LANGUAGE LEARNING AND DEVELOPMENT
https://doi.org/10.1080/15475441.2023.2285776

A Comparison of Perception-Based and Production-Based Training Approaches to Adults' Learning of L2 Sounds

Ying Li

ABSTRACT

While phonetic training in laboratory settings has been shown to be helpful for second language (L2) sounds learning in prior studies , it is still open to debate whether a perception- or a production-based approach can better help adults learn L2 sounds in classroom settings. This study aimed to fill this knowledge gap. The participants were three groups of adult Chinese college students who had difficulties in perceiving and producing English/ θ /-/s/ and/ð/-/z/. Group 1 and Group 2 were exposed to a perception-based and a production-based training approach respectively. Group 3 was the control group (no training). The training materials were "minimal pairs" embedding the target contrasts in various phonetic environments. Each session included explicit articulatory instructions and perception- or production-based practice activities. Individual participants' perception and production of the target contrasts were tested before (pretest), after (posttest), and one month after the training programme (delayed posttest). According to the results, (1) in comparison with the pretest, both the perception- and production-based groups showed significant perception and production improvement in the posttest, which remained in the delayed posttest; the control group had no significant perception or production changes across the three tests; (2) the production-based group had significantly better perception and production performance than the perception-based group both in the posttest and the delayed posttest. The overall results suggest that a production-based approach may be optimal for L2 sounds learning.

Introduction

Adult second language (L2) learners are frequently found to have difficulties in learning L2 sounds, particularly when a sound is not present in their native language (L1) (Flege, 1995a; Flege, 1995b; Best & Tyler, 2007). While some studies have explored the effectiveness of either perceptual or production training on L2 speech perception and production (e.g., Bradlow et al., 1997; Hazan et al., 2005; Leather, 1990; Li, 2016; Wang et al., 2003), none of the empirical studies draw on the comparison of the relative effectiveness of the two approaches in L2 sounds learning. Although teachers are unlikely to teach/practice L2 sounds either only through perception or only via production, figuring out which approach is superior would help teachers better design pedagogical activities and allocate the amount of time they spend on perception/production instruction/practice. To fill this gap and shed further light on this issue, this study targeted adult L1-Mandarin speakers and compared the effects of perception- and production-based approaches on their learning of English (L2) sounds.

Previous literature

According to Skill Acquisition Theory, practicing one skill (e.g., perception or production) is most effective for that skill (Dekeyser, 2007; for support, see Locke, 1988; Schwartz & Leonard, 1982;

CONTACT Ying Li liying_22@163.com

Zampini & Green, 2001 etc.). In fact, a number of experimental studies have revealed that perception training does not benefit the trainees' production proficiency, and vice versa (e.g., Sheldon & Strange, 1982; Strange & Dittmann, 1984; Smith, 2001). For example, Li and DeKeyser (2017) compared the effectiveness of perception and production practice on English-speaking adults' learning of Mandarin words. A group of native English speakers were first taught Mandarin tones and the target words. They were then engaged in a three-day perception or production practice session with the target words. Immediate posttest assessment showed that participants were more accurate and faster when being tested on the practiced skill (Li & DeKeyser, 2017). Likewise, in Lu, Wayland and Kaan (2015), native English speakers' ability to perceive lexical tone perception improved after both perception-only and perception-plus-production training. However, other experimental studies on phonetic training found that training on speech perception could improve trainees' production of the trained sounds, and vice versa (e.g., Hazan et al., 2005; Li & Somlak, 2019). For example, Wong (2016) trained Mandarin speakers' perception and production of English/ θ /-/s/and/ð/-/z/with auditory-only, audio-visual, auditory and explicit production, visual-only or explicit production-only approaches. Significant production improvement was only found when participants were trained in the auditory, audio-visual, explicit production and production-only conditions. On the whole, Wong (2016) indicated that direct production training could be more effective in improving trainees' production ability. These findings support the hypothesis that speech perception and production are two indivisible modalities, which are closely (Rvachew and Jamieson, 1989; Watkins, Strafella, & Paus, 2003; Williams & McReynolds, 1975) or even innately linked to each other (Liberman et al., 1967; Liberman & Mattingly, 1985; Liberman & Whalen, 2000).

Herd, Jongman, and Sereno (2013), however, point out that the effectiveness of a particular training paradigm depends on various factors beyond the pure production/perception dichotomy. They investigated the effectiveness of perception-based, production-based, and joint perception- and production-based training paradigms on English speakers' perception and production of Spanish/d, ɾ, r/. All three paradigms were found to be effective, whereas the degree of effectiveness depended on the nature of the target sounds and the modality of testing. In comparison, Liberman and Whalen's (2000) Motor Theory of speech perception indicates that speech production is intricately linked to perception, which might imply that there should be no difference in improvement across perception- and production-based training paradigms. On the whole, whether a production-based approach is more effective than a perception-based approach in segmental learning is still an open question.

A perception-based approach envisions learning as the comprehension of a target feature. Detailed auditory-to-articulatory mapping is used to conceptualize speech perception and production, and learners' production of L2 sounds reflects their mental representations, which are restricted by the perceived distinctive features of the phonemes (Shintani et al., 2013). Therefore, perception is a prerequisite of production. This view is supported by the Speech Learning Model, which hypothesizes that (1) speech perception precedes production and production ability cannot exceed perception ability; (2) learners' capacity to acquire language remains intact throughout their lives; (3) the perceived similarity or phonetic distance between L1 and L2 phonetic categories determines whether a new L2 category can be formed; and (4) the increase of L2 experience/input facilitates L2 sounds learning (Flege, 1995a, 1995b, 1999, 2002).

The Perception Assimilation Model for L2 also supports the notion that speech perception precedes production. Specifically, the Perception Assimilation Model for L2 hypothesizes that language learners assimilate unfamiliar non-native sounds to the most articulatorily-similar sound in their L1. Whether an L2 sound can be learned is dependent upon how it relates to the learner's L1 counterpart in terms of articulatory gestures: (1) only one phonetic category is permanently assimilated as an equivalent to learners' L1; (2) two L2 sounds are perceived in the same phonetic category, one a better exemplar than the other, such as English speakers' success in categorizing French/b/-/p/might be based on the extent to what they can distinguish the two sounds' aspiration as a corresponding distinction from English; (3) two L2 sounds are equally appropriate to be assimilated into a single L1 category, such as the misproduction of “fried rice” as “flied lice” by some Japanese speakers might be caused by the assimilation of English/l/-/r/into a single category/l/in their L1; (4) no assimilation occurs – an L2 sound is perceived without being assimilated to learners' L1. Moreover, in line with Speech Learning Model, Perception Assimilation Model for L2 hypothesizes that as L2 input increases, adult learners will be able to learn previously difficult L2 sounds (Best, 1995; Best & Tyler, 2007).

A perception-based approach has often been implemented with identification and/or discrimination tasks in prior studies. For instance, in a study by Lee et al. (2020), Japanese participants who were perceptually trained to learn English segmentals received two 30-minute treatment sessions per week for two weeks. Each session began with explicit instructions on the place and manner of the target sounds' articulation followed by identification practices. Specifically, participants were asked to listen to the instructor and choose the sound/word corresponding to the phonemes written on the blackboard by raising their left or right hand. In a 20-trial Phonetic Recognition Task, participants were then asked to listen to the instructor and select a phoneme/word from three to four items in a trial. At the end of the task, feedback was provided in a class discussion and a pronunciation demonstration was given by the instructor. Training results were tested by free-response questions, Japanese-English translations, and a read-aloud task. Participants' productions were assessed on a 9-point Likert scale by three raters. The overall findings suggested that the perception-based training programme significantly improved participants' production accuracy. In Bradlow et al. (1997), Japanese speakers were perceptually training English/r/-/l/with identification tasks. The learning phase lasted for three to four weeks (45 sessions), in which participants were asked to complete identification tasks and were given feedback on the validity of their responses. The stimuli were a large number of minimal pairs produced by various native English speakers. Both perception and production improved. Likewise, in a study by Rochet (1995), Mandarin speakers' production of French/bu/-/pu/displayed more native French-like voice onset time (VOT) after they were exposed to an identification training programme. Moreover, Sakai and Moorman (2018) conducted a meta-analysis of the studies of L2 perception training effects in production over the past 25 years and found that they led to a medium-sized improvement in perception and a small production improvement (e.g., Lambacher et al., 2005; Iverson & Evans, 2007; Iverson & Evans, 2009; Lengeris & Hazan, 2010).

