Hybrid perceptual training to facilitate the learning of nasal final contrasts by highly proficient Japanese learners of Mandarin

Yanping Li [1, 2], Catherine T. Best [2, 3], Chong Cao[1], Jinsong Zhang[1]

[1] Beijing Language and Culture University, Beijing, China
[2] Western Sydney University, NSW, Australia, [3] Haskins Laboratories, New Haven CT, USA
yanping.li@westernsydney.edu.au, c.best@westernsydney.edu.au
caochong2013@126.com, jinsong.zhang@blcu.edu.cn

ABSTRACT

Native speakers of Japanese experience challenges in differentiating Mandarin nasal finals, even after years of experience with Mandarin. We used a hybrid perceptual training approach with highly proficient Japanese learners of Mandarin to improve their ability to distinguish nasal final contrasts, which are not distinctive in Japanese. Eight learners participated in a 6-day adaptive and high-variability perceptual training procedure, including a pre-, mid-, and post-test on categorisation of nasal finals, whereas eight control participants received the same three tests without the intervening training. No significant pre-post performance changes were observed in the controls, whereas the trainees achieved an overall 13% improvement in identifying nasal final contrasts and better categorisation of nasal final continua. Additionally, they showed better generalisation to untrained nasal finals in both citation form and continuous speech. These findings suggest that hybrid adaptive and high-variability perceptual training helps facilitate highly proficient foreign learners' formation of non-native phonological representations.

Keywords: L2 learners, perceptual assimilation, non-native contrasts, adaptive/high-variability training.

1. INTRODUCTION

It may seem reasonable to assume that highly proficient second language (L2) users perceive L2 phonetic contrasts accurately. However, even for them some L2 perceptual difficulties do persist [1]. The degree of difficulty depends on the similarities and differences between the L2 learners' native (L1) phonemes and the L2 phones, and reflects persisting influences from the phonetic-phonological relations they have attuned to in their native language [2-3]. For example, highly proficient Japanese learners still have difficulty discriminating the Mandarin alveolar-velar nasal contrast /n/-/ŋ/ in final position (coda) [4]. This difficulty can hinder the learners' understanding of spoken Mandarin, as nasal finals (finals ending with nasal codas, i.e., [ən]/[əŋ]) occur in at least 46% of the most frequently used Mandarin monosyllables [5]. Japanese has a single "placeless" nasal final, the mora *N*, i.e., no place of articulation contrast. According to the Perceptual Assimilation Model (PAM) [2], Japanese-L2 listeners should perceptually assimilate both Mandarin nasal finals to their placeless *N* and show poor discrimination if they perceive both Mandarin nasals as equally good or bad exemplars of *N* (Single Category assimilation: SC). However, the nasal finals cause nasalisation of preceding vowels in both languages, and Mandarin listeners discriminate their nasal final contrast on the basis of the nasalised vowels [6-8]. As Japanese *N*-nasalised vowels have been reported to be more similar to the Mandarin velar than alveolar nasalised vowels [8], Japanese listeners may instead perceive the Mandarin nasal finals as a good versus a less-good *N*-nasalised vowel (Category Goodness difference assimilation: CG), which PAM [2] predicts to be easier to discriminate than an SC assimilation, and more amenable to L2 learning improvement [9].

This study focuses on aiding Japanese learners to overcome difficulties in discriminating Mandarin nasal final contrasts, using a hybrid perceptual training approach that utilises L2 phonetic variability as a tool for enhancing perceptual learning and generalisation. Phonetic variability is ubiquitous in verbal behaviour and perceivers use two complementary principles to detect more abstract phonological information from phonetic variability [10]: *phonological distinctiveness*, by which critical differences between phonetic segments distinguish similar-sounding words (i.e., c<u>a</u>ke from c<u>o</u>ke); and *phonological constancy*, which keeps word identity intact across lexically irrelevant variations such as talker differences. Exposing L2 learners to natural variations in non-native contrasts (high-variability perceptual training, HVPT) contributes to successful perceptual learning because phonetic variability helps L2 learners develop abstract representations that can accommodate a wider range of examples [11]. For example, [12] trained native speakers of Japanese with English /r/-/l/ distinctions produced by multiple speakers in variable word forms, resulting in better generalisation performance than [13], which trained

