A Signal Detection Theorem Analysis of Native Japanese Production of American English Vowels

Stephen G. Lambacher
School of Social Informatics
Aoyama Gakuin University

Abstract: The purpose of this paper was to investigate the effects of identification training on the ability of native Japanese to pronounce the five American English (AE) vowels /æ/, /ɑ/, /ʌ/, /ɔ/, /ɝ/. A production task, performed both before and after a six-week vowel identification training program, included an experimental group and a control group who produced words containing each of the five target vowels within a varied consonantal context. In a separate task, a group of native speakers of English performed an identification task that included a random sampling of the pre- and post-training vowel productions of the Japanese subjects from each of the two groups. A random sampling of vowel utterances as produced by the group of experimental Japanese talkers were then rated by three native English listeners in a two-interval, forced-choice identification task. A signal detection theorem-based correction for bias served to reveal a different pattern of production results and influenced the interpretation of the data, especially with regard to the /ʌ/ vowel. The results showed that the trained Japanese group's productions were more intelligible than the control group's productions for each of the five target AE vowels.

Keywords: signal detection theorem, identifcation training, L2 vowel production

— I —
Introduction

When learning to communicate in a foreign or second language (L2), adults naturally have difficulty mastering certain sound contrasts in the target language. Unfamiliar sounds tend to be perceived and produced according to the acoustic patterns of native language (L1) categories, which leads to the mispronunciation of certain sounds in the L2, and ultimately to miscommunication [1]. This problem may be due to equivalence classification, which is a process whereby two L2 sounds are linked or assimilated into the same native sound category as a result of the listener's native sound system "filtering out" phonetically important features of the L2 sounds [2]. Phonological filtering makes it more difficult to distinguish the acoustic differences between native and nonnative sounds for L2 learners. A learner's phonological filter acts like a "sieve" and passes only information useful to categorizing sounds in the L1 [3]. As a result, L2 learners have a tendency to ignore allophonic differences which distinguish L1 sounds, as well as the subcategorical phonetic differences which distinguish similar sounds in the native and foreign languages [4]. For phonological filtering to occur, an L2 learner must possess enough sensory knowledge to identify the overall physical similarity between the specific L1 and L2 sounds and then must

be able to categorize them as identical as a result of neglecting the physical differences that might serve to distinguish them [1].

Vowels are speech sounds in which there is no obstruction to the flow of air as it passes through the vocal tract from the larynx to the lips. Vowels can be classified according to their tongue height and their frontness or backness. Variation in L1 vowel production depends on a number of factors, including geographical region, social class, educational background, age, and gender. The number of vowels in English is large in comparison to Japanese. General American English (AE) has 14 or 15 vowels including diphthongs (/i/, /ɪ/, /eɪ/, /ɛ/, /æ/, /ɑ/, /ɔ/, /ʊ/, /oʊ/, /u/, /ʌ/, /ɝ/, /ai/, /aʊ/, /ɔɪ/) [5]. Japanese has two sets of five monophthongal vowels (/i/, /e/, /a/, /o/, /ɯ/) which contrast long and short (mono-moraic or bi-moraic), e.g., /ie/ 'house' versus /i:e/ 'no' [6]. Japanese long vowels are 2.4 to 3.2 times longer in duration than short vowels, a ratio much larger than the differences of long and short vowels in other languages [7].

The research examining L2 vowel production training is sparse, with most prior studies focusing primarily on the predominant cues used by learners to perceive and produce L2 vowels, as well as their degree of difficulty. Research has shown that L2 vowels are often perceived and produced in terms of L1 vowels, and that the degree of learning L2 vowels varies according to a number of factors, including the perceived phonetic dissimilarity of the L1 and L2 vowel(s) in question, the size of the L1 vowel inventory, the importance of phonetic features in the L1, such as vowel length, and the amount of prior target language experience [2], [8], [9].

Rochet [10], for example, compared the identification and production performance of native speakers of Portuguese and English of the French vowels /i/, /y/, /u/, and /a/. Results showed that native Portuguese often identify and produce French /y/ as /i/ as opposed to /u/, while English speakers do just the opposite. Polka [11] examined the discrimination of the two German vowel contrasts /y/–/u/ and /ʏ/–/U/ by native speakers of English and found that while most listeners could discriminate the tense vowel pair /y/–/u/ with nativelike performance, they failed to attain nativelike discrimination accuracy for the lax vowel pair /ʏ/–/U/. Flege [2] examined the perception of 14 AE vowel contrasts by native speakers of five L1 backgrounds (Portuguese, Spanish, German, Korean, and Arabic), and found that all the language groups had difficulty discriminating between many of the vowel contrasts. For example, all five groups had difficulty with the AE contrasts /ɛ/–/æ/ and /ɑ/–/ʌ/, and the /u/–/ʊ/ was difficult for the Spanish, Dutch, and Korean speakers.

