IDENTIFICATION AND DISCRIMINATION TRAINING YIELD COMPARABLE RESULTS FOR CONTRASTING VOWELS

Daniel T. J. Wee[1], Izabelle Grenon[1], Chris Sheppard[2], John Archibald[3]

[1]The University of Tokyo, [2]Waseda University, [3]The University of Victoria
wee.daniel19852000@gmail.com, grenon@boz.c.u-tokyo.ac.jp, chris@waseda.jp, johnarch@uvic.ca

ABSTRACT

This study compares the use of an identification task with the use of a discrimination task for training Japanese speakers on the English high front vowels. Seventeen Japanese speakers completed two sessions of identification training with feedback, using 'ship' and 'sheep' tokens that were manipulated to vary along the vowel duration and formant dimensions. Their results were compared with those of twenty Japanese speakers trained with an AX discrimination task. A two-alternative forced-choice identification task (without feedback) evaluated the participants' use of temporal and spectral information before and after training. The results indicate that both training paradigms led to comparable improvement in the use of temporal and spectral information. Hence, previous research suggesting that the identification task provides superior results to the discrimination task may be due to mislabeling issues, rather than learners' lack of improvement in the use of spectral information.

Keywords: Phonetic training, cue-weighing, vowel perception, English vowels, Japanese.

1. INTRODUCTION

Phonetic training paradigms typically use an identification task [e.g., 7, 12, 13], although a few training paradigms have tested the use of a discrimination task [e.g., 8, 9, 17]. An identification task consists of presenting the second language (L2) learners with an audio recording of a word such as *ship* and asking them to identify whether they heard the word 'ship' or 'sheep'. An AX discrimination task, on the other hand, consists of presenting the L2 learners with an audio recording of two words (e.g., *ship – sheep*) and asking them to identify whether the two words were the same or different. Hence, a possible advantage of the use of a discrimination task over the use of an identification task is that it could be used with populations who are not literate in the L2, such as elementary school children learning their heritage language. An important concern, therefore, is whether the discrimination task yields similar results to the identification task.

A few studies have compared the efficiency of discrimination training versus identification training. While some studies evaluating the acquisition of Thai tones by English speakers [18] and the acquisition of English coda consonants by Mandarin Chinese speakers [5] have concluded that both tasks are equally effective, studies with vowel contrasts have generally found that the identification task yields superior results [3, 4, 15]. A recent study, however, found that while the AX discrimination task was effective for increasing learners' sensitivity to contrasts along the spectral dimension, there were some instances of mislabeling issues where 25% of the Japanese listeners associated the vowel /i/ with the word 'ship' instead of 'sheep' post-training [9]. Crucially, mislabeling issues do not mean the learners failed to develop separate vowel categories along the spectral dimension. Accordingly, the current study evaluated whether identification training yields superior results to discrimination training with the high front vowel contrast, when mislabeling issues are disregarded.

While English speakers rely mainly on changes in the first (F1) and second (F2) formant frequencies to distinguish the high front vowels as in 'ship' and 'sheep' [2, 6, 10], Japanese speakers tend to rely instead on vowel duration [6, 14]. Hence, the current study looked at the use of temporal (vowel duration) and spectral (formant) information for identification of the English high front vowels by Japanese speakers before and after either identification or discrimination training.

2. METHOD

2.1. Participants

Thirty-seven right-handed native speakers of Japanese with no reported history of speech or hearing impairment and recruited at The University of Tokyo took part in the experiment. Twenty were assigned to the discrimination condition and seventeen to the identification condition. Those assigned to the discrimination training were between 18 and 27 years old (M=20) and had never spent more than 8 weeks (M = 1.7 week) in an English-speaking country. Their results were first reported in [9]. Those

assigned to the identification training were aged between 18 and 23 (M = 21) and had never spent more than 3 weeks in an English-speaking country (M = 1.25 week). The Japanese participants received a monetary compensation.

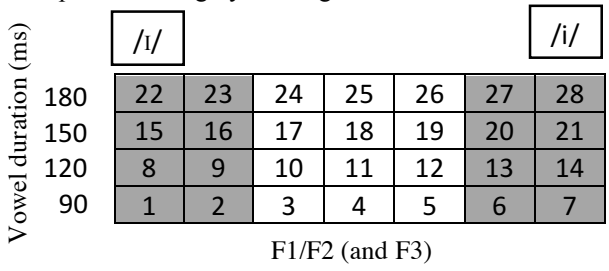
The results of a group of forty monolingual North American English speakers recruited at The University of Victoria in Canada aged 17-28 (M = 21), first reported in [9], are used as a baseline for comparison. They received course credit for their participation.

