THE INFLUENCE OF IDENTIFICATION TRAINING ON IDENTIFICATION AND PRODUCTION OF THE AMERICAN ENGLISH MID AND LOW VOWELS BY NATIVE SPEAKERS OF JAPANESE

Stephen LAMBACHER1, William MARTENS2, Kazuhiko KAKEHI3

1Center for Language Research, The University of Aizu lambach@u-aizu.ac.jp
2Graduate Department of Information Systems, The University of Aizu wlm@u-aizu.ac.jp
3Graduate School of Informatics, Nagoya University kakehi@info.human.nagoya-u.ac.jp

ABSTRACT

Vowel identification and production performance for an experimental group of 35 Japanese subjects was measured before and after a 6-week period within which subjects received identification training with feedback. A control group of 9 Japanese subjects was administered the same identification and pre- and post-test, separated by the same 6-week interval during which they did not receive the experimental vowel identification training that provided feedback regarding which vowel sound had been produced by the native American English talkers. The results showed that the experimental group’s identification performance (as measured by d’) for each of the five American English (AE) vowels /æ/, /ɪ/, /ʌ/, /ɜ/, /ə/ improved more than the control group’s did. Both before and after the 6-week identification-training period, recordings were made of the experimental group’s productions of a list of words in a varied [CVC] context, each containing one of the five target AE vowels, which were evaluated by 3 AE native listeners using a 2-interval, forced-choice identification task. The results showed the AE listeners preferred the post-test to pretest productions for four out of the five target AE vowels. Overall, the results indicate the feedback-based identification training had a positive effect on both the experimental group’s identification and production performance.

1. INTRODUCTION

When pronouncing non-native speech sounds, second language (L2) learners frequently do not perform as well as native (L1) speakers [1]. This means that L2 learners acquiring a foreign language will have difficulty mastering certain sound contrasts in the target language. There have been numerous cross-language training studies focusing on L2 acquisition of English /t/ and /l/ by native speakers of Japanese, with mixed results. For example, Strange and Dittmann [2] failed to produce long-term changes in Japanese perception of /t/ and /l/ using an AX discrimination task.

However, Lively et al [3] and Bradlow et al [4] were successful in training Japanese in the /t/-/l/ contrast using a 2AFC identification task. To our understanding, there have been no prior studies investigating the effects of identification training on the identification and production of American English (AE) vowels by native Japanese.

The main purpose of this study was to ascertain the extent to which training on identification of the AE mid and low vowels /æ/, /ɪ/, /ʌ/, /ɜ/, /ə/ influences native Japanese learners’ performance in both identifying and producing these same five vowels. Specifically, we wanted to determine if training in identification of AE mid and low vowels transfers to improvement in production of the same vowels by native Japanese. These five AE vowels were chosen as a result of a previous study, which showed that they were the most difficult for Japanese listeners to identify amongst a larger group of ten vowels in a ten-alternative, forced-choice task [5]. In contrast to a subsequent study, in which discrimination training was employed, listeners in this study were given identification training with feedback, since there was no indication that the discrimination training improved L2 identification performance [6].

This study is divided into four phases: (1) a pretest and post-test identification task in which data are collected from an experimental and control group of Japanese subjects; (2) a 6-week identification-training task in which the group of experimental Japanese subjects are trained to identify the English vowels /æ/, /ɪ/, /ʌ/, /ɜ/, and /ə/ and given feedback about the correctness of their responses; (3) a pretest and post-test production task in which data are collected from the same experimental group of Japanese subjects; and (4) an assessment of the pre- and post-test productions of the experimental group of Japanese subjects, which are evaluated by 3 native talkers of American English (AE) using a 2-interval, forced-choice identification task.

2. EXPERIMENT 1

2.1. Subjects

The experimental group consisted of 35 native speakers of Japanese, and the control group consisted of 9 native speakers of Japanese. All of the listeners were second year students majoring in computer science at the University of Aizu (average age, 20 years). As is common in Japan, all the listeners had six years of prior English instruction at the junior and senior high school levels, the focus of that instruction being on English reading and grammar. However, all listeners had completed one year of undergraduate studies in which they were required to take one course each in English conversation and listening, composition, grammar, and reading. None of the listeners had any reported history of speech or hearing impediments.

