

RESEARCH REPORT

L2 SPEECH LEARNING: PERCEPTION, PRODUCTION & TRAINING

Xinchun WANG  

Department of Linguistics – California State University, Fresno

ABSTRACT

Adult L2 learners have difficulties in perceiving and producing L2 speech sounds. In analyzing learners' L2 speech learning problems, this study provides research data from a series of studies on L2 speech perception, production, and training. Section 1 investigates how the L1 sound system influences L2 speech perception. A recent study shows that phonetic differences and distances between English and Mandarin consonants predicted the perceptual problems of Mandarin consonants by native English learners of Chinese. Section 2 explores the relationship between L2 speech perception and production and reports a subsequent study on Mandarin consonants that shows English learners of Chinese performed better in perception than production on Mandarin retroflex sounds but vice versa on palatal sounds. The lack of alignment between perception and production suggests the relationship between L2 speech perception and production is not straightforward. In Section 3, two training experiments are reported and compared to explore the effects of phonetic training on the learning of English vowel and Mandarin tone contrasts.

RESUMO

Os alunos adultos de L2 têm dificuldades em perceber e produzir sons de fala de L2. Ao analisar os problemas de aprendizagem de fala L2 dos alunos, este estudo fornece dados de pesquisa de uma série de estudos sobre percepção, produção e treinamento de fala de L2. A



OPEN ACCESS

EDITED BY

– Miguel Oliveira, Jr. (UFAL)
– René Almeida (UFS)

REVIEWED BY

– Adelaide H. P. Silva (UFPR)
– Ubiratã Alves (UFRGS)

DATES

– Received: 01/07/2020
– Accepted: 04/14/2020
– Published: 07/09/2020

HOW TO CITE

WANG, Xinchun (2020). L2 Speech Learning: Perception, Production & Training. *Cadernos de Linguística*, v. 1, n. 1, p. 01-22.

seção 1 investiga como o sistema de sons da L1 influencia a percepção da fala em L2. Um estudo recente mostra que as diferenças fonéticas e as distâncias entre consoantes em inglês e mandarim previram os problemas perceptuais das consoantes em mandarim por alunos nativos de inglês em chinês. A Seção 2 explora a relação entre percepção e produção da fala L2 e relata um estudo subsequente sobre consoantes do mandarim que mostra que os alunos de chinês (L1: inglês) tiveram melhor desempenho na percepção do que na produção dos sons retroflexos do mandarim, mas vice-versa nos sons palatais. A falta de alinhamento entre percepção e produção sugere que a relação entre percepção e produção de fala em L2 não é direta. Na Seção 3, dois experimentos de treinamento são relatados e comparados para explorar os efeitos do treinamento fonético no aprendizado de vogais em inglês e de contrastes de tom no mandarim.

KEYWORDS

Perception; Production; Phonetic Training; Training Methods.

PALAVRAS-CHAVE

Percepção; Produção; Treinamento Fonético; Métodos de Treinamento.

INTRODUCTION

There has been a general consensus among researchers over the past two decades that the goal of second language (L2) pronunciation teaching and learning is not to strive for “native-like” productions but to improve intelligibility and comprehensibility for effective communication (CELCE-MURCIA *et al.*, 2010; LEVIS, 2005; MUNRO & DERWING, 1995). The goal shift from “Nativeness” to “Intelligibility” is mainly because native-like pronunciation in L2 is unrealistic to achieve for most adult speakers nor is it necessary if both interlocutors are nonnative speakers of the target language (CELCE-MURCIA *et al.*, 2010; LEVIS, 2005). However, to improve intelligibility, L2 learners still need to comprehend and produce the target language speech sounds and prosodic features correctly for effective communication. Yet, it is well-known that adult speakers often have difficulties with the perception and production of L2 speech sounds.

Previous research has examined different factors that influence L2 speech learning. The single most important factor has been consistently found to be the learners’ age of arrival (AOA) in the target language environment, the beginning point of first-time, sudden, and massive exposure to the target language (FLEGE; MUNRO; MACKAY, 1995; MACKAY, FLEGE; IMAI, 2006; MOYER, 1999; PISKE; MACKAY; FLEGE, 2001). AOA strongly influences all aspects of L2 pronunciation: the degree of foreign accent, the accuracy in perceiving and producing L2 consonant and vowels (FLEGE, 2007).

L2 learning experience and length of residence in the target language environment have also been examined for its impact on L2 speech learning. The findings on the effect of LOR are not consistent partly because different gaps in LOR were examined that led to different results. For example, Trofimovich and Baker (2006) found that native Korean speakers with a mean of 10 years of LOR in the U.S. were rated by native English listeners to be as accented as those with three years of LOR. However, both groups were rated less accented than the inexperienced group with only three months of LOR. The results suggest that an extremely short LOR of three months in the native speaking environment is not sufficient for L2 speakers to reduce foreign accent. In contrast, Wang (2013) found native Mandarin speakers with 12 years’ length of residence (LOR) in the U.S. were rated to be accented as those with zero LOR on English sentence productions and spontaneous speech.

One problem with indexing the LOR only to L2 experience is that LOR may not truly reflect the real amount of L2 input, as learners may differ widely in the amount of use of the L2. To address this problem, in a recent study, Flege and Wayland (2019) weighted both factors to calculate the L2 “input” by multiplying learners’ years of LOR by their self-reported % use of English outside home. To examine the effect of L2 input on the learners’ discrimination and production of English consonants and vowels, 60 native Spanish

speakers who moved to the U.S. after the age of 16 formed three groups of 20 each based on their mean full-time input (LOR \times self-reported % L2 use outside home): Low input (0.2 year), Mid input (1.2) year, and High input (3.0). The results showed a null effect of increased input as there was no significant differences between the three groups on vowel production and consonant discrimination.

Learners' first language experience also plays an important role in L2 speech learning. L2 speakers' perception and production problems are closely related to their first language experience (FLEGE, 1995). The phonetic similarities and differences between L1 and L2 sound systems strongly and systematically influences L2 perception and predicts the perception difficulties (BEST, 1994; BEST; MCROBERTS; GOODELL, 2001; BEST; TYLER, 2007).

The study focuses on the influence of learners' L1 experience and the phonetic distances between L1 and L2 sound systems on L2 speech learning. A series of studies on L2 speech perception, production, and training are reported and discussed in terms of current L2 speech perception models. Different training paradigms and outcomes are compared for effective learning outcomes.

ADULT L2 LEARNERS' SPEECH PERCEPTION PROBLEMS

Adult L2 speakers' problems with the perception and production of nonnative speech sounds is closely related to their L1 experience. Evidence from infant speech perception studies suggests that infants are language-general perceivers of speech sounds at the phonetic level. At about 10 months old and later, this universal perceptual pattern undergoes a profound change due to increased experience with their ambient/first language (BEST, 1994; POLKA; BOHN, 1996; STRANGE, 1995; WERKER, 1994; WERKER; POLKA, 1993). It is the specific language experience that shapes the way infants categorize the speech sounds in a phonologically relevant way. Therefore, the shift from language-general to language-specific speech perception is viewed as attentional reorganization for speech functions, L1 speech learning, rather than as a "loss" of sensory abilities (WERKER, 1994).

Adult monolingual speakers are language-specific perceivers. One classical study on stops that manipulated VOT values along a continuum with equal steps showed that adult speakers identified stop categories according to their native language stop inventories (LISKER; ABRAMSON, 1964; 1970). Similarly, studies on L2 vowel perception using a synthesized vowel continuum also indicated that listeners labeled vowel categories according to their L1 vowel categories. Rochet (1995) found that monolingual Portuguese and English listeners showed different category boundaries while labeling /i/ and /u/ high vowels along a synthesized /i/-/u/ continuum differed in steps in F2 values. As expected, the French lis-

teners identified three categories /i/-/y/-/u/ along the same vowel continuum that matched French high vowel categories. The area that covered the French /y/ was heard as /u/ by English listeners but as /i/ by Portuguese listeners, although both English and Portuguese use the same phonetic symbols /i/ and /u/ for their high vowels.