Table 1. Vocal articulation for production of five AE vowels, including for each vowel its IPA symbol and an example word.

IPA symbol	Example word	Vocal Articulation		
		lips	tongue	mouth
æ	bad	unrounded	front	near-open
ɑ	bod(y)	unrounded	back	open
ʌ	bud	unrounded	mid-back	open
ɔ	bawd	rounded	mid-back	open
ɜ˞	bird	rounded	back	closed

The main goal of the present paper is to examine the effects of identification training on the ability of native Japanese to pronounce the AE mid and low vowel /æ/, /ɑ/, /ʌ/, /ɔ/, /ɝ/. These five vowels were chosen, as prior research has shown these particular vowels are among the most difficult among English vowels for native Japanese to pronounce intelligibly [12], [13]. The vowels of General American English are dealt with exclusively, which is the standard pronunciation used in the American national media and is spoken predominantly by people living in the metropolitan regions of the Midwestern United States. (See Table 1 for a description of the vocal articulation

of the five target AE vowels.) I was also interested in gaining a deeper grasp of which of the target AE vowels are initially more difficult for native Japanese to pronounce and whether their pronunciation improves more for some of the vowels than others as a result of training, including which vowels as spoken by the Japanese talkers are more difficult for AE listeners to identify.

—II— Methods

This study consists of four phases — (1) a six-week identification training task, (2) a pre- and post-training production task, (3) a five-alternative, forced-choice (5AFC) idenfication listening task, and (4) a two-interval, forced-choice (2IFC) listening task. The identification training task was administered once per week for six weeks, and it included 34 subjects from an experimental group who identified words containing each of the five target AE vowels within a varied consonantal context. The pre-/post-training production task included two groups of subjects (the 34 experimental subjects exposed to training and a control group of 20 subjects not exposed to training) who produced words containing each of the five target AE vowels within a varied consonantal context. In the third task, a group of 26 native AE listeners performed a 5AFC identification task that included a random sampling of the pretest and posttest productions of Japanese subjects in each of the two groups. In the fourth task, a random sampling of vowel utterances as produced by the group of experimental Japanese talkers (who received training) were rated by three AE listeners in a two-interval, forced-choice (2IFC) identification task.

All experiments of the present study, except for the identification task in Section 2.4, were carried out in 2004 at the University of Aizu in Fukushima prefecture using the following research facilities: Language Media Laboratory (LML), Recording Sound Studio, and Anechoic Chamber.

2.1 Subjects of the identification training task

The subjects of the identification training task included 34 native speakers of Japanese (all male, mean age=20.2 years) who were from various prefectures throughout Japan, none of whom had lived abroad in an English-speaking country. All of the subjects were second-year students majoring in computer science at the University of Aizu. As is common in Japan, all the listeners had six years of prior English instruction at the junior and senior high school levels, but the focus was on English reading and grammar. The listeners also had little or no prior exposure to training in English pronunciation. None of the participants reported any history of speech or hearing impediments.

2.2 Identification training task stimuli and procedure

The identification training task was administered once per week for a total of six weeks. During this six-week training period, the same five AE vowels /æ/, /ɑ/, /ʌ/, /ɔ/, /ɝ/ were presented in a 5AFC identification task. The stimuli were naturally spoken words recorded by five native speakers of American English. In the training task, the subjects heard 75 stimuli (five speakers x 15 trials), which were presented within a varied consonantal context, including two initial consonants /p/, /t/ and two final consonants /b/, /n/. These particular consonants were chosen to provide a reasonably balanced phonetic context in which to present the five target vowels. A majority of the stimuli were nonsense words (e.g., pob, pon, pern), but only a few were real words in English that the Japanese participants were familiar with (e.g., pan, pub, tub) (see Table 2).

During the training task, the CVC stimuli were presented to the subjects using a Sony TCD-D8

digital audio recorder at an average presentation level of 69 dB SPL via Sony HL-90 headphones in a noise-reduced language laboratory. An introduction to the five AE vowels was given to the participants beforehand, in the form of practice trials, to ensure they understood the correct labeling of each of the target vowels on the answer sheet. For this purpose, the training task administrator uttered sample words, containing one of the five AE vowels. These words were presented orthographically on a handout that was passed out to each participant as a reminder of which alphabet letters (A–E) corresponded to which vowel /æ/, /ɑ/, /ʌ/, /ɔ/, /ɝ/.

During training, immediate feedback was given to subjects through their headphones after each response informing them of the correct answer (either A, B, C, D, or E) for each stimulus. There was a pause of seven seconds between each stimulus, and an interval of five seconds between the presentation of each stimulus and the administration of the feedback.

