INVESTIGATING THE EFFECTIVENESS OF AUDITORY TRAINING ON CHINESE LISTENERS' PERCEPTION OF ENGLISH VOWELS

Jian Gong, Zhenzhen Yang, Xiaoli Ji, Feng Wang

Phonetics Lab, Jiangsu University of Science and Technology, China
j.gong@just.edu.cn, 2247822562@qq.com, jixiaoli2005@126.com, 1243849736@qq.com

ABSTRACT

Previous studies have demonstrated that high variability auditory training is generally effective. However more recent studies also revealed that the effectiveness of training can be task based. In the current study, a group of native Chinese learners of English underwent a 16-session identification training on English vowels. Two types of tasks, including category identification and discrimination were used in the pre- and post-tests. Although significant training effect were found for both tasks, learners' performance improvement in the discrimination task was very limited, comparing to a near thirty percent points improvement in the identification task. This result provides more evidence that the effectiveness of high variability training is task based, supporting the claim that learners' underlying fundamental L2 category processing mechanism may not be really changed by training.

Keywords: Auditory training, identification, discrimination, vowels.

1. INTRODUCTION

Acquiring the sounds in a second language (L2) is often difficult for adult language learners. However Speech Learning Model [4] claims that learners' L2 sound learning ability remains intact over the life span. Auditory training has been used for decades as a method to improve learners' L2 sound perception. Over the past 20 years, high-variability phonetic training (i.e., using multi-talker, multi-phonetic environment, natural speech etc.) has gained enormous attention and has been adopted in various kind of training tasks, including training learners' perception ability of consonant [14, 15, 2, 8], vowels [12, 15], tones [22], and their phoneme production ability as well [11, 21].

Generally, the high-variability training has been reported effective in most of the researches, however more and more studies have revealed that the effectiveness of training might not be as straightforward as expected. For example, [20] found perception training can improve the production while [9] found there was no significant production improvement after training. Results from some other studies demonstrated that high-variability training was not always more effective than low-variability training [3, 16, 19, 7].

Normally there are two kinds of perception tasks in the auditory training, namely identification and discrimination. It is claimed that identification task and discrimination task tap into different aspects of perception [5, 11]. Recently there were several studies looked into the differences of effectiveness between identification training and discrimination training, the results showed that both identification training and discrimination training can improve learners' L2 perception [17, 21], but no clear difference between the two methods was found [21].

Another specific issue is whether auditory training can improve learners' performance in both identification and discrimination. Recent studies revealed that the identification training may not be as effective in discrimination task as in identification task [11], indicating the effectiveness of identification training is task based. Therefore the focus of the current study is to investigate whether identification training has the same effects on both identification and discrimination. More specifically, native Chinese learners' identification and discrimination of English vowels were examined before and after high-variability identification training.

2. METHOD

The current study followed the pre-test—training—post-test paradigm. Two types of perception test, namely, category identification and category discrimination were included in the pre/post tests, while 16 sessions of identification training were conducted in between.

2.1. Subjects

24 native Chinese subjects, including 19 males and 5 females, were recruited in the current study. These subjects were undergraduate and postgraduate stu-

dents from Jiangsu University of Science and Technology, China, studying non English major courses, with the age ranged from 20 to 30 years. These subjects were relatively low level English learners, without any reported hearing problems. They were all from the central-east (Jianghuai) Mandarin dialect spoken region. These subjects were randomly divided into two groups, a control group and an experiment group, each contained 12 subjects. The subjects in the control group were only given the pre- and post-tests while the subjects in the experiment group finished the whole pre-test—training—post-test program.

2.2. Materials

Real English words with 15 English vowels (including /ɪ i: e æ ʌ ɑ: ɒ ɔ: ʊ u: eɪ aɪ ɔɪ əʊ aʊ/) were used as stimuli in the current study. The stimuli were recorded from 9 native British speakers (5 males and 4 females). They were university teachers in UK or English teachers in China. They were asked to produce all the stimuli with RP. The recordings took place in sound attenuated research labs in UK and China. Stimuli from 3 speakers (2 males and 1 female) were used in the pre/post tests, while stimuli from the rest 6 speakers were used in training.