2.2. Stimuli

Twenty-eight tokens were created by manipulating the vowel of a *ship* sample produced by a female American English speaker. The recording was done in a sound attenuated booth with a Shure SM10A microphone, and saved directly to computer using Praat [1]. The first (F1), second (F2) and third (F3) formants were manipulated in 7 equal steps on the Bark scale [21] using a script [20], to go from 'ship' to 'sheep'. The formant values of the 7 resulting vowels schematized in Figure 1 below are: token 1 (679/2087/2999), token 2 (631/2203/3041), token 3 (585/2326/3084), token 4 (540/2457/3128), token 5 (497/2596/3172), token 6 (456/2744/3218) and token 7 (415/2902/3264).

After manipulation of the vowel quality, the duration of the 7 vowels was manipulated from short to long (90ms, 120ms, 150ms and 180ms) in 4 equal steps of 30ms using a script [19], to yield 28 tokens. The duration of the onset consonant was fixed to 210ms, the closure of the stop consonant to 136ms and the release burst to 100ms, across all 28 tokens. The pitch pattern was altered to downward-rising to provide a more natural pitch contour.

Figure 1: Distribution of the 28 stimuli used for the pre- and post-tests. The 16 stimuli used for training are presented in grey shading.



Only 16 of the 28 manipulated tokens, presented in grey shading in Figure 1, were used for training. These 16 tokens were selected from both ends of the spectral continuum, and testing with English speakers confirmed that they were the tokens most often categorized as 'sheep' or 'ship' (that is, they represented clear exemplars of each vowel category).

2.3. Procedure

2.3.1. Pre-test and post-test

The pre-test and post-test were identical. The 28 tokens used as stimuli were presented randomly 4 times with the first round discarded from the analyses as a practice session. Each test was conducted in the form of a two-alternative forced-choice identification task. The tests were performed in a sound-attenuated room, with stimuli presented via high quality BOSE headphones. No feedback was provided during a test. The English participants completed the test only once and they did not do any training.

2.3.2. Identification training

After the pre-test, the seventeen Japanese listeners assigned to the identification condition went through one hour of training, which was divided into two 30-minute sessions held on separate days. For the identification training, participants underwent the same two-alternative, forced-choice identification task akin to the pre-test and post-test but with feedback (a written message indicating whether the choice was correct). The 16 training tokens were presented randomly 32 times each, for a total of 512 words heard during a session of identification training.

2.3.3. Discrimination training

The twenty Japanese listeners assigned to the discrimination condition went similarly through one hour of training, divided into two 30-minute sessions held on different days (their results were first reported in [9]). In the discrimination training, the 16 training tokens were presented through a 'same-different' AX discrimination task. The 16 training words were paired so that 16 combinations featured words that differed in terms of spectral quality, 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 spectral quality was the same, such as token 1 and token 16 (these should be labeled as 'same' by the participants). None of the words was paired with itself. Each word was presented 32 times, for a total of 512 words heard during a session of discrimination training.

3. RESULTS AND DISCUSSION

First, we confirmed that the participants improved their performance during training with both the discrimination task and the identification task. As

reported previously [9], the average scores on the discrimination task were 88.3% (std. dev.: 10.6) on the first training day, and increased to 93.6% on the second training day (std. dev.: 8.10), a significant improvement of 5.3% ($t(19) = 4.09$, $p < 0.001$). Participants assigned to the identification training paradigm also improved their performance from the first training day (93.2%, std. dev.: 6.57) to the second training day (97.2%, std. dev.: 2.70), a significant increase of 4% ($t(16) = 2.96$, $p < 0.01$).

The goal of the current study was to assess whether identification training leads to greater changes in cue-weighting than discrimination training for categorization of the vowel contrast in 'ship' and 'sheep' by native Japanese speakers. The following sections look, in turn, at changes in the use of vowel duration and at changes in the use of spectral information by participants in the two training conditions. Their performance post-training is also compared with that of native English speakers for reference.

