Discrimination training for learning sound contrasts

Izabelle Grenon[1], Chris Sheppard[2], John Archibald[3]

[1]The University of Tokyo, Japan, [2]Waseda University, Japan, [3]The University of Victoria, Canada

[1]*grenon@boz.c.u-tokyo.ac.jp*, [2]*chris@waseda.jp*, [3]*johnarch@uvic.ca*

Abstract

Recent studies have suggested that for training on second language (L2) vowels, the use of an identification task results in greater improvement, and generalization to new words than does the use of a discrimination task. The current study investigated why this may be the case.

Twenty native Japanese speakers received two thirty-minute sessions of discrimination training with the English high front vowels, which were modified to vary along two dimensions to go from 'ship' to 'sheep': Vowel quality (i.e., spectral cues), and vowel duration (i.e., temporal cues). Using a cue-weighting task, we found that while the L2 learners significantly improved their use of vowel quality to classify the English vowel contrast after training, some of them (25%) were unable to associate the vowels with their proper orthographic representations. We conclude that unlike what was suggested by previous studies, the discrimination task appears successful in helping L2 learners create new vowel categories along the spectral dimension. However, since the discrimination task does not provide information about phoneme-grapheme associations, additional instructions about how each vowel is represented by the orthography should be provided to the learners at some point.

Keywords: Discrimination training, Cue-weighting, English vowels, Japanese.

1. Introduction

Phonetic training can be a powerful tool for enhancing the perception of non-native (L2) contrasts. Different types of tasks can be used for phonetic training, but the most well-documented are the discrimination and identification tasks. In an AX discrimination task, language learners hear two words (e.g., *ship – sheep*) and are required to decide if the two words were the same or different. An identification task, on the other hand, consists of presenting the language learners with an audio recording of only one word, such as *ship*, and ask them to identify which word they heard, 'ship' or 'sheep'.

In one of the earliest attempts at phonetic training, the use of a discrimination training program with synthetic 'rock'-'lock' stimuli designed for Japanese speakers by Strange and Dittman [1] resulted in some improvement on trained tokens, but this improvement failed to generalize to natural words. Following this, Logan and colleagues [2, 3] designed a training program using an identification task instead with natural stimuli contrasting the English 'r' and 'l' in different contexts and produced by various speakers (i.e., featuring high variability) to train Japanese speakers. They found that that this program led to significant improvement on trained tokens as well as generalization to new tokens.

The large body of research that has followed this work has accordingly focused on the use of high variability combined with an identification task to test the validity of a variety of phonetic training programs targeting different L2 segmental and suprasegmental contrasts and different populations [e.g., 4, 5, 6, 7]. It is yet unclear, however, whether the identification task is actually superior to the discrimination task, and if so, why it is the case. One of the possible reasons for the discrepancy between identification and discrimination results may come from the method used for measurement. While the training paradigm may consist of a discrimination task, an identification task is typically used for assessing improvement. Hence, it is possible that the L2 learners properly learned to distinguish the sounds based on the relevant acoustic cue, but have not learned yet which phoneme corresponds to which grapheme. The goal of the current study was therefore to assess the validity of the discrimination task by looking at whether this type of task may: (1) fail to resolve mislabeling issues (that is, when the learners associate the phonemes with the wrong graphemes), or (2) fail to enable learners to create new categories along the critical acoustic dimension.

10.21437/ISAPh.2018-9

2. Identification vs. discrimination training

Some previous studies report comparable improvement when using a discrimination task compared to using an identification task when training with English final stops [8] and Thai tones [9]. For instance, Wayland and Li [9] trained English and Chinese listeners with Thai tones. Participants in each language group were assigned to either a two-alternative forced-choice identification procedure or to a same/different (AX) discrimination procedure. Both groups of listeners improved their perception of Thai tones significantly after training. Most importantly, the amount of improvement was comparable across training procedures, suggesting that discrimination training is as efficient as identification training for the learning of tone contrasts.