Table 2: The stimuli used during the identification training task.

Training Stimuli

æ	ɑ	ʌ	ɔ	ɝ
Paeb	pab	pub	paub	perb
paen	pan	pun	paun	pern
taeb	tab	tub	taub	terb

2.3 Production task stimuli and procedure

The production task included the same experimental group of 34 native Japanese who participated in the identification training program as well as a control group of 20 native Japanese with the same background but without training. All of the subjects recorded words, each containing one of the five target AE vowels /æ/, /ɑ/, /ʌ/, /ɔ/, /ɝ/. The five vowels were inserted in /k_d/, /k_p/, /t_d/, /t_k/ frames yielding 20 minimal pairs.

The production pretest recordings were carried out before the identification training to avoid the possible effects of exposure to the identification stimuli on production. An introduction to the production task was given to the subjects beforehand in the form of practice trials. For the two groups of Japanese subjects, a native speaker of American English uttered sample words containing each of the five AE vowels, so that subjects better understood the pronunciation of each of the vowels, but no explicit instruction in the articulation in any of the vowels was administered.

Both groups of Japanese subjects were recorded individually in a soundproof room using a Denon Model DTR-80P digital audio recorder at a 48 kHz sampling rate. Each participant read all of the stimuli from a list and was informed beforehand to produce them with a natural 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. All of the uttererances were normalized for peak frequency.

2.4 Identification of Japanese pre-/post-training productions of AE vowels (procedure)

To evaluate the transfer of the identifcation training to the experimental group's production of the five AE vowels, a group of native AE listeners performed a 5AFC identification task that included the pre- and post-training productions of the two groups of Japanese subjects. The purpose of this task was to determine if the group of native AE listeners could reliably identify each of the five AE vowels as produced by both groups of Japanese subjects.

Twenty-six native speakers of American English (17 female and nine male) served as the listeners. All of the listeners (mean age = 21 years) were undergraduates majoring in linguistics at West

Chester University in Pennsylvania who had received training in both phonetics and phonology and were familiar with IPA transcription. None of the AE listeners reported that they had a history of speech or hearing impediments.

The pre-/post-training stimuli produced by the experimental and control groups of Japanese subjects were presented in a set of four 5AFC identification tests to the 26 AE listeners. Vowel productions from the two groups of native Japanese subjects were included in the 5AFC task. It should be noted, however, that production data from a subset of 20 experimental subjects of the original group of 34 subjects was used. Each test was comprised of words in mixed consonantal contexts, each word containing one of the five target vowels, and blocked for each group: i.e., the trained subjects' pre-training productions; the trained subjects' post-training productions; the control subjects' pre-training productions; and the control subjects' post-training productions. From each of the 40 Japanese subjects (20 per group), a random sampling of eight out of the 20 vowel stimuli they produced were chosen for each of the four identification tasks presented to the 26 AE listeners.

Four tests were presented in a randomized order. In each test, the 26 AE listeners were presented 160 tokens (20 talkers x 8 stimuli), and each of the four tests was administered during a separate class meeting. Each stimulus was presented to the AE listeners at an average presentation level of 70 dB SPL via Sennheiser HD590 headphones in a noised-reduced room. In each test, there was a pause of five seconds between the presentation of each stimulus and a longer pause of 15 seconds after every 20 items to give the AE listeners a short break. Each of the four listening tests took approximately 30 minutes to complete.

—Ⅲ— Results

3.1 Identification of native Japanese AE vowel productions by AE listeners

The overall accuracy of the AE listeners' identification of the target vowels as produced by the experimental Japanese talkers in the pretest was not very high. The grand total of the correct pretest responses shows an overall identification rate of 38%. In the pretest, confusion was greatest between the /ɑ/–/ʌ/, /ɝ/–/ɔ/, and /ʌ/–/ɔ/ contrasts. For example, while only 25% of /ɑ/ tokens were actually heard as /ɑ/, 35% were heard as /ʌ/. Also, while only 21% of /ɝ/ tokens were heard as /ɝ/, 36% were heard as /ɔ/. In addition, 23.5% of /ɔ/ stimuli were heard as /ʌ/. In the posttest, the confusion between the /ʌ/–/ɑ/ and /ʌ/–/ɔ/ contrasts decreased by only 5% and 2%, respectively. However, the pretest percentage of /ɝ/ tokens identified as /ɔ/ decreased by 16% in the posttest.