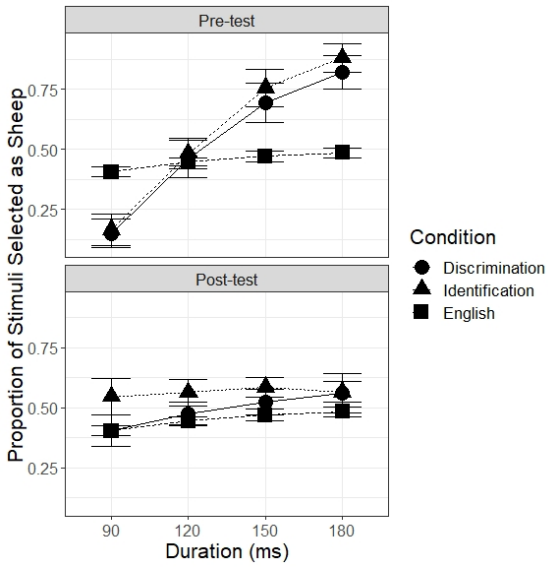
3.1. The use of temporal cues

Figure 2 shows the use of vowel duration by the discrimination group and identification group before (pre-test) and after training (post-test), compared with English speakers' performance. Both training groups used vowel duration to contrast the vowels on the pre-test, whereas English speakers did not. At post-test, both training groups have learned to ignore the vowel duration contrast.

The vowel duration data were analysed using a mixed-design ANOVA in R [16] with within-subject factors of 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 ($W = 0.32$, $p < .001$), therefore the degrees of freedom were corrected ($\varepsilon = 0.58$). Participants in the two training conditions behaved differently from pre-test to post-test, which was shown by the significant Time X Duration interaction; $F(3, 105) = 112.34$, $p < .001$, $\eta_p^2 = 0.43$. However, the Time X Condition X Duration interaction was not significant; $F(3,105) = 2.53$, $p = .095$, $\eta_p^2 = .016$, and so there was no differential effect for training: Participants in the discrimination group used vowel duration in a way comparable to the identification group both in pre-test and post-test.

Furthermore, the post-test performance on vowel duration of both training groups was comparable to that of the native speakers. A mixed-design ANOVA was performed with Duration as the within-subject and Condition (discrimination, identification, and English) as the between-subject factor. The data was not spherical ($W = 0.40$, $p < .001$), and so Greenhouse-Geisser estimate of sphericity ($\varepsilon = 0.51$) was used. The Condition X Duration interaction was not significant; $F(3, 105) = 0.52$, $p < .582$, $\eta_p^2 = .007$. Thus, the behaviour of both training groups was the same as that of the English native speakers after training.

Figure 2: Comparative results between discrimination and identification training involving the 28 test tokens across all vowel duration values for the pre-test and post-test. English data are provided on both graphs for reference.



3.2. The use of spectral cues

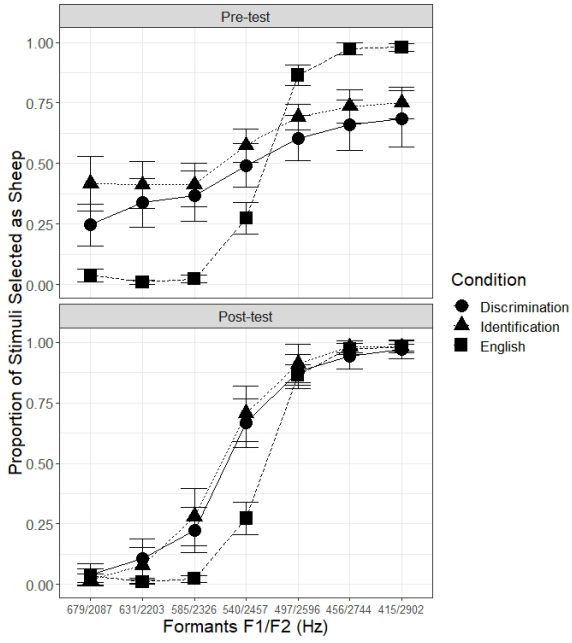
In the discrimination training study, some cases of mislabeling issues were found on the post-test, where 5 out of 20 Japanese speakers associated the vowel /i/ with the word 'ship' instead of the word 'sheep' [9]. No mislabeling issues were found with the identification training. As the focus in the current study is to compare the ability of Japanese speakers use of spectral information after discrimination versus identification training, the association between the vowel categories and their respective orthographic representations was disregarded. That is, the post-test data of the mislabeling participants in the discrimination condition were reversed by recoding as 'sheep' all instances labeled by the participant as 'ship', and vice versa. Hence, the data used for analyses for the discrimination task includes all 20 participants (whereas only 15 were used for the analyses in [9]).

As shown in Figure 3, the use of the spectral cues by the discrimination and identification training group were comparable on the pre-test, with less reliance on spectral information than native English speakers in order to classify the vowels. On the post-

test, however, both training groups had increased their use of spectral information towards that of native English speakers' performance. And the improvement observed with the discrimination task was equivalent to the improvement achieved with the identification task, though neither achieved native-like performance.

