Grenon, I., Sheppard, C., & Archibald, J. (2019). The effect of discrimination training on Japanese listeners' perception of the English coda consonants as in 'rose' and 'roads'. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Ames, IA, September 2018 (pp. 127-136). Ames, IA: Iowa State University.

PRESENTATION/POSTER

THE EFFECT OF DISCRIMINATION TRAINING ON JAPANESE LISTENERS' PERCEPTION OF THE ENGLISH CODA CONSONANTS AS IN 'ROSE' AND 'ROADS'

Izabelle Grenon, The University of Tokyo, Japan
Chris Sheppard, Waseda University, Japan
John Archibald, The University of Victoria, Canada

The efficiency of phonetic training via discrimination tasks has been questioned, as phoneme-grapheme correspondence is not transparent in discrimination training. Indeed, we showed in a previous study that some Japanese learners of English associated the vowels in 'ship' and 'sheep' with the wrong orthographic representation. The current study evaluated if mislabeling issues would occur when Japanese learners of English train with the contrast as in 'rose' and 'roads', and whether any improvement over time would be observed. Forty native English speakers from North America participated as the group of reference. Twenty Japanese speakers received two discrimination training sessions of about thirty minutes with the target contrast, with stimuli varying along two relevant dimensions (coda closure duration, and vowel duration) to go from 'rose' to 'roads'. The cue-weighting task administered before and after training revealed mislabeling issues that were present *before* training: The language learners associated a short vowel with the word 'rose' instead of the word 'roads'. However, the learners improved their use of both vowel duration and coda closure duration towards native-like performance post training. Hence, discrimination training was effective for altering Japanese speakers' use of the acoustic cues that contrast the English words 'rose' and 'roads'.

INTRODUCTION

Living in an English-speaking country for an extensive period of time is no guarantee that adult learners of English will learn to distinguish novel speech sounds (e.g., Oh et al., 2011). Conversely, few hours of phonetic training appear to lead to the creation of new speech categories for a considerable number of adult learners (Grenon, Kubota, & Sheppard, 2019). Most phonetic training programs feature an identification task consisting of presenting second language (L2) learners with one word (aurally), for instance *ship*, and asking them to identify which word they heard, 'ship' or 'sheep'. This type of training is designed to improve identification accuracy and has been shown to be effective in the training of a variety of L2 segmental (e.g., Logan, Lively, & Pisoni, 1991; Iverson, Hazan, & Bannister, 2005; Wang & Munro, 2004) and suprasegmental contrasts (e.g., Wang, Spence, Jongman, & Sereno, 1999).

However, it also presupposes that the learners are familiar with the L2 grapheme-phoneme correspondence. But what if they are not? The use of a discrimination task may serve as an alternative, for instance, for training beginner English learners of Russian or young Japanese learners of English with difficult L2 contrasts. In the AX discrimination task, the learner hears two

words, for instance *ship* and *sheep*, and has to decide if the two words are the 'same' or 'different'. Hence, this task does not require familiarity with the L2 orthographic system.

Comparisons between the use of an identification task and a discrimination task have yielded contradictory results. While some studies conclude that both types of tasks are equally effective (Flege, 1995; Wayland & Li, 2008), others suggest that the identification task is superior to the discrimination task (Carlet & Cebrian, 2015; Cebrian, Carlet, Gavaldà, & Gorba, 2018; Nozawa, 2015; Shinohara & Iverson, 2018). For instance, Spanish/Catalan speakers learning English as an L2 were trained to perceive English vowels using nonsense words produced by multiple talkers combined with either an identification task, AX discrimination task, or transcription task (control group) (Carlet & Cebrian, 2015). Training consisted of five 30-min sessions. Before and after training, learners' ability to identify English vowels was evaluated with the use of an identification task. The improvement in identification accuracy on both trained and new minimal-pairs was found to be higher for the identification training than the discrimination training group.

In a previous study we found that one hour of discrimination training with the English vowels in 'ship' and 'sheep' was sufficient for improving Japanese speakers' reliance on spectral information, as assessed through a cue-weighting task (Grenon, Sheppard, & Archibald, 2018). The cue-weighting task also revealed that twenty-five percent of the learners exhibited mislabeling issues, where they associated the vowel /i/ with the word 'ship' instead of 'sheep'. Considering that in most of the previous studies cited above, an identification task (using different minimal pairs) was used to assess improvement rather than a cue-weighting task (using one minimal pair manipulated along different acoustic dimensions), it is possible that mislabeling issues have contributed to lower the observed effect of the discrimination training paradigm. In turn, these results have led to the conclusion that discrimination training was not as efficient as identification training for improving the perception of new sound contrasts.