To a large extent, the language-specific nature of adult monolinguals' speech perception underlies the difficulties adult learners face in L2 speech learning. L2 Learners' perception problems with nonnative speech sounds are well documented on consonants (BRADLOW *et al.*, 1997; GUION *et al.*, 2000; MUNRO; DERWING; THOMSON, 2015; WANG; CHEN, 2019), and on vowels (EVANS; ALSHANGITI, 2018; MUNRO; DERWING, 2008; WANG, 1997; WANG; MUNRO, 2004), as well as on lexical tones (WANG, 2006; 2008; 2013). As the perception of L2 sounds is heavily influenced by the learners' L1 sound system, the phonetic differences and distances between learners' L1 and L2 sound systems play an important role in the degree of success in L2 speech perception. In searching for the nature of such cross linguistic influence in L2 phonetic learning, researchers have come up with different L2 perception models. The following section discusses the two most influential models: the Speech Learning Model (SLM) and Perceptual Assimilation Model (PAM).

L2 SPEECH PERCEPTION MODELS

Flege's (1995, 2007) Speech Learning Model (SLM) posits that a learner's L1 and L2 sound systems interact and exist in a common phonological space. The listeners' ability to perceive the phonetic differences between an L2 sound from the nearest L1 sound or the closest L2 sound will lead to the establishment of a new phonetic category. In contrast, "equivalence classification" of an L2 sound with the nearest L1 category blocks the formation of a new phonetic category. Flege claims that learners' ability to establish new phonetic categories remains intact throughout the life span and increases with their L2 experience. Perceptual learning will eventually lead to better production, although the alignment between perception and production may be partial only (FLEGE, 1999). Therefore, the SLM is a dynamic model that emphasizes learners' L2 experiences with the target language.

The Perceptual Assimilation Model (BEST, 1994; BEST *et al.*, 2001; BEST; TYLER) places more weight on the phonetic/gestural distances between the L1 and L2 sound systems in explaining the assimilation patterns. According to the PAM, several pairwise assimilation types are possible when two non-native phones are mapped on the L2 sound system. The pair of L2 phones may be assimilated to two different L1 phones, the Two Category (TC) type. The English /p/ and /b/ assimilated to Mandarin /p^h/ and /p/ can be an example of

the TC type. The pair of L2 phones may also be assimilated to a single L1 category equally poor or well, the Single Category type (SC). The English /l/ and /r/ to the single Japanese category /r/ is an example for the SC assimilation type. The two L2 sounds can also be assimilated to a single native category but one is a better fit than the other, the Category Goodness type (CG). An example of the CG type can be the French vowels /y/ and /u/ both mapped onto English /u/, but one might be considered to be a better fit than the other. The PAM model also predicts the degree of difficulties in discriminations of L2 sounds from the most to the least: SC > CG > TC (Best *et al.*, 2001). It is important to note that the PAM model better explains the naïve listeners' L2 perception assimilation patterns as it does not take into account the L2 learning experience.

With regard to the methodologies of assessing phonetic distances between L1 and L2 speech sounds, phoneme inventory comparison is often used but not sufficient as the IPA symbols do not provide the detailed phonetic properties of sounds across languages. For example, as discussed earlier, Rochet's (1995) findings on the phonetic differences between Portuguese and English /i/ and /u/ and their distances to the French /y/ sound provided strong evidence for the inaccuracy of direct phoneme comparisons across languages. Predictions based on phonetic/acoustic properties of phones across languages may not be sufficient either as such measurement does not always capture the most crucial phonetic cues of category distinctions. Cross-linguistic speech perception, that is, having the listeners identify the target L2 sounds as their L1 categories, adopted in recent L2 speech research, can be a more reliable method (GUION *et al.*, 2000; FLEGE; WAYLAND, 2019; WANG; CHEN, 2019). In a cross-linguistic perceptual study to assess phonetic distances between Japanese and English consonants, monolingual Japanese speakers identified English consonants using Japanese consonant categories. The subsequent experiment found that phonetic distances between Japanese and English consonants, as established by the cross-linguistic direct mapping experiment, predicted discrimination patterns of English consonants by Japanese learners of English with different L2 experience (GUION *et al.*, 2000).

L2 PERCEPTION PROBLEMS: EVIDENCE FROM MANDARIN CONSONANTS

In a recent study on native English learners of Chinese as a Foreign language, (CFL), Wang and Chen (2019) adopted a cross-linguistic direct mapping method to assess the phonetic distances and differences between English and Mandarin consonants. The participants were 16 native American English speakers with some training in linguistics but with no exposure to Chinese. They identified 10 Mandarin consonants (z /ts/, c /ts^h/, s /s/, j /tɕ/, q /tɕ^h/, x /ç/, zh /tʂ/, ch /tʂ^h/, sh /ʂ/, r /ʐ/ in C+/a/ syllables) using the closest English sounds in a ten-

way forced choice task followed by a goodness rating task along a scale of 1 (poor) -7 (good). The listening tasks were performed in a classroom equipped with an internal speaker system. The test stimuli were randomized and played back three times with an inter stimulus interval (ISI) of 7 seconds for identification and goodness rating tasks. The 10 English labels used for the identifications are “cha” “sha” “sa” “ra” “ja” “za” “ta” “da” “ea” and “ða” (‘tha’ is avoided because it could not distinguish the voice contrasts). The results of the identification and rating tasks are presented in Table 1.

Mandarin Sounds	English ID	% ID	Rating	Fit Idx	Match
r /z/	/ɹ/	100	6.3	6.3	Good
sh /ʃ/	/ʃ/	100	5.8	5.8	Good
ch /tʃ ^h /	/tʃ/	90	4.9	4.4	Fair
s /s/	/s/	77	5.6	4.3	Fair
z /ts/	/s/	79	4.9	3.9	Fair
j /tɕ/	/tɕ/	81	4.4	3.6	Poor
zh /tʂ/	/tʃ/	67	4.9	3.3	Poor
q /tɕ ^h /	/tʃ/	67	3.5	2.3	Poor
c /ts ^h /	/s/	46	3.8	1.7	Poor
c /ts ^h /	/t/	38	4.5	1.7	Poor
x /ɕ/	/ʃ/	46	2.6	1.2	Poor
x /ɕ/	/z/	27	3.7	1.0	Poor

Table 1*. Mean % ID, rating scores, and fit indexes of Mandarin to English sound mapping by English listeners.
* Adopted from Wang and Chen (2019, p. 251)

The fit index (% identification x rating score) was calculated for each cross linguistically classified category that received more than 25% of identification score (see Table 1). The numbers in boldface are the Modal classifications (the highest frequency of identifications of each Mandarin consonant as the English category). As seen in Table 1, there is a range of phonetic distances between the L1 and L2 sounds based on the fit indexes (1.0-6.3). The “poor” matching categories were x /ɕ/, c /ts^h/, q /tɕ^h/, zh /tʂ/, and j /tɕ/ whose fit indexes were below the mean (3.7, s.d.=1.7). The “fair” fitting categories were ch /tʃ^h/, s /s/, and z /ts/ whose fit indexes were at the mean. The “good” matching sounds were r /z/, and sh /ʃ/ whose fit indexes were 1s.d. above the mean. (WANG; CHEN, 2019).