In contrast, the production-based training approach views the articulation of target features as the learning source (Saito, 2018; Shintani et al., 2013). The Ontogeny Phylogeny Model and the Markedness Differential Hypothesis hypothesize that learners' success in L2 pronunciation is largely decided by the difference of L1-L2 markedness (Major, 1986; 2001; Eckman, 1991; Lee et al., 2020). Helping learners notice the markedness of an L2 sound, therefore, would facilitate its correct pronunciation. For example, speakers of some Asian languages (e g., Japanese, Korean, Mandarin) can correctly produce English/l/-/r/but fail to perceptually distinguish them in identification tasks (Raver-Lampman and Wilson, 2018). Yamada et al. (1994) examined the effects of immersion in English environment on Japanese speakers' perception and production of English/r, l, w/. Some speakers' production abilities exceeded their perception abilities, but not vice versa. Likewise, Tsukada et al. (2005) found that Korean children could better produce than perceptually discriminate the English vowels/i, ɪ, e, ɛ, æ, ɑ, ʌ, u/. Dutch learners of English also performed better when producing the English stops/b/-/p/than when perceiving them (Flege & Eefting, 1987).

A production-based training approach is usually carried out with articulatory demonstrations and/or imitation tasks. For example, the aforementioned Japanese speakers of the phonetic instruction group were first given explicit instructions of how to articulate the target sounds followed by imitation tasks among the entire class and in pairs (Lee et al., 2020). At the end of this task, individual pairs were asked to pronounce the sounds in front of the class, which was followed by the instructor's feedback and/or correction. Participants were allowed to recast until the target sound was correctly produced. Findings indicated that the production-based approach effectively improved the phonetic instruction group's pronunciation accuracy. In comparison, the production-based approach was found to be less effective than the perception-based approach in improving pronunciation accuracy. Nonetheless, due to the lack of perception tests in Lee et al. (2020), it was unclear whether the production-based approach could similarly enhance participants' *perception* accuracy.

To fill the above-mentioned gap, this study compared the effects of a perception-based approach – implemented mainly by providing participants with opportunities to observe the target sounds' articulatory gestures with matched sounds (following Li & Somlak, 2019) – with those of a production-based approach – implemented mainly through articulatory demonstration and imitation of an instructor's production of target sounds (following Lee et al., 2020). Specifically, our main research question was whether or how the perception-based vs. the production-based approach would help the Mandarin speakers develop their perception and production of L2-English sounds.

Experiment

Methods

Pilot study: participant and stimuli selection

A pilot study was conducted to select participants and target English sounds (consonants only; a further study focusing on vowels was carried out later). Four classes of first-year law majors (*n* = 162) at a university in southwest China were asked to do a read-aloud task in a quiet classroom. To maximize participants' active participation during the learning programme, only those who indicated a strong motivation to improve their English pronunciation were recruited. All the students were native Mandarin speakers who spoke English as a second language. They came from southwest Chinese cities so they spoke various Mandarin dialects in addition to standard Mandarin. They had been educated in Mandarin since kindergarten and used Mandarin for communication in daily life at the university. Moreover, as in Mandarin, the phonological inventories of all participants' dialects lacked interdentals (/ θ , ð/).

The stimulus was a 200-word English text – *The Boy Who Cried Wolf* by Deterding (2006) – which contained all the English consonants in various phonetic environments. Individual students' readings were recorded with a Roland-05 recorder (settings: 16-bit mono channel, 44.1 KHz), transferred to a laptop in WAV format, and then sent to three raters for evaluation via Dropbox. The raters were two females and one male who were born and raised in London and York respectively. They were phonetically trained with Ph.D. degrees in linguistics. They had done similar ratings jobs before the study. They were asked to listen to the recordings, note the incorrectly produced words, and transcribe the realizations of the words. The raters agreed with each other's assessment on incorrect realizations of consonants in the selected words (disagreements occurred in the transcriptions of vowels, which were not the focus of the study). The incorrectly produced consonants were found to be/l, n, , , , , θ , ð, f, t, /. In particular, 117 out of the 162 students (72.2%) were found to have replaced/ θ /with/s/in the production of *thought*, *threaten* and *third*. These 117 students and another 3 students (hence, 120 out of 162 students) also incorrectly produced/ð/as/z/when producing *there*, *the*, *with*, *this*, *them*, *they*, *bother*, *that* and *than.*

Considering that speech perception and production are likely to be linked (e.g., Liberman et al., 1967; Liberman & Mattingly, 1985; Liberman & Whalen, 2000), the 117 students who incorrectly produced/ θ , ð/as/s, z/were further tested on the perception of the two contrasts with an AXBtest[1] in a quiet room. During the test, after listening to a trial of three sounds, the students were asked to decide whether the sound they heard in the middle was more similar to the first or the third by mouse-clicking on the corresponding part on a laptop screen, which automatically triggered the following trial. The stimuli and procedure were adopted from Li and Somlak (2019). The stimuli were 60 nonce words auditorily recorded by a male and a female native English speaker. The words embedded the target contrasts in initial, medial, and final positions, which were counter-balanced in VC, VCV, and CV syllables with the vowel contexts/i, a, u/. To ensure that the responses were based on phonetic distinctions rather than on auditory discriminations, the interstimulus interval (ISI) was 1,000 ms, and

[1] An AXB test compares two choices of sensory stimuli to identify detectable differences between them. A subject is presented with two known samples (sample *A* and *B*) followed by one unknown sample *X* that is randomly selected from either A or B. He/She is then required to identify X as either A or B (Clark, 1982).

Table 1. Participants' background information.

Variables	Statistics	Perception-based group	Production-based group	Control group
Age (in years)	Mean	20	21	20
	Range	18 ~ 23	18 ~ 23	18 ~ 22
Onset age of English study (in years)	Mean	9.6	1.2	1.1
	Range	8 ~ 12	8 ~ 12	8 ~ 12
Years of English study	Mean	1.5	11.0	1.1
	Range	7 ~ 15	7 ~ 15	7 ~ 14
English proficiency (number of students)	Advanced	5	9	3
	Upper intermediate	6	3	7
	Intermediate	8	11	10
	Lower intermediate	10	5	10

L1 dialect		Chongqing (*n* = 16);
		Sichuan (*n* = 10);
		Yunnan (*n* = 2)
			Chongqing (*n* = 15);
			Sichuan (*n* = 12);
			Guizhou (*n* = 1)
				Chongqing (*n* = 18);
				Sichuan (*n* = 7);
				Yunnan (*n* = 3)

the intertrial interval was 3000 ms (Pisoni, 1973). Each stimulus word was repeated 3 times, yielding 108 stimuli in total for each contrast.

Eighty-four[2] students whose perception accuracy fell in the range of approximately 50% to 70% (/ θ /-/s/: accuracy range = 46.3%-69.4%, *M* = 59.0%, *S.E.*=0.80;/ð/-/z/: accuracy range = 45.4%-69.4%, *M* = 58.5; *S.E.*=0.70) were selected as the participants (see Table 1 for these participants' background information). Their perception test results in the pilot study (see "pre-test" results in Figures 4 and 5) were employed as the pretest results in the main study.

The pairs/ θ /-/s/and/ð/-/z/were selected as the target contrasts. Despite having low functional load (e.g., Suzukida & Saito, 2019), the two contrasts served well for the purpose of this study, because (1) compared with other less severe mistakes, training on/ θ /-/s/and/ð/-/z/allows relatively larger room for perception/production improvement and a larger number of "qualified" participants to be included; (2) alveolars/s, z/exist in both English and partly in Mandarin,[3] while interdentals/ θ , ð/ exist in English but not in Mandarin (Li, 2016). This according to the Speech Learning Model and the Perception Assimilation Model for L2, should make the contrasts difficult to learn by Mandarin learners of English.

