WHICH IS BETTER: IDENTIFICATION OR DISCRIMINATION TRAINING FOR THE ACQUISITION OF AN ENGLISH CODA CONTRAST

Iris Lok Gi Law[1], Izabelle Grenon[1], Chris Sheppard[2], John Archibald[3]

[1]The University of Tokyo, [2]Waseda University, [3]The University of Victoria
irislaw@link.cuhk.edu.hk, grenon@boz.c.u-tokyo.ac.jp, chris@waseda.jp, johnarch@uvic.ca

ABSTRACT

This study evaluates whether identification training yields superior results to discrimination training for improving the perception of a very difficult non-native contrast. Seventeen Japanese speakers received one hour of identification training with feedback with the English coda contrast /z/-/dz/ as in 'rose' and 'roads'. The 'rose'-'roads' stimuli were manipulated to vary in terms of the duration of the stop closure and the duration of the preceding vowel. These results were compared with that of twenty Japanese speakers who received one hour of AX discrimination training with the same contrast using the same stimuli. The cue-weighting pre-test and post-test revealed no improvement in the use of vowel duration for either group. However, the identification group improved their use of the closure duration to a slightly (and significant) greater extent than the discrimination group. Hence, identification training may provide superior results for the acquisition of very difficult L2 contrasts.

Keywords: Identification training, discrimination training, coda contrast, English, Japanese.

1. INTRODUCTION

Phonetic training often uses an identification task, where the second language (L2) learners are presented with one word (e.g., *rose*) and have to decide which word was heard (e.g., 'rose' or 'roads'). This type of training, especially combined with high variability, have been shown to be effective for improving the perception of L2 contrasts [e.g., 7, 10, 12]. While identification tasks require some level of literacy or, at least, some knowledge of phoneme-grapheme correspondence in the L2, discrimination tasks do not. An AX discrimination task, for instance, consists of presenting the learners with two words (e.g., *rose-roads*) and asking them to decide whether the two words were the same or different. Hence, a possible advantage of the use of a discrimination task for sound training is that it could be used with populations that are not literate in the L2, and therefore, could be used at the very onset of L2 learning. Some studies reported that both tasks were equally effective for training sound contrasts [5, 16], while others reported better improvement with identification training [3, 4, 13, 15]. Thus, it is still unclear which task is better and why.

The studies that reported better improvement with an identification task targeted vowels [3, 4, 13] or the /r/-/l/ contrast [15]. However, a recent study has demonstrated that the difference between identification training results and discrimination training results may be due to mislabeling issues: 25% of the Japanese speakers in the discrimination training condition associated the English vowel /i/ with the word 'ship' instead of 'sheep' in the post-test [8]. However, their improvement in the use of temporal and spectral information was comparable across both training conditions [17]. Hence, the discrimination task may be as effective as the identification task for helping learners to contrast categories. It is still possible that the lack of difference found between conditions may be due to a ceiling effect since the Japanese speakers were sensitive to both duration and spectral changes *before* training. Also, their average training scores were around 90% [17], thus, this contrast may not have been overly difficult for them to begin with.

The /z/-/dz/ contrast as in the English words 'rose' and 'roads' is also difficult for native Japanese speakers [6]. This is because the affricate /d^{z}/ and fricative /z/ are allophones in Japanese [1]. In a previous study [9], we found that the average training scores of Japanese speakers on this contrast was around 50% (that is, chance level) when using an AX discrimination task, suggesting that this contrast is more difficult than the 'ship'-'sheep' contrast previously investigated ([8], [17]). The Japanese speakers slightly improved their use of vowel duration and closure duration of the stop /d/ to categorize the words 'rose' and 'roads' with discrimination training [9]. The current study is interested in whether identification training may help Japanese speakers improve their use of the contrastive cues more than discrimination training.

At least two acoustic cues are used by native English speakers to distinguish the word 'rose' from 'roads': the duration of the vowel and the duration of the stop closure [9]. English speakers associate a longer vowel with the word 'rose', and generally perceive the word as 'rose' when there is no stop closure or it is short (around 10 ms). Conversely,

Japanese speakers associate a longer vowel with the word 'roads' instead of 'rose', and disregard changes in closure duration. After discrimination training, the Japanese speakers slightly improved their use of the stop closure duration, while starting to rely less on vowel duration [9]. Thus, the current study evaluated whether the cue-weighting of Japanese speakers can get closer to native performance with an identification training paradigm using the same set of manipulated stimuli as those used in the previously reported AX discrimination training paradigm [9].