Figure 1 summarizes the perceptual assimilation patterns of Mandarin consonants onto the closest English categories based on the fit indexes shown in Table 1. Applying L2 perception models in explaining the assimilation patterns observed in Figure 1, the Category Goodness (CG) type of the PAM model may account for the sh /ʃ/ and x /ɕ/ to English /ʃ/ match. Their fit indexes suggest that sh /ʃ/ (5.8) is a much better fit than /ɕ/ (1.2) to Eng-

lish /j/ match. The CG type may be expanded to include the 3 to 1 category mapping shown in Figure 1: ch /tʂʰ/ (4.4), zh /tʂ/ (3.3), and q /tɕʰ/ (2.3) to English /tʃ/; and s /s/ (4.3), z /ts/ (3.9), and c /tsʰ/ (1.7) to English /s/. In all these cases, the better fit sounds (among the 3 to 1 and 2 to 1 mappings) ch /tʂʰ/, s /s/, and sh /ʃ/ received higher fit index scores than the poorer matching categories.

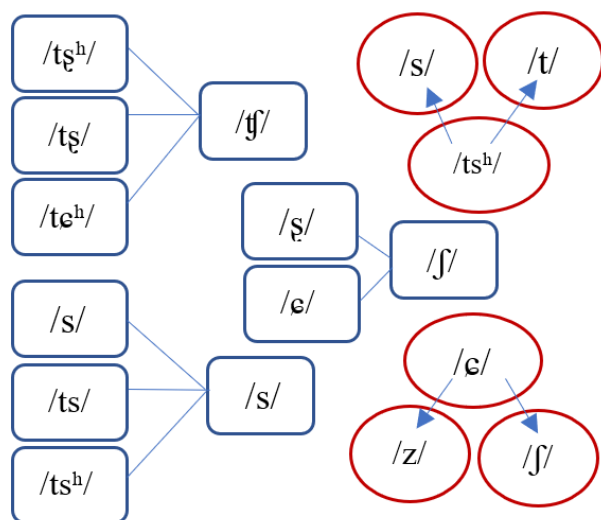


Figure 1*. Mandarin to English sound mapping patterns by English listeners. 3 to 1 and 2 to 1 mappings are in blue squares and 1 to 2 mappings are in red circles.

*Adopted from Wang and Chen (2019, p. 252).

In contrast, Mandarin c /tsʰ/ was mapped onto two English categories /s/ and /t/ with the same poor fit index score of 1.7. Similarly, x /ɕ/ was also categorized as English /tʃ/ (1.2) and /z/ (1.0) sounds with the lowest fit indexes. This one-to-two “split” mapping pattern might be considered the “reversed” Single Category type of assimilation in which the target sound is classified as two different L1 sounds.

Flege’s (2007) Speech Learning Model may also explain the assimilation patterns observed in Figure 1. For the English CFL learners, their English and Mandarin systems need to be reorganized in order to establish phonetic categories for those poor matching Mandarin consonants, especially those 3 to 1, and 2 to 1, as well as the 1 to 2 mapping sounds discussed in the above. For example, learners need to establish separate categories for c /tsʰ/, x /ɕ/, z /ts/, q /tɕʰ/ and others. On the other hand, “equivalence classification” of the SLM may be at work for the Mandarin r /z/ classified as English /ɹ/ (6.3), and sh /ʃ/ as /ʃ/ (5.8). While these sounds were the best categorized by the learners, they are phonetically very different from the L1 categories.

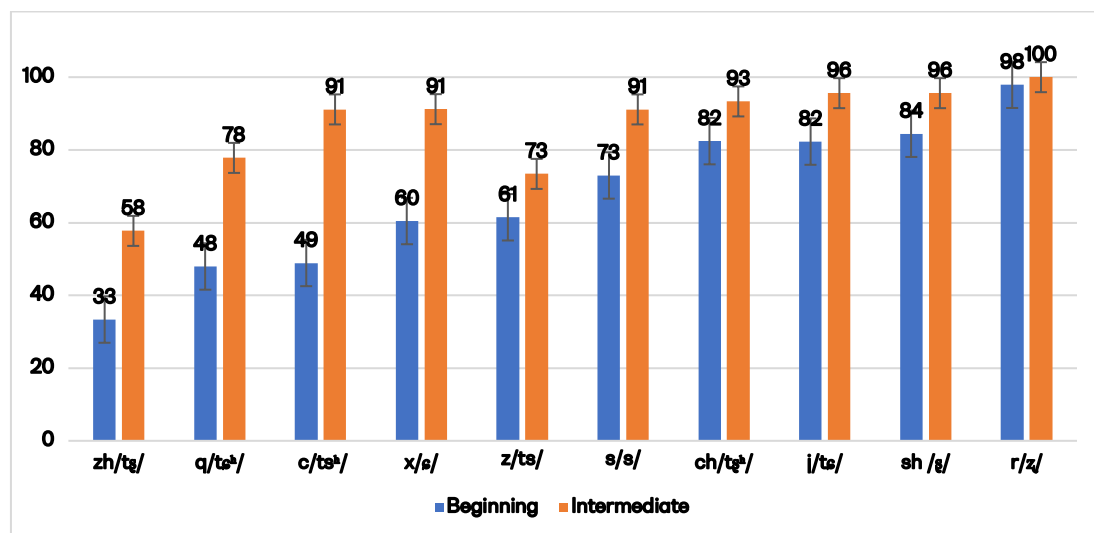


Figure 2. The % correct identifications of Mandarin consonants by English CFL learners.

To assess how these cross-linguistic assimilation patterns influence CFL learners' perceptual learning of Mandarin consonants, a subsequent experiment on English CFL learners' identifications of Mandarin consonants was carried out. A total of 47 English-speaking CFL learners with different proficiency levels at a US university participated as listeners. The beginning level group were 32 beginners (mean age = 18.9) who studied Chinese for one semester. The intermediate level group had 15 participants (mean age = 20.6) who studied Chinese for 3-5 semesters. The same 10 Mandarin consonants used in the cross-linguistic perceptual study were identified by the CFL learners in a ten-way forced choice task using the corresponding Pinyin labels of the target sounds. The listeners' mean percentage correct identifications of each Mandarin consonant are summarized in Figure 2.

Results show that zh /tʂ/, q /tɕʰ/, c /tsʰ/, and x /ɣ/ received the lowest % identification scores among the 10 sounds by the beginning level learners. These four sounds were the "poor", (also the worst) fitting categories in cross-linguistic mapping test (see Table 1). In contrast, the two best matching categories, r /z/, and sh /ʃ/, received the highest identification scores for both groups. It appears that the perceived phonetic distances between L1 and L2 consonants predicted the CFL learners' L2 Mandarin consonant perception problems for beginning level learners. It is important to note that z /ts/, which was one of the "fair" fitting categories in cross-linguistic mapping test (see Table 1), did not receive comparable scores as the other "fair" fitting categories ch /tʂʰ/ and s /s/. In particular, z /ts/ was also poorly identified by the intermediate group. One possible explanation for the learners' difficulties with Mandarin z /ts/ could be it competes with both s /s/ and c /tsʰ/ for English /s/ sound and the three-to-one assimilation pattern causes confusions.

A One-way ANOVA established significant differences between the groups: $F=17.146$, $p=.000$. A series of one-way ANOVAs revealed the differences ($p < .05$) were on zh /tʂ/, q /tʂʰ/, c /tsʰ/, x /ɕ/, and /s/. The intermediate level group performed significantly better than the beginning level group on 5 of the 10 Mandarin consonants, indicating increased L2 learning experience helped the learners in their identifications of mostly the “poor” fit categories. Therefore, the data partially support the SLM in that L2 learning experience is a predictor for learning.