Main study

In the main study, participants were exposed to either a perception- or a production-based training programme for the target sound contrasts (a control group received no training). Explicit instructions on the target contrasts' articulation and perception/production practice was provided in the training sessions. Participants' perception and production performance was tested before, after, and one month after training.

Materials

Materials used for training

Minimal pairs with the target contrasts embedded in initial (*n* = 20), medial (*n* = 20), and final (*n* = 20) positions of various phonetic environments were prepared (60 pairs per contrast, 120 pairs in total).

[2] According to the AXB test results, 89 students showed accuracy lower than 70% in perceiving both of the contrasts. Unfortunately, five of them dropped out, thus leaving 84 participants in total. Their levels of English proficiency shown in Table 1 were obtained with an Oxford Quick Placement Test, which tests English learners' knowledge of grammar and vocabulary, as well as how learners use that knowledge in order to understand communication. It is usually used to place students into an appropriate class for a language course, or as a quick measure of a student's general language ability (Li & Somlak, 2019).

[3] /z/is absent in Mandarin but occurs in most Mandarin dialects. As a common sound in most languages, there is no evidence showing that Mandarin speakers have difficulties in correctly perceiving/producing/z/(Li, 2016; Li & Somlak, 2019).

Due to the limited number of vocabulary items, there were nonce words among the minimal pairs (i e., *thirty* and *sirty* for/ θ /-/s/). To avoid causing confusion, participants were told that there might be some words new to them during the study.

Thirty of the minimal pairs were employed in the whole-class activity of both training approaches: 15 pairs in which/ θ /-/s/were embedded in initial ($n = 5$), medial ($n = 5$), and final ($n = 5$) positions; and 15 pairs in which/ð/-/z/were embedded in initial ($n = 5$), medial ($n = 5$), and final ($n = 5$) positions. The remaining 90 "minimal pairs" were used in the *pair-work activity* of the production-based approach and the *Two-alternative forced-choice task* of the perception-based approach.

For the *Two-alternative forced choice* task, six native English speakers who majored in linguistics auditorily recorded the 90 minimal pairs. The speakers had varied English accents (two females and one male were respectively from York, Manchester, and Newcastle England; one female was from Ottawa Canada; two males were from Miami America). They were asked to read the stimuli clearly but naturally, since exposing L2 learners to various native accents of naturally produced speech has been shown to facilitate L2 learning (Lively et al., 1993; Logan et al., 1991; Li & Somlak, 2019). Each pair was read twice by individual speakers: the second time in the reverse order (i e., *sink – think*; *think – sink*). In total, 180 trials were produced by each speaker, which were then distributed into separate trials, randomized, and entered into the software Psychopy for the tasks (ISI = 1000 ms; ITI = 3000 ms).

Materials used for testing

The materials used for perception tests were the same stimuli used in the pilot study. The materials used for the read-aloud task of the production tests were 12 English sentences adopted from Li (2016), which contained stimulus words embedding the target phonemes (/ θ , s, ð, z/) in initial, medial, and final positions. Each of the target phonemes occurred 15 times in total. The sentences were randomized in the three tests.

Procedure

The procedure is shown in Table 2. Participants were randomly divided into 3 groups of 28 students: the *perception-based* group was given perception-based training; the *production-based* group experienced production-based training; and the *control* group did not experience any training.

Both the perception- and the production-based training included six 45-minute sessions. The number of sessions was decided both for the convenience of participants and according to common phonetic training designs in prior studies (e.g., Sakai & Moorman, 2018). To examine participants' perception and production performance, a pretest (right before the programme was conducted; see pilot study), a posttest (the day after the programme was finished), and a delayed posttest (one month after the posttest) were carried out. During these tests, perception was evaluated with an AXB task and production with a read-aloud task. The same two tasks were administered in pretest, posttest and delayed posttest.

Training sessions and perception/production tests were conducted and recorded in a quiet classroom at the participants' university, where there were 32 desktops (monitor: Dell U2715H) equipped with high-quality headphones (Mpow H10). The instructor for both the perception- and production-based groups was a native Mandarin speaker who spoke English as a second language and had

Table 2. The schedule of each learning session.

Procedure	Perception-based group	Production-based group
Step 1 (10 minutes)	Explicit instructions on the target contrasts' articulation	Explicit instruction on the target contrasts' articulation
Step 2 (10 minutes)	Practice: whole-class activity	Practice: whole-class activity
Step 3 (25 minutes)	Practice: 2I2AFC discrimination task	Practice: pair-work activity

a Ph.D. degree in linguistics. She had studied and worked in English-speaking countries for more than 10 years. Her English proficiency was native-like.

Training phase

The training materials were displayed on participants' desktops, which were controlled by the investigator through a central computer in front of the classroom. To ensure participants' full understanding, instructions were given in Mandarin.

Step 1 The instructor explicitly taught the target contrasts' place and manner of articulation. Relevant graphs (e.g., Figure 1) were employed for assistance. After that, the instructor demonstrated the articulations in single phonemes and example words. Participants were asked to watch the demonstrations carefully. During this step, each student sat in front of a desktop computer, wore headphones, and adjusted the volume as needed.

Step 2 - Perception-based group: Following Lee et al. (2020), p. 30 minimal pairs were written on the whiteboard in front of the classroom – the phonemes/ θ , ð/and words containing/ θ , ð/were written on the left side, while the phonemes/s, z/and words containing/s, z/ were listed on the right side (see Figure 2). The instructor produced a phoneme or a word containing the target phoneme. Participants were asked to listen to the utterance and decide which phoneme/word was heard by raising their left or right hand, corresponding to the phoneme/word on the whiteboard. They were then asked to respond as soon as they heard the utterances and told there was no need to worry about making mistakes. The correct answer was immediately provided after they responded. Each "minimal pair" was practiced three times in contrast to each other.

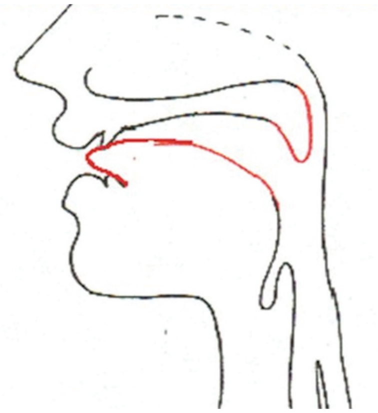
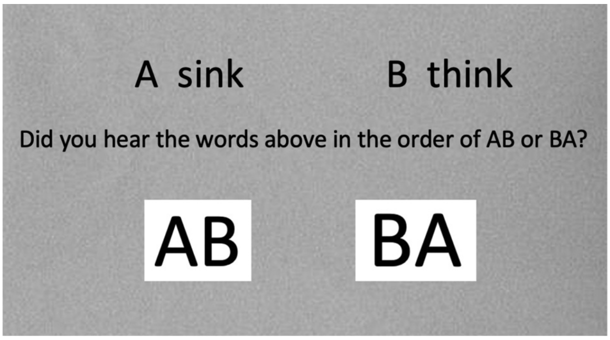


Figure 1. Graphic depiction of the pronunciation of interdental/ θ /.



A sink B think

Did you hear the words above in the order of AB or BA?

AB BA

Figure 2. A sample stimulus for the Two-alternative forced choice task presented with PsychoPy.

sink think

/s/ /θ/

Figure 3. An example of a "minimal pair" listed for pair-work practice.

Step 2 - Production-based group: The same 30 minimal pairs used by the perception-based group in *Step 2* were written on the whiteboard. Participants were asked to imitate the instructor's mouth movement and read word by word after the instructor. The instructor purposefully exaggerated the articulatory gesture of interdental and minimized the sound of her productions, so that the participants could focus on observing her articulatory gestures. Each member of a minimal pair was practiced three times in contrast to the other.

