# SECOND LANGUAGE VOWEL PERCEPTION TRAINING: EFFECTS OF SET SIZE, TRAINING ORDER, AND NATIVE LANGUAGE

*Kanae Nishi* [a), b)] *and Diane Kewley-Port* [b)]

[a)] Boys Town Nat'l Res. Hospital, Omaha, NE and [b)] Indiana University, Bloomington, IN, U.S.A.
nishik@boystown.org and kewley@indiana.edu

## ABSTRACT

This paper reports results of a series of vowel training studies. Study 1 trained two groups of Japanese learners of English on American English vowels and examined the effects of training set sizes (9 vs 3 more difficult vowels); Study 2 trained Korean learners of English and examined the efficiency of training protocols using both 9- and 3-vowel sets. Study 3 compared the Japanese and Korean results on the untrained materials.

Results suggested following: 1) vowel training works best when a large set of vowels, rather than a subset, is used; 2) training focusing on a smaller yet difficult vowels may have detrimental effects on later learning; and 3) improvement due to training on nonsense words may or may not carry over to untrained real word materials, possibly due to an interaction between native and non-native phonology.

**Keywords:** vowel perception, training, second language

## 1. INTRODUCTION

It has been shown that adult second language (L2) learners can learn to discriminate difficult L2 contrasts through structured training [2]. Previous vowel training studies reported that perception of difficult vowels improved after training, but that the untrained vowels did not change [1, 3]. It was suggested that the training set used may be responsible for this uneven learning [1, 3]. The present study trained L2 learners of English, Japanese and Korean, on American English (AE) vowels. Three experimental variables were examined for perceptual training of AE monophthongs. They were: training set sizes (9 vs 3 more difficult vowels), orders of training sets (9-first vs 3-first), and learners' native language (L1) (Japanese vs Korean).

## 2. STUDY 1: EFFECTS OF TRAINING SETS

The research questions asked in Study 1 were:

1. Is it possible to train a large number of vowels concurrently?
2. If yes, which protocol is better, one using a smaller vowel set but focusing on more difficult vowels or the other using a larger set of vowels covering the entire vowel space?

### 2.1. Method

#### *2.1.1. Listeners*

Seventeen Japanese learners of English were divided into 3 groups: Fullset (n = 6), Subset (n = 6), and Control (n = 5). All listeners resided in Bloomington, IN. None had lived outside Japan for more than 1 year.

#### *2.1.2. Materials*

Four native AE speakers of the Midland dialect recorded stimulus materials. Tokens from 2 speakers (M1, W1) were presented at tests and during training (Trained speakers), whereas the tokens by remaining 2 speakers (M2, W2) were presented only at tests (New speakers).

Stimulus materials were 9 AE vowels /i, ɪ, ɛ, æ, ɑ, ʌ, ɔ, ʊ, u/ embedded in 54 nonsense words (NSW, /bVbə, bVpə, dVdə, dVtə, gVgə, gVkə/) and 36 English monosyllabic real words (RW) that included more consonantal contexts than NSW. NSW were used in tests and in training. RW were used only in tests. All speakers recorded 2 tokens for each word.

#### *2.1.3. Procedures*

Tests and training were self-paced and were conducted in a sound-treated room. Listeners' performance was assessed before (pre-test) and after training (post-test) as well as 3 months after the completion of training (3mo-test). Times between pre- and post-tests varied among the listeners, and ranged between 2 weeks and 1 month.

Listeners identified the target vowel in each stimulus word choosing a button on a computer screen. Prior to pre-test, listeners were familiarized

with the IPA symbols for the 9 AE vowels and keywords for each symbol.

Tests presented a total of 432 NSW trials and 288 RW trials. Utterances were blocked by speaker and the orders of blocks were randomized among listeners.

There were nine 648-trial training sessions. The Fullset condition presented all 9 vowels. The Subset condition presented only the 3 vowels (3V, /ɑ, ʌ, ʊ/) that were determined as mutually confusable and less accurately perceived than the other vowels in a pilot study. In the Subset training, each token was repeated 3 times in a block to maintain an equal number of trials between training conditions.

Feedback was provided using text displays of "Correct" or "Incorrect." Only when the answer was wrong, in addition to the text "Incorrect," listeners were prompted to hear the sound of the correct or the incorrect stimulus up to 10 times in any order, or to proceed to the next trial.

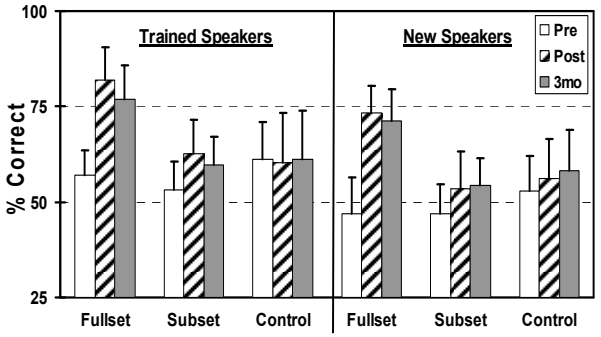
## 2.2. Results

This section presents only the results for the NSW. RW results are presented under Study 3.