Japanese learners with limited variability (*rock-lock* contrast produced by a single speaker). HVPT has been applied to the learning of different L2 phonetic contrasts, such as English /u:/-/ʊ/ by Catalan/Spanish bilinguals [14], and English /w/-/v/ by native German speakers [15]. We used an HVPT approach in our study, reasoning it could help Japanese learners to form abstract representations of nasal final contrasts.

Unlike naïve L2 learners, however, highly proficient L2 users have been familiarised with the target phonetic contrasts produced by multiple talkers in variable lexical representations [16]. They may thus benefit less from HVPT than naïve learners. On the other hand, they may have failed to learn the critical perceptual cues for difficult L2 contrasts [17] such as nasal finals, i.e., they may not have learned phonological distinctiveness for those contrasts. Thus, training highly proficient listeners to perceive critical perceptual cues calls for modifications of the HVPT paradigm. Adaptive perceptual training (APT) meets this challenge by exposing L2 learners to an acoustic continuum synthesised with the primary perceptual cues, which is expected to aid L2 learners in gradually forming non-native abstract representations of the critical cues to the contrast [18]. A hybrid perceptual training paradigm of APT followed by HVPT has been shown to be effective in improving Japanese learners' perception of Mandarin tones [19]. However, little attention has been paid to improve perception of Mandarin nasal final contrasts by highly proficient Japanese learners. This study thus used the hybrid perceptual training with highly proficient Japanese learners to train them on Mandarin nasal finals.

2. METHOD

2.1. Participants

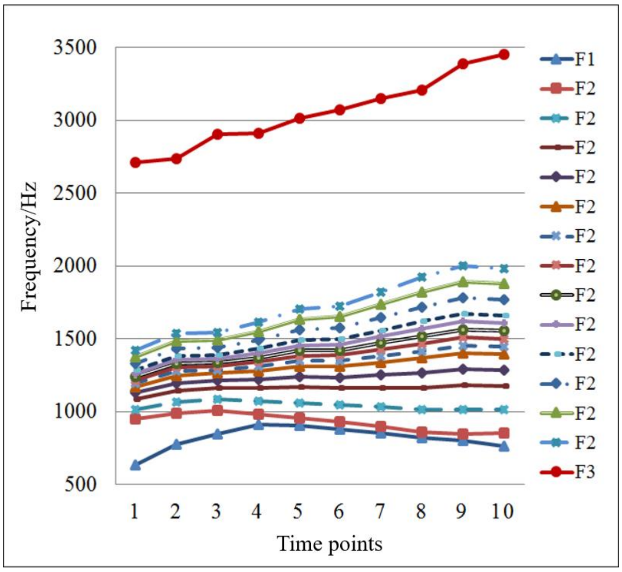
16 native speakers of Japanese were recruited, eight as trainees (2 males, 6 females; Mean = 23 years old) and eight as controls (3 males, 5 females; Mean = 23 years old). All were students at Beijing Language and Culture University who had learned Mandarin for 2 years, with a mean Mandarin proficiency of five (on a scale of 1-6, from beginner to advanced, where 5 refers to advanced), as certified by HSK (Hanyu Shuiping Kaoshi) [20].