While we have shown that mislabeling issues may need to be addressed when training with a vowel contrast using a discrimination task (Grenon, Sheppard, & Archibald, 2018), the current study investigated the effect of the same task when training with a new consonantal contrast. The same twenty Japanese participants who trained with the vowel contrast reported in the study above were trained with the coda consonant contrast as in 'rose' and 'roads' using an AX discrimination task. This contrast is of particular interest in providing information about the learning of new speech sounds, as the words 'rose' and 'roads' are generally thought to be homophonous by Japanese speakers because the fricative /z/ and the affricate /dᶻ/ are phones in free variation in Japanese (that is, unlike the vowel contrast, Japanese speakers are unaware that these words are contrastive in English). Hence, the current study was expected to shed further light on the underlying learning mechanisms involved when training with a discrimination task.

Research questions and hypotheses

The specific research questions addressed by this paper are: (1) do Japanese speakers encounter mislabeling issues when training with a discrimination task with the 'rose' and 'roads' contrast, and (2) do they improve their sensitivity to the critical acoustic cues that serve to distinguish those sounds. The results of the L2 trainees are compared with that of native English speakers to evaluate if any change in perception is moving towards native performance.

Based on previous research (e.g., Grenon, 2011), it was expected that before training the Japanese speakers' sensitivity to vowel duration to distinguish short and long vowels in their L1 would be used to contrast the 'rose' and 'roads' stimuli in the cue-weighting task, while they would not rely significantly on the closure duration. After training, it was expected that the Japanese speakers would be relying less on vowel duration while relying more on the closure duration.

METHOD

Participants

The participants were the same twenty native Japanese speakers who took part in the study by Grenon, Sheppard and Archibald (2018) (the results of an additional participant were excluded from analysis for intensive exposure to English during early childhood). They were all students at the University of Tokyo in Japan aged between 18 and 27 years old (M = 20) who had never stayed in an English-speaking country for more than 8 weeks (M = 1.7 week). They received a monetary compensation for their participation.

Fifty-four native English speakers from North America also participated in this experiment. The data of fourteen participants were discarded either because the participant had been exposed regularly to another language during early childhood, or he or she reported a history of speech or hearing impairment. The resulting forty participants were all students at the University of Victoria in Canada aged between 17 and 28 years old (M = 21). They received course credits for their participation.

All participants signed a consent form prior to their participation. The Japanese participants completed the pre-test, two discrimination training sessions and post-test all on different days over a 2 to 3-week period (note that half of the participants were trained on the 'ship' and 'sheep' vowel contrast discussed previously before training on the 'rose' and 'roads' contrast). The time elapsed between the last 'rose' and 'roads' training session and post-test ranged between 1 and 11 days (M = 5.05, St. dev. = 3.14). The English participants completed the pre-test only (the pre-test and post-test were identical).

Stimuli

Six 'rose' and six 'roads' samples produced by a female university student from the United-States were recorded with a Sony microphone (ECM-MS957) at 44,100Hz in a sound attenuated booth at the University of Tokyo directly to computer using Praat (Boersma & Weenink, 2017). The intensity of the initial recording was scaled to 70dB. The vowel duration in the six recorded 'rose' samples varied from 280ms to 306ms (M = 291ms). The vowel duration in the six recorded 'roads' samples were consistently shorter, and varied from 220ms to 267ms (M = 243ms), while the closure duration varied from 64ms to 88ms (M = 78ms). Although the /d/ may disappear in fluent speech in some English dialects (Roca & Johnson, 1999), it was present in all the recorded 'roads' samples.

From the six recorded 'rose', a clear exemplar was chosen (i.e., without any glottalization or other features that may interfere with the manipulations). Then, 40ms of closure duration was extracted

from a 'roads' sample and inserted between the vowel and the word-final fricative in the word 'rose'. The closure duration was then modified from 0ms to 60ms in 7 steps of 10ms using a script for making a duration continuum (Winn, 2014). Using the same script, the vowel duration of each of the seven tokens was modified to vary from 210ms to 300ms in 4 equal steps of 30ms. The 28 resulting tokens are schematized in Figure 1, with spectrograms of token 1 and 28 presented in Figure 2.