Step 3 - Perception-based group: Two-alternative forced choice task: Individual participants were asked to wear the headphones, listen to a minimal pair, and decide whether it was read in a certain order or the reverse by mouse-clicking on the corresponding part on the screen (Figure 4; 90 pairs in total). Feedback on the responses was provided immediately after a response was made; a correct response triggered a *ding* sound, and an incorrect response triggered a *buzz* sound. The following trial appeared after the feedback had been provided.

Step 3 - Production-based group: Pair-work activity. The same 90 "minimal pairs" used by the perception-based group in *Step 3* were presented on individual participants' laptop screens. Given that some of the stimuli were nonce words and all participants had basic knowledge of phonetic symbols,[2] the target phonemes were labeled under corresponding letter(s) colored in red (Figure 3). Participants were asked to work in pairs to read the "words" and correct each other's pronunciation if their partners made any mistakes in terms of articulatory gestures. Prior to that phase, the investigator asked a student to do the task with her in front of the class as a demonstration, including the corrections of tongue movement, lip movement, etc. The instructor walked around the classroom and provided help when needed. At the end of the class, 2–3 pairs of participants were selected to read the minimal pairs in front of the whole class. Different participants were selected at a session; thus, every participant had

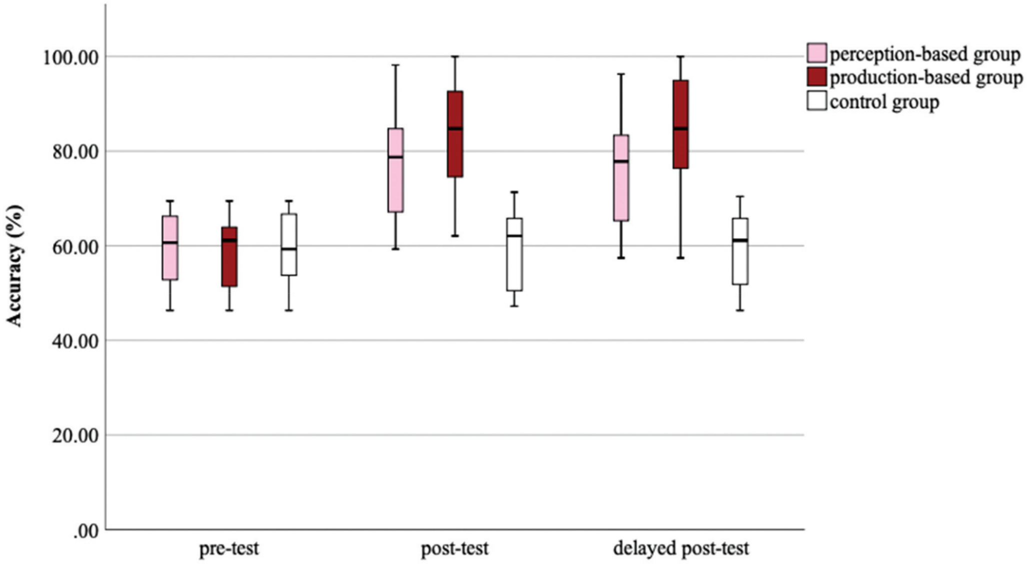


Figure 4. Boxplots of the three groups' accuracy in perceiving/ θ /-/s/.

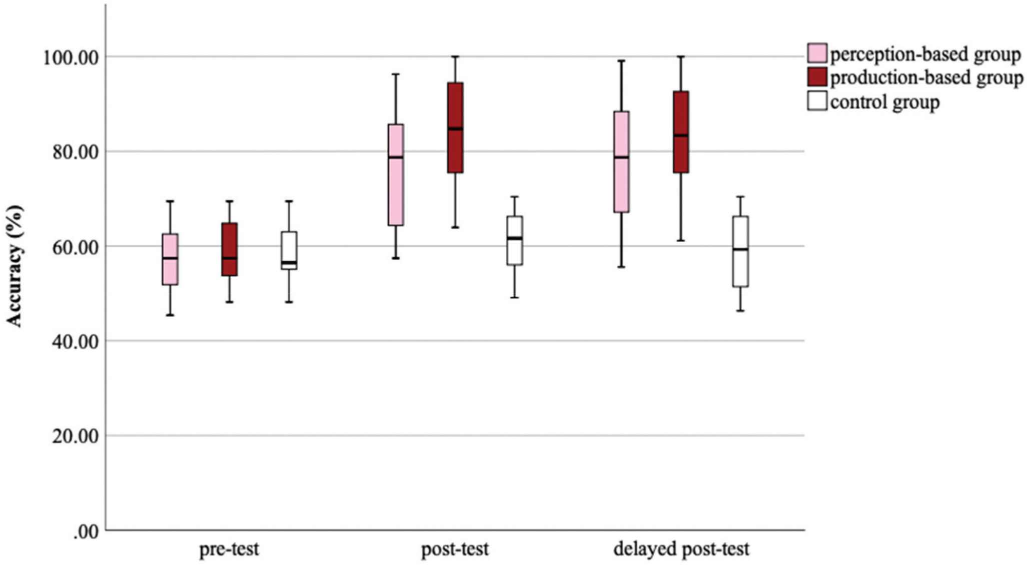


Figure 5. Boxplots of the three groups' accuracy in perceiving/ð/-/z/.

the chance to complete the task in front of the class and receive feedback from the instructor. Following Lee et al. (2020), any incorrect pronunciations were discussed with the whole class and recast until the correct pronunciations were realized.

Testing phase and evaluation

Participants were asked to do the same AXB perception test as in the pilot study. After that, they were asked to do the production (read-aloud) task individually. Their readings were audio-recorded and labeled with numbers. After the pre-, post-, and delayed posttest were completed, the recording clips were renamed with random numbers, so that the raters would be unlikely to know their recording orders. The renamed clips were then sent (via Dropbox) to one male and two female native English speakers for evaluation. The raters were neither told that the participants went through a training programme nor that the recordings were from different times of testing. The raters were linguistics majors and were trained previously to do the job. They were asked to evaluate the pronunciation accuracy of the two target contrasts on a 10-point Likert scale (0=sound totally unlike the target sound; 9=sound totally like the target sound). The interrater reliability was examined with Pearson's correlation coefficients. Positive and strong correlations among the three raters were revealed ($r = 0.91$, 0.83, and 0.92, respectively). The final score of each stimulus word's accuracy was the average score of the three raters' evaluations. For purposes of analysis, these ratings were converted into percentage scores.

Results

The perception task results are depicted in Figures 4 and 5. For each of the learning targets, a two-way ANOVA was conducted with Training approach (perception-based, production-based, and control) as the between-subjects variable, Test (pre-, post-, and delayed posttest) as the within-subjects variable, and individuals' perception accuracy as the dependent variable. Both ANOVAs revealed significant main effects of Test, Training approach, as well as an interaction between the two variables ($p < .05$) (see Table 3). A separate ANOVA carried out with individual participants' perception accuracy for each of the two contrasts as the dependent variable and Test (pre-, post-, and delayed posttest) as the independent variable confirmed that the perception-based group's accuracy in perceiving both/ θ /-/s/ ($F(1, 55) = 43.8$, $p<0.001$) and/ð/-/z/($F(1, 55) = 49.6$, $p<0.001$) was significantly higher in posttest than

Table 3. Two-way ANOVA results on participants' perception accuracy.