More recent studies, however, have reported that identification training leads to greater improvement than discrimination training, particularly when targeting English vowels [10, 11, 12] or the English /r/-/l/ contrast [13]. For instance, Carlet and Cebrian [10] trained Spanish/Catalan speakers with five difficult English vowels using either a five-alternative forced-choice identification task or an AX discrimination task. Participants received five training sessions of about thirty minutes each with the target vowels in one of the experimental conditions (identification or discrimination training) using nonsense words. Both tasks led to improvement on the words used for training, although the group assigned to the identification training condition improved significantly more than the group assigned to the discrimination training condition. Moreover, when tested for generalization to real words (not used during training), only the group assigned to the identification task improved significantly.

One may wonder why discrimination training seems to result in poorer improvement than identification training when training with vowel sounds. One possibility is that while listeners may learn to hear a difference between the target vowels, they may fail to associate the vowel sounds with the proper orthographic representations (e.g., associating the vowel /i/ with the vowel in 'ship' instead of the vowel in 'sheep'). Hence, the first objective of the current study was to investigate possible mislabeling issues among Japanese participants before and after discrimination training with the 'ship' and 'sheep' contrast using a cue-weighting task.

The second objective was to evaluate, using the same cue-weighing task, which cue the Japanese participants most relied on before and after training, and compare their performance with that of English speakers. Japanese speakers have been shown to rely on vowel duration to categorize the vowels /i/ and /ɪ/ as in 'sheep' and 'ship' [14, 15], whereas native English speakers rely primarily on vowel quality (i.e., on changes in the first and second formant frequencies) [14, 16, 17]. Possibly as a result of this difference in cue-weighting, Japanese speakers have difficulty properly categorizing the two vowels [18]. In a previous study, we found that when training with an identification task, about half of the Japanese listeners were able to redeploy their attention away from vowel duration, and towards the use of vowel quality (spectral information), and as a result, improve their ability to accurately categorize the high front vowels [4]. In the present study, we tested whether Japanese listeners can change their cue-weighing for the perception of the vowels as in 'ship' and 'sheep' when training instead with a discrimination task.

3. Method

3.1. Participants

Twenty native Japanese speakers (all students at the University of Tokyo, Japan) participated in the experiment (one other participant was excluded from the analyses because of spending a year in an English-speaking country during early childhood.) They received a monetary compensation for their participation.

A group of forty monolingual native English speakers from North America (all students at the University of Victoria, Canada) participated as the reference group (fourteen extra participants were discarded either because they had been exposed regularly to another language during early childhood, or they reported a history of speech or hearing impairment.) They received course credits for their participation.

The Japanese participants completed the pre-test (day 1), two sessions of discrimination training with the 'ship'-'sheep' contrast (day 2 and 3), two sessions of discrimination training with another contrasted (not reported here) (day 4 and 5) and the post-test (day 6) (half of the participants were trained on the other contrast *before* training on the 'ship'-'sheep' contrast, as this research is part of a larger study). The English participants completed the pre-test only. Note that the pre-test and post-test were identical.

3.2. Stimuli

First, 'ship' and 'sheep' samples produced by a female university student from the United-States were recorded in a sound attenuated booth at the University of Tokyo, and measurements of her vowel formants were used as references for the manipulations in Praat [19]. Second, a 'ship' sample

was manipulated resulting in 28 ‘ship’ and ‘sheep’ tokens varying along two dimensions as illustrated in Figure 1 below: vowel duration was manipulated from short to long (90ms, 120ms, 150ms, and 180ms), and the first (F1), second (F2) and third (F3) formant frequencies were manipulated in 7 equal steps on the Bark scale (from /ɪ/ to /i/) using Praat scripts [20, 21]. The critical formant frequencies for identification of the vowels are the F1 and F2, but the F3 was also manipulated because it resulted in more natural sounding tokens. The pitch pattern was also altered from relatively flat to high-low-rising for the same reason. The values of the F1, F2 and F3 (reported in Hz) for the 7 vowel qualities are as follow (the values reported were taken at mid-vowel in the filter used to create each vowel quality and may vary at other locations): 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). The F4 (4262 Hz), F5 (4378 Hz), the duration of the initial ‘sh’ sound (210 ms) and final ‘p’ (closure duration: 136 ms; release burst duration: 100 ms) were kept constant across the 28 test tokens.