For the pre/post test stimuli, the 15 vowels were put into a /hVd/ context (e.g., *hid, heed, head, had*). For the training stimuli, the 15 vowels were divided into 4 blocks (/e æ ʌ ɑ:/, /ɪ i: eɪ aɪ/, /ɒ ɔ: əʊ/, /ʊ u: ɔɪ aʊ/). Minimal pair words with various forms of consonant context were recorded for each block (e.g., *bed, bad, bard, bud* and *met, mat, mart, mutt* for /e æ ʌ ɑ:/, *feed, fid, fade, fide* and *sleet, slit, slate, slight* for /ɪ i: eɪ aɪ/) to form highly variable phonetic environments.

2.3. Procedure

The pre/post tests were carried out one day before/after the training program. Subjects first finished the category discrimination test, then the category identification test. In the category discrimination test, each of the 15 vowels was paired with another vowel to form 21 "different" contrasts, each contained 10 stimuli trials. And the 15 vowels also paired with themselves to form 15 "same" contrasts, each contained 14 trials. These particular "different" contrasts were among the most difficult contrasts for Chinese learners according to previous study's results. The "same" contrasts served as fillers. Subjects were asked to judge whether the two /hVd/ words they heard in each trial are the same, and click the button shown on a computer screen with the word "same" or "different". All together, there were 420 trials in the discrimination test. The two stimuli in each trial were always from different speakers, and were presented with a 300ms interval. In the category identification test, subjects were asked to classify the vowel in each /hVd/ word they heard into one of the 15 vowel categories, by clicking the corresponding button on the computer screen with the correct /hVd/ words. There were 270 tokens all together, 18 tokens for each vowel. The order of the stimuli presentation in both discrimination and identification tests were random. Subjects from both groups had the same time interval between pre and post tests.

The training had 16 sessions, each contained 4 blocks as described earlier. There were 10 tokens for each vowel, 150 tokens for each session all together. The presentation sequence of the 4 blocks in each session was random to each subject. Subjects were required to identify the vowel in the word they heard, and click the correspondent button on the screen from the 4 or 3 options according to the different blocks. In the training, although subjects heard various forms of words with the 15 vowels, they only saw the /hVd/ word on screen. If they made a wrong click, the correct answer would be highlighted and they had to click the right answer and listened again to proceed. The whole training was finished in 4 consecutive days, 4 sessions for each day.

3. RESULTS

3.1. Identification

Fig. 1 shows the identification accuracy of control group and experiment group before and after training. Repeated-measures of ANOVA confirmed that there was a significant main effect of test (pre-post) [$F(1,22) = 353.4, p < .001, \eta_p^2 = .941$], a significant difference between the two groups [$F(1,22) = 14.3, p < .01, \eta_p^2 = .394$], and a significant interaction between test and group [$F(1,22) = 314.5, p < .001, \eta_p^2 = .935$]. Further simple effect analysis with Bonferroni adjustment proved that there was no significant difference of identification accuracy between the two groups before training ($p > .05$), but a significant difference after training ($p < .001$). There was almost no change of identification performance for the control group after training (55.1% for pre-test and 55.9% for post-test, $p > .05$), while the experiment group had a significant 28 percentage points improvement ($p < .001$) from pre-test (51.8%) to post-tests (79.8%).

Figure 1: Mean identification accuracy% over all vowels for both groups.