2. METHODOLOGY

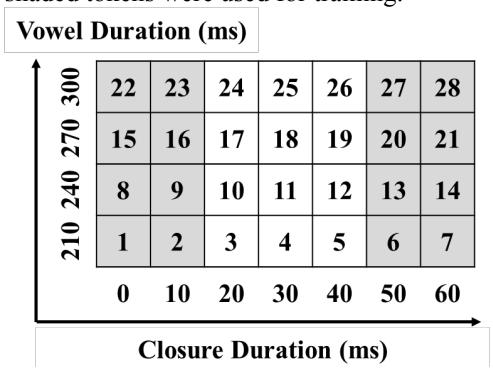
2.1. Participants

Seventeen native Japanese speakers who had not spent more than 3 weeks in an English-speaking country (mean = 1.25) took part in the identification training with the 'rose' and 'roads' contrast. Their results were compared with that of twenty native Japanese speakers who had not spent more than 8 weeks (mean = 1.7) in an English-speaking country and took part in the AX discrimination training with the same contrast. All participants were right-handed university students in Japan between the age of 18 and 23 (mean = 21) for the identification task and between 18 and 27 (mean = 20) for the discrimination task, with no reported history of speech or hearing impairment. Forty North American English university students recruited in Victoria, Canada and aged between 17 and 28 (mean = 21) took part in the pre-test.

2.2. Stimuli

Test tokens were created from the recordings of 'rose' and 'roads' samples by a female North American English speaker. The recordings at 44,100Hz were performed in a sound-proof room at The University of Tokyo with a Sony ECM-MS957 condenser microphone connected directly to a computer using Praat [2]. The recorded words were scaled to an intensity of 70 dB before manipulations. Forty milliseconds of stop closure were extracted from a clear 'roads' sample and inserted between the vowel and the coda fricative of a clear 'rose' sample. The closure duration was then modified from 0 to 60 ms in steps of 10 ms using a script [18]. The vowel duration of each of the resulting tokens was modified from 210 to 300 ms in steps of 30 ms using the same script, which resulted in a total of 28 test tokens as illustrated in Figure 1. While all 28 tokens were used in the identification pre-test and post-test, a subset of 16 tokens was used for training. The tokens chosen for training were those at the extreme ends of the closure duration, a cue that is used more categorically than vowel duration by English speakers to distinguish 'roads' from 'rose'. The tokens chosen for training correspond to the 8 tokens (tokens 1, 2, 8, 9, 15, 16, 22, 23) most often categorized as 'rose' by native English speakers and the 8 tokens (tokens 6, 7, 13, 14, 20, 21, 27, 28) most often categorized as 'roads'.

Figure 1: The 28 tokens used for the pre- and post-test were manipulated along the closure duration (x-axis) and vowel duration (y-axis) dimensions. The 16 shaded tokens were used for training.



2.3. Procedure

2.3.1. Pre- and post-test

The English participants completed the pre-test only (the pre-test and post-test were identical). Their results were reported in [9] and are used in the current study for comparison with native proficiency. All Japanese participants completed the pre-test, one hour of training, and the post-test. For the tests, participants were presented with one token at a time, and asked to identify whether the word was 'rose' or 'roads'. The 28 tokens were presented randomly 4 times. The results on the first set of 28 tokens were discarded from the analyses as a practice session. All the tests were carried out in a sound-attenuated room, with stimuli presented via high quality BOSE headphones. No feedback was provided during a test.

2.3.2. Identification training

After the pre-test and before the post-test, the seventeen Japanese participants assigned to the identification training condition received two 30 to 40-min. sessions of identification training with feedback using the 16 tokens in grey shading in Figure 1. For the training, participants were presented with one token at a time, and asked to identify whether the word was 'rose' or 'roads'. Feedback consisted of a written message indicating whether the choice was correct (the tokens were never repeated). The 16 training tokens were presented randomly 8 times per block, and there were 4 blocks per training

session. Hence, each of the 16 tokens was presented 32 times for a total of 512 tokens per training session.

