Computer Assisted Language Learning Electronic Journal, 22(1), 2021, 378-400

Self Adaptive Phonetic Training for Mandarin Nasal Codas

Ruining Yang (yangrn@hnu.edu.cn)
Hunan University, Changsha, China

Hiroaki Nanjo (nanjo@media.kyoto-u.ac.jp)
Kyoto University, Kyoto, Japan

Masatake Dantsuji (mdantsuji@media.kyoto-u.ac.jp)
Kyoto University, Kyoto, Japan

Abstract

This study proposed a new perceptual training paradigm called Self Adaptive Phonetic Training (SAPT) for Japanese listeners' perception of nasal codas in Mandarin, which dynamically provided a customized training plan for each participant. 17 participants were divided into two groups: the SAPT group and the High Variability Phonetic Training (HVPT) group. They were required to conduct 12 training sessions in a self-controlled environment with the training software. The results of a perception and production assessment showed that participants could benefit from both training paradigms and the effect size of SAPT was a little larger than HVPT. The training effect was transferred to new stimuli and new talkers in SAPT and HVPT. However, it did not transfer to new phonetic environments in both training paradigms. Moreover, perceptual training turned out to be effective in the production domain in both paradigms. These findings suggest that, in general, SAPT is an effective training paradigm for improving listeners' perceptual and production ability to identify nasal codas in Mandarin. Notably, it reduces the time spent on unnecessary items and raises the efficiency of training. It sheds new light on constructing the perceptual training paradigm and may have implications for phonetic training in other second language learning.

Keywords: Self Adaptive Phonetic Training (SAPT), High Variability Phonetic Training, Mandarin nasal coda, Japanese listener

Introduction

Learning a new language as a second language or foreign language (L2) is always challenging and learning a new language's phonetic system can be especially difficult. Flege's Speech Learning Model (SLM) situates L2 learners as tending to refer to their first language (L1) when perceiving L2 sounds (Flege, 1995). Such "perceived phonetic similarity" between L2 and L1 phonetic categories decides how learners identify L2 sounds. When an L2 sound has no counterpart in an L1 inventory, it is relatively easy for listeners to establish a new phonetic category. Meanwhile, when the phonetic dissimilarity between the L2 sound and the L1 sound is small, it is difficult for L2 learners to perceive the L2 sound. However, SLM also indicates that L2 learners can establish a new phonetic category for an L2 sound if they are given adequate input (Flege, 1995). Inspired by this, perceptual training has been conducted to help learners establish a new phonetic category.

One well-known training paradigm is High Variability Phonetic Training (HVPT). HVPT is a perceptual training paradigm that contains multiple stimuli in different phonetic environments produced by various native talkers (Lively et al., 1993). More specifically, HVPT considers three factors: talker variability, stimulus variability, and phonetic context variability (Shport, 2016). HVPT is based on the principle that exposing L2 learners to varied phonetic input allows them to establish a more robust L2 sound system, allowing them to generalize to sounds in new words and words uttered by unfamiliar speakers (Thomson, 2018).

The effectiveness of HVPT has been verified for vastly contrasting vowels and consonants (Hazan et al., 2005; Rato, 2014) and even suprasegmental features, such as Mandarin tones (Wang et al., 1999) and the Japanese pitch accent (Shport, 2016). The training, which encourages a long-term modification of listeners' phonetic perception, has also proved effective with new stimuli and new talkers' voices (Lively et al., 1994). Furthermore, previous research has also shown its ability to facilitate improvements in the production domain (Bradlow et al., 1997; Lambacher et al., 2005; Thomson & Derwing, 2016). Although the research on HVPT's effect is numerous, most of them targeted English vowels and English /r/ - /l/, which have vowel-like characteristics (Thomson, 2018). For L2 consonants, to form general conclusions on the effectiveness, more research is needed.

In addition, some issues of HVPT remain unaddressed. For example, individual differences may impact the effectiveness of HVPT (Fuhrmeister & Myers., 2020; Perrachione et al., 2011; Sadakata & McQueen., 2014); thus, canonical HVPT's nonadaptive approach may prevent it from serving all learners. Moreover, although some attempts have been made to conduct the HVPT in non-laboratory settings (Thomson, 2011, 2012), little research has been done on its applications—as well as the applications of computer-assisted language learning (CALL) materials—in non-laboratory settings. Accordingly, much remains unknown about whether HVPT can be applied to a self-paced learning environment on a larger scale. Besides, although it has long been understood that normally in L2 speech development perception precedes production (e.g., Cardoso, 2011; Flege, 1995), not all research suggests a unidirectionality of this relationship. (e.g., Linebaugh & Roche, 2015; Olson, 2014). A meta-analysis suggests a small- to medium-sized relationship between perception and production gains via perceptual training; however, the relationship was not significant (Sakai & Moorman, 2018).

In this study, we developed a new phonetic training software that uses a Self Adaptive Phonetic Training (SAPT) paradigm to help Japanese listeners identify nasal codas in Mandarin. Notably, SAPT considers individual differences and accordingly asks listeners to focus on those items with which they have difficulty. Moreover, this paper also discusses the transfer effect to new stimuli, new talkers, and new phonetic environments. Furthermore, we report the influence of perceptual training on the production domain and discuss the relationship between speech perception and production. These findings contribute to the field of perceptual training for L2 learning by providing a new phonetic training paradigm that takes into account individual differences. Notably, the training software trains learners in a practical and convenient, self-paced learning environment.

Literature Review

High Variability Phonetic Training (HVPT)

The SLM argues that L2 learners can modify their existing phonetic categories and develop new ones for L2 sounds if they are provided with enough naturalistic input or targeted phonetic training. In recent years, there has been an increasing interest in the research of HVPT. In a review of HVPT, Barriuso and Hayes-Harb (2018) present a brief history of HVPT and discuss its use. Logan et al. (1991) is considered a seminal study on HVPT. In this study, 68 minimal pairs differing only in /r/ and /l/, which were produced by six native speakers were applied as the training items. Japanese native speakers were asked to conduct a forced-choice identification task with feedback during the training phase. A pre-test and a post-test were set to verify the training effect. The training was successful because of the application of multiple phonetic environments and various speakers in the training. Later studies verified the generalization effect (Lively et al., 1993) and retained effect (Lively et al., 1994). These studies prove that using high variability in training helps learners to not only improve their perception of nonnative sounds but also help them generalize their learning to new stimuli and talkers as well as helps them retain their learning.

Moreover, When L2 learners develop new perceptual representations for segmentals and suprasegmentals, this allows for increasingly fluid access to these systems to produce the L2 (Saito, 2018). Therefore, perceptual training with HVPT has also proved to be capable of transferring to the production domain and showed great promise in increasing listeners' production ability of L2 sounds (e.g., Bradlow et al., 1997; Iverson et al., 2012; Lambacher et al., 2005; Wiener et al., 2020).