	F		Sig.		Partial Eta Squared	
Variable	/ θ /-/s/	/ð/-/z/	/ θ /-/s/	/ð/-/z/	/ θ /-/s/	/ð/-/z/
Test	72.8	71.4	<0.001	<0.001	0.16	0.13
Training approach	62.9	60.7	<0.001	<0.001	0.15	0.14
*Test * Training approach*	16.4	16.8	<0.001	<0.001	0.11	0.09

in pretest, but was not significantly different between posttest and delayed posttest ($p > .05$). Likewise, the production-based group displayed significantly higher perception accuracy in posttest than in pretest (/ θ /-/s/: ($F(1, 55) = 93.2$, $p<0.001$);/ð/-/z/: ($F(1, 55) = 101.9$, $p<0.001$), but no significant difference between posttest and delayed posttest ($p > .05$). Moreover, the production-based group's performance was relatively better than the perception-based group in terms of showing higher mean accuracies both in posttest (/ θ /-/s/: 83.5% vs. 77.8%;/ð/-/z/: 83.8% vs. 76.4%) and delayed posttest (/ θ /: 84.4% vs. 76.5%;/ð/-/z/: 83.4% vs. 77.4%), though they had similar accuracy means in pretest (θ /-/s/: 58.5% vs. 59.7%;/ð/-/z/: 59% vs. 57.2%). In contrast, the control group displayed similar accuracy means across the three tests (/ θ /-/s/: 58.9%, 59.4% and 59.2% respectively;/ð/-/z/: 58.7%, 60.7%, and 59% respectively).

Moreover, post hoc tests (see Table 4) adjusted through a Bonferroni correction revealed that, both in the post- and delayed posttest, the production-based group had significantly higher mean accuracy than both the perception-based group and the control group in perceiving the target sounds ($p < .001$). The perception-based group had significantly higher mean accuracy than the control group in perceiving the target sounds ($p < .001$).

As for the production task results, according to the three raters, no participant had any difficulty in correctly producing/s, z/, both were assessed with an accuracy of 100%. Therefore, only the production results of/ θ , ð/are reported (Figures 6 and 7). For each of these two learning targets, a two-way ANOVA was conducted with Training approach (perception-based, production-based, and control) as the between-subjects variable, Test (pre-, post-, and delayed posttest) as the within-subjects variable, and individuals' production accuracy as the dependent variable. Each of the ANOVAs revealed significant main effects of Test, Training approach, as well as an interaction between the two variables ($p < .05$) (see Table 5). Follow-up analyses showed that the interactions mirrored the participants'

Table 4. Two-way ANOVA results on participants' production accuracy.

	F		Sig.		Partial Eta Squared	
Variable	/ θ /	/ð/	/ θ /	/ð/	/ θ /	/ð/
Test	16.4	19.8	<0.001	<0.001	0.12	0.14
Training approach	21.1	21.6	<0.001	<0.001	0.15	0.15
*Test * Training approach*	5.8	6.1	<0.001	<0.001	0.09	0.09

Table 5. Post hoc tests results – three groups' perceptual in the post- and delayed post-test.

			Mean Difference (I-J)		S. E.		Sig.	
Target contrast	Group (I)	Group(J)	post-test	delayed posttest	post-test	delayed posttest	post-test	delayed posttest
/ θ /-/s/	perception-based	production-based	−5.7	−7.9	2.8	2.9	0.002	0.001
		control group	18.4	17.3	2.8	2.9	<0.001	<0.001
	production-based	control group	24.1	25.2	2.8	2.9	<0.001	<0.001
/ð/-/z/	perception-based	production-based	−7.4	−6.0	2.7	2.8	0.001	0.002
		control group	15.7	18.4	2.7	2.8	<0.001	<0.001
	production-based	control group	23.1	24.4	2.7	2.8	<0.001	<0.001

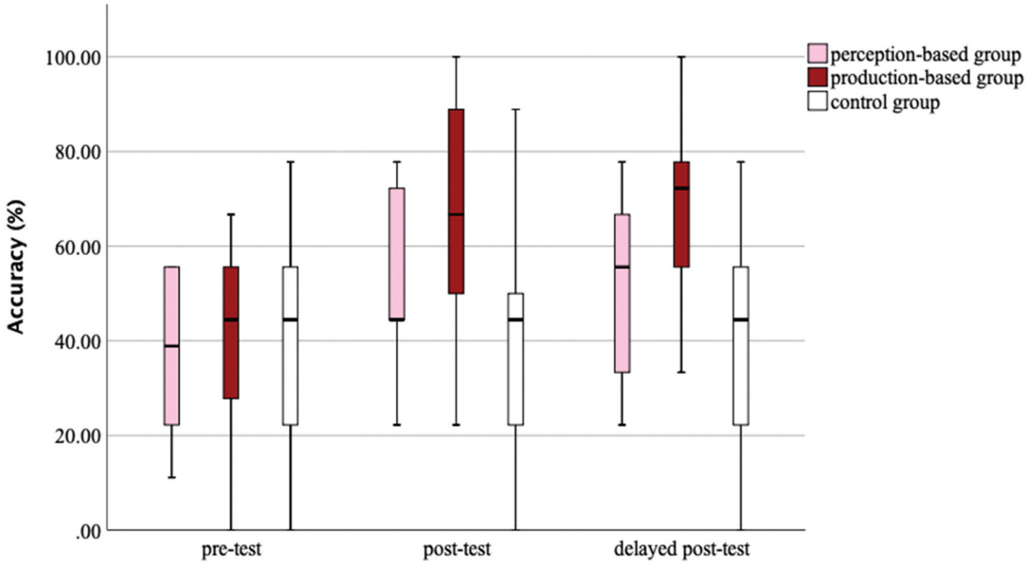


Figure 6. Boxplots of the participants' production of/ θ /.

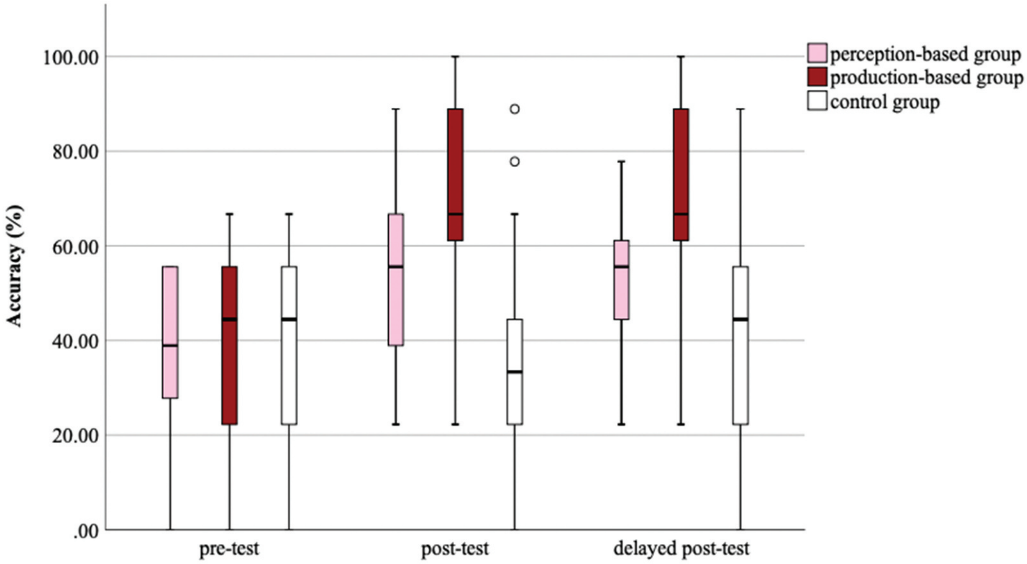


Figure 7. Boxplots of the participants' production of/ð/.

perceptual performance (see Figures 4 and 5). Both the perception- and the production-based group had significantly higher accuracy in the posttest than in the pretest in producing both/ θ /(perception-based group: $F(1,\ 55) = 46.3$, $p < .001$); production-based group: ($F(1,\ 55) = 95.8$, $p<0.001$) and/ð/ (perception-based group: $F(1,\ 55) = 52.5$, $p < .001$); production-based group: ($F(1,\ 55) = 104.4$, $p<0.001$), while their performance in the posttest and the delayed posttest did not differ ($p > .05$). The control group, however, did not show significant changes in producing/ θ /or/ð/across the three tests ($p > .05$) – most participants' production accuracy was between 20% and 60% in the three tests.