1. THE RELATIONSHIP BETWEEN L2 SPEECH PERCEPTION AND PRODUCTION

Perception learning cannot be completely evaluated without examining its relationship with production, as the goal of L2 speech learning is the success in both perception and production. As discussed in the above section, the current L2 speech learning models are perceptually based. Both the PAM and SLM place much emphasis on the perceptual assimilation or dissimilation of the L2 sound categories to the L1 sounds. The question then arises: are the problems in production perceptually based? If so, will perceptual learning of L2 sounds lead to better production? Previous research on L2 speech perception and production has led to different findings. For example, native Portuguese and English speakers’ productions of French /ɣ/ as /i/ and /u/ respectively reflected their perceptual problems (ROCHET, 1995). Similarly, Mandarin speakers’ production problems with French voiced stops was related to their faulty perception of the voiced stops that do not exist in the Mandarin sound system (ROCHET, 1995). In a study on English front vowels /i ɪ eɪ ε æ/, Wang (1997) found that Mandarin speakers had problems with both the perception and production of /i ε æ/, but they performed better in perception than in production on these three vowels. In contrast, they performed better in production than in perception on English /i eɪ/ categories. The discrepancies between the perception and production of the English front vowels by Mandarin speakers suggest that learners may have used different cues or strategies in their perception and production of the L2 vowels. Flege (1999) also reported a series of studies that showed partial alignment between L2 perception and production.

To further explore the relationship between L2 speech perception and production, the author recently conducted another study on the Mandarin consonants that included both the perception and production tests. The participants were 30 beginning level English speaking CFL learners enrolled in a first semester CFL class at a US university. The data were collected at the end of the 16-week semester. Each participant first read a list of target words in a carrier sentence followed by identifying the target sounds in a forced choice task. The recordings were made in a sound booth on a MacBook Pro computer us-

ing Praat speech software. Individual perception identification tasks were then performed also on a MacBook Pro computer using Praat ExperimentMFC identification test design. The eight test stimuli in the C+/a/ syllables (z /ts/, c /ts^h/, j /tɕ/, q /tɕ^h/, x /ɕ/, zh /tʂ/, ch /tʂ^h/, sh /ʂ/) were selected for analysis in this study.

The participants' productions of the target sounds were evaluated by three native Mandarin speakers in an identification test using Praat ExperimentMFC identification test design. A reliability test assessing the interrater variability showed a high degree of agreement among the three raters. The average measures ICC was .821 with a 95% confidence interval from .788 to .849 (F (399, 798) = 5.605, p<.001). Therefore, the mean group production score for each consonant was calculated by taking the average of the three listeners' identification scores. The mean production and perception scores of the eight consonants are presented in Figure 3.

The CFL learners have difficulties with both the perception and production of the target consonants. The range of accuracy scores were 34% to 79% for perception and 26% to 84% for production. To examine the relationship between the learners' performances in perception and production of the eight consonants, Pearson Coefficients Correlation tests were performed on the percentage correct perception and production scores of the eight consonants. The results showed no significant relationship between the perception and production scores for the eight consonants. The strength of correlation ranged from none (r(30) = .000, p= 1) for j /tɕ/ to very weak (r(30) = .264, p= .159) for /ts/. The lack of correlations between the production and perception scores provided further evidence that perception does not always lead production and the relationship between perception and production of L2 consonants is not straightforward.

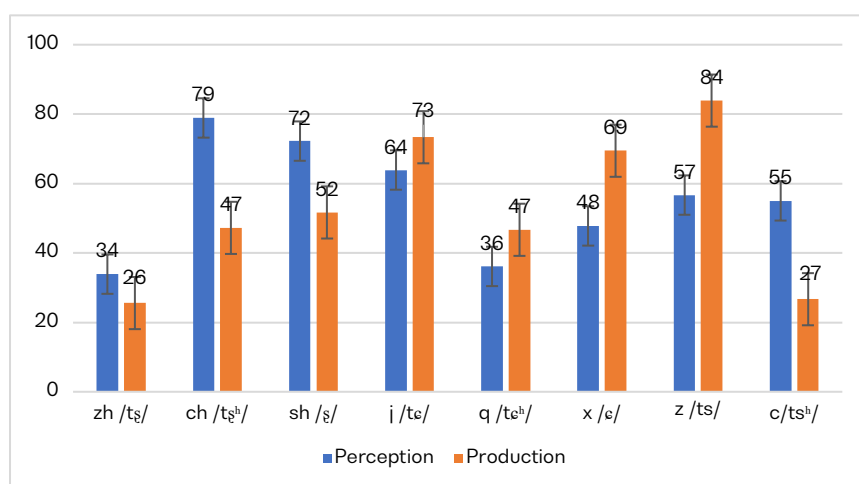


Figure 3. Mean % perception and production scores of Mandarin consonants by CFL learners (N=30)

Although both the retroflex and palatal fricatives/affricates pose problems for the English CFL learners in perception and production, the degree of difficulties appeared to be different. Overall, there is the tendency that the participants performed better on the retroflex sounds zh /tʂ/, ch /tʂʰ/, and sh /ʂ/ in perception than in production but vice versa on palatal sounds j /tɕ/, q /tɕʰ/, and x /ç/. The results on dental sounds z /ts/, and c /tʃ/ were mixed across the two domains. One possible explanation for the discrepancies in perception and production could be that different mechanisms were involved in perception and production of the sounds. Future studies need to examine the detailed confusion patterns of mis-perceived and mis-produced sounds for each consonant to explore the patterns of errors.

L2 phonetic training studies have also examined the relationship between perception and production when assessing the outcomes of the training. While there is evidence that perceptual training only led to improvement in both perception and production of L2 consonants and lexical tones (BRADLOW *et al.*, 1997; WANG, 2008; WANG; JONGMAN; SERENO, 2003), such improvement in production of L2 vowels after perceptual training was not always observed (WANG, 2002). In a meta-analysis study of 30 perception training studies on L2 segments conducted in the past 25 years, Sakai and Moorman (2018) found that perception training only led to small-sized gains in productions of the target sounds. Their subsequent statistical analysis based on 18 out of the 30 studies led to the conclusion the production gains are larger on obstruents than on sonorants and vowels. Correlation tests suggest there is a small to medium-sized but statistically nonsignificant relationship between gains in perception and production. Sakai and Moorman's (2018) study analyzed training studies on L2 segments only by excluding studies on L2 suprasegmental features. The following section presents training studies on L2 vowels and lexical tones to further explore the relationship between perception and production and to evaluate the effects of phonetic training using different training methods.

2. TRAINING AND L2 SPEECH LEARNING

There is sufficient evidence that intensive laboratory-based training helps learners to establish non-native segmental and tonal contrasts. (BRADLOW *et al.*, 1997; HARDISON, 2003; 2004; HUENSCH; TREMBLAY, 2015; JAMIESON; MOROSAN, 1986; 1989; KINGSTON, 2003; LIVELY; LOGAN; PISONI, 1993; LOGAN; LIVELY; PISONI, 1991; 1993; LOGAN; PRUITT, 1995; SAKAI; MOORMAN, 2018; WANG; 2002; 2008; 2012; 2013; WANG; MUNRO, 1999; 2004). Questions about optimal training methods for better outcomes remain to be addressed. First, what training methods are more suitable and efficient for learning which type of L2 phonetic contrasts, such as L2 vowels and tones? Second, to what ex-

tent will the effect of perceptual learning be retained? Lastly, will the training in perception lead to better production of the target L2 contrasts? To answer these questions, this section introduces and evaluates the commonly used training methods and present training studies on Mandarin lexical tones (WANG, 2008) and English vowel contrasts (WANG, 2002; WANG; MUNRO, 2004). Furthermore, the relationship between perception and production is examined when evaluating the outcomes of training in perception and its effect on production.

