203–220. doi: 10.1037/h0087426

Mok, P. K. (2010). Language-specific realizations of syllable structure and vowel-to-vowel coarticulation. *J. Acoust. Soc. Am.* 128, 1346–1356. doi: 10.1121/1.3466859

Munro, M. J., and Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Lang. Learn.* 45, 73–97. doi: 10.1111/j.1467-1770.1995.tb00963.x

Munro, M. J., Derwing, T. M., and Morton, S. L. (2006). The mutual intelligibility of L2 speech. *Stud. Second Lang. Acquisit.* 28, 111–131. doi: 10.1017/S0272263106060049

Nam, H., Goldstein, L., and Saltzman, E. (2009). "Self-organization of syllable structure: a coupled oscillator model," in *Approaches to Phonological Complexity*, eds F. Pellegrino, E. Marsico, and I. Chitoran (Berlin: Mouton de Gruyter), 299–328.

Nazzi, T., and Bertoncini, J. (2009). Phonetic specificity in early lexical acquisition: new evidence from consonants in coda positions. *Lang. Speech* 52, 463–480. doi: 10.1177/0023830909336584

Nittrouer, S. (2004). The role of temporal and dynamic signal components in the perception of syllable-final stop voicing by children and adults. *J. Acoust. Soc. Am.* 115, 1777–1790. doi: 10.1121/1.1651192

Ohala, J. J. (1996). Speech perception is hearing sounds, not tongues. *J. Acoust. Soc. Am.* 99, 1718–1725. doi: 10.1121/1.414696

Pickett, J. M., Bunnell, H. T., and Revoile, S. G. (1995). Phonetics of intervocalic consonant perception: retrospect and prospect. *Phonetica* 52, 1–40. doi: 10.1159/000262027

Pisoni, D., and Lively, S. (1995). "Variability and invariance in speech perception: a new look at some old problems in perceptual learning," in *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues*, ed. W. Strange (Timonium, MD: York Press), 433–459.

Polka, L., and Strange, W. (1985). Perceptual equivalence of acoustic cues that differentiate /r/ and /l/. *J. Acoust. Soc. Am.* 78, 1187–1197. doi: 10.1121/1.392887

Pols, L. C., and Schouten, M. E. (1978). Identification of deleted consonants. *J. Acoust. Soc. Am.* 64, 1333–1337.

Praamstra, P., Hagoort, P., Maassen, B., and Crul, T. (1991). Word deafness and auditory cortical function. A case history and hypothesis. *Brain* 114, 1197–1225.

Qin, L., Chimoto, S., Sakai, M., Wang, J., and Sato, Y. (2007). Comparison between offset and onset responses of primary auditory cortex ON-OFF neurons in awake cats. *J. Neurophysiol.* 97, 3421–3431. doi: 10.1152/jn.00184.2007

Rauber, A., Escudero, P., Bion, R. A. H., and Baptista, B. O. (2005). "The interrelation of perception and production of english vowels by native speakers of brazilian portuguese," in *Proceedings of Interspeech 2005*, (Lisbon), 2913–2916.

Redford, M. A., and Diehl, R. L. (1999). The relative perceptual distinctiveness of initial and final consonants in CVC syllables. *J. Acoust. Soc. Am.* 106, 1555–1565. doi: 10.1121/1.427152

Rochet, B. L. (1995). "Perception and production of second-language speech sounds by adults," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, ed. W. Strange (Baltimore, MD: York Press), 379–410.

Rogers, C. L., and Dalby, J. (2005). Forced-choice analysis of segmental production by Chinese-accented English speakers. *J. Speech Lang. Hear. Res.* 48, 306–322. doi: 10.1044/1092-4388(2005/021)

Samuel, A. G. (1989). Insights from a failure of selective adaptation: syllable-initial and syllable-final consonants are different. *Percept. Psychophys.* 45, 485–493. doi: 10.3758/BF03208055

Sheldon, A., and Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: evidence that speech production can precede speech perception. *Appl. Psychol.* 3, 243–261. doi: 10.1017/S0142716400001417

Singh, R. (1985). Prosodic adaptation in interphonology. *Lingua* 67, 269–282. doi: 10.1016/0024-3841(85)90001-4

Sinnott, J. M., Brown, C. H., and Borneman, M. A. (1998). Effects of syllable duration on stop-glide identification in syllable-initial and syllable-final position by humans and monkeys. *Percept. Psychophys.* 60, 1032–1043. doi: 10.3758/BF03211937

Snow, C. E. (1983). "Age differences in second language acquisition: research findings and folk psychology," in *Second Language Acquisition Studies*, eds K. M. Bailey, M. H. Long, and S. Peck (MA: Newbury House), 141–150.