It has to be pointed out, however, that not all research suggests a unidirectionality of this relationship: Linebaugh and Roche (2015), for example, report some evidence that L2 production training can enhance perception. Other studies have shown that L2 production might have an articulatory basis rather than a perceptual basis—that is, after articulatory training, certain sounds can be produced quite well without being accurately perceived (Olson, 2014; Smith, 2001). Moreover, finally, Nagle (2018) showed that the perception-production link might not be as synchronous as we thought. Rather, we should expect a time lag and individual differences for gains in perception transferring to gains in production. As a result, learning to perceive L2 sounds may not always help learners to produce those sounds (Nagle, 2020). Data about the relationship between perceptual training and production gains are limited and more discussion is expected.

Concerning the paradigm of HVPT, one issue that remains unsolved is individual differences. Golestani and Zatorre (2009) verified that there were considerable individual differences in perceptual training performance. Perrachione et al. (2011) indicate that individual differences in perceptual abilities and the design of the training paradigm interact with the training effect. Not all participants likely benefit from HVPT, and listeners with weaker perceptual abilities may find low variability training paradigms more effective. Along these lines, Yang et al. (2018b) found that only some participants show gains from the training. Fuhrmeister and Myers (2020) also verified that individual aptitude predicted overnight improvement after training on discrimination task.

What is important to take away here for our purposes is that the nonadaptive training paradigm may fail to account for individual differences and thus limit its effectiveness. Addressing this issue, an adaptive training approach is suggested to apply. Levis (2007) and Munro et al. (2015) emphasize the importance of an adaptive approach, which takes individual differences into account. Providing learners with their difficulties helps learners achieve more progress from training. For example, Qian et al. (2018) consider individual differences in conducting perceptual training that uses an adaptive high variability training paradigm. This training paradigm maintained a high variability of training stimuli and, at the same time, allowed participants to target their problematic sounds. If the phonemic contrast accuracy at the pre-test phase exceeded 80%, participants did not need to complete training on this item. However, if the accuracy was lower than 80%, the system automatically presented the participant with a series of training exercises until their accuracy reached 80%. Because training duration and training items varied from person to person, this training was adaptive. Nevertheless, they did not mention the criterion of the cutoff score (80%). This makes phonemic contrast with an accuracy higher than 80% lose the opportunity to be trained.

Furthermore, some other training paradigms based on HVPT provide more insights into the constitution of the perceptual training and improve the effectiveness. For example, training in adaptive adverse conditions (Leong et al., 2018); temporal acoustic exaggeration in HVPT (Cheng et al., 2019). The suggested approaches showed more improvement than traditional HVPT. This kind of “adaptive” training that provides more possibilities for perceptual training should get more attention.

Another issue is the application of HVPT in pedagogic. As Wang and Munro (2004, p.540) point out, "there is a significant gap between some of the key research findings of laboratory studies from the past two decades and techniques that have been put into practice." Despite technological advancements enabling the development of more and more CALL materials, few educational software programs exist for phonetic contrast identification. In the majority of studies involving laboratory training, participants were subjected to rigid research paradigms with strict schedules, and training sessions were relatively long. These conditions differ from those of more realistic learning environments. Along these lines, Barriuso and Hayes-Harb (2018) observe that few studies directly verify the application of HVPT in non-laboratory settings; Thomson (2018, p.209) states that "HVPT remains unfamiliar to most language instructors and learners." Therefore, HVPT must be developed in ways that allow it to be applied in different settings to meet learner needs. If we can develop more applications like English Accent Coach (Thomson, 2012), the effectiveness of HVPT can be better verified in practice and HVPT can benefit more language learners.

Nasal Codas in Mandarin

In Mandarin, there are two types of nasal codas, "n [n]" and "ng [ŋ]." Two words that differ only at the position of nasal coda may demonstrate different meanings. The two nasal codas differ physiologically in the place of articulation: "-n [n]" is an alveolar sound that is produced with the constriction of the tongue blade while "-ng [ŋ]" is a velar sound that produced with the constriction of the tongue body. In Japanese, /ɴ/ is a nasal sound appearing at the end of a syllable. Pronunciation varies depending on the articulation position of the following articulation. Although both [n] and [ŋ] exist in Japanese, they are not used to discriminate between the meanings of different words since they are two allophones of phoneme /ɴ/ in Japanese.

According to SLM, the category formation for Mandarin [n] and [ŋ] may be blocked by equivalence classification because Japanese speakers subsume [n] and [ŋ] within the existing perceptual representation of /ɴ/ in Japanese. As a result, they usually have difficulty identifying "-n" and "-ng" in Mandarin (Wang, 2002; Yang et al., 2016), and many Mandarin teachers find it difficult to teach Japanese Mandarin learners to identify Mandarin nasal codas.

In Mandarin, as with nasal codas, "n" and "ng" usually occur with vowels and come at a syllable's final position. The high vowel /i/, mid vowel /ə/, and low vowel /a/ can precede both "n" and "ng" (Mou, 2006). In this study, we only concentrate on these six items: "-an", "-ang", "-en", "-eng", "-in", "-ing". Previous research has reported that the vowel before the nasal codas is an important cue for Mandarin native speakers to identify "-n" and "-ng" from different perspectives, such as nasalization, formant transitions of vowels, duration of vowels. (Chen, 2000; Zhang & Wang, 2017). Similarly, Wang (2002) also found that in perception, Japanese listeners rely on the preceding vowels to distinguish "-n" and "-ng". The more similar the phonetic values of the preceding vowels are, the less distinctive the two nasal codas are. Thus, it is more difficult for Japanese listeners to distinguish them. These studies indicate that there is a relationship between preceding vowels and the identification of nasal codas. Nasal codas with different preceding vowels may change differently through perceptual training.

The evidence presented in this section suggests that Mandarin nasal codas are difficult for Japanese Mandarin learners to acquire. Few solutions to their perceptual confusion on "-n" and "-ng" have been found. Although some attempts have been made to help Japanese listeners to identify nasal codas (Li et al., 2019; Yang et al., 2018b), the deeper point here is that these studies highlight the need for a perceptual training application for Mandarin nasal codas.

This paper proposes a new training paradigm—SAPT—to teach Japanese Mandarin learners to identify "-n" and "-ng". SAPT considers individual differences and asks listeners to focus on those items they have difficulty with and always conduct the training on their weak items. This training paradigm is implemented through training software developed by authors. A pre- and post-test design with a comparison of SAPT group and HVPT group was applied to address the following research questions:

(1) How effective are SAPT and non-adaptive HVPT in a nonlaboratory and self-controlled setting?
(2) Does the training effect generalize across perceptions of new stimuli, new talkers, and new phonetic environments in SAPT and HVPT?
(3) Does the effect of perceptual training transfer to production improvement in SAPT and HVPT?

Method

Procedure

The experiment was divided into two parts: a listening task and a speech task. In order to investigate accuracy before training, participants were asked to conduct speech pre-task (SPRE) and listening pre-task (LPRE) activities. Prior to the training, the authors explained the usage of the software and some details to which the participants should pay attention during the training period. After the explanation, we distributed USB flash drives with training software to participants so that the participants could conduct the training anywhere at any time. Participants were required to complete 12 training sessions and, subsequently, to answer a questionnaire and conduct speech post-task (SPOST) and listening post-task (LPOST) activities.