2.1. DISCRIMINATION VS. IDENTIFICATION TRAINING

The two most commonly used methods in perceptual training of L2 speech sounds are discrimination and identification tasks with immediate feedback. In a discrimination task, two training stimuli are presented in a sequence such as light/right or right/right, and the trainee is asked to tell whether the initial sounds they heard are the same or different. This type of discrimination task is also known as the AX (same/different) task. The other type of discrimination training task is the ABX task in which three stimuli such as light/right/light are presented in a single trial and the “odd” sound is selected.

In an identification task, the stimuli are presented one by one in a forced choice paradigm. Trainees identify each stimulus by labeling it among the choices they are given, whether it is a two-way, three-way, four-way or more forced choice task depending on how many phonetic contrasts are involved in the training. For example, in a two-way forced choice task involving the English /i/ and /ɪ/ contrast, the listeners identify each stimulus by choosing either /i/ or /ɪ/ as they hear it. Identification training on the four Mandarin lexical tones would use the four-way forced choice task in which the trainee is forced to choose one of the four tones each time a tone stimulus is presented.

The perceptual fading technique, first used by Jamieson and Morosan (1986, 1989) to train Canadian francophone adult speakers on the English /e/-/ɛ/ contrast, is a special type of identification task. In this approach, synthesized speech is normally used and trainees begin by identifying the most extreme tokens at the end points of a synthetic continuum. Gradually less clear exemplars along the continuum are presented to the trainee to widen the range of the phonetic representations of the target sounds. The fading technique focuses on the key acoustic properties that distinguish the category contrasts under training. For example, the F1 and F2 values that distinguish the English vowels /i/-/ɪ/ contrast can be manipulated with different steps so the best and most extreme exemplars are introduced first followed by those less extreme samples gradually (WANG, 2002; WANG; MUNRO, 2004).

Though both the discrimination and identification training are commonly used in training, the AX task discrimination training may not be optimal for phonetic category formation (JAMIESON; MOROSAN, 1986; 1989; LOGAN; PRUITT, 1995; WANG; MUNRO, 2004). Forced-choice identification tasks, on the other hand, have the advantage of directing learners' attention to the specific characteristics of a speech sound, the key acoustic/phonetic cue that makes it different from the other sound(s) under training (WANG; MUNRO, 2004).

2.2. THE HIGH VARIABILITY PERCEPTUAL TRAINING (HVPT)

Speaker variability is another important factor to take into consideration when designing training tasks. If trainees are exposed to a single speaker's voice during the training, they may develop sensitivity to that particular voice but not to unfamiliar voices. What the trainees learned from one speaker's phonetic contrast may not be generalized to the abstractness of phonetic contrast of the target sounds. This problem can be dealt with by adopting the High Variability Perceptual Training (HVPT) paradigm that uses multiple speakers' voices in training (WANG; MUNRO, 2004). The HVPT can also embed the target sounds in different phonetic environments so the learners are exposed to different versions of the same sounds. For example, for the English /ɪ/ and /I/ contrast, the training stimuli may appear at the initial, final, and medial positions of the syllable or word: light/right, peel/peer, borrow and below.

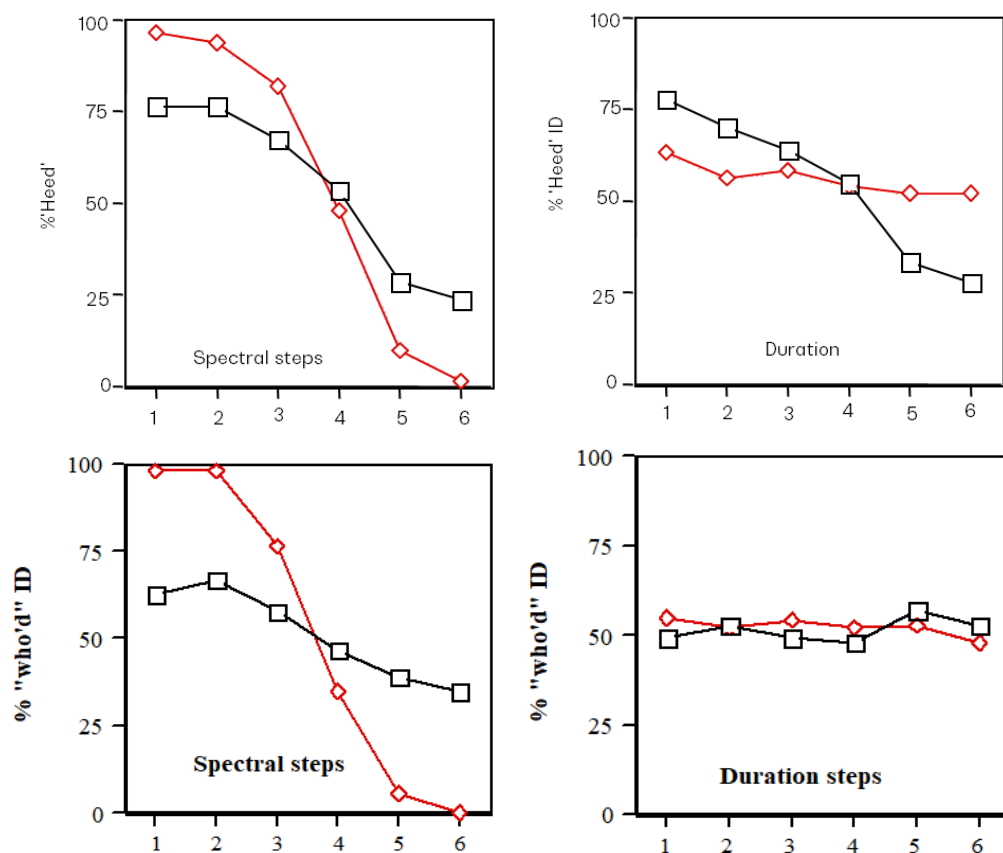
2.3. PRODUCTION TRAINING WITH AUDIO AND VISUAL INPUT

The training methods discussed above are for perceptual training only. Training in both the perception and production modes with audio and visual input and feedback has also been reported (HARDISON, 2004). Using Sona Speech II, which displays the pitch contour of the target sounds on the screen in real time, Hardison (2004) trained learners on L2 French stress patterns. Such training allows the trainees to imitate, record, and display their own productions to be compared with the target sounds and sentences on the screen. The results showed that native English speakers' productions of French stress and intonation improved significantly after receiving 3 weeks of production training.

2.4. TRAINING ON ENGLISH VOWEL CONTRASTS

Motivated by the findings of an earlier study on Mandarin speakers' problems with English vowels (1997), Wang (2002) conducted a subsequent training study on English vowel contrasts. The 21 participants (17 Mandarin and 4 Cantonese speakers) were advanced ESL speakers studying for their degrees in a Canadian university. Sixteen trainees completed 4-6 weeks of training (2-4 hours per week) on English /i/-/ɪ/, /ʊ/-/u/ and /ɛ/-/æ/ contrasts. A control group of five participants took the pre- and post tests without training.

Training began with the fading technique using synthesized vowel continua with six spectral and six duration steps. The stimuli at spectral end points (steps 1 & 6) were first used followed by those less extreme tokens at steps 2 & 5 in two-way forced choice tasks with immediate feedback. In the subsequent High Variability Perceptual Training (HVPT) task, natural stimuli of minimal pairs of the target vowels produced by four native English speakers were presented for identifications. Results of perceptual tests on synthesized vowel pairs that differed in six spectral and six duration steps are presented in Figure 4.