### *2.2.1. Tokens by Trained Speakers*

Fig. 1 presents the performance of the 3 listener groups on Trained (left) and New speakers (right) at the 3 tests.

**Figure 1:** Average percent-correct scores for 3 groups on the tokens from Trained and New speakers at pre-, post- and 3mo-tests.




No difference was found among the 3 groups at pre-test ($F(2, 14) = 1.64$, $p = .23$). The Control group's performance did not change across the tests ($F(2, 8) = .08$, $p = .92$). However, the 2 training groups significantly improved after training ($F(2, 18) = 52.25$, $p < .001$), and the Fullset group improved more than the Subset group ($F(1, 9) = 11.98$, $p < .01$).

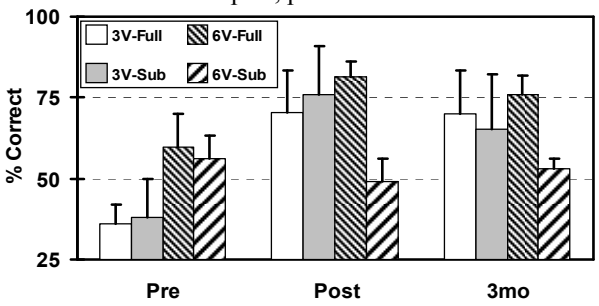
### *2.2.2. Tokens by New Speakers*

No difference was found among the groups at pre-test on the tokens from New speakers ($F(2, 14) = 1.10$, $p = .36$). Comparison between the 2 training groups revealed the Fullset group's post- and 3mo-test scores were significantly higher than pre-test ($p < .001$ for both). However, the Subset group did not show improved performance until 3mo-test ($p < .005$).

### *2.2.3. Difference between Training Sets*

Fig. 2 presents average percent-correct scores for 3V and the other 6 vowels (6V) obtained for Fullset and Subset groups at the 3 tests.

**Figure 2:** Average percent-correct scores for Fullset and Subset groups separately calculated for 3V and 6V at pre-, post- and 3mo-tests.




At post- and 3mo-tests, the Fullset group improved on both 3V and 6V ($p < .002$ for both). For the Subset group, improvement was observed for the trained 3V ($p < .001$ for both) but not for 6V ($p = 1.00$ for both). In fact, their performance for 6V became slightly worse after training.

## 2.3. Discussion

The results obtained for the Subset group suggest that improvement observed for the trained vowels did not generalize to the untrained vowels, while the Fullset group improved on all vowels and maintained improvement over 3 months. Results suggested that a vowel training works best when many vowel categories are trained concurrently.

# 3. STUDY 2: EFFECTS OF TRAINING ORDER

Two research questions were asked in Study 2:

1. Is the Fullset protocol also effective for listeners with different L1?
2. Is training combining Fullset and Subset protocols (hybrid) more efficient than Fullset only?

### 3.1. Method

*3.1.1. Listeners*

Fifteen native speakers of Korean were divided into 3 groups (n = 5 each). All listeners resided in Bloomington, IN. None had lived outside South Korea for more than 1 year.

*3.1.2. Materials*

The same as Study 1.

*3.1.3. Procedures*

Listeners were trained using 1 of the 3 protocols: Fullset-only (9V-9V), Fullset-first (9V-3V), or Subset-first (3V-9V). The 9V-9V group received the Fullset training for 9 days; the 9V-3V group received the Fullset training for the first 6 days then the Subset training for 3 days; training schedule was reversed for the 3V-9V group. Listeners' performance was assessed before (pre-test) and after training (post-test) as well as at mid-training test (probe) when the 2 hybrid groups switched training sets. The 9V-9V group used the same probe schedule as the 9V-3V group. Other than these modifications to the training protocols, the procedures were the same as in Study 1.

### 3.2. Results

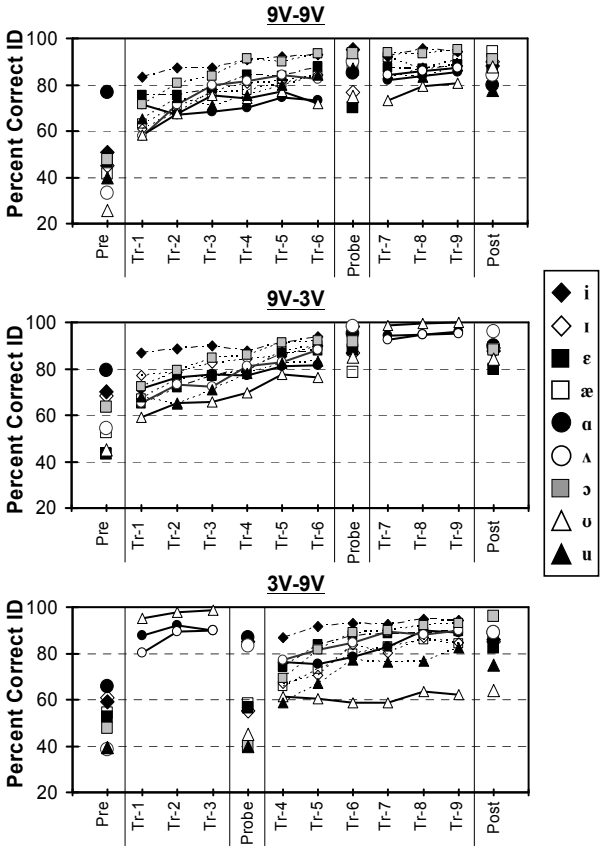
Average percent correct performance observed for the three groups are presented as Fig. 3. Each data point represents a score for a vowel averaged across 5 listeners during training and for 3 tests. 3V are connected by heavy lines.