Participants

We recruited 21 native Japanese speakers as our participants. They were students studying Mandarin as a second foreign language at a university in Japan. None of the participants reported a history of hearing and speech impairments. Participants were randomly separated into two groups: the SAPT group (11) and the HVPT group (10). However, only nine participants in SAPT and eight participants in HVPT accomplished all phases of the experiments. As a result, our analysis was based on the data of 17 participants. The average age of the participants was 20.7 (*SD* = 2.7) years old. All the participants had more than one year of previous Mandarin learning experience (one year to six years), except for one participant in the HVPT group, whose previous experience was only one month. During the first month of the Mandarin lessons, Pinyin (the official romanization system for standard Mandarin Chinese) and pronunciation were the main tasks. As a result, none of the participants had any problems reading Pinyin.

Stimuli

The stimuli used in the experiment were syllables ending with nasal codas in four tones in Mandarin. The stimuli were digitally recorded using a TASCAM HD-P2 recorder in a sound-proof room with a sampling rate of 44,100 Hz.

Stimuli for training

We selected nine pairs of monosyllables ending with nasal codas with preceding vowels of “a /a/,” “e /ə/,” “i /i/” with four tones for our training stimuli[1]. Four native Mandarin speakers (2 males, 2 females) from areas where “n” and “ng” are well distinguished produced the stimuli. As a result, total stimuli amounted to 4 (talker) × 4 (tone) × 3 (vowel) × 2 (type: n/ng) × 9 (item) = 864. One hundred and twenty stimuli were randomly extracted from the 864 items in each training session. In this way, the diversity of stimuli met the needs of HVPT.

Stimuli for listening task

Seventy-six monosyllables ending with nasal codas were selected as the stimuli for LPRE and LPOST. These stimuli were composed of 38 syllables that were part of the training (old stimuli) and 38 syllables that were not part of the training (new stimuli). The voices of two more speakers were included in the listening task to investigate whether the training effect transferred to the new talkers’ voices. In addition, 24 disyllables with nasal codas in their preceding syllable and latter syllable were also included as testing items to verify whether the training effect transferred to the new phonetic environment.

Training Software

The training software applied in our study was introduced in Yang et al. (2018a); below, we provide only a brief overview of the training software. There are three modes in this training software: a learning mode, a test mode, and a practice mode. In the learning mode, participants listened to the correct model sound of monosyllables beginning with vowels and ending with a nasal coda. The times that participants listened to the model sound were recorded when the participant clicked the button.

In the test mode, as described in the stimuli session, all items were divided into four labels: talker, tone, vowel, and type. Participants were asked to conduct a test, which included 96 questions to verify perceptual ability before training. After the test, the score was presented to the listener.

In the practice mode, the system calculated the accuracy of each stimulus label in the test mode. In SAPT, the system picked up the lowest and the second-lowest accuracy in each label and generated the training items. The items for training were randomly selected from the lowest and second-lowest accuracy folders, that is, from 2 (talker) × 2 (tone) × 2 (vowel) × 2 (type) × 9 (item) = 144 items. In this way, we realized the SAPT. However, there was no adaptive conception in HVPT. The training item was selected from the system randomly, which is from 4 (talker) × 4 (tone) × 3 (vowel) × 2 (type) × 9 (item) = 864 items. Audio feedback was given after the listener made a selection on each question. After listeners pressed the button to choose "-n" or "-ng," a warning sound was presented to help listeners immediately judge if they had made the correct decision. If the listeners chose correctly, then a crisp sound played. If the listeners chose incorrectly, then a heavy sound played while the correct answer was presented in Pinyin. After that, the stimulus sound automatically played one more time to help the listener better understand the item.

Speech Task

The recording environment was the same as that in which the stimuli were recorded. During their orientation, participants were asked to be specific, enter the recording room one-by-one, pronounce the item presented, and try repronouncing the item if they were not satisfied with their first pronunciation. The test items were identical to the items in the listening task. Seventy-six items were presented to the participants three times in random order; hence, 228 items were included. While recording, a break occurred every 40 items.

Three Mandarin speakers judged participants' pronunciations. Before the judging, the judges conducted the listening task to investigate whether they could correctly identify the nasal codas. All judges demonstrated a good ability to identify the codas, evident in their full scores. They were required to listen to the pronunciations and write them down in Pinyin. Isbell (2016) indicates that there is an inconsistency between the measurement of perceptual ability and the measurement of productional ability since the chance rate is usually included in the design of the perceptual task, while it is not included in the design of the productional task. As a result, our judges were required to write the end part of the articulation with "n" or "ng" to facilitate the method by which the listening task was judged, namely, a two-alternative forced-choice identification task. Moreover, the productions of SPRE and SPOST were randomly reordered to prevent prejudice. The judges were allowed to listen to the pronunciations multiple times.

The summary of the experimental setup is shown in Table 1:

Table 1

Summary of the experimental setup

Procedure	# of participants	# of stimuli	# of talkers	Duration
LPRE	SAPT 11 HVPT 10	100 (monosyllables: 76, disyllables: 24)	6	10 mins
SPRE	SAPT 11 HVPT 10	Monosyllables: 76×3=228		50 mins
Training with Software: **Placement test**	SAPT 11 HVPT 10	Monosyllables: 96	4	10 mins
Training with Software: **Training**	SAPT 11 HVPT 10	Monosyllables: 120	4	20~30 mins
LPOST	SAPT 9 HVPT 8	Same as LPRE	6	10 mins
SPOST	SAPT 9 HVPT 8	Same as SPRE		50 mins

Data Analysis

All analyses were carried out using R (R Core Team, 2019), version 3.6.1. We applied nonparametric statistics to verify whether any differences in accuracy before and after training were significant in different groups because of the small sample size. In our study, significance was set at the 5% level. Moreover, effect size was used to verify the size of the relationship between or among the variables. For nonparametric statistics, the *r* value is applied to compare the means of two groups. The *r* value is classified as small ($.25 \leq r < .40$), medium ($.40 \leq r < .60$), or large ($.60 \leq r$) (Plonsky & Oswald, 2014). When comparing the means of three groups, we assume that ANOVA is sufficiently robust against the normality assumption since there is no corresponding nonparametric test. In this case, the effect size partial eta-squared (η_p^2) was used. η_p^2 is classified as small ($.01 \leq \eta_p^2 < .06$), medium ($.06 \leq \eta_p^2 < .14$), or large ($.14 \leq \eta_p^2$) (Cohen, 1988).

Results

Training Effectiveness with SAPT and HVPT in Perception

The means of two groups in LPRE and LPOST are shown in Figure 1. The mean accuracy in the SAPT group increased from 62.1% to 78.9%. A Wilcoxon signed-rank test was conducted, and the result showed that the difference was significant, and the effect size was large (Z = 2.549, p = .011, r =.601). The mean accuracy in the HVPT group increased from 55.1% to 68.9%. A Wilcoxon signed-rank test was conducted, and the result showed that the difference was also significant, and the effect size was medium (Z = 2.316, p = .021, r = .579). There was no significant difference between the mean accuracy of the two groups at LPRE (Z = 1.373, p = .170, r = .333), which means the two groups' perceptual ability had no significant difference before training. The participants in both groups demonstrated improved perceptual ability to identify nasal codas in Mandarin. Considering the effect sizes of the within-group comparisons, the SAPT group improved slightly more than the HVPT group.