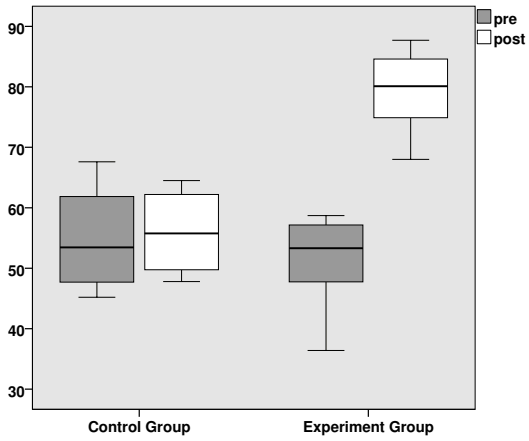

Table. 1 shows the experiment group's identification accuracy for individual vowels before and after training. Paired-samples t-tests indicated that only 3 vowels' improvements were not reached statistical significance. Among these 3 vowels, /ɔɪ/ had reached 100% accuracy while the other two, /e/ and /æ/, were among the most difficult sounds to identify for Chinese learners after training. In fact, Chinese learners had a lot of mutual confusions between these two vowels in both pre- and post-tests. Another noteworthy result is that most of the lax vowels such as /ɪ, e, æ, ɒ, ʊ/ were worse identified than their tense counterparts and than most of the diphthongs, especially after training.

Table 1: Identification correct% for individual vowels in pre- and post-tests for experiment group. P-values were from separate paired-samples t-tests performed for each vowel.

Vowels	pre	post	improve	sig.
ɪ	50	64	14	$p < .05$
iː	69	98	29	$p < .05$
e	53	65	12	$p = .139$
æ	38	52	14	$p = .119$
ʌ	16	73	57	$p < .05$
ɑː	59	85	26	$p < .05$
ɒ	54	75	21	$p < .05$
ɔː	30	75	45	$p < .05$
ʊ	32	62	30	$p < .05$
uː	47	84	37	$p < .05$
eɪ	43	90	47	$p < .05$
aɪ	81	99	18	$p < .05$
ɔɪ	89	100	11	$p = .06$
əʊ	43	81	38	$p < .05$
aʊ	72	95	23	$p < .05$
mean	51.7	79.9	28.2	$p < .05$

3.2. Discrimination

The mean discrimination accuracy across all 36 vowel contrasts for both control and experiment groups before and after training are shown in Fig. 2. Repeated-measures of ANOVA indicated that there was a significant test effect [$F(1,22) = 18.9, p < .001, \eta_p^2 = .462$]. However, not like in the identification test, there was no significant group difference [$F(1,22) = 3.3, p > .05, \eta_p^2 = .131$]. There was a significant test*group interaction [$F(1,22) = 9.7, p < .01, \eta_p^2 = .305$]. Further simple effect analysis with Bonferroni adjustment revealed that there was no significant difference between control group (62%) and experiment group (62.6%) before training ($p > .05$). After training, the experiment group's performance slightly but significantly improved (66.7%, $p < .001$) while no significant change was found for the control group (62.6%, $p > .05$). The experiment group's overall discrimination accuracy was slightly but significantly better than the control group's after training ($p < .01$).

Figure 2: Mean discrimination accuracy% over all vowel contrasts for both groups.

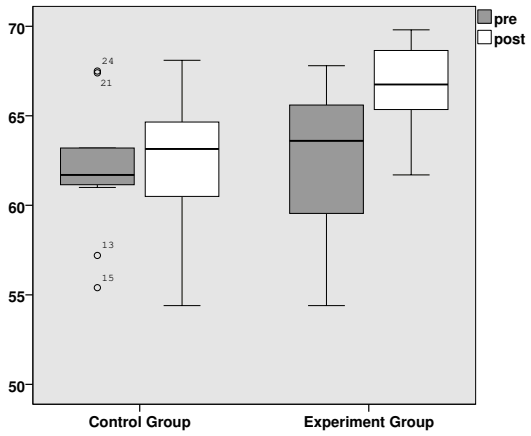

Table 2 lists 10 vowel contrasts with the lowest discrimination accuracy in post-test for the experiment group, and their discrimination accuracy in pre-test as well. The table also provides the mean discrimination accuracy for "same" vowel contrasts, "different" vowel contrasts and overall mean across all contrasts separately. It can be seen that learners' discrimination accuracy for "same" vowel contrasts didn't change much after training, at a quite high level in both pre- and post-tests (over 90%). On the contrary, learners' discrimination accuracy for "different" vowel contrasts remained at a low level even after training, with a significant but rather small 6.7 percentage points improvement from pre- to post-