2.3.2. AX Discrimination training

The twenty Japanese participants assigned to the discrimination training condition were also exposed to the same 512 tokens during a training session. Two training sessions were provided in total, each lasting around 30 minutes. The 16 training tokens (shaded in Figure 1) were presented 32 times each in a 'same'-'different' AX discrimination task with an inter-stimulus-interval (ISI) of 1500 ms, where tokens were paired as 'same' (e.g., token 1 and 16) or 'different' (e.g., token 1 and 7) in terms of closure duration category. No token was paired with itself. The results on the AX discrimination task was reported in [9].

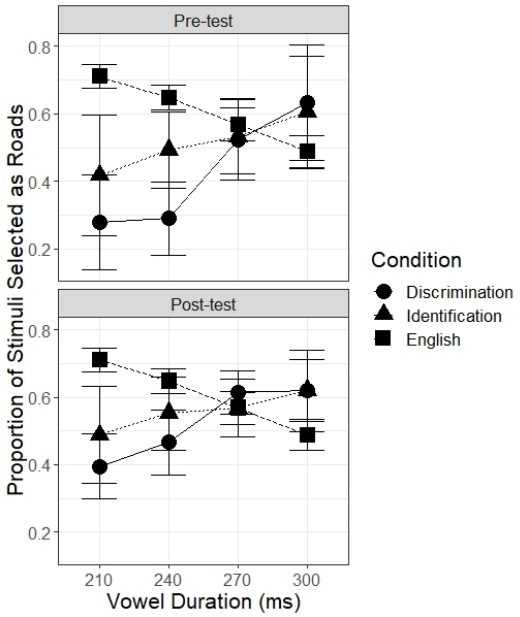
3. RESULTS AND DISCUSSION

In order to assess whether any improvement occurred during identification training, we looked at the average correct training scores, which were 65.08% on the first day, and 68.77% on the second day, a significant improvement of 3.69% ($t(16)$: 3.69, p = .039). In comparison, the training scores on the discrimination task were 52.40% on the first day, and 56.02% on the second day, an improvement of 3.62% that was also significant ($t(20)$ = 2.49, p = .025). Hence, a small but significant increase in task performance was observed with both identification and discrimination training. Accordingly, the research question addressed by the current paper was which training paradigm yielded the best results for a change in the use of vowel duration and closure duration for categorization of the English 'rose' and 'roads' contrast by Japanese speakers. The statistical analyses revealed that neither training group showed a significant improvement on the use of vowel duration from pre-test to post-test, but that the identification training group exhibited slightly better (and significant) results for the use of the stop closure duration.

3.1. Vowel duration

A look at the vowel duration data, presented in Figure 2, indicates that while English speakers generally associated a short vowel with the word 'roads', the Japanese participants in both training conditions associated a short vowel instead with the word 'rose', both before and after training. The vowel duration data were analysed using a mixed-design ANOVA in R [14] with a within-subject factor of Vowel Duration and Time (pre-test and post-test) and a between-subject factor of Condition (discrimination and identification). The package "ez" was used for the analysis [11]. Mauchly's test indicated that the assumption of sphericity had been violated (W = 0.18, p < .001), therefore degrees of freedom were corrected (ε = 0.49). Both training groups did not improve from pre-test to post-test, which was shown by the non-significant Time X Vowel Duration interaction; $F(3, 102) = 0.98$ p = .326, η_p^2 = 0.008. The Time X Condition X Vowel Duration interaction was also non-significant; $F(3,102) = .304$, p = .67, η_p^2 = .002. Hence, although a slight change in the use of vowel duration may be observed in Figure 2, neither the identification nor the discrimination training group experienced a significant improvement in the use of vowel duration for categorisation of the manipulated 'rose' and 'roads' tokens in the current study.

Figure 2: The proportion of items identified as 'roads' in the pre-test and post-test for the discrimination and identification groups by changes in vowel duration. English results on pre-test are added to both graphs for comparison.