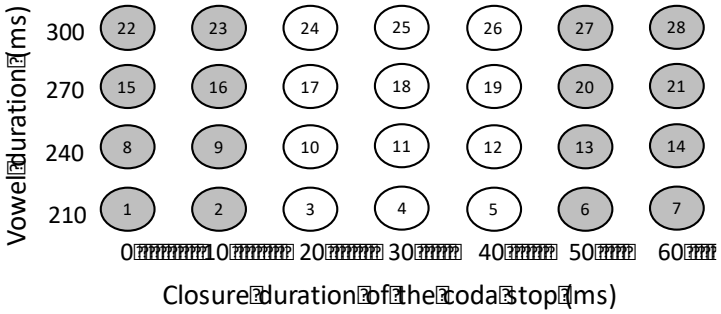

Figure 1. The 28 manipulated tokens used for the pre- and post-test were varied in terms of duration of the coda stop closure (x-axis) and vowel duration (y-axis). The 16 tokens used for training are presented in grey shading.

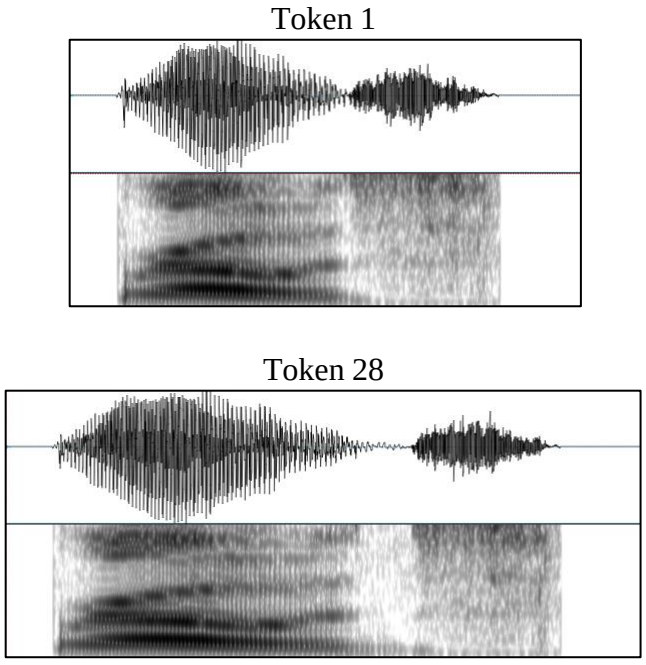

Figure 2. Spectrogram of the manipulated stimulus with no stop closure and a vowel duration of 210ms (token 1 in Figure 1). Spectrogram of the manipulated stimulus with 60ms of stop closure and a vowel duration of 300ms (token 28 in Figure 1).

The 28 resulting words were used in the identical pre- and post-test. A subset of 16 tokens were used for training. The tokens chosen for training were situated at the extreme ends of the closure duration continuum and are identified with grey shading in Figure 1. The 16 tokens were paired

for the AX discrimination training task so that 16 combinations featured words that differed in terms of closure duration, such as token 2 in Figure 1 followed by token 6 (these should be labeled as 'different' by the participants), and 16 pairs featured words that may have different vowel duration, but the closure duration was within the same category, such as token 1 and token 16 (these should be labeled as 'same' by the participants). None of the words was paired with itself. The resulting 32 pairs were also presented in reverse order, for a total of 64 training pairs, presented randomly 4 times, for a total of 512 words heard during one training session.

Procedure

The pre- and post-test done in a sound-attenuated room were meant to evaluate the weighting of each acoustic cue manipulated. Before reading the set of instructions for the task in their respective L1, participants were required to wear BOSE AE2 headphones and adjust the sound level to the most comfortable setting. For the task, the participants were presented with a red cross in the middle of the computer screen for 1000ms, then heard one of the 28 manipulated tokens, and had to decide if the word was 'rose' or 'roads' (the written words appeared on the screen) by pressing the appropriate key on the response pad. No feedback was provided during the tests, the words were never repeated, and the learners were requested to respond as quickly as possible. The 28 manipulated tokens were presented randomly four times during a test for a total of 112 test tokens (the first round of 28 words, considered a practice session, was discarded from the analyses). A test lasted 5 to 10min, with no break.