2.2. Stimuli/procedure

The listeners were presented a total of 150 tokens (5 talkers x 30 stimuli) in a five-alternative, forced-choice (5AFC) identification task. The control group was administered the same identification
pre- and post-test, separated by a 6-week interval in which they received no vowel identification training. In the identification task, the 5 AE vowels /æ/, /a/, /æ/, /ɛ/ and /ɜ/ were presented in a varied [CVC] context. The CVC utterances consisted of all combinations of the three initial consonants /b/ m p and two final consonants /p/ f for a total of 30 different consonant contexts. These stimuli were produced by five AE native speakers who spoke with a Midwestern accent and were able to maintain the /æ/-/ɛ/ distinction in their spontaneous speech. The majority of the stimuli were natural words in English (e.g., hop, bop, mud), although some were nonsense words (e.g., hup, hau).

During the 6-week training period, the same 5 AE vowels /æ/, /a/, /æ/, /ɛ/ and /ɜ/ were presented to the experimental listeners in stimuli that were not included in any of the pre-/post-tests within three different CVC contexts in a 5AFC identification task. These were also naturally spoken words recorded by the same 5 AE native speakers. In the training task, listeners heard a total of 75 stimuli (5 speakers x 15 trials), which included three initial consonants /p/ m b and three final consonants /f/ b k. The “goodness” (or prototypicality) of the stimuli recorded for all of the identification tasks were verified during a preliminary pilot study. If a stimulus was regarded as a poor exemplar of a target vowel it was discarded and a new recording was made to ensure that each of the tokens were as intended. During the identification training, after each response, listeners were given feedback regarding the correctness of their responses. There was no feedback given during the pre-/post-identification tests.

An answer sheet containing printed single letters in regular orthography was passed out to each of the listeners, and they identified each word they heard as containing either /æ/, /a/, /ɛ/ or /ɜ/ by marking one of the five possible responses for each stimulus. An introduction to each of the five vowels was given to the Japanese listeners beforehand, in the form of practice trials, so that they understood the correct labeling of each of the vowels on their answer sheets. The stimuli were presented to the listeners with a pause of seven seconds between each stimulus.

All of these speech samples were recorded binaurally in a large anechoic chamber at the University of Aizu. The talker was positioned at ear level, but slightly off the median plane of the receiver, at an azimuth angle of approximately 15 degrees and at a distance of 2 meters. Synthetic reverberation generated by the Yamaha REV-500 was added to the dry binaural signal. Listening to these binaural speech samples via headphones confirmed that this configuration gave a clearly externalized auditory image. This avoided the rather unnatural sensation associated with monaural headphone listening. The samples were recorded at a 48 kHz sampling rate using a Denon Model DTR-80P digital audio recorder, and were reproduced at an average presentation level of 69 dB SPL via Sony HL-90 headphones in a noised-reduced language laboratory.

3. RESULTS

3.1. Pretest/post-test

For the experimental listeners, the identification responses were pooled across the two experimental variables (consonant and speaker) to form an overall summary. The grand total of the correct pretest responses shows an overall recognition rate of 55%, much higher than chance would allow. The identification scores ranged from a maximum of 63% for all the /h/ stimuli to 44% for all the /h/ stimuli. The overall correct response rates for the remaining three vowels were: /a/ (61%), /a/ (56%), and /h/ (52%). In the pretest, confusion was greatest between the three vowels /a/, /a/, and /h/. For example, 31% of /a/ stimuli were identified as /a/; 22% of /a/ stimuli were identified as /a/; and 25% of /a/ stimuli were identified as /h/. A comparison of the post-test identification results showed that the overall improvement of the experimental listeners who were exposed to identification training was greater (71% vs. 57%) than that of the control listeners who had no such training. Also, in the post-test, the experimental listeners exhibited the greatest improvement in identifying the /h/ and /h/ vowels, with performance increases of 23% and 19%, respectively.

![Fig. 1](image)

Fig. 1. A comparison of the pre-and post-test identification rates of both groups of listeners using SDT (as measured by d’). Note that the roman letters /æ/ /æ/ /æ/ /æ/ /æ/ were presented in a varied CVC context and /æ/ /æ/ /æ/ /æ/ /æ/ at the bottom of the figure correspond to the IPA symbols /æ/, /æ/, /æ/, /æ/, /æ/, respectively.