2.2. Stimuli

2.2.1. Nasal final continua

Four nasal final contrasts ("fan/fang", "shan/shang", "fen/feng", and "shen/sheng") produced by a female speaker were chosen from Chinese Inter-Language Corpus of Beijing Language and Culture University for Computer Assisted Pronunciation Training (CAPT) as the original materials for creation of synthetic stimuli for APT training. Eight nasal final continua ("fan-fang/fang-fan", "shan-shang/shang-shan", "fen-feng/feng-fen", and "shen-sheng/sheng-shen") were synthesised by varying the second formant (*F*2) of the vowel part in nasal finals, which is the critical perceptual cues for nasal finals [21]. Figure 1 shows the synthetic continuum in "fang-fan" group, where the x-axis shows 10 equal time points of the stimulus syllable and the y-axis displays frequency values. The blue line at the bottom with triangle markers is the first formant (*F*1), and the red line on the top with circle markers is the third formant (*F*3), for all stimuli. *F*2 varied from "fang" (lowest *F*2) to "fan" in 13 equal steps, resulting in 13 stimuli (13 *F*2 contours). There were 104 stimuli (13 stimuli × 8 continua) used in APT and 312 stimuli (13 stimuli × 8 continua × 3 repetitions) used in the categorisation tests.

Figure 1: Synthetic "fang-fan" continuum



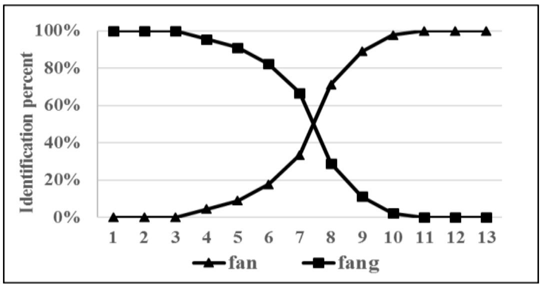
2.2.2. Categorisation of nasal final continua

Fifteen Chinese listeners (5 males, 10 females; Mean = 25 years old) born in the northern China provided native baseline perceptual data on the continua. They heard 312 trials in random order via E-prime 2.0 professional [22], blocked by 78 trials, in a two-alternative identification task pressing *F* for alveolar nasal finals and *J* for velar nasal finals. The inter-trial interval was 3000ms. It took each listener one hour to complete categorisation tests. Figure 2 shows the categorisation of "fang-fan" continuum. All eight nasal final continua were perceived categorically by Chinese listeners, and served as Chinese Perceptual Models (CPMs). Therefore, the continua were used as training materials in APT.

2.2.3. Nasal final words

1216 real Mandarin words (152 words × 8 female speakers) were chosen from CAPT with various syllabic structures to ensure context variability for use in HVPT training. All had nasal finals. There were 76 stimuli in each of the pre-test, mid-test, and post-test and 30 stimuli in daily progress tests. 60 stimuli produced by a trained and a novel speaker were prepared for the word and speaker generalisation tests, respectively. In addition, 60 sentences consisting of nasal final words in word generalisation test were recorded by the novel speaker in speaker generalisation test with the carrier sentence "yí gè X zì" ("X is a Chinese character") [23] for the sentence generalisation test.

Figure 2: Identification functions for "fang-fan" continuum



2.3. Procedures

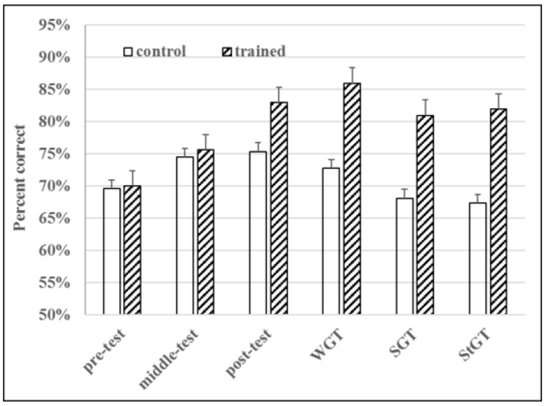
Following a pre-test on categorisation and identification of nasal finals, a 2-day APT training was conducted with the 8 Japanese learners using the nasal final continua. After finishing the mid-test, they started a 4-day HVPT procedure, followed by a post-test and three generalisation tests. There was a 5-minute daily progress test at the end of each training session. No more than one training session was run in a day. The controls only completed the pre-, mid-, and post-tests without training. The procedures for testing were similar to the categorisation tests of nasal final continua by native Chinese listeners.