After the pre-test and before the post-test, the Japanese listeners completed 2 training sessions of about 30min on two different days in a sound-attenuated room, for a total of 1h of discrimination training. For the training, the learner would hear two words with an inter-stimulus-interval of 1500ms (e.g., *rose – roads*), and had to decide if the words were the 'same' or 'different' (only the words 'same' and 'different' were written on the screen) by pressing the appropriate key on the computer keyboard. Each trial was followed by feedback (a written message) indicating whether the choice was correct. The next trial was presented after an inter-trial-interval of 2000ms added after a participant's response.

RESULTS AND DISCUSSION

Training

To assess any improvement during training, the raw training scores (percentages of correct responses) were computed for each participant. The average score on the first training session was near chance level, that is 52.40% (St. dev. = 11.07), meaning the contrast of interest was indeed very difficult for them. The Japanese participants slightly improved their performance on the second training session, with an average score of 56.02% (St. dev. = 16.22). Although small, this improvement was significant (t (20) = 2.483, $p < .05$, d = 0.53).

Mislabeling

To evaluate mislabeling issues, the proportion of tokens identified as 'roads' by native Japanese speakers were compared with that of native English speakers on each dimension of interest: vowel duration and closure duration of the stop coda consonant.

Mislabeling issues were found to occur *before* training on the vowel duration dimension as shown in Figure 3 below. That is, while English speakers associated a shorter vowel with the word 'roads' and a relatively longer vowel with the word 'rose', the Japanese trainees did the opposite, associating a shorter vowel with the word 'rose' instead. Given that the Japanese language distinguishes between short and long vowels, it is not surprising that Japanese participants are sensitive to vowel duration and may use it for distinguishing foreign words.

In the current case, the word 'rose' is generally the one with a longer vowel than the word 'roads' in English, and the results show that the English speakers in the study are sensitive to the vowel duration when contrasting the two words. For Japanese speakers, it is possible that the use of two vowel letters in the word 'r**oa**ds' has led them to mistakenly presume that this word contains a long vowel. Hence, an influence of orthography may be possible in this case. That being said, there was a change in the Japanese speakers' use of vowel duration between pre-test and post-test, where they started to rely less on vowel duration post training, as reported in the next subsection.

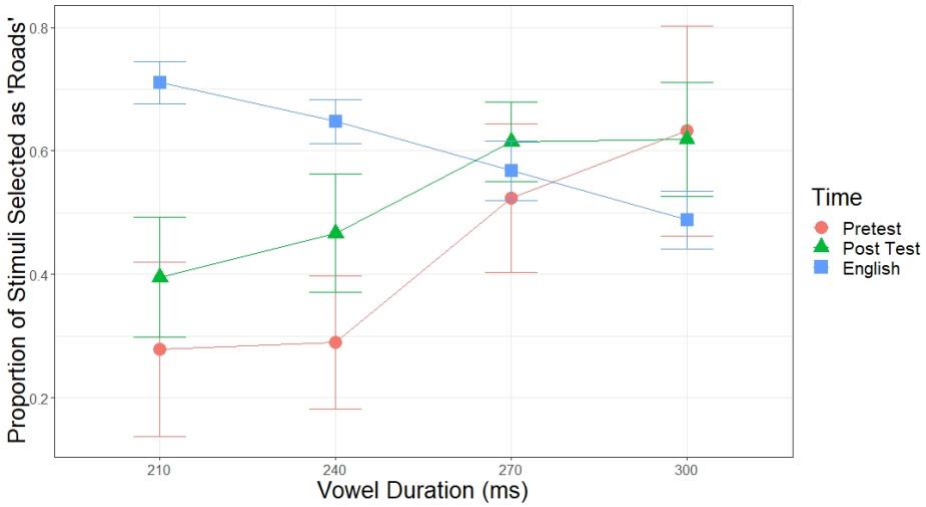

Figure 3. Results of the 28 test tokens across all vowel duration values for the Japanese listeners before (red circles) and after (green triangles) training compared with the results of the native English listeners (blue squares).

Cue-weighting

The results of a multi-level linear model analysis confirmed that the change in the use of vowel duration between pre-test and post-test was significant (X^2 (2) = 4.54, p < .05, d = 0.535). A follow up analysis comparing the Japanese native speakers' post-test performance with that of the native speakers found that the behavior of the Japanese speakers did not attain the same level as the native speakers as represented by a significant effect of Group (X^2 (2) = 6.65, p = .001, d = 0.714) as well

as a significant Group x Vowel duration interaction (X^2 (4) = 92.11, p < .0001, effect sizes for the planned comparisons: d = 2.01-2.80).