The identification responses of the group of 26 AE listeners obtained in four different testing sessions were submitted to a three-way ANOVA in which the factor termed "Group" served as a between-subjects factor (trained vs. untrained subjects). Two within-subject factors were termed "Time of Test" (pre- vs. post-training) and "Vowel" (five stimulus vowel sounds). Of course, terming the "Time of Test" levels as "pre" vs. "post" refers to the training that only the experimental group subjects received. The ANOVA yielded significant main effects of Group, $F(1,54)=14.31$; $p < .01$, Test of Time, $F(1,54)=11.42$; $p < .01$, and Vowel, $F(4,216)=138.49$; $p < .01$. There were significant two-way interactions between variables Group and Time of Test, $F(1,54)=14.33$; $p < .01$, and between Group and Vowel, $F(4,216)=2.28$; $p < .01$, as well as a significant three-way Group x Time x Vowel interaction, $F(4,216)=6.26$; $p < .01$. Post-hoc t tests revealed the identifiability of the

trained group's productions improved from the pretest (mean = 39%) to posttest (mean = 47%), $p < .01$, while the identifiability of the control (untrained) group's productions did not increase from pretest (mean = 36%) to posttest (mean = 35.5%), $p = .844$.

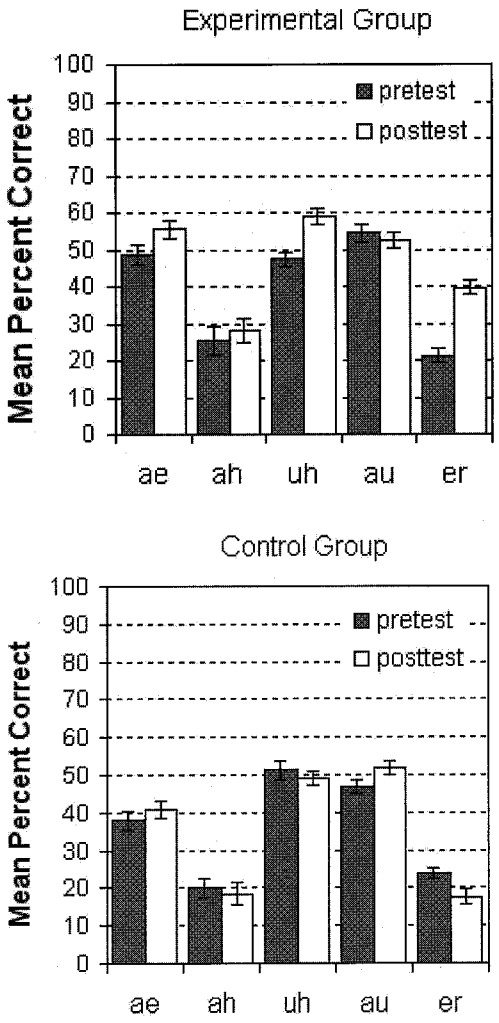

Figure 1. Mean percentage of correct responses by the AE listeners during the pre- and post-training identification tasks of the experimental group's (top panel) and control group's (bottom panel) productions of the 5 AE vowels. Standard error bars are shown for each. Note: The letters ae, ah, uh, au, er in both figures refer to the AE vowels /æ/, /ɑ/, /ʌ/, /ɔ/, /ɝ/, respectively.

Post hoc *t* tests revealed the two-way interaction between variables Time of Test and Vowel reached significance for the trained group, $p < .01$, but not for the untrained group, $p > .05$. The AE listeners identified the trained group's pretest productions of /ɔ/ (mean = 55%), /æ/ (mean = 49%), and /ʌ/ (mean = 47%) more accurately than the other two vowels /ɑ/ (mean = 25%) and /ɝ/ (mean = 21%). The AE listeners thus showed minimal ability to identify the experimental group's pre-training versions of the /ɝ/ and /ɑ/ stimuli, since random guessing would lead the listeners to exhibit a 20% correct identification rate. However, the ability of the AE listeners to identify the trained group's productions of /ɝ/ (+19%) increased considerably more in the posttest than their ability to identify the other four vowels /ʌ/ (+ 11%), /æ/ (+ 7%), /ɑ/ (+ 4%), /ɔ/ (-2%) (See Figure 1).

—Ⅳ—
Two-Interval, Forced-choice (2IFC) identification task procedure

To further evaluate the transfer of the identification training to the experimental group's production of the five AE vowels, three AE listeners performed a 2IFC identification task that included a random sampling of the pretest and posttest productions of the experimental group. The intention was to determine whether the experimental group's production accuracy of the five target AE vowels improved as a result of the identification training.

The task examined whether the AE speakers could reliably distinguish the better of two productions, one produced before the identification training, and the other, after. Since the listeners were not informed regarding which production was which, the test provided an unbiased means to verify whether the talkers had improved their production of the five target AE vowels after training. The three AE listeners, all professors from the University of Aizu, were from the Midwestern United States who were familiar with /ɑ/–/ɔ/ distinction (mean age = 42 years old). None of the three listeners reported any history of speech or hearing impediments.