The formant data were analysed using, once again, a mixed-design ANOVA with within-subject factors of Formant 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 ($W = 0.05$, $p < .001$), therefore degrees of freedom were corrected using Greenhouse-Geisser estimate of sphericity ($\varepsilon = 0.45$). Both of the training conditions changed their behaviour over time, which was shown by the significant Time X Formant interaction; $F(6, 210) = 55.51$, $p < .001$, $\eta_p^2 = .34$. However, the Time X Condition X Formant interaction was not significant; $F(6, 210) = 0.85$, $p = .53$, $\eta_p^2 = .008$, indicating there was no differential effect for training: both training groups exhibited comparable results.

Figure 3: Comparative results between discrimination and identification training involving the 28 test tokens along the spectral continuum for the pre-test and post-test. English data are provided on both graphs for reference.



The post-test formant data of both training groups were then compared with that of the English native speakers with Formant 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 ($W = 0.054$, $p < .001$), therefore Greenhouse-Geisser estimate of sphericity ($\varepsilon = 0.58$) was used. The Condition X Formant interaction was significant; $F(12, 444) = 16.83$, $p < .001$, $\eta_p^2 = .26$. Both groups did not attain the same behaviour as native speakers.

Understandably, the effect of training versus the effect of exposure to the stimuli need to be disentangled in the future by running a control group. Keeping this in mind, the results so far suggest that discrimination training and identification training may yield comparable results when looking at changes in the use of vowel duration and spectral information. Hence, previous studies demonstrating that identification training yields better results than discrimination training for the learning of vowel categories (e.g., [3], [4]) may have encountered mislabeling issues, which may have affected the results of the group assigned to the discrimination condition. As discrimination training does not provide information about phoneme-grapheme associations, mislabeling issues may be expected with this kind of training if no specific instruction in this regard has been provided. The long-term benefits of discrimination training as well as generalization to new tokens and talkers still need to be investigated by taking into consideration mislabeling issues.

4. CONCLUSION

The current study compared the use of an identification task with the use of an AX discrimination task for training with the English vowels in the words 'ship' and 'sheep' by native Japanese speakers. A cue-weighting task evaluated the use of temporal and spectral information by the two groups before and after one hour of training. Their performance was also compared with that of native English speakers. It was found that the learners assigned to the discrimination group and those assigned to the identification group equally reduced their reliance on vowel duration while equally increasing their sensitivity to spectral information towards native speakers' performance. Participants in both training conditions performed like native speakers on the use of vowel duration after one hour of training, but still differed from native speakers on their use of spectral information, though their use of this cue became more categorical.

5. ACKNOWLEDGMENTS

A sincere thank you to our research assistants, participants and Jim Tanaka for their help with the current 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] Boersma, P., & Weenink, D. 2016. Praat: Doing phonetics by computer (ver. 6.0.18) [computer program]. Retrieved from <http://www.praat.org/>.

[2] Bohn, O.-S. 1995. Cross-language speech perception in adults: First language transfer doesn't tell it all. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research* (pp. 279-304). Timonium, MD: York Press.

[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 Pronunciation in Second Language Learning and Teaching 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. 2012. The bi-level input processing model of first and second language perception (Doctoral dissertation, U. of Victoria, Canada). *Dissertation Abstracts International, 72* (8), 285.

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

[9] 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.

[10] Kondaurova, M. V., Francis, A. L. 2008. The relationship between native allophonic experience with vowel duration and perception of the English tense/lax vowel contrast by Spanish and Russian listeners. *JASA, 124* (6), 3959-3971.

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

[12] Lively, S. E., Logan, J. S., & Pisoni, D. B. 1993. Training Japanese listeners to identify English /r/ and /l/ II: The role of phonetic environment and talker variability in learning new perceptual categories. *JASA, 94* (3), 1242-1255.

[13] 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.

[14] Morrison, G. S. 2002. Effects of L1 duration experience on Japanese and Spanish listeners' perception of English high front vowels. Unpublished master's thesis, Simon Fraser University, Canada.

[15] Nozawa, T 2015. Effects of training methods and attention on the identification and discrimination of American English Coda Nasals by native Japanese listeners. *JASA,, 138* (3), 1947-1947.

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

[17] Strange, W., Dittmann, S. 1984. Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception & Psychophysics, 36* (2), 131-145.

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

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

[20] Winn, M. 2016. Make formant continuum [Praat script]. Version July 2016, retrieved May 29, 2017 from http://www.mattwinn.com/praat.html.

[21] Zwicker, E. 1961. Subdivision of the audible frequency range into critical bands (Frequenzgruppen). *JASA, 33*(2), 248.