After training, the L2 listeners also improved their sensitivity to the most critical acoustic cue that serves to distinguish the coda contrast. As shown in Figure 4, native English speakers rely heavily on the closure duration to categorize the word 'rose' and 'roads', whereas native Japanese listeners mostly ignored this cue before training. After training, however, their use of this cue has increased, and is starting to resemble the native speakers' performance.

The results of a multi-level linear model analysis confirmed that the identification behavior along the closure duration dimension of the L2 listeners changed significantly from pre-test to post-test (X^2 (2) = 4.94, p < .05, d = 0.521), although their performance was still different from native listeners with a significant effect of Group (X^2 (2) = 5.96, p < .05, d = 0.714) and Group x Closure duration interaction (X^2 (4) = 159.31, p < .0001, effect sizes of the planned comparisons: d = 1.35-2.57).

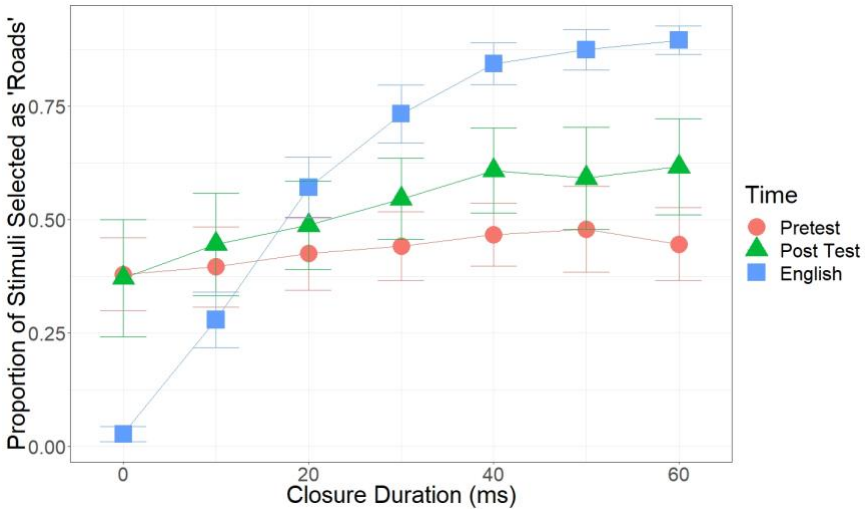

Figure 4. Results of the 28 test tokens across all closure duration values for the Japanese listeners before (red circles) and after (green triangles) training compared with the results of the native English listeners (blue squares).

Summary of results and general discussion

In sum, L2 listeners could change their cue-weighting for categorization of the 'rose' and 'roads' contrast towards native speakers' performance. After training, the Japanese listeners improved their use of both vowel duration and closure duration of the coda consonant for categorizing the coda contrast. In the case of the novel consonant contrast, mislabeling issues existed prior to the start of training, where listeners mistakenly associated a short vowel with the word 'rose' rather than with the word 'roads'. A possible effect of orthography was mentioned, where the use of one letter in the word 'rose' may have misled the L2 learners to assume that this word features a short vowel, whereas the word 'roads' features a long vowel. The possible effect of orthography could be ruled out by testing another minimal-pair featuring a single letter for the vowel sound in both words, such as 'cars' versus 'cards'.

A question that remains, however, is whether a change in the reliance of relevant acoustic cues would improve more with identification training. This possibility was tested using the same manipulated tokens presented the same number of times across two training conditions (identification vs. discrimination), with both the 'rose' and 'roads' contrast as well as with the 'ship' and 'sheep' contrast. The two tasks yielded comparable change in cue-weighting for the 'ship'-'sheep' contrast (Wee et al., in press) when disregarding mislabeling issues. Hence, the results of previous training studies with vowel sounds indicating that identification training is superior to discrimination training (Carlet & Cebrian, 2015; Cebrian et al., 2018; Nozawa, 2015) may have been affected by mislabeling issues.

However, for the 'rose'-'roads' contrast, identification training provided marginally superior results than discrimination training for the reliance on closure duration (Law et al., in press). Hence, while the two types of training were equally effective for the vowel contrast ('ship'-'sheep'), they were not as equally effective in the case of the consonantal contrast ('rose'-'roads'). A tentative explanation for this discrepancy, is that prior sensitivity to the critical acoustic cue may be required for discrimination training to be as effective as identification training. Also, in our studies, only one minimal-pair produced by one speaker was used. It remains to investigate if the use of high-variability combined with more training time could encourage further changes in cue-weighting.