The three listeners judged the full set of pre-/post-test productions of all 34 Japanese subjects. The stimuli consisted of 680 pre-/post-word pairs (34

talkers x 5 vowels x 4 words). The stimuli were presented in five 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 words containing only one of the five AE vowels, and each trial contained two productions of the same word (pretest and posttest, in a randomized order) by a single Japanese talker. During each block of trials, the three AE listeners knew which of the five target vowels they were judging. The three 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 AE 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 AE listeners were urged to use the whole scale and had to choose even if they were uncertain of how to respond.

4.1 Results (2IFC)

The results of the 2IFC task are presented in Figure 2 in the form of the hit rates for posttest productions of the trained talkers (i.e., if the posttest 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 34 trained talkers. The bars in Figure 2 show the average hit rates for the three 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 Figure 2, the three AE listeners preferred the posttest productions of the trained Japanese talkers for four out of five target AE vowels. The "post" hit rates were highest for the /æ/ and /ɝ/ stimuli (.75). The posttest versions of /ɔ/ (.63) and /ɑ/ (.60) stimuli were also preferred, although only two of three AE listeners were above the level expected by chance alone (.57 at $p < .05$). The three judges showed no ability to distinguish between the Japanese talkers' pre- and post-test versions of the /ʌ/ stimuli.

The results of the two production evaluation tests (5AFC and 2IFC identification tasks) demonstrated significant improvement in the experimental group's production of each of the five AE vowels as a result of the identification training. In the 5AFC identification task, the experimental group's productions of each of the five vowels were more accurately identified by the AE listeners in the posttest. In contrast, there was no proof that the control group's productions of the target vowels improved. Additionally, in the 2IFC identification task, the AE judges preferred the trained group's posttest to pretest versions of each of the AE vowels, except for the /ʌ/ vowel.

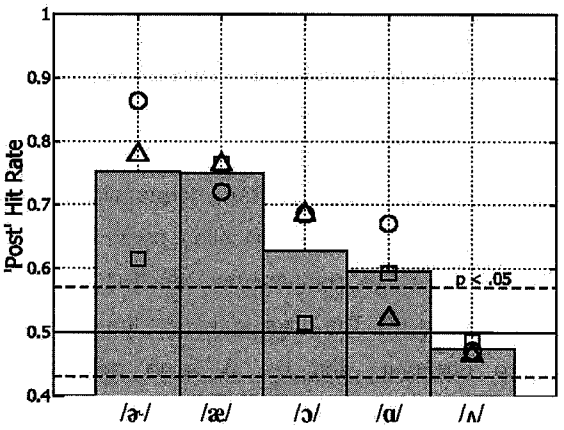

Figure 3: Average hit rate for each of five AE vowels for the 34 Japanese trained talkers' posttest productions as judged by the three AE listeners. The symbols in the figure for the vowels represent the 'posttest' hit rates by each of the three judges, respectively. Note that out of 140 trials, there is a p = < .05 probability of reaching a .57 hit rate by chance alone.

—V—
Signal-Detection-Theoretic (SDT) Analysis (d')

It was perplexing to that the hit rates (percent correct) for the /ʌ/ vowel in the identification task by AE listeners appeared inflated in comparison to other target AE vowels. Recall in the 2IFC task that the three phonetically trained AE judges were unable to distinguish between the pretest and posttest training productions of the /ʌ/ vowel. It was decided that additional analyses were necessary to examine the identification data more carefully to determine if perhaps an underlying bias had elevated the hits rates in the 5AFC task by the group of 23 AE listeners for the /ʌ/ vowel stimuli as produced by the trained Japanese talkers. Subsequently, the confusion rates from the identification tasks were submitted to signal detection theory (SDT) analysis.

Signal Detection Theory (SDT) provides an analytic technique that finds more in identification data than simple recognition rates. The application of SDT to the analysis of speech intelligibility is of course not new [14], and although SDT has not been used so frequently in L2 speech and language research, it has been fruitfully employed in recent studies [15]. What is the benefit of SDT for perception studies? The correction for bias can serve to change the pattern of results and can influence the interpretation of the data. SDT represents identification performance by separating out two components: sensitivity (d') and response bias (β). The d' value reflects the sensitivity of the observer, and the β value reflects the observer's criterion for choosing one response over another (capturing the tradeoff between goals of indicating a signal is present when it is present, versus not reporting presence when the signal is absent). The importance of SDT is that it allows the examination of these two factors underlying observed identification rates (and these separate measures are based on all of the identification data rather than just percent correct) [14], [16]. Thus, when one response is simply more commonly produced than another, regardless of the stimulus (i.e., higher false alarm rates), the sensitivity measure can be calculated with a compensation for this bias, which is expressed as a tendency to produce elevated hit rates.