3.2. Signal detection theory

A signal-detection-theoretic (SDT) analysis of the confusion data was performed. The goal of SDT is an attempt to determine the perceptual structure underlying observed identification rates by using all of the identification data rather than just the percent correct data. This analysis generates a more meaningful representation of perceptual structure by separating identification performance into sensitivity, d’, and response bias. Fig. 1 shows the resulting bias-corrected identification scores for both groups of subjects in pre- and post-test. Though the rank ordering of vowel recognition rates agrees closely with the ordering of vowel identifiability measured by d’, there is one noteworthy exception. In the pretest, due to the fact that experimental listeners were biased in favor of the /a/ response relative to /a/ response, the percentage recognition rate for /a/ (56%) slightly dominates that for /a/ (52%). However, this does not mean that /a/ was more easily identifiable than /a/. On the contrary, SDT-based analysis showed that /a/ was somewhat less identifiable than /a/ (d’ values of 1.24 vs. 1.35, respectively). In effect, examining recognition rates without paying attention to variations in response bias can be misleading. (For another example with
4. EXPERIMENT 2

4.1. Stimuli/procedure

The 35 experimental listeners also participated in a production experiment. All of the Japanese talkers recorded nonsense words, each containing one of the target AE vowels /æ/, /a/, /i/, /u/, and /r/. These 5 vowels were inserted in /kæl/, /kæl/, /kæl/, /kæl/, frames yielding 20 minimal pairs. These were words that the talkers had not encountered in any of the identification tasks. The pretest recordings were carried out before the identification pretest to evade the possible effects of exposure to the identification stimuli on production. An introduction to the production task was given to the speakers beforehand in the form of practice trials. A list of example CVCs, each containing one of the target AE vowels printed in regular orthography, was passed out to the Japanese talkers before the experiment. The purpose was to provide the talkers with a model of how to pronounce the entire CVC not just the individual vowels, so that they better understood the pronunciation of each of the five target AE vowels. The 44 talkers were recorded individually in a soundproof room using a Denon Model DTR-80P digital audio recorder at a 48 kHz sampling rate. Each talker read from a list of words and was told to produce the CVC stimuli with the same loudness and speaking rate. The individual recordings were later stored on the hard disk of an SGI workstation and reproduced using the standard SGI audio hardware; each sound file was then presented to 3 native speakers of English for their judgments at an average presentation level of 69 dB SPL via Sennheiser HD590 headphones in a noise-reduced room.

4.2. Blind forced-choice selection

To evaluate the transfer of the Japanese talkers’ identification training to their production, 3 native speakers of American English (AE) performed a 2-interval, forced-choice (2IFC) identification task that included the pretest and post-test productions of the trained Japanese talkers. The purpose of this task was to determine if the native AE speakers could reliably distinguish the better of two productions, one produced before the identification training with feedback, and the other, after. This provided an unbiased means to verify whether the Japanese talkers had improved their production of the 5 target AE vowels after training. The 3 native speakers were trained phoneticians from the Midwestern United States who were familiar with /æ/-/i/ distinction. All of the AE listeners reported no history of speech or hearing impediments.

The 3 listeners judged the full set of pre- and post-test productions of all 35 experimental talkers. The stimuli consisted of 700 pre-/post-word pairs (35 talkers x 5 vowels x 4 words). The stimuli were presented in 5 counterbalanced blocks of 140 stimuli with a fixed 1.0-sec interval between the pair of stimuli and a 5-sec interval between each response and presentation of the next stimulus. Each block of stimuli included CVCs with only 1 of the 5 AE vowels, and each trial contained two productions of the same word (pretest and post-test, in a randomized order) by 1 Japanese talker. During each block of trials, the 3 AE listeners knew which of the five target vowels they were judging. The 3 AE listeners were told to decide whether the first or second version of the word was superior (i.e., had a clearer and more comprehensible pronunciation). The listeners rated the vowels in each pre-/post-word pair using a 4-point scale by marking one of four possible responses: (1) “the first version was considerably better than the second version,” (2) “the first version was somewhat better than the second version,” (3) “the second version was somewhat better than the first version,” and (4) “the second version was considerably better than the first version.” The listeners were urged to use the whole scale and had to choose even if they were uncertain of how to respond.

5. RESULTS

The results of the 2IFC task are presented in Fig. 2 in the form of the hit rates for post-test productions of the trained Japanese talkers (i.e., if the post-test production was chosen as the better of the two). These judgments of the AE listeners were collapsed across consonant context, and so the figure summarizes results for each of the five AE vowels as produced by the 35 trained talkers. The bars in Fig.2 show the average hit rates for the 3 AE listeners. The circles, triangles, and squares shown in the figure each represent the ‘post’ hit rates of each AE listener, individually. As shown in Fig. 2, the 3 AE listeners preferred the post-test productions of the trained talkers for 4 out of 5 target AE vowels. The “post” hit rates were highest for the /æ/ and /r/ stimuli (.75). The post-test versions of /æ/ (.63) and /r/ (.60) and had to choose even if they were uncertain of how to respond. The 3 judges showed no ability to distinguish between the pre- and post-test versions of the /æ/ stimuli.