We also listed the difference between each participant's LPRE accuracy and LPOST accuracy in Table 2. In the SAPT group, the difference was positive for eight participants and negative for one participant. In the HVPT group, the difference was positive for seven participants and negative for one participant.

We divided all the syllables with nasal codas into six types according to the vowel types: "-an [an]," "-ang [ɑŋ]," "-en [ən]," "-eng [ʌŋ]," "-in [in]," and "-ing [iŋ]" (Chao, 2011). To investigate the change during the perceptual training, we calculated the accuracy of nasal coda in each vowel type.

Figure 1

The mean accuracy of LPRE and LPOST in SAPT and HVPT in perception.

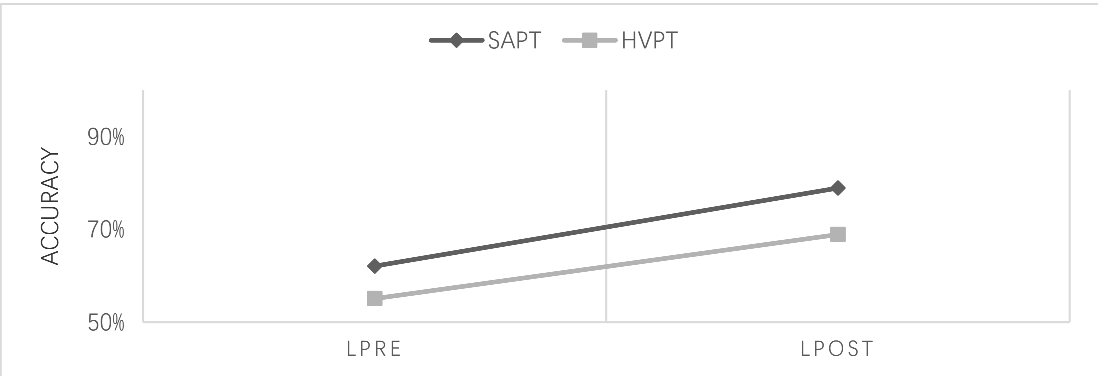


Table 2
Perceptual performance of participants in each group.

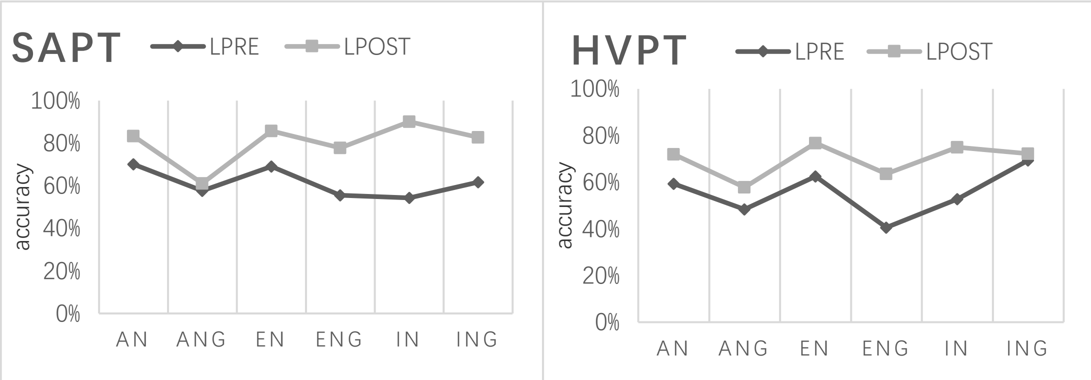
SAPT				HVPT			
Participants	LPRE	LPOST	Difference	Participants	LPRE	LPOST	Difference
SA01	78.9	86.8	+7.9	HV01	46.1	60.5	+14.5
SA02	48.7	82.9	+34.2	HV02	52.6	73.7	+21.1
SA03	71.1	84.2	+13.1	HV03	64.5	77.6	+13.2
SA04	63.2	88.2	+25.0	HV04	55.3	75.0	+19.7
SA05	71.1	77.6	+6.5	HV05	64.5	63.2	-1.3
SA06	59.2	77.6	+18.4	HV06	48.7	68.4	+19.7
SA07	52.6	77.6	+25.0	HV07	63.2	85.5	+22.4
SA08	57.9	52.6	-5.3	HV08	46.1	47.4	+1.3
SA09	56.6	82.9	+26.3				

In the SAPT group, we carried out a repeated-measures ANOVA with training (LPRE and LPOST) and vowel ("-an", "-ang", "-en", "-eng", "-in", "-ing"). The analysis showed a training × vowels interaction [F (5, 40) = 2.525, p = .045, η_p^2 = .240]. Simple main effect tests were conducted to verify whether there was a significant difference before and after the training in each vowel type[2]. The difference was only significant for "-in" [F (1, 8) = 18.899, p = .002, η_p^2 = .703]. However, it was not significant for "-an" [F (1, 8) = 11.108, p =.010, η_p^2 = .581], "-ang" [F (1, 8) =.357, p =.567, η_p^2 = .042], "-en" [F (1, 8) = 4.780, p = .060, η_p^2 = .374], "-eng" [F (1, 8) =6.095, p = .039, η_p^2 = .432], and "-ing" [F (1, 8) = 4.219, p = .074, η_p^2 = .345]. Although the difference of "-an", "-ang", "-en", "-eng", and "-ing" was not statistically significant, descriptive statistics as shown in Figure 2 (left) and the effect size suggest that pronunciations of "-an", "-en", "-eng", and "-ing" also improved throughout the SAPT.

In the HVPT group, the same repeated-measures ANOVA was conducted and the result revealed that there was no interaction between training and vowels [F (5, 35) = .826, p = .540, η_p^2 = .106]. To verify the change after the training, we carried out simple main effect tests on different vowel types[2]. The result showed that the difference was only significant for "-an" [F (1, 7) = 12.444, p = .001, η_p^2 = .640]. However, it was not significant for "-ang" [F (1, 7) = 1.400, p = .275, η_p^2 = .167], "-en" [F (1, 7) = 5.600, p = .050, η_p^2 = .444], "-eng" [F (1, 7) =11.29, p = .012, η_p^2 = .617], "-in" [F (1, 7) = 4.480, p = .072, η_p^2 = .390], and "-ing" [F (1, 7) = .043, p = .842, η_p^2 = .006]. Although the difference of "-ang", "-en", "-eng", "-in", and "-ing" was not statistically significant, descriptive statistics as shown in Figure 2 (right) and the effect size suggest that pronunciations of "-ang", "-en", "-eng", and "-in" also improved throughout the HVPT.

Figure 2

Perceptual accuracy of different vowels with nasal coda in SAPT (left) and HVPT (right).



Above all, it is apparent that "-ang" displayed little improvement in SAPT and "-ing" almost no improvement in HVPT. These results indicate that different training paradigms may lead to different changes in items.