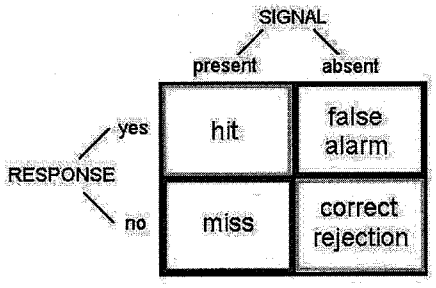

Figure 4: SDT Model

SDT computations are based upon examination of the number of hits (correct detections) versus the number of false alarms (incorrect detection responses). A subject's response bias, (β), is expressed as a likelihood ratio for different responses, while d' is computed by converting hit rate (H) and false alarm rate (F) to z-scores and then subtracting, as follows: $d' = z(H) - z(F)$, where H is the proportion of hits relative to the number of trials in which a signal is present, and F is the proportion of false alarms relative to the number of trials in which no signal is present. See Figure 4 for a model of SDT.

5.1 Results (SDT analysis)

Figure 5 shows the resulting bias-corrected identification scores (d') for the AE listeners. From the post-training results, it is clear that the AE listeners could more sensitively identify the experimental 'trained' group's post-training productions of each of the AE vowels than they could the control 'untrained' group's productions of the same vowels (except for the /ɔ/ vowel). Though the rank ordering of vowel identification hit rates agrees fairly close with the ordering of vowel identifiability measured by d', there are a number of noteworthy

exceptions. In the identification pretest, due to the fact that AE listeners were biased to give more /ʌ/ responses (beta=0.78) than /ɜ/ responses (beta=0.40), the percentage identification hit rate for /ʌ/ (47%) clearly dominates that for /ɜ/ (22%). However, this does not mean that /ʌ/ was more easily identified than /ɜ/. On the contrary, the SDT-based analysis showed that the /ʌ/ vowel, as produced by the trained talkers, was less identifiable than /ɜ/ (with d' values of 0.64 vs. 0.77, respectively). Since, in the posttest, the AE listeners were biased in favor of the /ʌ/ response (beta=0.75) relative to /ɜ/ responses (beta=0.44), the percentage identification rate for /ʌ/ (58%) dominates that for /ɜ/ (40%). The SDT analysis clarified this potentially confusing result, that /ʌ/ was clearly less identifiable than /ɜ/ (d' values of 0.98 vs. 1.04, respectively), despite the fact that bias had elevated the percentage identification rate for /ʌ/. Additionally, there was an inconsistency between identification hit rates and sensitivity values when comparing the /ʌ/ and /æ/ vowels. In the posttest, the AE listeners were biased in favor of the /ʌ/ response (beta=0.75) relative to /æ/ responses (beta=0.27), so the percentage identification rate for /ʌ/ (58%) dominates that for /æ/ (55.4%); even though SDT revealed that /ʌ/ was considerably less identifiable than /æ/ (d' values of 0.98 vs. 1.77, respectively).

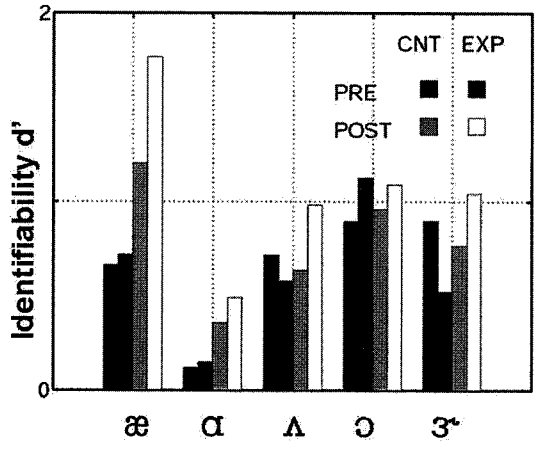

Figure 5: A comparison of the pre-/post-training identification rates of both groups of AE listeners using SDT (as measured by d').

Uncertainties about the elevated hit rates by the group of 23 AE listeners for the /ʌ/ vowel stimuli as produced by the trained Japanese talkers were thus confirmed by the results of the SDT analysis, which revealed that the AE listeners had an underlying bias for the /ʌ/ vowel. It should be reiterated how easy it is to be fooled by looking at identification hit rates without paying attention to variations in response bias.