During the APT training, trainees were presented with CPMs in eight slides via Powerpoint 2013 that labelled the alveolar and velar nasal final categories. For each continuum, they distinguished nasal final contrasts by pressing them audibly. A "test+study" mode was used for the HVPT training, using E-prime 2.0. On each trial, participants were asked to make a two-alternative identification judgment with immediate feedback. They could continue to listen to the target word until they mastered it. There were 304 trials (152 words × 2 speakers) in each training session blocked by 76 trials. Speakers were blocked and there was a short break after each block. The inter-trial interval was 3000ms. Each training session took an hour to complete.

3. RESULTS

The correct identification scores for the trained and control groups at pre-test, mid-test, post-test, word generalisation, speaker generalisation, and sentence generalisation are shown in Figure 3. The trainees showed improvements from the pre-test (70% correct identification) to the mid-test (76%), to the post-test (83%), a substantial 13% increase in nasal final identification accuracy. This increase was also observed in the three generalisation tests (86% correct identification in WGT; 81% in SGT; and 82% in StGT). In contrast, the control group exhibited a decreasing tendency across the three generalisation tests (73% in WGT; 68% in SGT; 67% in StGT), although they started at the same level as the trainees in the pre-test (70% correct identification) and exhibited a little improvement in the mid-test (75%), but no further improvement in the post-test (75%).

Figure 3: Mean percent correct identification of nasal finals for trained and control groups at the pre-test, mid-test, and post-test, word generalisation (WGT), speaker generalisation (SGT), and sentence generalisation (StGT) test. Error bars indicate the standard error of the mean.



The results were analysed using two two-way ANOVAs of test and group (trained, control), with test as the repeated measure. There was a significant main effect of test (pre-, mid-, versus post-test) [F (2, 28) = 7.278, p = .009], and significant main effects of test (WGT, SGT, versus StGT) [F (2, 28) = 4.374, p = .027] and group [F (1, 14) = 5.805, p = .030]. To further investigate these effects, we conducted two one-way ANOVAs with test (pre-, mid-, versus post-test) as the repeated measure. There was no significant main effect of test [F (2, 14) = 3.260, p = 0.069] for the control group, whereas a significant main effect of test [F (2, 14) = 4.426, p = 0.031] for the trained group, and post hoc comparisons showed that there was a significant difference between pre-test and post-test (t = -2.368, p = 0.05). These results indicate that trainees significantly improved their identification of nasal final contrasts. Another two one-way ANOVAs were conducted with test (WGT, SGT, versus StGT) as the repeated measure. There was significant main effect of test [F (3, 21) = 3.383, p = 0.037] for the control group, whereas no significant main effect of test [F (3, 21) = 0.972, p = 0.425] for the trained group. These results suggest that the trainees could generalise the trained words and speakers to novel ones, whereas the control's daily Mandarin input failed to help them distinguish nasal final contrasts.

Figure 4: Learning progress in the trained group. Error bars indicate the standard error of the mean.

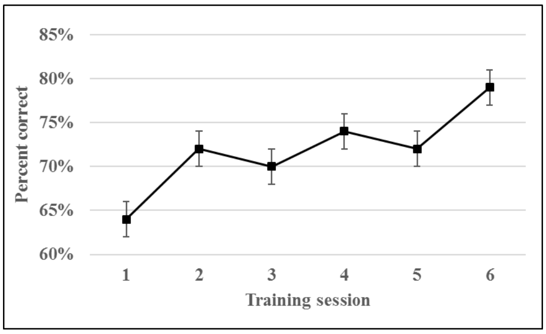


Figure 4 shows the learning progress in the trained group during the 6 training sessions. There was an improvement from training session 1 (64% correct identification) to training session 2 (72% correct identification), which was the end of the APT phase. The trainees experienced a fluctuating improvement in the following four training sessions, achieving a 79% correct identification score in training session 6. The overall results of daily tests were submitted to a one-way ANOVA test with the test session as the repeated measure. Results showed that there was significant main effect [F (5, 35) = 3.267, p = 0.016] and post hoc comparisons showed that there were significant differences between training sessions 1 and 2 (t = -3.317, p = 0.013), which indicates that trainees may master the critical perceptual cues for nasal final contrasts. There was significant difference between training sessions 1 and 6 (t = -3.349, p = 0.012), but no significant differences between the training sessions 2 and 6 (t = -1.623, p = 0.149), which is consistent with perceptual results between the pre- and post-test and between the mid and post-test.