All groups significantly improved after training ($F(2, 12) = 154.96$, $p < .001$). No differences were found among groups, indicating that, overall, the hybrid protocols were not more efficient than the Fullset protocol. Interestingly, despite the early focused training on the more difficult 3V, the 3V-9V group did not improve significantly on /ʊ/ at probe (+6%), whereas /ʌ/ (+45%) and /ɑ/ (+21%) improved. Further analysis for 3V during the 6 Fullset sessions indicated a significant vowel × group interaction ($F(4, 24) = 4.08$, $p < .05$). That is, on average, during the Fullset training sessions, only the 3V-9V group performed less accurately on vowel /ʊ/ than on /ɑ, ʌ/ ($p < .002$).

### 3.3. Discussion

This experiment demonstrated that there was no advantage for the hybrid protocols over the Fullset protocol. Moreover, results for the 3V-9V group indicated that early focused training on a smaller and more difficult vowel set may have had detrimental effects on the subsequent learning of L2 vowels.

**Figure 3:** Average percent-correct vowel scores for three Korean groups during training and at pre-, probe, and post- tests.




## 4. STUDY 3: EFFECTS OF L1

This section examines the effects of L1 on L2 vowel perception training. The Japanese Fullset (J-Full) group and Korean 9V-9V (K-Full) group were compared in terms of their performance. Specifically, preliminary analyses were conducted to determine the generalization of vowel perception training from the NSW to the untrained RW materials.

### 4.1. Method

Data from the 6 listeners in the J-Full and 5 listeners in K-Full groups were analyzed. Pre- and post-test results for the 9 AE vowels in NSW and RW were analyzed.

### 4.2. Results

Fig. 4 presents the results for J-Full (left panels) and K-Full (right panels) on the 9 AE vowels in

NSW (top panels) and RW (bottom panels). Post-test NSW scores (J-Full = 78%; K-Full = 84%) were not different for the 2 groups ($F(1, 9) = 1.44$, $p < .26$). However, for the vowels in RW, a significant L1 effect ($F(1, 9) = 39.73$, $p < .001$) and a significant L1 × vowel interaction ($F(1, 9) = 21.44$, $p < .001$) were found. A further analysis revealed that K-Full's post-test performance on /ʊ, u/ was significantly worse than the other vowels ($p < .001$), but no such vowel differences were observed for J-Full.

### 4.3. Discussion

Results suggested that after the training using the same material and the same protocol, the 2 L1 groups improved to a similar level on the trained NSW materials. However, generalization of improvement to untrained RW materials revealed a strong influence of listeners' native language.

## 5. CONCLUSIONS

The results of the 3 studies suggest the following:

- Vowel perception training should present many vowel categories from the beginning of training.
- Training focusing on a more difficult vowel subset may not only have any advantage over training that uses many categories concurrently, but it may also cause uneven learning of some vowels.
- Even though vowel perception training using nonsense materials can improve non-native vowel perception by learners from different native languages to similar levels, generalization of improvement to untrained real-word materials may be strongly influenced by learners' native language. Results suggest incorporating some real-word training occasionally throughout the more rigorous nonsense-word vowel training.

## 6. REFERENCES

[1] Akahane-Yamada, R., Strange, W., Kubo, R. 1997. Training Japanese listeners to identify American English vowels. *Proc. Fall Meeting Acoust. Soc. Japan.* 379-380.

[2] Bradlow, A.R. in press. Training non-native language sound patterns: Lessons from training Japanese adults on the English /r/-/l/ contrast. In: Hansen, J., Zampini, M. (eds), *State-of-the-art issues in second language phonology.*

[3] Sperbeck, M.N., Strange, W., Ito, K. 2005. Training L2 learners to perceive difficult American vowel contrasts. *J. Acoust. Soc. Am.* 117, 2400.

## 7. ACKNOWLEDGMENTS

This research was supported by NIH postdoctral fellowship (F32 DC-006313) to K. Nishi and NIH research grant to D. Kewley-Port (R01 DC-02229).

**Figure 4:** Average percent-correct scores for 9 AE vowels in trained nonsense words (NSW) and untrained real words (RW) observed for Japanese Fullset group (J-Full) and Korean 9V-9V (K-Full) group.