—VI— Discussion

The overall results revealed that identification training also had a positive effect on the experimental 'trained' group's production of the target AE vowels /æ/, /ɑ/, /ʌ/, /ɔ/, /ɜ/. These results support earlier findings that training using identification tasks are typically more effective than discrimination tasks in improving the L2 pronunciation skills [17], [18]. Identification tasks enable L2 learners to focus on higher-order linguistic processing of speech stimuli and to develop more robust categories of the target stimuli, which are stored in long-term memory, as opposed to discrimination tasks which focus only on tapping into short-term sensory memory [18]. Identification tasks have also been shown to be more difficult than discrimination tasks because they require listeners to sort stimuli into separate categories. They thus impose more echoic memory loads, which enable listeners to focus on more relevant linguistic processing of the speech signal [17]. Discrimination tasks, on the other hand, only emphasize within-category acoustic differences and may also focus listeners' attention at the boundaries of phonetic categories rather than help to define category centers [19].

The results would appear to support L2 speech theories in their claim that learners assimilate L2 segments to L1 categories based on the perceived phonetic similarities between L1 and L2 sounds

[2],[20]. The auditory and phonetic distinctness of the AE vowels /ɝ/ (a rounded, closed mid vowel), /æ/ (a near-open, low front vowel), and /ɔ/ (a rounded mid-back vowel) from the native Japanese /a/ vowel (an unrounded, open, low mid vowel) could have assisted trained subjects to become better attuned to the acoustic representations of these three AE vowels, thus making it easier for them to establish new phonetic categories of these vowels. The vowel /ɝ/, in particular, is articulated with round lips and a raised tongue, resulting in a unique character associated with its low third formant [5], which could have aided the subjects in acquiring it. In contrast, subjects had the most difficulty with both /ɑ/ and /ʌ/, which differ only slightly from Japanese /a/ in terms of phonetic and acoustic quality [5]. It is possible that the subjects may have equated the /ɑ/ and /ʌ/ vowels with Japanese /a/. Once assigned to the native /a/ category, all stimuli containing /ɑ/ and /ʌ/ were produced with the /a/ category and then heard by AE listeners as sounding like their own /ɑ/. The difference in duration between AE /ɑ/ and /ʌ/ could have also played a role in the confusion by the native Japanese subjects, since duration is phonemic and the primary perceptual cue to vowel length in Japanese [21].

According to Kuhl [20], during the course of acquiring an L1, the learner develops acoustic prototypes for native categories which involves general auditory processes that process acoustic and not specific phonetic information. The perceptual space surrounding a prototypical vowel is shrunk, making it harder for the L2 learner to perceive variant sounds located nearby, than it is to perceive sounds located near nonprototypes or poor exemplars. It is extremely difficult, therefore, for learners to develop prototypes of L2 sounds due to a lack of acoustic experience with the sounds. L2 vowels that are in a space close to L1 vowel prototypes are more difficult to perceive for L2 listeners than vowels that are not close to a prototype vowel. In this present study, this would perhaps suggest that since the perceptual space surrounding the native /a/ vowel is shrunk, the AE /ɑ/ and /ʌ/ vowels were harder for the native Japanese to produce in comparison to other target "nonprototypical" vowels such as /ɝ/, /æ/, and /ɔ/, since both /ɑ/ and /ʌ/ are perceptually closer to the "prototypical" native /a/ and can thus be classified as "deviant" vowel sounds.

It was expected that the use of a highly variable identification training procedure with multiple tokens and speakers (instead of a more restricted environment) would improve the experimental group's production ability, as the results confirm previous findings that the use of stimulus variability is important when training in L2 speech contrasts [2], [10]. Some researchers have suggested that the use of a variable identification procedure, because of its highly diverse nature, has the effect of shifting the selective attention weights of L2 learners to enable them focus on the stimulus dimensions that serve to characterize L2 categories [18]. Nosofsky describes this psychological process as "stretching" and "shrinking" of the vowel space of the listener. Through the identification training, vowels that belong to different categories become less similar to each other as representations from them are "stretched along category acquisition," and thus made easier to pronounce [22]. In the present study, the use of multiple exemplars of vowel tokens produced by multiple talkers might have sensitized subjects to the acoustic cues that are more generally useful in distinguishing nonnative contrasts [20]. As the training progressed, the listeners' attention weights were adapted to fit those of the new vowel contrasts. [10]. Such a process could account for the superior performance of the experimental subjects in producing the /ɝ/ vowel after their exposure to the identification training.