The overall improvements in the trained group could not identify the effects of the APT training on abstract representation formation. An error score proposed by [17] was therefore introduced to observe differences in categorisation of nasal final continua between native Chinese listeners and Japanese learners, which is defined in (1):

(1) $$Z_j = \sqrt{\frac{1}{13}\sum_{i=1}^{13}(S_{s,i} - S_{j,i})^2}$$

where j is the number of subjects (j=1-8), i is the number of stimuli in the continua (i=1-13), S_s is the perceptual results of native Chinese listeners, S_i is Japanese learners, and Z_j refers to error score between Chinese and Japanese learners in terms of categorical perception. The lower the error score, the smaller the differences between the trainees and native listener, indicating better training efficiency. The error scores, for example, on the "fang-fan" continuum, decreased from 0.43 in the pre-test to 0.38 in the middle-test and 0.39 in the post-test.

4. DISCUSSION

The perceptual results in the pre-test are consistent with the conclusion in [4] that highly proficient Japanese learners have difficulty in distinguishing Mandarin nasal final contrasts. Moreover, the control failed to improve by only receiving daily input in the non-native Mandarin language environment, which provides them with variable phonetic variations of the contrasting categories, suggesting that they fail to detect the critical perceptual cues for the phonological distinctiveness of the nasal final contrast. This problem was solved by the design of APT in this study. Because the trainees' categorisation of nasal final continua improved after APT and stayed approximately unchanged in the post-test. These results suggest that APT helped Japanese learners grasp the crucial perceptual cues for nasal final distinction, paving the way for forming phonological representations of nasal final contrasts.

It is interesting that there was a significant difference in correct identification between training sessions 1 and 2, but no significant difference between the pre- and mid-tests. We noticed that the correct identification decreased after the first APT training session, which may be that participants have difficulty in switching from synthetic to natural stimuli [13], because the synthetic stimuli lack speech information on phonological constancy [10]. However, the successful generalisation to novel words, novel speakers, and even continuous speech supports the idea that the trainees formed phonological representations of nasal final contrasts after HVPT, which provides L2 learners with information not only on phonological distinctiveness, but phonological constancy, making up the failure of APT in generalising to natural speech.

The hybrid perceptual training paradigm in this study appears to facilitate the formation of phonological representations for highly proficient Japanese learners, which proves that native language influence on perception of non-native contrasts is not absolute or permanent [24] as long as learners get trained with appropriate perceptual training methods. In addition, the paradigm can also be extended from L2 supra-segmental [19] to segmental perception.

5. REFERENCES

[1] Broersma, M., Cutler, A. 2008. Phantom word activation in L2. *System* 36 (1), 22–34.

[2] Best, C. T. 1995. A direct realist view of cross-language speech perception. In: Strange, W. (ed), *Speech Perception and Linguistic Experience: Issues in cross-language research*. Baltimore, Md.: York Press, 171-204.

[3] Flege, J. E. 1995. Second language speech learning: theory, findings, and problems. In: Strange, W. (ed), *Speech Perception and Linguistic Experience: Issues in cross-language research.* Baltimore, Md.: York Press, 233-277.

[4] Wang, Y. J. 2002. An experimental study on the perception and production of nasal finals by Japanese students. *Chinese Teaching in the World* 60 (2), 47-60.

[5] Dong, Y. G. 1997. Teaching Japanese students to learn nasal finals. *Chinese Teaching in the World* 42 (4), 66-70.

[6] Wang, Z. Y., Zhang, J. S. 2014. Influences of vowels on perception of nasal codas in Mandarin for Japanese learners and Chinese. *Proc. 9th ISCSLP* Singapore, 433-438.