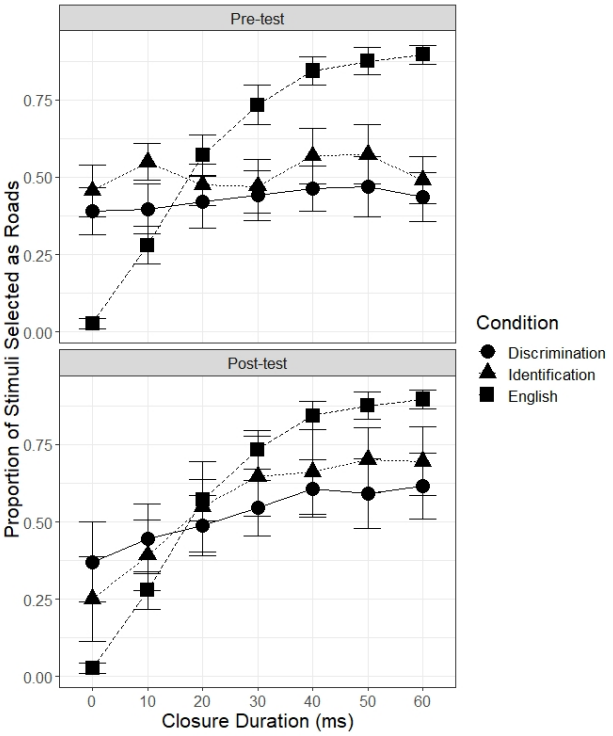


Furthermore, the post-test vowel duration data of both training groups were compared with the native English speakers' test results with a mixed-design ANOVA with Vowel Duration as the within-subject and Condition (identification, discrimination, and English) as the between-subject factor. The data was not spherical (W = 0.34, p < .001), and so Greenhouse-Geisser estimate of sphericity (ε = 0.58) was used. The Condition X Vowel Duration interaction was significant; $F(6, 222) = 15.48$, p < .001, η_p^2 = .18. Thus, the behaviour of both training groups was different to that of the English native speakers.

3.2. Closure duration

The closure duration data are presented in Figure 3. While both groups of Japanese participants mostly disregarded changes in closure duration in the pre-test, they used this cue significantly more in the post-test. The Closure Duration data were analysed using a similar mixed-design ANOVA with a within-subject factor of Closure Duration and Time (pre-test and post-test) and a between-subject factor of Condition (discrimination and identification). Mauchly's test indicated that the assumption of sphericity had been violated ($W = 0.09$, $p < .001$), therefore degrees of freedom were corrected using Greenhouse-Geisser estimate of sphericity ($\varepsilon = 0.47$). Both of the training conditions changed their behaviour over time, which was shown by the significant Time X Closure Duration interaction; $F(6, 204) = 6.74$, $p < .001$, $\eta_p^2 = .05$. The Time X Condition X Closure Duration interaction was also significant; $F(6, 204) = 2.44$, $p = .026$, $\eta_p^2 = .02$, indicating that there was a marginal effect benefitting the identification group.

Figure 3: The proportion of items identified as 'roads' in the pre-test and post-test for the discrimination and identification groups by changes in stop closure duration. English results on pre-test are added to both graphs for comparison.



The post-test closure duration data of both training groups were then compared with that of the English native speakers with Closure Duration as the within-subject factor, and Condition (discrimination, identification, and English) as the between-subject factor. Again, Mauchly's test indicated a violation of sphericity ($W = 0.081$, $p < .001$), therefore Greenhouse-Geisser estimate of sphericity ($\varepsilon = 0.506$) was used. The Condition X Closure Duration interaction was significant; $F(12, 444) = 14.67$, $p < .001$, $\eta_p^2 = .21$, meaning that neither group behaved in the same way as native speakers, although their behaviour has changed in the direction of native speakers' behaviour.

3.3. Discussion

The two training groups significantly improved their use of the stop closure duration—the primary cue used by native English speakers to distinguish the word 'roads' from 'rose'—with the identification group improving slightly but significantly more than the discrimination group. The question that remains is why. The identification task presumably taps into phonological processing, whereas the AX discrimination task into phonetic/acoustic processing when the ISI is short (e.g., 250 ms). However, the ISI used in the current discrimination task was long (1500ms), which should similarly tap into phonological processing. Thus, other cognitive aspects must be at play, such as the role of the proper grapheme-phoneme association, which was required only by the identification training task.

