The effect of high variability phonetic training on the production of English vowels and consonants

Hyosung Hwang, Ho-Young Lee

Department of Linguistics, Seoul National University
hshwang415@gmail.com, hylee@snu.ac.kr

ABSTRACT

This study evaluated Korean early learners' production of English vowels and consonants before and after high variability phonetic training (HVPT), for the purpose of investigating the effects of perceptual training on speech production. Korean learners aged 11-12 were trained to discriminate multiple sets of English vowel and consonant contrasts, and their pre- and post- training recordings were evaluated by native speakers of English. Although the trainees showed significant improvements for some sounds, the overall results demonstrated that the effect of perceptual learning on speech production is not significant and the improvements in the two domains (i.e., perception and production) are not correlated.

Keywords: phonetic training, L2 perception and production

1. INTRODUCTION

It is well known that learners of a second or foreign language commonly have foreign accents and have difficulty in mastering the perception and production of certain phonemic contrasts. Several studies have shown that HVPT (High Variability Phonetic Training) is highly effective for foreign learners to quickly improve their ability to discriminate different sounds [4, 5]. The improvement in perception was proven to be successfully transferred to production [1].

Most previous researches on the functional relationship between perception and production in L2 acquisition focused on the learning of limited sets of sound contrasts (e.g., English /r/ and /l/) of adult learners of English [1, 3], or of children with articulation disorders [6]. Current study aimed to investigate whether perception training using the HVPT method can improve the production of multiple sets of sound contrasts by early learners with different English proficiency. We examined whether production of certain sounds can be improved more than others, and whether the sounds display different hierarchy of learnability.

For these purposes, we trained Korean elementary school students to perceptually distinguish large sets of English vowel and consonant contrasts using the HVPT method (identification task with immediate feedback). The recordings of individual learners' production of English vowels and consonants were made before and after the perceptual training, which were then evaluated by native speakers of English for production accuracy (identification task) and goodness (rating task).

2. EXPERIMENT

2.1. Subjects

Sixty Korean elementary school students (29 boys and 31 girls, mostly aged 12 years) participated in the experiment. They have regular exposure to English both at school and the private institute for a similar amount of time, as is common for Korean learners of English. However, none of them reported that they had stayed in an English-speaking country for more than six months. The subjects were divided into three groups; twenty subjects were randomly assigned to the control group, and the remaining forty subjects were evenly divided into the lower intermediate group and the upper intermediate group based on their pre-test scores.

Six native speakers of English (2 males and 4 females) participated as listeners. They were undergraduate and graduate students at Seoul National University (mean age = 26.7 years, *SD* = 4.8). All had the general American English accent, and their average length of residence in Korea was 1.7 (*SD* = 1.3) years. None of the participants reported any history of speech or hearing problems.

2.2. Stimuli

A large set of English vowels and consonants were tested. The pre- and post- test stimuli for the vowel comprised 12 English b/pVC words (i.e., *beat* /iː/, *bit* /ɪ/, *bait* /eɪ/, *bet* /ɛ/, *bat* /æ/, *bought* /ɔː/, *boat* /oʊ/, *pool* /uː/, *pull* /ʊ/, *pot* /ɑː/, *putt* /ʌ/, *bird* /ɝ/). The consonant stimuli consisted of 24 words (i.e., *pat* /p/, *fat* /f/, *hat* /h/, *ban* /b/, *van* /v/, *tie* /t/, *thigh* /θ/, *day* /d/, *they* /ð/, *cap* /k/, *gap* /g/, *map* /m/, *nap* /n/, *bang* /ŋ/, *lay* /l/, *ray* /r/, *way* /w/, *sigh* /s/, *shy* /ʃ/, *chew* /ʧ/, *zoo* /z/, *Jew* /ʤ/, *you* /j/, *measure* /ʒ/), covering all

the consonant sounds in English. The stimuli were produced by 2 native speakers of American English (1 male and 1 female) for the identification test.

The training corpus contained four sets of minimal pairs or triplets for each mutually confusable contrast (e.g., *meal-mill* /i:-ɪ/, *pair-fair-hair* /p-f-h/). All the stimuli were new words that were not used in the pre- and post-test. The corpus was recorded by 4 new native speakers (2 males and 2 females).

2.3. Procedure

The subjects completed 18 sessions of online perceptual training (5 times a week for 4 weeks, each lasting approximately 15 min). The recordings of the subjects' production were made before pre- and post- identification tests. A randomized list of the pre- and post-test stimuli was printed out to be presented, and the subjects were asked to read out the list once at their natural speaking rate. They were allowed to repeat and correct their pronunciation whenever needed. The recordings were made in quiet classrooms with SHURE WH30 condenser microphones, Marantz PMD 661 and Sound Device 722.