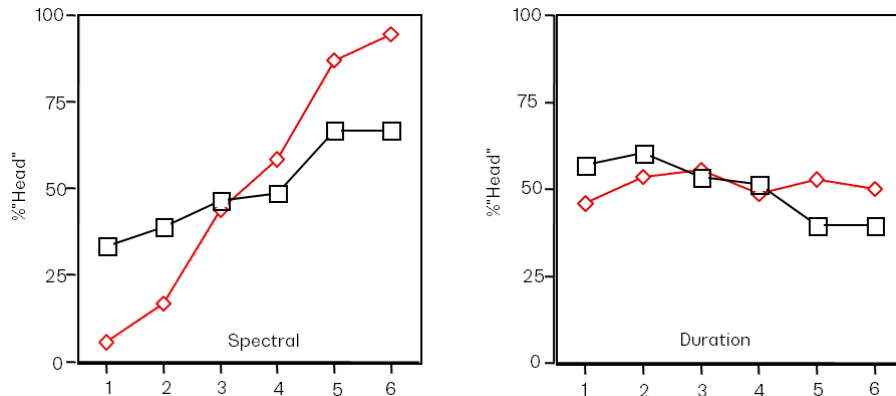


Figure 4. % identifications of “heed” (top), “who’d” (mid), and “head” (bottom) by spectral steps (left) and duration steps (right) at pre-test (black line) and at post test (red line).

At pretest, Mandarin and Cantonese speakers who had problems with the /i/-/ɪ/ contrast responded to duration cues. At post test, the reliance on duration cues was suppressed as the trainees learned to use spectral cues for the /i/-/ɪ/ contrast. Those who had problems with the /u/-/ʊ/ and /ɛ/-/æ/ pairs did not rely on duration cues but demonstrated confusions in their perceptual patterns, indicating their lack of two-category distinctions for these contrasts. Training helped them to establish separate categories for the /u/-/ʊ/ and /ɛ/-/æ/ pairs by attending to the spectral cues.

Figure 5 presents the trainees’ percentage perception scores of natural vowels (left) and their vowel production scores (right). The trainees’ identification scores for natural tokens increased significantly from pre-test to post test. Perceptual learning also generalized to new talkers and new stimuli and was retained three months later after training was completed. The trainees’ improvement in perception was not matched by the control group. There was also a small-sized gain in production of the target vowels, as judged by the native English speakers. However, the improvement in production did not reach the level of significance as compared with the control group.

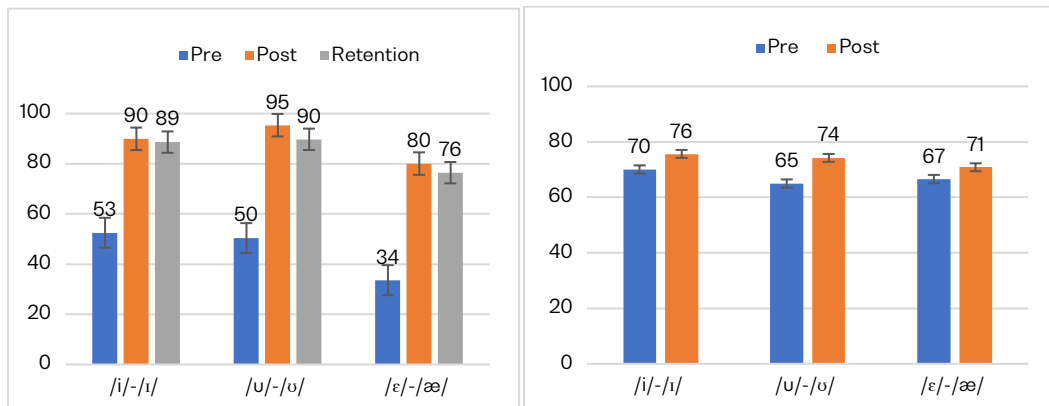


Figure 5. Trainees’ mean % perception scores of the natural vowels at pre, post, & retention test (left) & mean % production scores at pre & post test (right).

3.5. PERCEPTION VS. PRODUCTION TRAINING ON MANDARIN LEXICAL TONES

Wang (2008) used two training paradigms, audio only and audio visual to train beginning level CFL learners with different L1 experience on the four Mandarin lexical tones. The training stimuli were 160 syllables/words (40 minimal quadruplets) produced by four native Mandarin speakers. The audio only group (A Group) of 10 trainees took the four-way forced choice identification training with immediate feedback. The audio-visual group (AV Group) of eight trainees received perception and production training using Sona Speech II software which displays the pitch contours of the target tones on the screen in real time. During the training, the trainee opened and played back (through a pair of headphones) each training stimulus with real time display of the pitch contour in the top window of Screen A on the computer screen. The trainee then repeated the target tonal syllable and recorded his/her own production of the target tone by speaking into the microphone. The pitch contour of the trainee's production was instantly displayed in the bottom window of Screen B. The trainee could then compare his/her own sound with the target sound by playing them back repeatedly (auditory input). The trainee could also visually compare the two tones by overlaying the pitch contour of the target tone on that of his/her own production in different colors while alternately playing them back for auditory comparisons. All trainees completed 6-8 hours of training within three to four weeks. The trainees read and identified 40 words in pinyin at pre- and post test to provide the perception and production data and a control group (C Group) of 10 took the perception tests without the training. All three groups also took a perceptual generalization test in which they identified tones produced by new speakers.

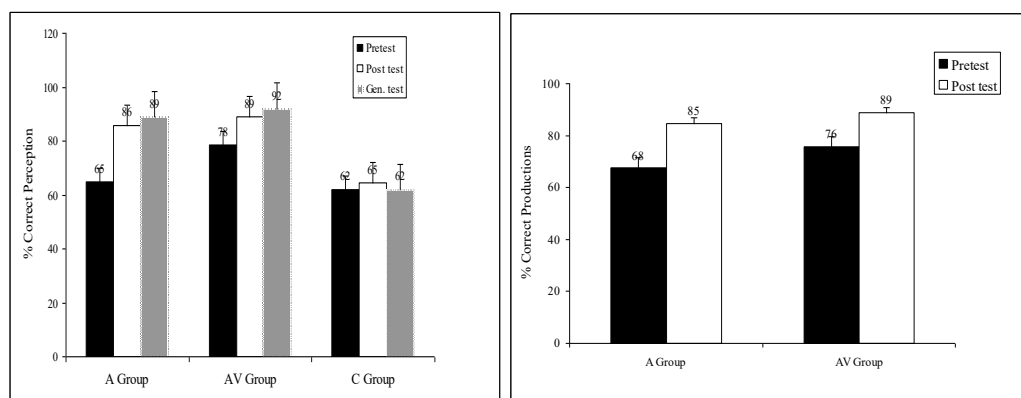


Figure 6*. Mean perception (left) and production (right) scores of Mandarin tones and standard errors for each group at pretest, post test, and generalization test.

*Adapted from Wang (2008).

The mean percentage correct identification scores of the four Mandarin tones by the training and control groups are presented in Figure 6 (left). A series of one-way ANOVAs

established significant differences between the groups at post test [$F(2,27) = 7.656$, $p = .003$] and at generalization test [$F(2,27) = 7.750$, $p = .002$] but not at pretest [$F(2,27) = 2.732$, $p = .085$]. Post hoc (Tukey HSD) pairwise comparisons ($p < .05$) revealed the differences were between the AV and the C groups, and between the A and the C groups, but not between the two training groups.

The 18 trainees' productions of the Mandarin tones at pretest and post test were mixed and evaluated by two native Mandarin speakers in four-way forced choice identification tasks. Inter-rater correlation ($r = .840$) was high for the two listeners' identification scores. The mean production scores (pooled over the two listeners) of the A Group and AV Groups at pretest and post test are presented in Figure 6 (right).