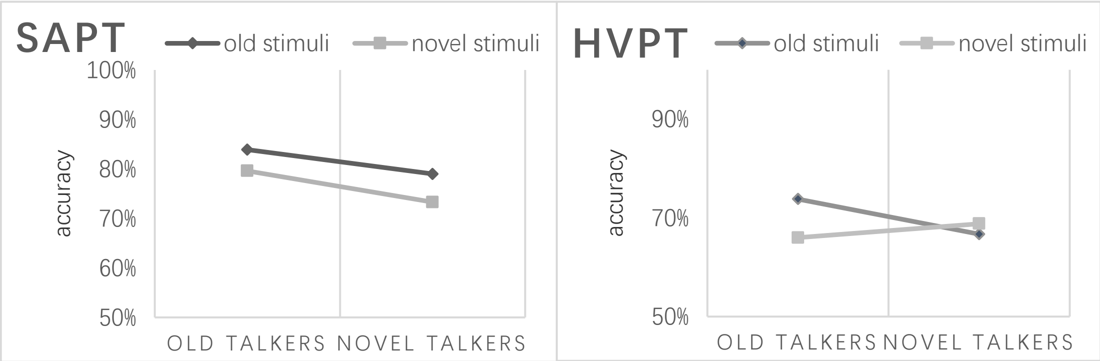
Transfer to new stimuli, new voices, and new phonetic environment

The accuracy of old stimuli, new stimuli, old talkers, and new talkers in LPOST in SAPT and HVPT (which can be compared in Figure 3) reveals the degree to which the training effect transferred from trained stimuli (old stimuli) to untrained stimuli (new stimuli) and from stimuli produced by talkers whose voices were included in the training (old talkers) to the stimuli produced by talkers whose voices were not included in the training (new talkers). A two-way ANOVA was applied to analyze the relationship between talkers and stimuli. For SAPT, there was no interaction between talkers and stimuli [F (1, 8) = .049, p = .830, η_p^2 = .006]. The main effect of the talker [F (1, 8) = 3.24, p = .110, η_p^2 = .288] was not significant. The main effect of stimuli [F (1, 8) = 2.059, p =.189, η_p^2 =.205] was not significant. Based on this result, we concluded that the training effect transferred from old talkers to new talkers and transferred from old stimuli to new stimuli in SAPT. In HVPT, there was no interaction between talkers and stimuli [F (1, 7) = 1.296, p = .292, η_p^2 =.156]. The main effects of the talkers [F (1, 7) = .617, p = .458, η_p^2 = .081] and stimuli [F (1, 7) = .354, p = .571, η_p^2 = .048] were not significant. These results indicate that the training effect did transfer from the old talkers and old stimuli to the new talkers and new stimuli in HVPT.

We also designed a disyllable test to verify whether the training effect was transferred to the new phonetic environment. In SAPT, the mean disyllable accuracy before training was 49.1%; after training, it increased to 54.2%. However, the increase was not significant since the p-value was over the significance level, and the effect size was small in the Wilcoxon signed-rank test (Z =.837, p = .405, r =.198). In the HVPT, the same trend was observed. The mean accuracy rose from 38.5% to 42.7%. The Wilcoxon signed-rank test showed that the increase was not significant (Z = 1.275,

Figure 3

The accuracy of old stimuli, new stimuli, old talkers, and new talkers (left: SAPT, right: HVPT).



p = .203, r =.319). These results suggest that the training effect did not transfer from the monosyllable environment to the disyllable environment in either the testing paradigm.

Transfer to Production

In this section, we report the production results of SAPT and HVPT. In speech pre-task (SPRE) and speech post-task (SPOST), the participants' pronunciation was judged by three native Mandarin speakers. When two or more judges considered a pronunciation correct, the item was judged as correct, and a circle was marked. The proportion of the circle to the total represents the accuracy of production. Since we have three judges, Kendall's W is more proper than Cronbach's alpha for the interrater reliability (Hove et al., 2018). Kendall's W was .785 ($p < .001$) for the SAPT group and .650 ($p < .001$) for the HVPT group; this confirms the reliability of the judgment results. Table 3 shows the change in production before and after training in two groups. The mean accuracy of production in the SAPT group increased from 63.7% to 70.5%. A Wilcoxon signed-rank test was conducted, and the result showed that the difference was significant (Z = 2.490, p = .013, r = .587). The mean accuracy in the HVPT group increased from 54.8% to 63.2%. A Wilcoxon signed-rank test was conducted, and the result showed that the difference was also significant (Z = 2.366, p = .018, r = .592). For the SAPT group, the difference was positive for eight participants and negative for one participant. For the HVPT group, the difference was positive for seven participants, and the accuracy stayed unchanged for one participant. These results suggest that the effect of perceptual training has a positive influence on production.

Table 3

Speech production performance of participants in each group.

SAPT				**HVPT**			
Participants	SPRE	SPOST	Difference	Participants	SPRE	SPOST	Difference
SA01	99.6	100	+0.4	HV01	50.0	56.6	+6.6
SA02	50.4	50.0	-0.4	HV02	53.1	53.1	0
SA03	59.2	74.6	+15.4	HV03	71.1	73.2	+2.1
SA04	57.0	69.7	+12.7	HV04	48.2	70.2	+22.0
SA05	95.6	97.8	+2.2	HV05	51.8	55.3	+3.5
SA06	53.5	63.2	+9.7	HV06	49.6	71.9	+22.3
SA07	54.4	57.9	+3.5	HV07	61.8	75.9	+14.1
SA08	49.6	56.6	+7.0	HV08	53.9	54.4	+0.5
SA09	54.4	64.5	+10.1				

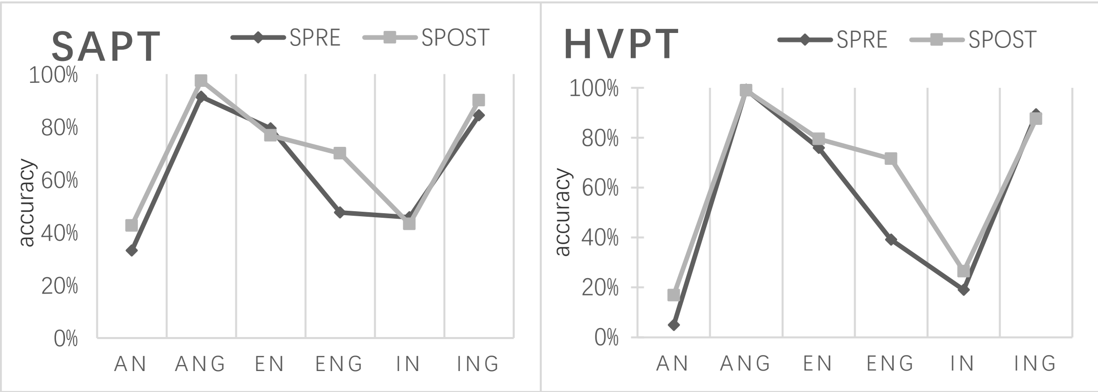
In addition, we investigated the accuracy of the nasal codas with different vowels. Figure 4 presents the mean accuracy of each vowel with a nasal coda in the SAPT group and the HVPT group. We carried out a repeated-measures ANOVA with training (SPRE and SPOST) and vowels ("-an," "-ang," "-en," "-eng," "-in," "-ing") in two groups. The result revealed that there was no interaction between training and vowels [F (5, 40) = 2.355, p = .058, η_p^2 = .227] in the SAPT group. To verify the change after the training, we also carried out simple main effect tests on different vowel types[2]. There was no significant difference for any vowel. However, the effect size of "-an" [F (1, 8) = 2.441, p = .157, η_p^2 = .234], "-ang" [F (1, 8) = 3.397, p = .103, η_p^2 = .298], and "-eng" [F (1, 8) = 7.780, p = .024, η_p^2= .493] was large.