Among the 60 subjects, 4 subjects in the control group who failed to participate in the post-test recording were excluded from the experiment. Remaining 56 subjects successfully completed the recording of 36 stimulus words (i.e., 12 vowel stimuli and 24 consonant stimuli) at both tests, yielding 4,032 stimuli (56 subjects * 36 stimuli * 2 tests) in total.

The native judges were first instructed to identify each token, and then they were asked to rate the token for goodness on a 7-point scale ranging from "1 (not at all native-like)" to "7 (native-like)". They were asked to focus on evaluating the realizations of target sounds in the word rather than the general pronunciation of the word. Vowel tokens were presented in two mutually confusable sets; the first set containing 5 front vowel tokens (i.e., *beat, bit, bait, bet, bat*), and the second set containing 7 mid and back vowel tokens (i.e., *bought, boat, pool, pull, pot, putt, bird*). Consonant tokens were presented in 11 different sets, each containing confusable minimal pairs or triplets (e.g., *ban-bang* /n-ŋ/, *sigh-thigh-shy* /s-θ-ʃ/, *zoo-chew-Jew-you* /z-ʧ-ʤ-j/, *measure-major-masher* /ʒ-ʤ-ʃ/). The stimuli *major* and *masher* were only dummy choices without sound that were brought in to form a set with *measure*. The order of the tokens were randomized automatically across the sound sets and then presented to the listeners. Judgment task was designed using Praat's ExperimentMFC, and each listener completed the 13 sets of identification and rating task individually over one week

3. RESULTS

3.1. Vowel production and goodness rating

Average scores for the 13 vowels appeared similar across the groups and tests; all three groups maintained virtually the same scores in the pre-test and post-test (67% for the control group, 62% for the lower intermediate group, and 66% for the upper intermediate group). A repeated-measures ANOVA on arcsine transformed scores demonstrated that there was no significant main effect of training ($p > .05$), indicating that overall vowel production performance hardly improved after the training. There was no significant main effect of group, and neither the interaction between training and group was significant ($p > .05$).

Table 1: Native speakers' identification and goodness judgment results for English vowels produced by Korean learners. Post-test results are presented in parentheses. (*: $p < .05$)

Rank	Vowel	Common ID	Percentage of ID	Goodness rating
1	/ɝ/	/ɝ/	90.8 (97.9)	4.8 (4.6)
2	/oʊ/	/oʊ/	87.5 (86.7)	5.0 (5.1)
3	/eɪ/	/eɪ/	86.7 (97.1*)	5.5 (5.5)
4	/ʌ/	/ʌ/	75.4 (61.3*)	2.2 (2.7)
		/ɑ:/	16.3 (34.6)	4.0 (5.2)
5	/i:/	/i:/	65.0 (70.4)	5.2 (5.2)
		/ɪ/	31.7 (27.9)	4.7 (4.8)
6	/ʊ/	/ʊ/	61.7 (54.6)	4.4 (4.0)
		/u:/	35.0 (40.4)	4.4 (4.3)
7	/ɑ:/	/ɑ:/	60.4 (73.3*)	5.4 (5.4)
		/ʌ/	29.6 (19.2)	4.2 (4.3)
8	/ɛ/	/ɛ/	56.3 (58.8)	5.2 (4.9)
		/æ/	30.0 (30.4)	5.3 (4.8)
9	/ɪ/	/ɪ/	38.3 (35.8)	4.7 (4.8)
		/i:/	56.7 (61.3)	5.2 (5.2)
10	/u:/	/u:/	37.2 (38.8)	4.4 (4.4)
		/ʊ/	55.5 (56.3)	4.1 (4.0)
11	/æ/	/æ/	34.6 (27.5)	5.3 (5.4)
		/ɛ/	60.8 (62.1)	5.4 (5.0)
12	/ɔ:/	/ɔ:/	33.3 (38.3)	3.5 (3.3)
		/oʊ/	62.9 (59.6)	5.1 (5.0)