Moreover, post hoc tests (see Table 6) adjusted through a Bonferroni correction revealed that, both in the post- and delayed posttest, the production-based group had significantly higher mean accuracy

Table 6. Post hoc tests results – three groups' production in the post- and delayed post-test.

			Mean Difference (I-J)		S. E.		Sig.	
Phoneme	Group (I)	Group(J)	post-test	delayed posttest	post-test	delayed posttest	post-test	delayed posttest
/ θ /	perception-based	production-based	−17.5	−18.3	5.9	5.2	<0.001	<0.001
		control group	13.1	12.3	5.9	5.2	<0.001	<0.001
	production-based	control	30.6	30.7	5.9	5.2	<0.001	<0.001
/ð/	perception-based	production-based	−15.8	−16.7	5.7	5.0	<0.001	<0.001
		control group	15.9	13.5	5.7	4.9	<0.001	<0.001
	production-based	control group	31.0	30.2	5.7	5.0	<0.001	<0.001

than both the perception-based group and the control group in producing the target sounds. The perception-based group had significantly higher mean accuracy than the control group in producing the target sounds.

In sum, participants who were exposed to either the perception- or the production-based approach had significant improvements in perceiving and producing the target contrasts. Importantly, participants who experienced the production-based approach had significantly better perception and production performance than those who were exposed to the perception-based approach.

Additional ANOVA tests found that the phonetic position of the target sounds in the stimulus words (initial, medial, or final), participants' age, onset age of English study, gender and years of English study did not have a significant effect on production or perception accuracy ($p > .05$).

Discussion

One of the key findings in the study was that both training approaches significantly improved participants' accuracy in perceiving and producing the target contrasts. This is at odds with the assumption of Skill Acquisition theory, which predicts that practicing one skill is most effective for that skill (Dekeyser, 2007). It is also at odds with some previous studies that revealed null effects of perception training on trainees' production ability, and vice versa (Sheldon & Strange, 1982; Strange & Dittmann, 1984; Smith). However, this finding is in line with studies that revealed a transferred effect of production training on trainees' perception accuracy (e.g., Li & Somlak), and theories that predict the close relationship between speech perception and production (e.g., Motor Theory of speech perception, Speech Learning Model and Assimilation Model).

The second finding was that the production-based approach was more effective than the perception-based approach. This finding is at odds with that in Lee et al. (2020), who reported that the perception-based training approach is better able to help L2 segmental learning. The training procedures and practice tasks of the present study were similar to those used in the phonetic training groups in Lee et al. (2020): explicit instructions on the pronunciation of the target phonemes to both perception- and production-based groups, followed by teacher-led and/or pair-work activities for practice. There might be two reasons for the discrepancy between the findings of this study and Lee et al. (2020). First, in the practice activities, the perception-based group in Lee et al. (2020) received extra articulatory demonstrations in addition to feedback about the correctness of their responses in the identification tasks, whereas those in the present study were only provided with the correctness of their responses during the tasks. Second, participants in Lee et al. (2020) only received four 30-minute treatment sessions in total, while participants of this study were exposed to a relatively longer training programme (six 45-minute sessions). As predicted by both the Speech Learning Model and the

Perception Assimilation Model for L2, the amount of L2 experience (input) plays a key role in the L2 learning of speech sounds (Best & Tyler, 2007; Flege, 1995a; Flege, 1995b).

Moreover, this finding is consistent with the rationale and theoretical basis of production-based approaches, such as the Ontogeny Phylogeny Model, the Markedness Differential Hypothesis (Major, 1986, 2001), and the Critical Period Hypothesis (Eckman, 1977, 1991), which predict that L1-L2 differences pose difficulties for learning L2 sounds. On the contrary, the Speech Learning Model and the Perception Assimilation Model for L2 posit that L1-L2 similarities rather than differences inhibit L2 learning (Best & Tyler, 2007; Flege, 1995a; Flege, 1995b; Flege, 1999; Flege, 2002). Findings from the pilot study also support the former view – almost half of the participating students displayed severe difficulties in perceiving and producing the target contrasts. As discussed in the Introduction,/ θ , ð/ and/s, z/have marked differences in articulatory and acoustic features. Particularly, the interdental articulatory features of/ θ , ð/do not exist in Mandarin. According to the Perception Assimilation Model for L2, participants might have assimilated the English interdentals/ θ , ð/into the Mandarin alveolars/s, z/, the articulatory gestures of which are close to each other in terms of place of articulation.

Contrary to the view that speech perception precedes production (e.g., the hypotheses of the Speech Learning Model and the Perception Assimilation Model for L2), this finding seems to be in support of the view that speech production might precede perception (Raver-Lampman and Wilson, 2018). However, recall that, compared with the pretest, both the perception- and the production-based group's perception and production accuracy significantly increased in the posttest, and the improvement was maintained in the delayed posttest. If production does precede perception, the perception-based group would not have had a significant perception/production improvement, as members of the group were not exposed to a production-based approach. This was also at odds with the hypothesis of the Ontogeny Phylogeny Model and the Markedness Differential Hypothesis according to which learners' success in L2 sounds learning is independent from their comprehension ability (Major, 2001; Eckman, 1991). Moreover, instead of being independent from each other, the finding that learning from one modality benefited the other might support the view that speech perception and production are interrelated (e.g., Bradlow et al., 1997; Bradlow et al., 1999; Hardison, 2003; Hazan et al., 2005; Inceoglu, 2016). This is consistent with findings in some prior phonetic training studies (Bradlow et al., 1997; Hazan et al., 2005; Lambacher et al., 2005; Iverson & Evans, 2007; Iverson & Evans, 2009; Lengeris & Hazan, 2010; Rochet, 1995), in which perception training improved participants' production capacity and/or vice versa.

In addition, participants in the production-based group mainly received feedback from peers regarding their pronunciation in the "pair-word" activity of step 3. Although all of them were asked to "perform" in front of the class and were provided feedback by the instructor at the end of the activity, it was still possible that peer students provided the wrong feedback, not noticing their peers' errors, or might not have corrected the errors even if they noticed them. In contrast, the perception-based group always received automated correct feedback on every single item by hearing a *buzz* or a *ding* sound in the Two-alternative forced choice task of step 3. Unexpectedly, the production-based group outperformed the perception-based group in the post- and delayed posttest, which may have further confirmed the relative effectiveness of production-based approach in L2 sounds learning. Nevertheless, the production-based programme involved a live social interaction both among the participants and the teacher (practice and performance in pairs; feedback provided by the teacher), whereas the perception-based programme involved a computer-based task that lacked social interactions. According to Kuhl (2007), social interaction facilitates language learning by "gating" the computational mechanisms involved in human language learning. Hence, the production-based group's higher performance might, to some extent, be attributed to its interactions with peers and the teacher. The perception-based programme was also less controlled and involved more components than the production-based programme, which might also have contributed to the production-based group's relatively better testing results. Moreover, learners in the production-based group might have also heard the sound during instruction and practicing tasks, despite that the instructor had purposefully lowered her voice during demonstration and participants were told to focus on observing the articulatory gestures. These features might have also explained the processing advantage of the production-based group.

As predicted by the Speech Learning Model and the Perception Assimilation Model for L2, even adult learners could eventually learn L2 sounds with increased L2 input/experience, despite the absence of the sounds from their L1. This finding is at odds with Critical Period Hypothesis, which claims that L2 knowledge is unlearnable in adulthood (Lenneberg, 1967; Oyama, 1976). Given the facilitating effects of L2 input/experience in L2 sounds learning (e.g., Perception Assimilation Model for L2, Speech Learning Model) and the prediction that language learners' capacity in speech learning remains intact throughout their life (see Speech Learning Model), we speculate that given a longer period of perception-based and/or production-based learning, further improvement might be observed among participants.

Finally, factors relating to the participants' background did not have a significant effect on their perception or production performance across the three tests. This was similar to Li and Somlak (2019), who also reported that factors such as age, gender, years of English study, etc. did not play a significant role in Mandarin participants' perception or production of English sounds. Moreover, as displayed in Table 1, most participants were from neighboring cities/provinces (e.g., Chongqing, Sichuan) in southwest China, which all lack interdentals, hence dialectal variance is unlikely to explain their performance.