A two-way repeated measures ANOVA with Test (pre- and post) as within subject factor and Group (A and AV) as between subject factor revealed a significant effect of Test [$F(1,16) = 25.486$, $p = .000$] but no effect of Group [$F(1,16) = 2.204$, $p = .157$]. There was no Test \times Group interaction [$F(1,16) = .470$, $p = .503$] either. Therefore, both training paradigms were effective and comparable for learning Mandarin tones. One important finding of this study was that perception training only was as effective as production training with audio and visual feedback on the production Mandarin tones. HVPT training in the perception mode only led to improvement in the production of Mandarin tones. The same transfer effect was not observed on perceptual training on L2 English vowels reported in the above.

3.6. REFLECTIONS ON TRAINING PARADIGMS AND EFFECTIVENESS

Based on the English vowel training study (Wang 2002) reported in the above section, identification training with fading technique using synthesized stimuli proved to be effective to shift learners' focus from duration cues to spectral cues for English /i/-/ɪ/contrast. This technique is especially suitable for L2 vowel trainings as the "fading" can easily be carried out by manipulating the spectral and duration features that characterize the target vowel contrasts. Trainees' attentions were redirected towards spectral differences for vowel contrasts not existent in their L1 vowel system. This acquired sensitivity to vowel spectral differences needs to be transferred to perceive natural speech that reflects a range of variabilities through the subsequent HVPT training. Therefore, training with both the fading technique and HVPT paradigm was helpful for trainees to establish two category contrasts for the /u/-/ʊ/ and /ɛ/-/æ/ pairs and to use spectral cues of the /i/-/ɪ/contrast. As the perceptual gains on these vowel contrasts did not lead to significant gains in production, these perceptual training only paradigms may not be sufficient for production improvement on L2 vowel contrasts. By comparison, the same

identification training on the Mandarin tones (Wang 2008) using multiple speakers' tokens, (the HVPT paradigm), was found effective for gains in both the perception and production modes. The different outcomes of training for L2 vowel and lexical tone contrasts are not surprising as Sakai and Moorman's (2018) meta-analysis of 18 perception training studies concluded that perceptual training only led to a small-sized gains in productions of the target sounds and the production gains are larger on obstruents than on sonorants and vowels. It may require simultaneous production training in addition to the two perceptual training paradigms for better results in production of L2 vowels.

4. SUMMARY AND CONCLUSIONS

Studies on infant and adult monolingual speakers' speech perception indicate that infants are language-general perceivers of speech sounds while adult monolingual speakers are language-specific perceivers. To a large extent, the nature of the language-specific perception of speech sounds underlines adult learners' L2 speech problem. The differences and distances between L1 and L2 sound systems play an important role in L2 speech perception. Both the PAM and SLM hypothesize how the sound systems of L1 and L2 interact and how L2 categories are assimilated to the nearest L2 sounds.

The commonly used phoneme inventory comparisons with phonetic descriptions is not the optimal method for assessing phonetic distances between L1 and L2 sounds. Instead, the cross-linguistic category mapping test can be a direct measurement of L1 and L2 phonetic distances when the listeners identify L2 sounds with L1 categories. The findings of Wang and Chen's (2019) study provided evidence that native English CFL learners' perception problems with Mandarin consonants are predicted by the phonetic differences and distances between Mandarin and English consonants as assessed by monolingual English speakers in a cross-linguistic mapping test.

Evidence from training studies suggest there is a direct link between L2 speech perception and production as perceptual training only led to gains in both perception and production of L2 consonants and tones (BRADLOW *et al.*; WANG, 2008). The current findings of the lack of correlation between English CFL learners' perception and production scores on Mandarin consonants suggest the relationship between perception and production of L2 speech sounds is not straightforward.

In terms of training methods, the HVPT training on Mandarin tones led to improvement in both perception and production of tones and the results are comparable to perception with production training with audio and visual input (WANG, 2008). The same HVPT paradigm along with the fading technique on English vowel contrasts did not lead to comparable success in the production of the target vowel contrasts (WANG, 2002). These findings

suggest that different L2 phonetic contrasts may need different training methods to obtain better training results.

5. ACKNOWLEDGMENTS

The author thanks the two reviewers for their insightful comments on an earlier version of the paper. Thanks also go to all the participants for all the studies reported.

REFERENCES

- BEST, C.T. "The Emergence of Native-Language Phonological Influences In Infants: a Perceptual Assimilation Model". In: GOODMAN, J.; NUSBAUM, H. (Eds.), *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words*. Cambridge: MIT Press, 1994: 167-244.
- BEST, C. T.; MCROBERTS, G. W.; GOODELL, E. (2001). "Discrimination of Non-Native Consonant Contrasts Varying In Perceptual Assimilation to the Listener's Native Phonological System". *The Journal of the Acoustical Society of America*, v. 109(2): 775-794.
- BEST, C. T.; TYLER, M. D. (2007). "Nonnative and Second-Language Speech Perception: Commonalities and Complementarities". *Language Experience In Second Language Speech Learning*. In Honor of James Emil Flege, 1334: 1-47.
- BRADLOW, A. R.; PISONI, D. B.; AKAHANA-YAMADA, R.; TOHKURA Y. (1997). "Training Japanese Listeners to Identify English /r/ and /l/: IV. Some Effects of Perceptual Learning on Speech Production". *Journal of the Acoustical Society of America*, 101: 2299-2310.
- CELCE-MURCIA, M.; BRINTON, D. M.; GOODWIN, J. M. *Teaching Pronunciation Hardback with Audio CDs (2): a Course Book and Reference Guide*. Cambridge University Press, 2010.
- EVANS, B. G.; ALSHANGITI, W. (2018). "The Perception and Production of British English Vowels and Consonants by Arabic Learners of English". *Journal of Phonetics*, 68: 15-31.
- FLEGE, J. E.. "Second Language Speech Learning: Theory, Findings, and Problems". In: Strange, W. (Ed.). *Speech Perception and Linguistic Experience: Issues In Cross-Language Research*. Baltimore: York Press, 1995: 233-277.
- FLEGE, J. E.. "The Relation Between L2 Production and Perception". In: OHALA, J.; HASEGAWA, Y.; OHALA, M.; GRANVILLE, D.; BAILEY, A. *Proceedings of the XIVth International Congress of Phonetics Sciences*, 1999, August: 1273-1276.
- FLEGE, J.. "Language Contacts In Bilingualism: Phonetic Systems Interactions". In: COLE, J.; HAULDE, J. (Eds.), *Laboratory Phonology 9*. Berlin: Mouton de Gruyter, 2007: 353-380
- FLEGE, J. E.; MUNRO, M. J.; MACKAY, I. R. (1995). "Factors Affecting Strength of Perceived Foreign Accent In a Second Language". *The Journal of the Acoustical Society of America*, 97(5): 3125-3134.
- FLEGE, J. E.; WAYLAND, R. (2019). "The Role of Input In Native Spanish Late Learners' Production and Perception of English Phonetic Segments". *Journal of Second Language Studies*, 2(1): 1-44.
- GUION, S. G.; FLEGE, J. E.; AKAHANE-YAMADA, R.; PRUITT, J. C. (2000). "An Investigation of Current Models of Second Language Speech Perception: The Case of Japanese Adults' Perception of English Consonants". *The Journal of the Acoustical Society of America*, 107(5): 2711-2724.