Table 1 shows the pre- and post-test production accuracy scores and goodness rating scores of each vowel for the two experimental groups. Vowels are ranked in ascending order of pre-test scores. In the pre-test, /ɔ:/, /æ/, /u:/, /ɪ/ showed the lowest scores ranging from 33% to 38%, and these vowels were most frequently perceived as /oʊ/, /ɛ/, /ʊ/, /i:/ respectively by the native listeners. On the other hand, /eɪ/, /oʊ/, /ɝ/ were mostly identified correctly with almost 90% of accuracy. In the post-test, the learners showed the biggest improvement for /ɑ:/, /eɪ/, and /ɝ/. More specifically, the lower intermediate group learners showed notable improvement in producing /eɪ/ in the post-test, with 20.8% of increased accuracy (from 75.8% to 96.7%). This was followed by /ɝ/, with 14.2% of improvement (from 85.8% to 100%) and /ɑ:/, with 6.7% of improvement (from 61.7% to 68.4%). The most improved vowel for the upper intermediate group learners was /ɑ:/, showing 19.2% of increased accuracy (from 59.2% to 78.3%). This was followed by /ɔ:/, with 10% of improvement (from 32.5% to 42.5%) and /i:/, with 9.2% of improvement (from 70.8% to 80%). Wilcoxon signed-rank test revealed that the learners in the two experimental groups showed significant improvement only in /eɪ/ [Z = -4.35, p < .001] and /ɑ:/ [Z = -3.58, p < .001].

The vowels that received the lowest accuracy scores in the pre-test (i.e., /ɪ/, /æ/, /ɔ:/, /u:/) were hardly improved in the post-test. This indicates that the native listeners classified Korean learners' production of /i:-ɪ/, /ɛ-æ/, /ɔ:-oʊ/, /u:-ʊ/ pairs into single English vowel categories, in both pre- and post-test, thus demonstrating no effect of training. Pearson correlation on arcsine transformed scores indicated that production accuracy of the subjects were significantly related to the identification accuracy before the training (r = .37, p = .02), but not after the training (r = .17, p > .05). The improvements in production (i.e., pre-test score – post-test score) were also not significantly related to the improvements in perception (r = .04, p > .05).

3.2. Consonant production and goodness rating

The average scores and improvements for consonant production were generally bigger than those of vowel production, although they were still minor compared to the improvements in identification performance (which will be discussed in a separate paper). The control group maintained the same score for both tests (82%), and the lower intermediate group and upper intermediate group showed respectively 1.9% and 2.1% of improvement in the post-test respectively (76.5% to 78.4% for the lower intermediate group and 83.4% to 85.5% for the upper intermediate group). A repeated-measures ANOVA on arcsine transformed scores demonstrated that there was no significant main effect of training (p > .05), indicating that consonant production performance did not improve significantly after the training. However, there was a significant main effect of group, [F (2, 53) = 6.4, p = .003], and Tukey's HSD test revealed that the upper intermediate group showed significantly higher production performance than the lower intermediate group (p = .002). No significant interaction was found between training and group.

Table 2: Native speakers' identification and goodness judgment results for English consonants produced by Korean learners. Post-test results are presented in parentheses. (*: p < .05)

Rank	Cons	Common ID	Percentage of ID	Goodness rating
1	/j/	/j/	99.6 (99.2)	5.8 (5.8)
2	/h/	/h/	97.9 (96.7)	5.1 (4.8)
3	/g/	/g/	97.5 (98.3)	4.9 (5.2)
4	/t/	/t/	96.7 (93.3)	5.5 (5.7)
5	/ʃ/	/ʃ/	95.4 (96.3)	5.6 (5.7)
6	/f/	/f/	94.6 (97.1)	5.1 (5.2)
7	/m/	/m/	92.9 (95.4)	5.3 (5.4)
7	/k/	/k/	92.9 (97.5)	5.3 (5.3)
9	/ŋ/	/ŋ/	92.1 (93.3)	5.2 (5.4)
9	/n/	/n/	92.1 (88.3)	5.1 (5.0)
11	/p/	/p/	91.7 (95.4)	4.4 (4.5)
12	/w/	/w/	87.1 (90.8)	5.1 (5.4)
13	/r/	/r/	85.8 (83.8)	5.1 (5.4)
14	/l/	/l/	82.1 (80.4)	5.5 (5.5)
15	/ʧ/	/ʧ/	77.9 (80.4)	5.0 (5.2)
16	/d/	/d/	75.4 (75.0)	5.1 (5.0)
17	/b/	/b/	74.6 (80.0)	4.5 (4.6)
18	/θ/	/θ/	69.3 (81.3*)	3.7 (4.4)
		/t/	30.7 (18.7)	5.4 (5.3)
19	/s/	/s/	67.1 (61.3)	5.2 (5.2)
		/θ/	24.2 (28.7)	3.8 (4.5)
20	/ʤ/	/ʤ/	63.3 (72.1*)	4.7 (4.8)
		/z/	29.6 (22.5)	3.9 (3.9)
21	/ʒ/	/ʒ/	62.5 (59.6)	3.9 (3.8)
		/ʤ/	28.3 (32.9)	4.2 (4.3)
22	/v/	/v/	55.4 (65.4*)	4.6 (4.7)
		/b/	44.6 (34.6)	4.4 (4.8)