[7] Zhang, J. S., Wang, H., Wang, Z. Y., Cao, W. 2013. The influence of vowel segments on Japanese learners' perception of Chinese nasal codas. *Proc.12th NCMMSC* Guiyang, China.

[8] Wang, Z. Y. 2014. The acoustic differences of nasal codas in Mandarin and Japanese and their influences on the perception of Mandarin nasal finals. Masters dissertation, Beijing Language and Culture University.

[9] Best, C. T., Tyler, M. D. 2007. Nonnative and second-language speech perception: Commonalities and complementarities. In: M. J. Munro, O.-S. Bohn (eds.), *Second language speech learning*: *The role of language experience in speech perception and production*. Amsterdam: John Benjamins, 13-34.

[10] Best, C. T. 2015. Devil or angel in the details? Perceiving phonetic variation as information about phonological structure. In: Romero, J. and Riera, M. (eds.), *Phonetics-Phonology Interface*: *Representations and Methodologies*, 3-31.

[11] Sadakata, M., McQueen, J. M. 2013. High stimulus variability in nonnative speech learning supports formation of abstract categories: Evidence from Japanese geminates. *The Journal of the Acoustical Society of America* 134 (2), 1324–1335.

[12] 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.

[13] Strange, W., Dittmann, S. 1984. Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception & Psychophysics* 36 (2), 131–145.

[14] Aliaga-García, C., Mora, Joan C. 2009. Assessing the effects of phonetic training on L2 sound perception and production. In: Watkins, M. A., Rauber, A. S., Baptista, B. O. (eds.), *Recent Research in Second Language Phonetics/Phonology: Perception and Production*. Newcastle upon Tyne, UK: Cambridge Scholars Publishing, 2-31.

[15] Iverson, P., Ekanayake, D., Hamann, S., Sennema, A., Evans, B. G. 2008. Category and Perceptual Interference in Second-Language Phoneme Learning: An Examination of English /w/-/v/ Learning by Sinhala, German, and Dutch Speakers. *Journal of Experimental Psychology: Human Perception and Performance* 34 (5), 1305–1316.

[16] Eger, N. A., Reinisch, E. 2017. The role of acoustic cues and listener proficiency in the perception of accent in non-native sounds. *Studies in Second Language Acquisition*, 1-22.

[17] Wong, J. W. S. (2015). The impact of L2 proficiency in vowel training. In J. A. Mompean and J. Fouz-González (eds.), *Investigating English pronunciation: Current Trends and Directions*. Hampshire: Palgrave Macmillan, 219-239.

[18] Lintern, G., Gopher, D. 1978. Adaptive training of perceptual-motor skills: issues, results, and future directions. *International Journal of Man-Machine Studies* 10 (5), 521-551.

[19] Feng, X. L., Sun, Y., Zhang, J. S., Xie, Y. L. 2014. A study on the long-term retention effects of Japanese C2L learners to distinguish Mandarin Chinese Tone 2 and Tone 3 after perceptual training. *Proc. 9th ISCSLP* Singapore, 599-603.

[20] Hanyu Shuiping Kaoshi (2010). Retrieved from https://en.wikipedia.org/wiki/Hanyu_Shuiping_Kaoshi.

[21] Li, Y. P., Xie, Y. L., Feng, L. D., Zhang, J. S. 2016 The perceptual cues for nasal Finals in Standard Chinese. *Proc. 10th ISCSLP* Tianjin, China.

[22] E-prime (Version 2.0 professional) [computer software]. Pittsburgh, PA: Psychology Software Tools.

[23] Hallé, P. A., Chang, Y.-C., Best, C. T. 2004. Identification and discrimination of Mandarin Chinese tones by mandarin Chinese vs. French listeners. *Journal of Phonetics* 32 (3), 395-421.

[24] Best, C. T. 1994. The emergence of native-language phonological influences in infants: A perceptual assimilation model. In: Goodman, J., Nusbaum, H. C. (eds.), *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words*. Cambridge MA: MIT Press,167-224.