In the HVPT group, the repeated-measures ANOVA was conducted and the result revealed that there was a training × vowels interaction [F (5, 35) = 4.275, p = .004, η_p^2 = .379]. The results of simple main effect tests showed that there was no significant difference in any vowel type[1]. However, the effect size of "-an" [F (1, 7) = 9.010, p = .020, η_p^2 = .563] and "-eng" [F (1, 7) = 9.054, p = .020, η_p^2 = .564] was large.

The items "-an" and "-eng" showed more improvement than other items in both SAPT and HVPT. Moreover, concerning the contrasts "-an, -ang," and "-in, -ing," participants were inclined to pronounce "-ng" instead of "-n." However, for the "-en, -eng" contrast, this trend was not observed.

Figure 4

Production accuracy of different vowels with nasal coda in SAPT (left) and HVPT (right).



Discussion

This study proposed a new perceptual training paradigm, SAPT, and compared the training effect with HVPT in a nonlaboratory and self-controlled setting. Regarding the first research question, "how effective are SAPT and non-adaptive HVPT in a nonlaboratory and self-controlled setting?", the results revealed that both training paradigms were effective for the participants. The effect size of training in SAPT was a little greater than that in HVPT, while the difference between the two effect size was not big. A possible explanation for the success of SAPT might be that SAPT chose the items that the participants were not good at and gave participants customized training. Prior studies note the effectiveness of HVPT from many perspectives. While HVPT may help some listeners to improve their perceptual ability, it sometimes ignores diversity—accordingly, it may not be effective for every listener. In the canonical HVPT research, participants were trained ground up regardless of the participants' level of learning and whether they have mastered some types of training targets. This makes training usually be long, and participants had to waste time on the items they mastered. SAPT, in our study, solved this problem and gave each participant customized training. Participants only need to concentrate on the items with which they were struggling. Unlike the adaptive training paradigm in Qian et al. (2018), our adaptive training paradigm picks up the lowest and the second-lowest accuracy items. It generates the training items in practice mode, having no concern with a cutoff score. The two items with low accuracy are updated with each training session. Because the two kinds of items with the lowest accuracy are always chosen, even if the accuracy is very high, the training can continue to achieve the goal that all the items can be trained eventually.

Besides, we observed that it was challenging for many listeners to improve their perceptual ability in "-in" and "-ing" via HVPT—this trend was verified by previous research (Yang et al., 2018b) as well as the HVPT experiment in this study. However, the accuracy of "-in" and "-ing" increased via SAPT. This result may be explained by the fact that different training paradigm prompts increase with different items. For example, since SAPT adapted to participant abilities and intensified training on the participants' weakest items, "-in" and "-ing" were the most trained items in SAPT. This proves that the perceptual ability of "-in" and "-ing" could be improved through adequate training.

Ultimately, both HVPT and SAPT help listeners to generate new perceptual categories for phonetic contrasts in L2. Wang et al. (1999) indicate that HVPT allowed listeners to contact the acoustic features of phonetic contrasts to the full range, making listeners generate new categories. Lively et al. (1994) note that selective attention to acoustic cues was crucial in improving perceptual ability. We wager that establishing a new perceptual category involves the transfer of explicit knowledge to implicit knowledge. The stimuli heard during training and feedback for each choice were given to the listeners as explicit knowledge. In the simple identification task, the listeners only needed to focus on the acoustic features of the stimuli and judge the types of nasal codas since the meanings of the stimuli in monosyllables had many possibilities, and it is possible to notice the acoustic cues from various stimuli produced by various talkers. Finally, the acoustic cues were gradually stored as implicit knowledge. The listeners who had successfully identified the nasal codas completed this process; meanwhile, the listeners who had not identified the nasal codas possibly failed in transferring explicit knowledge to implicit knowledge.

Regarding the second research question, the generalization of training effect to the perception of new stimuli, new talkers, and the new phonetic environment, in both SAPT and HVPT, the training effect generalized to the perception of new stimuli and new talkers. Concerning the new phonetic environment, both training paradigms did not show a positive transfer effect.

Looking at the SAPT from another angle: SAPT is a low variability phonetic training (LVPT) relative to HVPT. The discussion of appropriateness between HVPT and LVPT is long-standing. Some studies emphasize the importance of high variability in phonetic training (e.g., Lively et al., 1993; Wong, 2015; Bu et al., 2020). This may be because different talkers have different vocal tract sizes, glottal size functions, and speaking rates, and these factors help listeners get accustomed to the characteristics of new talkers (Wang et al., 1999). Others consider that LVPT may also contribute to forming perceptual categories for phonetic contrasts (Giannakopoulou et al., 2017; Sadakata & McQueen, 2014) and HVPT may not benefit children's L2 learning (Brekelmans, 2020). In this study, we designed a new LVPT in which the system picked up two items that listeners were not good at each level. In this way, the variability of stimuli and talkers decreased. Notably, both SAPT and HVPT helped participants develop their perceptual ability. This result suggests that LVPT may also improve listeners' perceptual ability when training items are targeted. Meanwhile, the transfer effect was verified in new talkers and new stimuli.

Further, little research has been done on this issue in terms of a new phonetic environment. Disyllables are more complex than monosyllables, and since both syllables in a disyllable end with nasal codas, listeners had to give great attention to judging the type of nasal coda. This may result in listeners experiencing difficulty in identifying the nasal codas.

The last research question examined the transfer effect of the perceptual training to production. The perceptual training effect has successfully transferred to the production domain in both SAPT and HVPT under the monosyllable environment. This result proves that perceptual training is effective in the perception domain and contributes to the production domain. This finding was also reported by many previous papers (Bradlow et al., 1997; Lambacher et al., 2005). To be sure, many studies attempted to verify the relationship between perception and production. Sakai and Moorman (2018) indicate two ways to probe this issue: 1) investigate the correlation analyses between perception and production of learners and 2) give perceptual training to learners and test the gains in the production domain. They carried out a meta-analysis of studies in which perceptual training was conducted, and gains in the production were tested in the past 25 years. The result revealed that perceptual training had a medium effect on the improvement of perceptual ability and a small effect on the gains of the production domain. The relationship between perception and production gains was small to medium.