test. In fact, the 10 most difficult contrasts in post-test were all "different" vowel contrasts, and 9 of them were also among the 10 most difficult contrasts in pre-test. Table 2 also shows that half of these 10 vowel contrasts' discrimination accuracy actually didn't improve significantly after training, indicating a rather limited training effect for many of the tense-lax or lax-lax vowel contrasts.

3.3. Relation between identification and discrimination

A series of Pearson's correlation coefficient analyses were carried out on experiment group to investigate the relation between learners' identification and discrimination performances. The results showed that there was significant medium correlation between listeners' identification accuracy and discrimination accuracy in pre-test ($r = .641, p < .05$) but no significant correlation in post-test ($r = .466, p = .127$). Further analysis demonstrated that there was no significant correlation between learner's identification improvement and discrimination improvement from pre to post-test ($r = .555, p = .061$). These results thus indicated that the effectiveness of auditory training may not be the same for different perceptual tasks.

Table 2: Discrimination correct% for the 10 most difficult contrasts in pre- and post-tests for experiment group. P-values were from paired-samples t-tests performed for each contrast.

contrasts	pre%	post%	% impr.	sig.
ɒ-ʌ	22	28	6	$p = .239$
e-æ	34	32	-2	$p = .600$
uː-ʊ	25	33	8	$p = .166$
æ-ʌ	33	41	8	$p = .082$
ɔː-ɒ	26	43	17*	$p < .05$
ʌ-ɑː	30	44	14*	$p < .05$
iː-ɪ	23	45	22*	$p < .001$
ɪ-eɪ	42	48	6*	$p < .05$
ɪ-e	39	48	9*	$p < .05$
əʊ-ɒ	45	49	4	$p = .096$
mean overall	62.7	66.7	4*	$p < .001$

4. DISCUSSION

Current study investigated the effectiveness of high variability auditory training on Chinese learners' perception of English vowels. The results showed that, comparing to a significant near 30 percentage points performance improvement in category identification, learners' discrimination accuracy improvement was quite limited, suggesting that identification training may not be as effective in improving learners' discrimination ability as it does to identification ability. This result is in line with [11], that native French speakers' discrimination ability of English vowels didn't improve much after 8 sessions of identification training. It is claimed that discrimination task taps into different aspects of perception from category identification task [5, 11]. Thus the current study provides more evidence that the effect of auditory training is task based.

On the level of individual vowel, the results showed that learners still had great difficulty in distinguishing some English tense/lax and lax/lax contrasts even after training (e.g., /ɒ-ʌ/, /e-æ/ and /uː-ʊ/). This was not a surprising result because compared to the crowed English vowel space which has 12 monophthongs [18], the Mandarin Chinese only has 6 normal monophthongs [13]. Chinese monophthongs are more similar to English tenses vowels, thus the Chinese listeners may treat the English lax vowels as allophones to their tense counterparts, and the confusions are predictable [1, 4].

Consistent with [11], no significant correlation between learners' identification improvement and their discrimination improvement was found in the current study. Again this result suggests that the effect of training is task based, which means what learners benefited from the identification training might not be suit for discrimination task. [10] argued that training can not change learners' low level of processing such as cue weighting. [11] suggested that training can improve learners' ability of applying existing category knowledge to high variability speech rather than changing the category representations. Current study's results, especially the discrimination test results, indicated that the identification training was not really successfully in helping Chinese learners to acquire the crucial vowel quality differences between English tense/lax contrasts, supporting the claim that learners' underlying fundamental L2 category processing mechanism may not be really changed by training.