6. DISCUSSION

Taken as a whole, the results appear to indicate that the overall identification and production performance of the experimental group improved from the pretest to post-test, and that the identification training with feedback had a positive effect on their produc-
tion of four out of the five AE vowels, even though no explicit articulatory instruction was given about the correct pronunciation of each of the vowels. We can speculate as to why the training method employed in this study was helpful in improving the experimental group’s identification and production performance. Previous studies have, in fact, shown that identification tasks are effective in improving L2 acquisition of nonnative speech contrasts, as they enable L2 learners to develop more robust categories of the target stimuli, which are stored in long-term memory, as opposed to discrimination tasks that focus only on mapping into short-term sensory memory [3][4]. Another important factor to consider in this study is that during training, feedback was administered after each trial informing listeners about the correctness of their responses, which may have helped to accelerate their L2 learning. Additionally, multiple AE talkers were used during training instead of a single talker, which may have resulted in listeners generalizing from a large number of tokens stored in memory. Prior research in categorization and memory has shown that stimulus variability is effective in developing robust generalization, as L2 learners are motivated to focus on general and not token-specific properties in the stimuli. Whereas the use of training stimuli with a single talker and little variability has shown to be unsuccessful in improving the perceptual spaces of Japanese listeners by extending them for nonnative contrasts [3][8]. Lastly, the identification training was long-term (6 weeks), which may have provided sufficient time for the trained subjects to improve their vowel identification and production performance.

The results of the 2IFC identification task indicated that the experimental talkers did not improve their production of /æ/, even after their exposure to the identification training with feedback. In L2 speech learning, it is expected that certain sounds will be more difficult for L2 learners to acquire than others. L2 sounds are presumably identified and pronounced according to articulatory similarities or differences of L1 sounds [9]. Flege has hypothesized that the greater the perceived distance of an L2 vowel from the closest L1 vowel, the greater likelihood that it will be perceived and produced more accurately and a new category will be established for it [1]. The auditory distinctness of /r/ and /l/ from the nearby Japanese /a/ vowel may have helped the Japanese learners to become better attuned to the acoustic representations and articulatory gestures of these two AE vowels. AE /r/ and /l/, in particular, has a unique spectral structure due to its low third formant, which may have aided the Japanese learners in acquiring it. Whereas the shorter duration of /l/ and its close proximity to /l/ might have contributed to the confusion between these two sounds, as well as to the lack of improvement in production of /l/ by the trained talkers. Japanese learners of English must learn to produce the /l/ sound as distinct from the nearby /l/ sound, which are both likely to be assimilated to Japanese /a/ [10]. In general, then, if multiple L2 vowels are assimilated to a single native vowel sound, those nonnative vowels should be more difficult to identify and produce for L2 learners than other vowels; and therefore, these vowels merit more rigorous attention in the L2 speech classroom.

The effectiveness of identification training with feedback was tested using a pretest and post-test design that included a production test. The SDT-based correction for bias served to reveal a different pattern of results and did influence the interpretation of the data, as seen in the example of the /æ/ and /l/ contrast. Although the overall accuracy of the experimental listeners was greater than that of the control listeners in the identification post-test for all five AE vowels, particularly the /r/ and /l/ sounds, they still had difficulty identifying /l/.

In addition, the production results showed the 3 AE judges preferred the post-test to pretest versions for all of the vowels except /æ/ and /i/. Taken as a whole, the results would appear to indicate that the identification training with feedback had a positive effect on the trained Japanese talkers’ production performance.

7. CONCLUSION

The effectiveness of identification training with feedback was tested using a pretest and post-test design that included a production test. The SDT-based correction for bias served to reveal a different pattern of results and did influence the interpretation of the data, as seen in the example of the /æ/ and /l/ contrast. Although the overall accuracy of the experimental listeners was greater than that of the control listeners in the identification post-test for all five AE vowels, particularly the /r/ and /l/ sounds, they still had difficulty identifying /l/.

In addition, the production results showed the 3 AE judges preferred the post-test to pretest versions for all of the vowels except /æ/ and /i/. Taken as a whole, the results would appear to indicate that the identification training with feedback had a positive effect on the trained Japanese talkers’ production performance.

8. REFERENCES