Conclusions

This study examined the effects of perception- and production-based training approaches on adult Mandarin speakers' learning of L2-English sounds. The overall results indicated that, although both perception- and production-based approaches significantly improved participants' perception and production performance, the production-based group significantly outperformed the perception-based group. Therefore, a production-based approach might be optimal for adults' learning of L2 sounds, even though some caveats remain (see below).

The above findings have pedagogical implications. Teachers may want to arrange relatively more time on production-based than perception-based teaching/practice activities, such as imitating the instructor's/native speakers' pronunciation (whether this is experienced in person or through video recordings). During the practice activities, learners may also benefit from feedback on the accuracy of their pronunciations, which has been found to be beneficial for L2 sound learning (e.g., Saito & Lyster, 2012). Although the production-based group in the present study also received feedback during the pair-work activities, most of the feedback was provided by their peers. As mentioned above, the comments from peers might not be totally reliable. Therefore, teachers may need to consider offering learners more reliable feedback, such as by providing the feedback directly themselves. Moreover, recasting – a reformulation of a learner's errors (Gooch et al., 2016) – is helpful to refine learners' pronunciation accuracy. This can be implemented in terms of whole-class/pair-work discussions as in the present study, or through other forms of spontaneous production tasks employed in prior studies —e.g., picture description tasks used in Gooch et al. (2016) and Saito (2013). Providing the students with explicit instruction on the articulation of L2 sounds can also be helpful, particularly on those absent from their L1. When a qualified instructor (e.g., a native speaker of a target language) is not available, pictures or multimedia resources (e.g., videos that demonstrate a sound's pronunciation) could be used, since providing learners with explicit visual demonstrations of the articulation of target sounds has been found to facilitate the learning process (e.g., Hazan et al., 2005; Li & Somlak, 2019). In addition, encouraging the learners to speak out without worrying about making mistakes during a task is critical because a teacher could only provide clarified feedback when a pronunciation error is heard.

It must be acknowledged that this study had some limitations. First, due to the lack of available words containing the target contrasts, nonce words were included in the teaching and testing materials. This may have had a negative impact on the results. For instance, according to some prior studies on the influence of cognitive factors on speech perception (e.g., Face, 2006; Waltermire, 2004), listeners might intend to choose the most frequently occurring/familiar sound (word) while doing a multiple-choice perceptual task. The participants might have selected a real word in a trial when that word was paired with a nonce one in the AXB task, because the real word occurred more frequently than the nonce word. Second, the teaching programme only lasted for six days. Even though significant perception and production improvement was observed, it was unclear whether given a longer period of teaching, participants' perception/production accuracy would improve continuously or stop/fluctuate at certain point. Moreover, participants within a group were trained together, hence some pro-active participants might have been more involved in whole-class tasks than others, despite the fact that they all indicated that they were strongly motivated to improve their English pronunciation. It remains to be seen whether the results would generalize with a more diverse sample. These factors could be taken into consideration in future studies.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Ying Li http://orcid.org/0000-0003-1783-9083

References

Best, C. T. (1995). A direct realist perspective on cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171–204). York Press.

Best, C., & Tyler, M. (2007). Nonnative and second–language speech perception. In O. Bohn & M. Munro (Eds.), *Language experience in second language speech learning: In honour of James Emil Flege* (pp. 13–34). John Benjamins.

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. I. (1999). Training Japanese listeners to identify English/r/and/l: Long-term retention of learning in perception and production. *Perception & Psychophysics*, *61*(5), 977–985. https://doi.org/10.3758/BF03206911

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. I. (1997). Training Japanese listeners to identify English / r / and / l /: IV. Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America*, *101*(4), 2299–2310. https://doi.org/10.1121/1.418276

Clark, D. (1982). High-resolution subjective testing using a double-blind comparator. *Journal of the Audio Engineering Society*, *30*(5), 330–338.

Dekeyser, R. (2007). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 97–113). Lawrence Erlbaum Associates, Inc.

Deterding, D. (2006). The north wind versus a wolf: Short texts for the description and measurement of English pronunciation. *Journal of the International Phonetic Association*, *36*(2), 187–196. https://doi.org/10.1017/S0025100306002544

Eckman, F. R. (1977). Markedness and the contrastive analysis hypothesis. *Language Learning*, *27*(2), 315–330.

Eckman, F. R. (1991). The structural conformity hypothesis and the acquisition of consonant clusters in the interlanguage of ESL learners. *Studies in Second Language Acquisition*, *13*(1), 23–41. https://doi.org/10.1017/S0272263100009700

Face, T. L. (2006). Cognitive factors in the perception of Spanish stress placement: Implications for a model of speech perception. *Linguistics*, *44*(6), 1237–1267. https://doi.org/10.1515/LING.2006.040

Flege, J. E. (1995a). Second language speech learning: Theory, findings and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). Balti more: York Press.

Flege, J. E. (1995b). Two procedures for training a novel second language phonetic contrast. *Applied Psycholinguistics*, *16*(4), 425–442. https://doi.org/10.1017/S0142716400066029

Flege, J. E. (1999). Age of learning and second-language speech. In E. Birdsong (Ed.), *New perspectives on the critical period hypothesis for second language acquisition* (pp. 101–132). Lawrence Erlbaum.

Flege, J. E. (2002). Interactions between the native and second-language phonetic systems. In P. Burmeister, T. Piske, & A. Rohde (Eds.), *An integrated view of language development: Papers in honor of Henning Wode* (pp. 217–244). Wissenschaftlicher Verlag.

Flege, J. E., & Eefting, W. (1987). Production and perception of English stops by native Spanish speakers. *Journal of Phonetics*, *15*(1), 67–83. https://doi.org/10.1016/S0095-4470(19)30538-8

Gooch, R., Saito, K., & Lyster, R. (2016). Effects of recasts and prompts on L2 pronunciation development: Teaching English/ɹ/to Korean adult EFL learners. *System*, *60*, 117–127. https://doi.org/10.1016/j.system.2016.06.007

Hardison, D. M. (2003). Acquisition of second-language speech: Effects of visual cues, context, and talker variability. *Applied Psycholinguistics*, *24*(4), 495–522. https://doi.org/10.1017/S0142716403000250

Hazan, V., Sennema, A., Iba, M., & Faulkner, A. (2005). Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English. *Speech Communication*, *47*(3), 360–378. https://doi.org/10.1016/j.specom.2005.04.007

Herd, W., Jongman, A. & Sereno, J. (2013). Perceptual and production training of intervocalic/d, ɾ, r/in American English learners of Spanish. *The Journal of the Acoustical Society of America*, *133*(6), 4247–4255.

Inceoglu, S. (2016). Effects of perceptual training on second language vowel perception and production. *Applied Psycholinguistics*, *37*(5), 1175–1199. https://doi.org/10.1017/S0142716415000533

Iverson, P. & Evans, B. G. (2007). Learning English vowels with different first-language vowel systems: Perception of formant targets, formant movement, and duration. *The Journal of the Acoustical Society of America*, *122*(5), 2842–2854.

Iverson, P., & Evans, B. G. (2009). Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakers. *The Journal of the Acoustical Society of America*, *126*(2), 866–877. https://doi.org/10.1121/1.3148196

Kuhl, P. K. (2007). Is speech learning 'gated' by the social brain?. *Developmental science*, *10*(1), 110–120.

Lambacher, S. G., Martens, W. L., Kakehi, K., Marasinghe, C. A., & Molholt, G. (2005). The effects of identification training on the identification and production of American English vowels by native speakers of Japanese. *Applied Psycholinguistics*, *26*(2), 227–247. https://doi.org/10.1017/S0142716405050150

Leather, J. (1990). Perceptual and productive learning of mandarin lexical tone by Dutch and English speakers. In J. Leather & A. James (Eds.), *New Sounds 90: Proceedings of the Amsterdam Symposium on the Acquisition of Second Language Speech*. Amsterdam, The Netherlands: University of Amsterdam.