- HARDISON, D. (2003). "Acquisition of Second-Language Speech: Effects of Visual Cues, Context, and Talker Variability". *Applied Psycholinguistics*, 24: 495-522.
- HARDISON, D. (2004). "Generalization of Computer-Assisted Prosody Training: Quantitative and Qualitative Findings". *Language Learning & Technology*, 8: 34-52.
- HUENSCH, A.; TREMBLAY, A. (2015). "Effects of Perceptual Phonetic Training on the Perception and Production of Second Language Syllable Structure". *Journal of Phonetics*, 52: 105-120.
- JAMIESON, D.; MOROSAN, D. (1986). "Training Non-Native Speech Contrasts In Adults: Acquisition Of The English /e/and /ɔ/ Contrast By Francophones". *Perception and Psychophysics*, 40: 205-215.
- JAMIESON, D., &. MOROSAN, D. (1989). "Training New, Nonnative Speech Contrasts: a Comparison of the Prototype and Perceptual Fading Techniques". *Canadian Journal of Psychology*, 43: 88-96.
- KINGSTON, J. (2003). "Learning Foreign Vowels". *Language and Speech*, 46: 295-349.
- LEVIS, J. M. (2005). "Changing Contexts and Shifting Paradigms In Pronunciation Teaching". *Tesol Quarterly*, 39(3): 369-377.
- LISKER, L.; ABRAMSON, A. S. (1964). "A Cross-Language Study of Voicing In Initial Stops: Acoustical Measurement". *Word*, 20: 384-422.
- LISKER, L.; ABRAMSON, A. S. "The Voicing Dimension: Some Experiments on Comparative Phonetics". *Proceedings of the 6th International Congress of Phonetic Sciences*. Prague: Academia, 1970.
- LOGAN, J. S.; LIVELY, S. E.; PISONI, D. B. (1993). "Training Japanese Listeners to Identify English /r/ and /l/: II. The Role of Phonetic Environment and Talker Variability In Learning New Perceptual Categories". *Journal of the Acoustical Society of America*, 94: 1242-1255.
- LOGAN, J. S.; LIVELY, S. E.; PISONI, D. B. (1991). "Training Japanese Listeners to Identify English /r/ and /l/: a First Report". *Journal of the Acoustical Society of America*, 89: 874-886.
- LOGAN, J. S.; LIVELY, S. E.; PISONI, D. B. (1993). "Training Listeners to Perceive Novel Phonetic Categories: How Do We Know What Is Learned?". *Journal of the Acoustical Society of America*, 94: 1148-1151.
- LOGAN, J. S.; PRUITT, J. S. "Methodological Issues In Training Listeners to Perceive Non-Native Phonemes". In: STRANGE, W. (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Baltimore: York Press, 1995.
- MACKAY, I. R.; FLEGE, J. E.; IMAI, S. (2006). "Evaluating the Effects of Chronological Age and Sentence Duration on Degree of Perceived Foreign Accent". *Applied Psycholinguistics*, 27(2): 157-183.
- MOYER, A. (1999). "Ultimate Attainment In L2 Phonology: The Critical Factors of Age, Motivation, and Instruction". *Studies in second language acquisition*, 21(1): 81-108.
- MUNRO, M. J.; DERWING, T. M. (1995). "Foreign Accent, Comprehensibility, and Intelligibility In the Speech of Second Language Learners". *Language learning*, 45(1): 73-97.
- MUNRO, M. J.; DERWING, T. M. (2008). "Segmental Acquisition In Adult ESL Learners: a Longitudinal Study of Vowel Production". *Language Learning*, 58: 479-502.
- MUNRO, M. J.; DERWING, T. M.; THOMSON, R. I. (2015). "Setting Segmental Priorities for English Learners: Evidence from a Longitudinal Study". *International Review of Applied Linguistics in Language Teaching*, 53(1): 39-60.
- PISKE, T.; MACKAY, I. R.; FLEGE, J. E. (2001). "Factors Affecting Degree of Foreign Accent In an L2: a Review". *Journal of Phonetics*, 29(2): 191-215.
- POLKA, L.; BOHN, O.-S. (1996). "A Cross-Language Comparison of Vowel Perception In English-Learning and German-Learning Infants". *Journal of the Acoustical Society of America*, 100: 577-592.
- ROCHET, B. L. "Perception and Production of Second-Language Speech Sounds by Adults". In: STRANGE, W. (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Baltimore: York Press, 1995: 379-410.

SAKAI, M.; MOORMAN, C. (2018). "Can Perception Training Improve the Production of Second Language Phonemes? A Meta-Analytic Review of 25 Years of Perception Training Research". *Applied Psycholinguistics*, 39(1): 187-224.

STRANGE, W. "Cross-Language Studies of Speech Perception: a Historical Review". In: STRANGE, W. (Ed.), *Speech perception and linguistic experience: Issues in cross-language research*. Baltimore: York Press, 1995: 3-45.

TROFIMOVICH, P.; BAKER, W. (2006). "Learning Second Language Suprasegmentals: Effect of L2 Experience on Prosody and Fluency Characteristics of L2 Speech". *Studies in second language acquisition*, 28(1): 1-30.

WANG, X.. *The Acquisition of English Vowels by Mandarin ESL Learners: a Study of Production and Perception*. Unpublished MA Thesis, Simon Fraser University, 1997.

WANG, X. *Training Mandarin and Cantonese Speakers to Identify English Vowel Contrasts: Long-term retention and effects on production*. Tese (Doctor of Philosophy) – Department of Linguistics, Simon Fraser University, 2002.

WANG, X. (2006). "Perception of L2 Tones: L1 Lexical Tone Experience May Not Help". In: HOFFMANN, R.; MIXDORFF, H. (Eds.). *The Proceedings of the Third International Conference on Speech Prosody*. Dresden: TUD Press, 2006: 85-88.

WANG, X.. "Training for Learning Mandarin Tones". In: ZHANG, F.; BARBER, B. (Eds.). *Handbook of Research on Computer-enhanced Language Acquisition and Learning*. Hershey, PA Information Science Reference (IGI Global), 2008: 259-274.

WANG, X. (2012). "Auditory and Visual Training on Mandarin Tones: A Pilot Study on Phrases and Sentences". *International Journal of Computer-Assisted Language Learning and Teaching (IJCALLT)*, 2(2), 16-29. DOI: 10.4018/IJCALLT.2012040102

WANG, X. (2013). "Perception of Mandarin Tones: The Effect of L1 Background and Training". *The Modern Language Journal*, 97: 144-160. DOI: 10.1111/J.1540-4781.2013.01386.X

WANG, X.; CHEN, J. (2019). "English Speakers' Perception of Mandarin Consonants: The Effect of Phonetic Distances and L2 Experience". In: CALHOUN, Sasha; ESCUDERO, Paola; TABAIN, Marija; WARREN, Paul (Eds.) *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019* (pp. 250-254). Canberra, Australia: Australasian Speech Science and Technology Association Inc.

WANG, X.; MUNRO, M. J. (1999). "The Perception of English Tense-Lax Vowel Pairs by Native Mandarin Speakers: The Effect of Training on Attention to Temporal and Spectral Cues". *Proceedings of the 14th international congress of phonetic sciences: Vol. 3*. Berkeley, CA: University of California, 1999: 125-128.

WANG, X.; MUNRO, M. (2004). "Computer-Based Training for Learning English Vowel Contrasts". *System*, 32, 539-552.

WANG, Y.; JONGMAN, A.; SERENO, J. A. (2003). "Acoustic and Perceptual Evaluation of Mandarin Tone Productions Before and After Training". *Journal of the Acoustical Society of America*, 113, 1033-1043.

WERKER, J. F.; POLKA, L. (1993). "Developmental Changes In Speech Perception: New Challenges and New Directions". *Journal of Phonetics*, 21, 83-101.

WERKER, J. F. "Cross-Language Speech Perception: Developmental Change Does Not Involve Loss". In: GOODMAN, J. C.; NUSBAUM, H. C. (Eds.). *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words*. Cambridge, Massachusetts: The MIT Press, 1994: 93-120.