The large set of vowels and multi-talker tokens used in the current study might be among the reasons why learners' discrimination improvement was limited. High variability may cause more cognition load for learners, and they couldn't generalize the crucial distinctive features between L2 contrasts in a short intensive training. It is demonstrated in [6] that explicit training was more effective than implicit training on Iranian learners' identification of English vowels. Further study is required to examine whether high-variability auditory training with explicit instruction can improve both identification and discrimination.

5. ACKNOWLEDGMENTS

This study was supported by grants from the Humanities and Social Science Fund of the Ministry of Education of China (16YJC740020, 15YJC740034), the Jiangsu Social Science Fund (18YYB009) and the Jiangsu Provincial Postgraduate Innovation Project (KYCX18_2293).

6. REFERENCES

[1] Best, C. T. 1995. A direct realist view of cross-language speech perception. In: Strange, W., (ed), *Speech perception and linguistic experience: Theoretical and methodological issues*. Baltimore: York Press 171–204.

[2] Bradlow, A. R. 2008. Training non-native language sound patterns: lessons from training japanese adults on the english. *Phonol. Second Lang. Acquis* 36, 287–308.

[3] Clopper, C. G., Pisoni, D. B. 2004. Effects of talker variability on perceptual learning of dialects. *Language and Speech* 47(3), 207–238.

[4] Flege, J. E. 1995. Second-language speech learning: Theory, findings, and problems. In: Strange, W., (ed), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research*. Baltimore: York Press 233–277.

[5] Flege, J. E. 2003. Methods for assessing the perception of vowels in a second language. *Issues in clinical linguistics* 19–44.

[6] Ghorbani, M. R., Neissari, M., Kargozari, H. R. 2016. The effect of explicit pronunciation instruction on undergraduate efl learners' vowel perception. *Language and Literacy* 18(1), 57–70.

[7] 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.

[8] Gong, J., Lecumberri, M. L. G., Cooke, M. 2017. Ab initio perceptual learning of foreign language sounds: Spanish consonant acquisition by chinese learners. *System* 66, 142–155.

[9] Hwang, H., Lee, H.-Y. 2015. The effect of high variability phonetic training on the production of english vowels and consonants. *Proceedings of the 18th International Congress of Phonetic Sciences edited by Maria Wolters, Judy Livingstone, Bernie Beattie, Rachel Smith, Mike MacMahon, Jane Stuart-Smith and Jim Scobbie. The Scottish Consortium for ICPhS*.

[10] Iverson, P., Hazan, V., Bannister, K. 2005. Phonetic training with acoustic cue manipulations: A comparison of methods for teaching english/r/-/l/to japanese adults. *The Journal of the Acoustical Society of America* 118(5), 3267–3278.

[11] 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.

[12] 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.

[13] Lin, T., Wang, L. 2013. A course in phonetics.

[14] 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.

[15] Nishi, K., Kewley-Port, D. 2007. Training japanese listeners to perceive american english vowels: Influence of training sets. *Journal of Speech, Language, and Hearing Research* 50(6), 1496–1509.

[16] Perrachione, T. K., Lee, J., Ha, L. Y., Wong, P. C. 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.

[17] Rato, A. 2014. Effects of perceptual training on the identification of english vowels by native speakers of european portuguese. *Proceedings of the international symposium on the acquisition of second language speech* volume 5 529–546.

[18] Roach, P. 2004. British English: Received Pronunciation. *Journal of the International Phonetic Association* 34(2), 239–245.

[19] Sadakata, M., McQueen, J. M. 2014. Individual aptitude in mandarin lexical tone perception predicts effectiveness of high-variability training. *Frontiers in psychology* 5, 1318.

[20] 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.

[21] Shinohara, Y., Iverson, P. 2018. High variability identification and discrimination training for japanese speakers learning english/r/–/l. *Journal of Phonetics* 66, 242–251.

[22] 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.