The better performance in producing AE /æ/ and /ɝ/ could be interpreted as evidence that the Japanese subjects did so by modifying previously established articulatory patterns. Perhaps, during the process of learning a second language, the learner becomes attuned to the distinct properties of the phones that are salient to the L2 and are able to shift their

attention weights to those properties and establish new categories for vowels. Such a process could account for the better performance of the Japanese subjects in both identifying and producing the AE /æ/ and /ɜ/ vowels. In the case of AE /ɑ/ and /ʌ/, it could be that Japanese L2 learners require additional identification training before they can begin to successfully shift their attention weights to attend to the salient acoutic properties that distinguish the /ɑ/ and /ʌ/ vowels, and thus begin to establish new categories for them.

—Ⅶ— Conclusion

The results from the 5AFC and 2IFC identification tasks showed that the experimental 'trained' group's productions were more intelligible than the control 'untrained' group's productions for each of the five target AE vowels. An SDT-based correction for bias served to reveal a different pattern of production results and influenced the interpretation of the data, especially in the case of the /ʌ/, /ɜ/, and /æ/ vowels. By correcting for such biases, future studies could benefit from examining the effects of identification training on L2 acquisition of speech contrasts, as it would allow for a deeper understanding of the underlying basis for the observed identification rates, rather than simply relying on the percent correct data. The present study thus deepens our understanding of the effects of auditory training in facilitating the production performance of nonnative vowels by a group of adult L2 learners, even without the administration of any explicit articulatory training.

Bibliography

[1] Flege, J.E. (1988). The production and perception of foreign language speech sounds. In H. Winitz (Ed.), *Human communication and its disorders* (pp. 224-401). New Jersey: Ablex Publishing.

[2] Flege, J.E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), Speech *perception and linguistic experience: Issues in cross-language research* (pp. 233-277). Baltimore: York Press.

[3] Trubetzkoy, N. (1939/1969) *Principles of Phonology*. Translated by C. A. Baltaxe, Berkley: University of California Press.

[4] Jenkins, J. (1998). Which pronunciation norms and models for English as an international language? *ELT Journal*, 52 (2), 119-126.

[5] Ladefoged, P. (2001). *Vowels and consonants: An introduction to the sounds of languages*. Oxford: Blackwell Publishers.

[6] Vance, T.J. (1987). *An introduction to Japanese phonology*. New York: State University of New York Press.

[7] Lehiste, I. (1970). *Suprasegmentals*. Cambridge, MA: The MIT Press.

[8] Flege, J.E., Munro, M., & Fox, R. (1994). Auditory and categorical effects on cross-language vowel perception. *Journal of the Acoustical Society of America*, *95*, 3623-3641.

[9] Polka, L. & Bohn, O. (1996). A cross-language comparison of vowel perception in English-learning and German-learning infants. *Journal of the Acoustical Society of America*, 100, 577-592.

[10] Rochet, B.L. (1995). Perception and production of second-language speech sounds by adults. In W.Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 379-410). Baltimore: York Press.

[11] Polka, L. (1995). Linguistic influences in adult perception of non-native vowel contrasts. *Journal of the Acoustical Society of America*, 97, 1286-1296.

[12] Nishi, K., Strange, W., Akane-Yamada, R., Kubo, R., & Trent-Brown, S. (2008). Acoustic and perceptual similarity of Japanese and American English vowels. *Journal of the Acoustical Society of America*, 124 (1), 576-588.

[13] Strange, W., Akane-Yamada, R., Kubo, R., Trent, S., Nishi, K., & Jenkins, J. (1998). Perceptual assimilations of American English vowels by Japanese listeners. *Journal of the Acoustical Society of America*, 26, 311-344.

[14] Macmillan, N. A., Kaplan, H. L., & Creelman, C.D. (1977). The psychophysics of categorical perception. *Psychological Review*, 84, 452-471.

[15] Takagi, N. (1995). Signal detection modeling of Japanese listeners' /r/–/l/ labeling behavior in a one-interval identification task. *Journal of the Acoustical Society of America*, 97, 563-574.

[16] Shepard, R.N. (1980). Multidimensional scaling, tree-fitting, and clustering. *Science*, 210, 390.

[17] Strange, W. & Dittman, S. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English.

Perception & Psychophysics, 36, 131-145.

[18] Lively, S.E., Logan, J.S., & Pisoni, D.B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The Role of phonetic environments and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America*, 94, 1242-1255.

[19] Repp, B. (1984). Categorical perception: Issues, methods, findings. In N. Lass (Ed.), *Speech and language: Advances in basic research and practice*, vol. 10 (pp. 243-335). New York: Academic.

[20] Kuhl, P.K. & Iverson, P. (1995). Linguistic experience and the "perceptual magnet effect." In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 121-154). Baltimore: York Press.

[21] Ingram, J. & Park, S.G. (1997). Cross-Language vowel perception and production by Japanese and Korean learners of English. *Journal of Phonetics*, 25, 343-370.

[22] Nosofsky, R. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 13, 87-108.