23	/z/	/z/	48.3 (47.5)	4.3 (4.5)
		/ʤ/	48.8 (45.4)	4.9 (4.6)
24	/ð/	/ð/	29.2 (37.9*)	4.8 (5.0)
		/d/	70.8 (62.1)	4.8 (4.9)

Table 2 presents the mean scores of pre- and post-test production accuracy and goodness rating of each consonant for the two experimental groups. /ð/ received the lowest accuracy score of 29.2%, and was perceived as /d/ by native listeners over 79% of the instances. /z/ and /v/ also showed low accuracy scores, due to the poor performance of the low intermediate learners (38% for /v/ and 39% for /z/). Other problematic consonant sounds were /ʒ/, /ʤ/, /s/, /θ/, which showed relatively low scores (62-69%) before the training. On the other hand, /p/, /t/, /k/, /g/, /m/, /n/, /ŋ/, /f/, /h/, /ʃ/ and /j/ were mostly identified correctly by the native listeners with 90% to 100% accuracy before and after the training. In the post-test, the experimental group learners showed improved production accuracy for 14 consonants out of 24 consonants, and some notable improvements were observed for /θ/, /ʤ/, and /ð/. The learners in the lower intermediate group showed the biggest improvement for /θ/; the accuracy increased by 17.8% in the post-test and it was identified with 79.2% of accuracy by the native listeners. The production of /v/ was also highly improved by 14.2% in the post-test. The upper intermediate group learners showed the biggest improvement in producing /ʤ/, with 9.2% of increased accuracy (from 65.8% to 75%). The production accuracy for /ð/ was also increased by 9% in the post-test for both experimental groups. Wilcoxon signed-rank test revealed that the consonant production improvement was significant for /θ/ [$Z = -4.22$, $p < .001$], /v/ [$Z = -2.83$, $p = .005$], /ð/ [$Z = -2.33$, $p = .02$], and /ʤ/ [$Z = -2.28$, $p = .023$]. Among the consonants that received the lowest production accuracy scores in the pre-test (i.e., /ð/, /v/, /z/, /ʒ/, /ʤ/), /ʒ/ and /z/ hardly improved in the post-test. Korean learners' production of /ʒ/ was identified as /ʤ/ for 33% of the time by the native listeners even after the training, and goodness rating stayed as low as 3.8. /z/ was perceived as /ʤ/ for 45.4% of the time with 4.6 of goodness rating. There was a positive correlation between production and identification accuracy both at the pre-test ($r = .48$, $p = .002$) and post-test ($r = .46$, $p = .003$). However, the degree of improvements in production was not significantly related to that in identification ($r = .03$, $p > .05$).

4. DISCUSSION & CONCLUSION

One of the main purposes of this study was to examine whether the learning effects gained from HVPT can be successfully transferred to speech production. However, the results of the current study show that the overall improvement for the experimental groups was too little that it is hard to say that the high variability phonetic training (HVPT) improved the production performance of the learners. Production of English minimal pairs and triplets by Korean early learners barely improved after the training regardless of the proficiency level of the subjects, except for few contrasts. This may have resulted from relatively smaller amount of perceptual training (approximately 4.5 hours in total) compared to the prior studies. Moreover, each session contained multiple sets of contrasts that practically cover the entire vowel and consonant systems of English. Intensive training focusing on small sets of problematic sounds would be far more effective to observe the transfer from perceptual training to speech production.

It is to be noted that some vowel and consonant productions did show some improvements and were more accurately identified by the native listeners after the perceptual training, exhibiting a possible correlation between perceptual learning and speech production. However, not only the correlation was not strong, but the degrees of improvement in the two domains were not significantly related. This results from the substantial individual differences across learners in terms of production performance, which has been pointed out in previous studies [1]. Another possible explanation for the discrepancies in the improvements of perception and production can be found in more recent studies, which assume that the two processes of speech may utilize different learning mechanisms [2].

Meanwhile, the identification and goodness rating data obtained from the native listeners showed an overall view of the intelligibility of Korean learners' pronunciation. The learners had particular difficulty in producing identifiable tokens for vowel pairs of /i:-ɪ/, /ɛ-æ/, /ɔ:-oʊ/, /u:-ʊ/ and consonant pairs of /d-ð/, /b-v/, /s-θ/, /z-ʒ-ʤ/ even after the training, and this result largely coincides with the trends found in the results of perceptual training task. It is suggested that perceptual training should be combined with relevant production training on problematic sounds to maximize the effect of pronunciation teaching, as production performance cannot be sufficiently improved only with perceptual training in the case of foreign/second language acquisition.

5. REFERENCES

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

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] Rvachew, S. 1994. Speech perception training can facilitate sound production learning. *Journal of Speech, Language, and Hearing Research*, *37*(2), 347-357.