4. CONCLUSION

The current study aimed at evaluating whether identification training yields superior results than AX discrimination training when training with a difficult consonant contrast. The contrast investigated was the English /z/-/dz/ as in 'rose' and 'roads', which is particularly difficult for Japanese speakers. We looked at improvement in the use of vowel duration and closure duration — two acoustic cues that are used by native English speakers to categorize the target contrast. While neither training paradigm yielded any significant change in the use of vowel duration after one hour of training, the identification training paradigm provided superior results in the use of closure duration. Therefore, the identification task may provide better results than an AX discrimination task for the training of some L2 contrasts.

5. ACKNOWLEDGMENTS

Thanks to our research assistants, participants and Jim Tanaka for their help with this study. This research was supported by a research grant by the Japan Society for the Promotion of Science, JSPS (16K02915), to Izabelle Grenon.

6. REFERENCES

[1] Akamatsu, T. 1997. *Japanese phonetics: Theory and practice*. Newcastle: Lincom Europa.

[2] Boersma, P., Weenink, D. 2017. Praat: doing phonetics by computer [Computer program]. Version 6.0.29, retrieved from http://www.praat.org/.

[3] Carlet, A., Cebrian, J. 2015. Identification vs. discrimination training: Learning effects for trained and untrained sounds. *Proceedings of 18th ICPhS.* Glasgow, Scotland, Aug. 10-14.

[4] Cebrian, J., Carlet, A., Gavaldà, N., Gorba, C. 2018. Effects of perceptual training on vowel perception and production and implications for L2 pronunciation teaching. Paper presented at *PSLLT 2018.* Ames, Iowa, Sept. 7-8.

[5] Flege, J.E. 1995. Two procedures for training a novel second language phonetic contrast. *Applied Psycholinguistics 16*, 425-442.

[6] Grenon, I. 2008. *The Acquisition of English Soun[dz] by Native Japanese Speakers: A Perceptual Study*. Germany: VDM Verlag.

[7] Grenon, I., Kubota, M., Sheppard, C. 2019. The creation of a new vowel category by adult learners after adaptive phonetic training. *J. Phonetics 72*, 17-34.

[8] Grenon, I., Sheppard, C., Archibald, J. 2018. Discrimination training for learning sound contrasts. *Proceedings of the 2nd International Symposium on Applied Phonetics (ISAPh)*, 51-56.

[9] Grenon, I., Sheppard, C., Archibald, J. in press. The effect of discrimination training on Japanese listeners' perception of the English coda consonants as in 'rose' and 'roads'. *Proceedings of Pronunciation in Second Language Learning and Teaching (PSLLT) 2018.* Ames, Iowa, Sept. 7-8.

[10] Iverson, P., Hazan, V., Bannister, K. 2005. Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r-l/ to Japanese adults. *JASA 118*(5), 3267-3278.

[11] Lawrence, M.A. 2016. ez: Easy analysis and visualization of factorial experiments. R package version 3.0-0.http://CRAN.R-project.org/package=ez.

[12] Logan, J.S., Lively, S.E., Pisoni, D.B. 1991. Training Japanese listeners to identify English /r/ and /l/: A first report. *JASA 89*(2), 874-886.

[13] Nozawa, T. 2015. Effects of attention and training method on the identification of American English vowels and coda nasals by native Japanese listeners. *Proceedings of 18th ICPhS.* Glasgow, Scotland, Aug. 10-14.

[14] R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org.

[15] Shinohara, Y., Iverson, P. 2018. High variability identification and discrimination training for Japanese speakers learning English /r/-/l/. *J. Phonetics 66*, 242-251.

[16] Wayland, R.P., Li, B. 2007. Effects of two training procedures in cross-language perception of tones. *J. Phonetics 36* (2), 250-267.

[17] Wee, D. T. J., Grenon, I., Sheppard, C., Archibald, J. 2019. Identification and discrimination training yield comparable results for contrasting vowels. *Proceedings of 19th ICPhS.* Melbourne, Australia, Aug. 5-9.

[18] Winn, M. 2014. Make duration continuum [Praat script]. Version August 2014, retrieved April 14, 2017 from http://www.mattwinn.com/praat.html.