The present study also compared the gains in perception and production in the same participant. Table 2 and Table 3 list the perceptual change and production change from before to after the training. The mean accuracy makes clear that for some participants, gains in perceptual ability led to gains in production ability (SA03, SA08, HV05); on the other hand, some participants' production ability preceded their perceptual ability (SA01, SA05, HV03). Moreover, we conducted a Spearman's rank correlation test between the difference in perceptual accuracy before and after the training and the difference in production accuracy before and after the training. The result suggested a small relationship between the two differences in SAPT ($r = .017$, $p = .996$), and a small relationship in HVPT ($r = .299$, $p = .471$). This result indicates that the participants who made great progress on the perceptual ability did not necessarily improve in the production domain. This result was not consistent with the theory that perception leads to production in SLM. Although it cannot be denied that there may be a ceiling effect in two participants in the SAPT group, both groups have improved their perceptual and production ability. However, the improvement in production is not as good as that in perception. We also cannot conclude that production accuracy rises with the increase of perceptual accuracy.

This contradiction is not uncommon. Bradlow et al. (1997) argue that individual differences may explain this phenomenon and stated that "the two processes proceeded at different rates within individual subjects." Flege (1999) insists that it takes time to realize the conversion to move the listener from the perceptual to the productive domain. It points that one possibility for this is that "not all subjects who have learned to perceive an L2 sound accurately will update segmental production to conform to their new, or modified long-term memory representation for the L2 sound" (Flege, 1999, p.1275). In a study investigating the link of perception and production, Nagle (2018) reported that the link might be not synchronous but is a time-lagged relationship, wherein gains in perception transfer to production at a later stage. Casillas (2016) also proved that perception and production improved at different phases and showed a time-lagged relationship. These findings may help us to understand that the relationship between perception and production is complicated. Various factors can influence the two processes, such as L1, motivation, L2's experience, perceptual aptitude, and imitation ability. These factors and longitudinal studies should be considered in detail in future work.

Conclusion

This study set out to examine the effectiveness of a new training paradigm of applied SAPT. In this training paradigm, listeners only needed to concentrate on items with which they were struggling. To be more specific, the software calculated the accuracy in a placement test and picked up the lowest and the second-lowest accuracy syllables with nasal codas from talker, tone, vowel, and type respectively. To compare this pedagogy with traditional HVPT, 17 participants completed the 12 training sessions in a self-controlled environment with the training software. Participants were divided into the SAPT group and the HVPT group. The experiment results showed that both SAPT and HVPT were effective in improving perceptual ability on nasal codas. The effect size of SAPT was a little larger than HVPT in perception. Moreover, the training effect was transferred to new stimuli and new talkers in SAPT and HVPT. Another major finding was that the training effect was also transferred to the production domain. These findings suggest that, in general, SAPT is an effective training paradigm for improving listeners' perceptual and production abilities to identify nasal codas in Mandarin. The new training paradigm reduces the time spent on unnecessary items and raises the efficiency of training. It provides a new possibility for learners who did not show progress from training with HVPT. This study also provides a new medium for self-regulated learning. Different from traditional classroom learning, the perceptual software which can help learners to conduct self-regulated learning anywhere at any time is developed. It can also be used as the carrier of the flipped classroom to provide a reference for the teaching mode reform of phonetic teaching. This study extends the focus from group learning effect to individual ability improvement, which meets the new requirements for L2 learning under the background of CALL with the characteristics of individualization and fragmentation.

Although we observed a change in the accuracy of the perception and production of nasal codas with different preceding vowels between the pre-test and post-test, the small sample size did not allow us to conclude that the different training paradigm led to the improvement of different items. Future work would do well to use a larger sample size. In addition, several questions remain unanswered. For example, a delayed post-test should be conducted to verify whether the training effect would be sustained after stopping the training. A longitudinal study should be considered in the future research design to investigate the relationship between perception and production. Besides native speakers' judgment, acoustic features should be analyzed to observe any changes before and after training. Further research on these questions would be a useful way to perfect this SAPT.

Notes

1. We selected nine pairs because only nine consonants can precede "-in."
2. *P*-value adjustment was made via the Bonferroni method for the simple main effect tests. Here, we defined a'=.008 (significance level 0.05 divided by 6 comparison pairs) as the significance level.

References

Barriuso, T. A., & Hayes-Harb, R. (2018). High Variability Phonetic Training as a Bridge From Research to Practice. *The CATESOL Journal*, *30*(1), 177–194.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (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

Brekelmans, G. (2020). *Phonetic vowel training for child second language learners : The role of input variability and training task.* [Unpublished doctoral dissertation]. University College London.

Bu, L., Nagono, Marisa., Harel, Daphna., & McAllister, Tara. (2020). Effects of practice variability on second-language speech production training. *Folia Phoniatrica et Logopaedica, Online First Article.* https://doi.org/ 10.1159/000510621

Cardoso, W. (2011). The development of coda perception in second language phonology: A variationist perspective. *Second Language Research*, *27*(4), 433–465. https://doi.org/10.1177/0267658311413540

Casillas, J. V. (2016). *The longitudinal development offine phonetic detail in late learners of Spanish*. [Unpublished doctoral dissertation]. The University of Arizona.

Chao, Y. R. (2011). *A Grammar of Spoken Chinese*. The Commercial Press.

Chen, M. Y. (2000). Acoustic analysis of simple vowels preceding a nasal in Standard Chinese. *Journal of Phonetics*, *28*(1), 43–67. https://doi.org/10.1006/jpho.2000.0106

Cheng, B., Zhang, X., Fan, S., & Zhang, Y. (2019). The role of temporal acoustic exaggeration in high variability phonetic training: A behavioral and ERP study. *Frontiers in Psychology*, *10*(MAY), 1–16. https://doi.org/10.3389/fpsyg.2019.

01178

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Flege, J. E. (1995). 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). York Press.

Flege, J. E. (1999). The Relation Between L2 Production and Perception. *Proceedings of the XIVth International Congress of Phonetics Sciences, 2*, 1273–1276.

Fuhrmeister, P., & Myers, E. B. (2020). Desirable and undesirable difficulties: Influences of variability, training schedule, and aptitude on nonnative phonetic learning. *Attention, Perception, & Psychophysics, 82,* 2049–2065. https://doi.org/10.3758/s13414-019-01925-y

Giannakopoulou, A., Brown, H., Clayards, M., & Wonnacott, E. (2017). High or low? Comparing high and low-variability phonetic training in adult and child second language learners. *PeerJ, 5,* e3209. https://doi.org/10.7717/peerj.3209

Golestani, N., & Zatorre, R. J. (2009). Individual differences in the acquisition of second language phonology. *Brain and Language*, *109*(2–3), 55–67. https://doi.org/10.1016/j.bandl.2008.01.005

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*, 360–378. https://doi.org/10.1016/j.specom.2005.04.007

Hove, D., Jorgensen, T. D., & van der Ark, L. A. (2018). On the usefulness of interrater reliability coefficients. In M. Wiberg, S. Culpepper, R. Janssen, J. González, & D. Molenaar (Eds.), *Quantitative Psychology: The 82nd Annual Meeting of the Psychometric Society, Zurich, Switzerland, 2017* (pp. 67-75). Springer. https://doi.org/10.1007/978-3-319-77249-3_6

Isbell, D. (2016). The Perception-Production Link in L2 Phonology. *MSU Working Papers in SLS*, *7*(March), 57–67.