Stevens, J., and Zhang, Y. (2014). Brain mechanisms for processing co-speech gesture: a cross-language study of spatial demonstratives. *J. Neurolinguist.* 30, 27–47. doi: 10.1016/j.jneuroling.2014.03.003

Stockman, I. J., and Pluut, E. (1999). Segment composition as a factor in the syllabification errors of second-language speakers. *Lang. Learn.* 49, 185–209. doi: 10.1111/0023-8333.49.s1.6

Strange, W. (2011). Automatic selective perception (ASP) of first and second language speech: a working model. *J. Phonet.* 39, 456–466. doi: 10.1016/j.wocn.2010.09.001

Swingley, D. (2009). Onsets and codas in 1.5-year-olds' word recognition. *J. Mem. Lang.* 60, 252–269. doi: 10.1016/j.jml.2008.11.003

Tarone, E. (1978). "The phonology of interlanguage," in *Understanding Second and Foreign Language Learning*, ed. J. Richards (Rowley, MA: Newbury House).

Tsukada, K., Birdsong, D., Mack, M., Sung, H. Y., Bialystok, E., and Flege, J. (2004). Release bursts in English word-final voiceless stops produced by native English and Korean adults and children. *Phonetica* 61, 67–83. doi: 10.1159/000082557

Underbakke, M. (1993). Hearing the difference: improving Japanese students' pronunciation of a second language through listening. *Lang. Q.* 31, 67–89.

Ur, P. (1996). *A Course in Language Teaching: Practice and Theory*. Cambridge: Cambridge University Press.

Van Leussen, J.-W., and Escudero, P. (2015). Learning to perceive and recognize a second language: the L2LP model revised. *Front. Psychol.* 6:1000. doi: 10.3389/fpsyg.2015.01000

Wang, W. S.-Y. (1991). *Languages and Dialects of China*, Vol. 3. Berkeley: Project on Linguistic Analysis, University of California.

Wang, Y., Behne, D., Jongman, A., and Sereno, J. (2004). The role of linguistic experience in the hemispheric processing of lexical tone. *J. Appl. Psycholinguist.* 25, 449–466.

Werker, J. F., and Tees, R. C. (2005). Speech perception as a window for understanding plasticity and commitment in language systems of the brain. *Dev. Psychobiol.* 46, 233–251. doi: 10.1002/dev.20060

Whalen, D. H. (1989). Vowel and consonant judgments are not independent when cued by the same information. *Percept. Psychophys.* 46, 284–292. doi: 10.3758/BF03208093

Zamuner, T. S. (2006). Sensitivity to word-final phonotactics in 9- to 16-month-old infants. *Infancy* 10, 77–95. doi: 10.1207/s15327078in1001_5

Zhang, Y., and Cheng, B. (2011). "Brain plasticity and phonetic training for English-as-a-second-language learners," in *English as a Second Language*, ed. Alonso (Hauppauge, NY: Nova Science Publishers), 1–50.

Zhang, Y., Koerner, T., Miller, S., Grice-Patil, Z., Svec, A., Akbari, D., et al. (2011). Neural coding of formant-exaggerated speech in the infant brain. *Dev. Sci.* 14, 566–581. doi: 10.1111/j.1467-7687.2010.01004.x

Zhang, Y., Kuhl, P. K., Imada, T., Iverson, P., Pruitt, J., Stevens, E., et al. (2000). "Neural plasticity revealed in perceptual training of a Japanese adult listener to learn American /l-r/ contrast: a whole-head magnetoencephalography study," in *Proceedings of the 6th International Conference on Spoken Language Processing*, Vol. 3. (Minneapolis), 953–956.

Zhang, Y., Kuhl, P. K., Imada, T., Iverson, P., Pruitt, J., Stevens, E., et al. (2009). Neural signatures of phonetic learning in adulthood: a magnetoencephalography study. *Neuroimage* 46, 226–240. doi: 10.1016/j.neuroimage.2009.01.028

Zhang, Y., Kuhl, P. K., Imada, T., Kotani, M., and Tohkura, Y. (2005). Effects of language experience: neural commitment to language-specific auditory patterns. *Neuroimage* 26, 703–720. doi: 10.1016/j.neuroimage.2005.02.040

Zhang, Y., and Wang, Y. (2007). Neural plasticity in speech learning and acquisition. *Bilingualism* 10, 147–160. doi: 10.1017/S1366728907002908

Zheng, Y., and Cheng, L. (2008). Test review: college english test (CET) in China. *Lang. Test.* 25, 408–417. doi: 10.1177/0265532208092433

Zou, S. (2003). The alignment of language teaching syllabus and language testing: TEM8 test development and administration. *Foreign Lang. World* 98, 71–78.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.