Lee, B., Plonsky, L., & Saito, K. (2020). The effects of perception-vs. production-based pronunciation instruction. *System*, *88*, 102185. https://doi.org/10.1016/j.system.2019.102185

Lengeris, A., & Hazan, V. (2010). The effect of native vowel processing ability and frequency discrimination acuity on the phonetic training of English vowels for native speakers of Greek. *The Journal of the Acoustical Society of America*, *128* (6), 3757–3768. https://doi.org/10.1121/1.3506351

Lenneberg, E. H. (1967). *Biological foundations of language*. Wiley.

Li, Y. (2016). Audiovisual training effects on L2 speech perception and production. *International Journal of English Language Teaching*, *3*(2), 14–23. https://doi.org/10.5430/ijelt.v3n2p14

Liberman, A. M., Cooper, F. S., Shankweiler, D. P. & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*(6), 431.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*(1), 1–36. https://doi.org/10.1016/0010-0277(85)90021-6

Liberman, A. M., & Whalen, D. H. (2000). On the relation of speech to language. *Trends in Cognitive Sciences*, *4*(5), 187–196. https://doi.org/10.1016/S1364-6613(00)01471-6

Li, M., & DeKeyser, R. (2017). Perception practice, production practice, and musical ability in L2 mandarin tone-word learning. *Studies in Second Language Acquisition*, *39*(4), 593. https://doi.org/10.1017/S0272263116000358

Li, Y., & Somlak, T. (2019). The effects of articulatory gestures on L2 pronunciation learning: A classroom-based study. *Language Teaching Research*, *23*(3), 352–371. https://doi.org/10.1177/1362168817730420

Lively, S. E., Logan, J. S., & 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. *The Journal of the Acoustical Society of America*, *94*(3), 1242. https://doi.org/10.1121/1.408177

Locke, J. L. (1988). Variation in human biology and child phonology: A response to goad and ingram. *Journal of Child Language*, *15*(3), 663–668. https://doi.org/10.1017/S0305000900012617

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English/r/and/1/. *The Journal of the Acoustical Society of America*, *89*(2), 874–886. https://doi.org/10.1121/1.1894649

Lu, S., Wayland, R., & Kaan, E. (2015). Effects of production training and perception training on lexical tone perception–A behavioral and ERP study. *Brain Research*, *1624*, 28–44. https://doi.org/10.1016/j.brainres.2015.07.014

Major, R. C. (1986). The ontogeny model: Evidence from L2 acquisition of Spanish r. *Language Learning*, *36*(4), 453–504. https://doi.org/10.1111/j.1467-1770.1986.tb01035.x

Major, R. C. (2001). *Foreign accent: The ontogeny and phylogeny of second language phonology*. Lawrence Erlbaum.

Oyama, S. (1976). A sensitive period for the acquisition of a nonnative phonological system. *Journal of Psycholinguistic Research*, *5*(3), 261–285. https://doi.org/10.1007/BF01067377

Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics*, *13*(2), 253–260. https://doi.org/10.3758/BF03214136

Raver-Lampman, G., & Wilson, C. (2018). An acceptable alternative articulation to remediate mispronunciation of the English/l/sound: Can production precede perception? *TESOL Journal*, *9*(1), 203–223. https://doi.org/10.1002/tesj.319

Rochet, B. L. (1995). Perception and production of second-language speech sounds by adults. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 379–410). MD: York Press

Rvachew, S., & Jamieson, D. G. (1989). Remediating speech production errors with sound identification training. *Journal of Speech-Language Pathology and Audiology*, *16*, 201–208. https://doi.org/10.1044/jshd.5402.193

Saito, K. (2013). The acquisitional value of recasts in instructed second language speech learning: Teaching the perception and production of English/ɹ/to adult Japanese learners. *Language Learning*, *63*(3), 499–529. https://doi.org/10.1111/lang.12015

Saito, K. (2018). Advanced segmental and suprasegmental acquisition. In P. Malovrh & A. Benati (Eds.), *The handbook of advanced proficiency in second language acquisition* (pp. 282–303). Wiley Blackwell.

Saito, K., & Lyster, R. (2012). Effects of form-focused instruction and corrective feedback on L2 pronunciation development of/ɹ/by Japanese learners of English. *Language Learning*, *62*(2), 595–633. https://doi.org/10.1111/j.1467-9922.2011.00639.x

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. https://doi.org/10.1017/S0142716417000418

Schwartz, R. G., & Leonard, L. B. (1982). Do children pick and choose? An examination of phonological selection and avoidance in early lexical acquisition. *Journal of Child Language*, *9*(2), 319–336. https://doi.org/10.1017/S0305000900004748

Sheldon, A., & Strange, W. (1982). The acquisition of/r/and/l/by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics*, *3*(3), 243–261. https://doi.org/10.1017/S0142716400001417

Shintani, N., Li, S., & Ellis, R. (2013). Comprehension-based versus production-based grammar instruction: A meta-analysis of comparative studies. *Language Learning*, *63*(2), 296–329. https://doi.org/10.1111/lang.12001

Smith, L. C. (2001). L2 acquisition of English liquids: Evidence for production independent from perception. In X. Bonch-Bruevich, W. J. Crawford, J. Hellermann, C. Higgins, & H. Nguyen (Eds.), *The past, present and future of second language research: Selected proceedings of the 2000 Second Language Research Forum* (pp. 3–22). Somerville, MA: Cascadilla Press.

Strange, W., & Dittmann, S. (1984). Effects of discrimination training on the perception of/rl/by Japanese adults learning English. *Perception & Psychophysics*, *36*(2), 131–145. https://doi.org/10.3758/BF03202673

Suzukida, Y., & Saito, K. (2019). Which segmental features matter for successful L2 comprehensibility? Revisiting and generalizing the pedagogical value of the functional load principle. *Language Teaching Research*, *25*(3), 431–450. Advance Online Publication https://doi.org/10.1177/1362168819858246

Tsukada, K., Birdsong, D., Bialystok, E., Mack, M., Sung, H., & Flege, J. (2005). A developmental study of English vowel production and perception by native Korean adults and children. *Journal of Phonetics*, *33*(3), 263–290. https://doi.org/10.1016/j.wocn.2004.10.002

Waltermire, M. (2004). The effect of syllable weight on stress in Spanish. In T. Face (Ed.), *Laboratory Approaches to Spanish Phonology* (pp. 171–191). Berlin: Mouton de Gruyter.

Wang, Y., Jongman, A., & Sereno, J. A. (2003). Acoustic and perceptual evaluation of mandarin tone productions before and after perceptual training. *The Journal of the Acoustical Society of America*, *113*(2), 1033–1043. https://doi.org/10.1121/1.1531176

Watkins, K. E., Strafella, A. P., & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia*, *41*(8), 989–994. https://doi.org/10.1016/S0028-3932(02)00316-0

Williams, G. C., & McReynolds, L. V. (1975). The relationship between discrimination and articulation training in children with misarticulations. *Journal of Speech and Hearing Research*, *18*(3), 401–412. https://doi.org/10.1044/jshr.1803.401

Wong, J. W. S. (2016). The effects of visual and production components in training the perception and production of English consonant contrasts by mandarin speakers. *The Journal of the Acoustical Society of America*, *140*(4), 3336–3336. https://doi.org/10.1121/1.4970642

Yamada, R., Strange, W., Magnuson, J., Pruitt, J., & Clark, W. (1994). The intelligibility of the Japanese speakers' production of American English/r/,/l/, and/w/, as evaluated by native speakers of American English. *International Conference on Spoken Language Processing*, Yokohama (Vol. 94, pp. 2023–2026). Acoustical Society of Japan.

Zampini, M. L. & Green, K. P. (2001). The voicing contrast in English and Spanish: the relationship between perception and production. In J. L. Nicol (Ed.), *One mind, two languages: Bilingual language processing* (pp. 23–48). Malden, MA: Blackwell.