Iverson, P., Pinet, M., & Evans, B. G. (2012). Auditory training for experienced and inexperienced second-language learners: Native French speakers learning English vowels. *Applied Psycholinguistics*, 33(1), 145-160. https://doi:10.1017/S0142716411000300

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

Leong, C. X. R., Price, J. M., Pitchford, N. J., & van, Heuven, W. J. B. (2018). High variability phonetic training in adaptive adverse conditions is rapid, effective, and sustained. *PLoS ONE*, *13*(10), 1–24.

Levis, J. (2007). Computer technology in teaching and researching pronunciation. *Annual Review of Applied Linguistics, 27*, 184–202. https://doi.org/10.1017/S0267190

508070098

Li, Y., Best, C. T., Cao, C., & Zhang, J. (2019). Hybrid perceptual training to facilitate the learning of nasal final contrasts by highly proficient Japanese learners of Mandarin. *The 19th International Congress of Phonetic Sciences ICPhS 2019*.

Linebaugh, G., & Roche, T. (2015). Evidence that L2 production training can enhance perception. *Journal of Academic Language & Learning*, *9*(1), 1–17.

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), 1242–1255. https://doi.org/10.1121/1.408177

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y., & Yamada, T. (1994). Training Japanese listeners to identify English /r/ and /l/. III. Long-term retention of new phonetic categories. *The Journal of the Acoustical Society of America*, *96*(4), 2076–2087. https://doi.org/10.1121/1.410149

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *The Journal of the Acoustical Society of America*, *89*(2), 874–886. https://doi.org/10.1121/1.406962

Mou, X. (2006). *Nasal codas in Standard Chinese-a study in the framework of the distinctive feature theory*. [Unpublished doctoral dissertation]. Massachusetts Institute of Technology.

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.

Nagle, C. L. (2018). Examining the Temporal Structure of the Perception – Production Link in Second Language Acquisition : A Longitudinal Study. *Language Learning*, *68*(1), 234–270. https://doi.org/10.1111/lang.12275

Nagle, C. L. (2020). Revisiting Perception–Production Relationships: Exploring a New Approach to Investigate Perception as a Time-Varying Predictor. *Language Learning*, 1-37. https://doi.org/10.1111/lang.12431

Olson, D. J. (2014). Benefits of visual feedback on segmental production in the L2 classroom. *Language, Learning and Technology*, *18*(3), 173–192.

Perrachione, T. K., Lee, J., Ha, L. Y. Y., & Wong, P. C. M. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *The Journal of the Acoustical Society of America*, *130*(1), 461–472. https://doi.org/10.1121/1.3593366

Plonsky, L., & Oswald, F. L. (2014). How Big Is “Big”? Interpreting Effect Sizes in L2 Research. *Language Learning*, *64*(4), 878–912. https://doi.org/10.1111/lang.12079

Qian, M., Chukharev-Hudilainen, E., & Levis, J. (2018). A system for adaptive high variability segmental perceptual training: implementation, effectiveness, and transfer. *Language Learning & Technology*, *22*(1), 69–96. https://doi.org/10125/44582

Rato, A. (2014). Effects of perceptual training on the identification of English vowels by native speakers of European Portuguese. *Concordia Working Papers in Applied Linguistics*, *5*, 529–546.

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Sadakata, M., & McQueen, J. M. (2014). Individual aptitude in Mandarin lexical tone perception predicts effectiveness of high-variability training. *Frontiers in Psychology*, *5*(NOV), 1–15. https://doi.org/10.3389/fpsyg.2014.01318

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

Saito, K. (2018). Advanced Second Language Segmental and Suprasegmental Acquisition. In P. A. Malovrh & A. G. Benati (Eds.), *The Handbook of Advanced Proficiency in Second Language Acquisition* (pp. 282–303). John Wiley & Sons.

Shport, I. A. (2016). Traning English listeners to identify pitch-accent patterns in Tokyo Japanese. *Studies in Second Language Acquisition*, *38*(4), 739–769. https://doi.org/10.1017/S027226311500039X

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). Cascadilla Press.

Thomson, R. I. (2011). Computer Assisted Pronunciation Training : Targeting Second Language Vowel Perception Improves Pronunciation. *CALICO Journal*, *28*(3), 744–765.

Thomson, R. I. (2012). Improving L2 Listeners' Perception of English Vowels: A Computer-Mediated Approach. *Language Learning*, *62*(4), 1231–1258. https://doi.org/10.1111/j.1467-9922.2012.00724.x

Thomson, R. I. (2018). High Variability [Pronunciation] Training(HVPT) A proven technique about which every language teacher and learner ought to know. *Journal of Second Language Pronunciation*, *4*(2), 208–231.

Thomson, R. I., & Derwing, T. M. (2016). Is phonemic training using nonsense or real words more effective. *Proceedings of the 7th Pronunciation in Second Language Learning and Teaching Conference*, 88–97.

Wang, X., & Munro, M. J. (2004). Computer-based training for learning English vowel contrasts. *System*, *32*, 539–552. https://doi.org/10.1016/j.system.2004.09.011

Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *The Journal of the Acoustical Society of America*, *106*(6), 3649–3658. https://doi.org/10.1121/1.428217

Wang, Y. (2002). An experimental study on the perception and production of nasal codas by Japanese Learners of Chinese Putonghua. *Chinese Teaching in the World*, *60*, 47–60.

Wiener, S., Chan, M. K. M., & Ito, K. (2020). Do Explicit Instruction and High

Variability Phonetic Training Improve Nonnative Speakers' Mandarin Tone Productions? *Modern Language Journal*, *104(1)*, 152–168. https://doi.org/10.1111/modl.12619

Wong, J. W. S. (2015). The effects of high-variability phonetic training on Cantonese ESL learners' production of English /ɪ/-/iː/contrast-an acoustics analysis. *Proceedings of the Phonetics Teaching and Learning Conference (PTLC2015)*, 107–111.

Yang, R., Nanjo, H., & Dantsuji, M. (2016). Nihonjin gakushūsha niyoru chūgokugo bion imbono chikaku ni okeru tokuchō-senkō shin ga haretsu shiin no baai- (The Perceptual Characteristics of Mandarin Chinese Nasal Codas by Japanese Speakers -with the condition of preceding consonants are plosives-). *Journal of the Japan Society for Speech Science*, *17*, 115–132.

Yang, R., Nanjo, H., & Dantsuji, M. (2018a). Development of Perceptual Training Software for Realizing High Variability Training Paradigm and Self adaptive Training Paradigm. *Proceedings of 32nd Pacific Asia Conference on Language, Information, and Computation*, 781–787.

Yang, R., Nanjo, H., & Dantsuji, M. (2018b). Training Japanese Speakers to Identify Nasal Codas of Mandarin Chinese. *Journal of Language Teaching and Research*, *9*(1), 7–15.

Zhang, J., & Wang, Z. (2017). Influences of vowels on the perception of nasal codas in Mandarin for Japanese and Chinese natives. *Journal of Tsinghua University (Sci&Technol)*, *57*(2), 164–169.