CONCLUSION

The current study investigated whether a training program featuring a discrimination task could help native Japanese learners of English improve their sensitivity to the acoustic cues used by native English speakers to distinguish the coda consonants as in 'rose' and 'roads'. The results revealed that *before* training the Japanese trainees generally ignored the closure duration of the stop consonant, a cue that English speakers rely heavily on to contrast the word 'roads' from 'rose'. The L2 trainees relied instead on vowel duration to contrast the two words, but they mistakenly associated a short vowel with the word 'rose' instead of the word 'roads'. After training, however, their use of vowel duration developed towards native performance. Similarly, their use of the closure duration of the coda stop consonant improved after one hour of discrimination training, also developing towards native-like performance. Hence, discrimination training appears successful for improving the use of acoustic cues related to the perception of a consonant contrast when using a discrimination task.

ACKNOWLEDGMENTS

We are grateful to our many research assistants and participants without whom this research would not have been possible, and thankful also to Jim Tanaka. This research has received a grant-in-aid to scientific research by the Japan Society for the Promotion of Science, JSPS (16K02915) granted to Izabelle Grenon.

ABOUT THE AUTHORS

Izabelle Grenon is an associate professor in the Department of Language and Information Sciences at the University of Tokyo in Japan. Her research interests are in experimental phonetics

related to first or second language acquisition. Her current research focuses on testing different phonetic training paradigms for improving the perception of non-native speech sounds.

E-mail: grenon@boz.c.u-tokyo.ac.jp

Chris Sheppard is an associate professor at the Center for English Language Education in Science and Engineering at Waseda University in Tokyo, Japan. His current research projects focus on fostering autonomy in English for Specific Purposes, the development of critical thinking through second language education, the acquisition of rhythm by learners of Japanese as a second language, and the role of feedback in second language acquisition.

E-mail: chris@waseda.jp

John Archibald is a professor of linguistics at the University of Victoria in Canada. He specializes in second language phonology. He has two current projects in the domain of intelligibility, one focusing on modified input, the other on psycholinguistic models of word recognition.

E-mail: johnarch@uvic.ca

REFERENCES

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

Carlet, A., & Cebrian, J. (2015). Identification vs. discrimination training: Learning effects for trained and untrained sounds. *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS)*, Glasgow, Scotland, Aug. 10-14.

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 *Pronunciation in Second Language Learning and Teaching (PSLLT)*, Ames, Iowa, Sept. 6-8.

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

Grenon, I. (2011). Dissociable levels of speech processing in second language perception. *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS)*, Hong Kong, Aug. 17-21.

Grenon, I., Kubota, M., & Sheppard, C. (2019). The creation of a new vowel category by adult learners after adaptive phonetic training. *Journal of Phonetics, 72*, 17-34.

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.

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. *Journal of the Acoustical Society of America, 118* (5), 3267-3278.

Law, I. L. G., Grenon, I., Sheppard, C., & Archibald, J. (in press). Which is better: Identification or discrimination training for the acquisition of an English coda contrast. *Proceedings of the 19th International Congress of Phonetic Sciences (ICPhS).* Melbourne, Australia.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America, 89* (2), 874-886.

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 the 18th International Congress of Phonetic Science (ICPhS),* Glasgow, Scotland, Aug. 10-14.

Oh, G. E., Guion-Anderson, S., Aoyama, K., Flege, J. E., Akahane-Yamada, R., & Yamada, T. (2011). A one-year longitudinal study of English and Japanese vowel production by Japanese adults and children in an English-speaking setting. *Journal of Phonetics, 39*, 156-167.

Roca, I., & Johnson, W. (1999). *A Course in Phonology*. Malden, MA: Blackwell Publishers.

Shinohara, Y., & Iverson, P. (2018). High variability identification and discrimination training for Japanese speakers learning English /r/-/l/. *Journal of Phonetics, 66*, 242-251.

Wang, X., & Munro, M. J. (2004). Computer-based training for learning English vowel contrasts. *System, 32*, 539-552.

Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *Journal of the Acoustical Society of America, 106* (6), 3649-3658.

Wayland, R. P., & Li, B. (2008). Effects of two training procedures in cross-language perception of tones. *Journal of Phonetics, 36* (2), 250-267.

Wee, D. T. J., Grenon, I., Sheppard, C., & Archibald, J. (in press). Identification and discrimination training yield comparable results for contrasting vowels. *Proceedings of the 19th International Congress of Phonetic Sciences (ICPhS).* Melbourne, Australia.

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