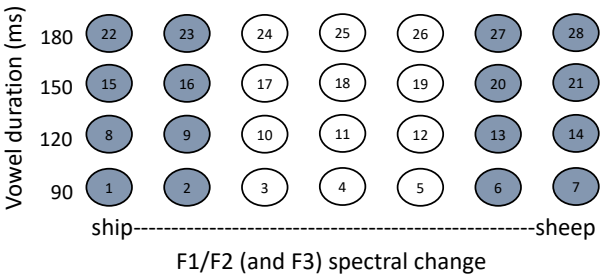

Figure 1: The 28 manipulated tokens used for the pre- and post-tests were varied in terms of F1, F2 and F3 (x-axis) and vowel duration (y-axis). The 16 tokens used for training are presented in grey shading.

The 28 resulting words were used in the pre- and post-test (the tests were identical). A subset of 16 tokens were used for training. The tokens chosen for training were situated at the extreme ends of the vowel quality 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 vowel 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 vowel 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. The resulting 32 pairs were also presented in reverse order, for a total of 64 training pairs, presented 4 times, for a total of 512 words heard during a training session.

3.3. Procedure

For the pre-test (and similarly for the post-test), the 28 tokens were presented randomly four times (we discarded the first round of 28 words from the analyses, considering it a practice session). These tests used an identification task, so that the learner would hear the word ‘ship’ and had to decide if the word was ‘ship’ or ‘sheep’ by pressing the appropriate key on the response pad. No feedback was provided during the tests.

After the pre-test and before the post-test, the Japanese listeners underwent one hour of discrimination training (2 sessions of about 30 minutes each, performed on different days). For the training, the learner would hear two words with an ISI of 1500ms and had to decide if the words were the same or different by pressing the appropriate key on the computer keyboard. Each trial was followed by feedback indicating whether the choice was correct (but the words were never repeated to keep the number of words heard constant across participants). The next trial was presented after an inter-trial-interval of 2000 ms added after the response.

4. Results and discussion

First of all, to confirm that the training had some effect, the training scores were compiled for each training session. The average score during the first training session was 88.3% (St.dev. 10.6), and increased to 93.58% (St.dev. 8.10) during the second training session. Hence, a significant improvement of 5.28% was noticeable after only one hour of exposure to the training stimuli through the discrimination task (t (19) = 4.09, $p < 0.001$).

The research question addressed by this study was why discrimination training was not as effective as identification training in the earlier studies using training with vowel contrasts. One possibility is that some learners fail to associate the vowels with the proper letters on the identification test after training. To evaluate this, the responses of each individual to the 28 tokens (repeated 3 times) on the post-test were compiled and the identification patterns were then visually inspected for possible mislabeling issues. Five individuals out of twenty (25%) were identified as associating the vowel /i/ with the word ‘ship’ instead of ‘sheep’ after training. The average scores along the formant dimension of those 5 participants (MisLab group) were then calculated and compared with the average scores of the other 15 participants (CorLab group) and reported in Figure 2 below.

The top panel of Figure 2 shows the pre-test scores and one can see that the correct labeling (CorLab) group was able to use formant information correctly to some extent even before training, and their use of

the spectral cues improved after training (bottom panel). Conversely, the mislabeling (MisLab) group did *not* use formant information prior to training, and hence, were presumably not aware of the correct association between vowel sound and letter before the start of the training sessions, and ultimately associated the vowels with the incorrect letters post training. Hence, these results show that a fair number of learners (25% in this case) may fail to associate the trained vowels with the proper orthographic representation after discrimination training. Importantly, this phenomenon may have impacted the resulting average post-test scores as used in previous studies comparing the use of an identification versus discrimination task for phonetic training with vowel sounds [10, 11, 12], leading to the conclusion that discrimination training is less effective than identification training when training with vowel sounds. But, is it? The answer depends on the criterion used to judge effectiveness, and what is the ultimate goal of the training sessions. If the goal is to enable L2 listeners to rely on the critical acoustic cue that native speakers are using, the next question should be can discrimination training help change the cue-weighting pattern towards native norms. To address this question, we looked at how the group who correctly labeled the vowels (CorLab, N=15) used temporal and spectral information before and after training.

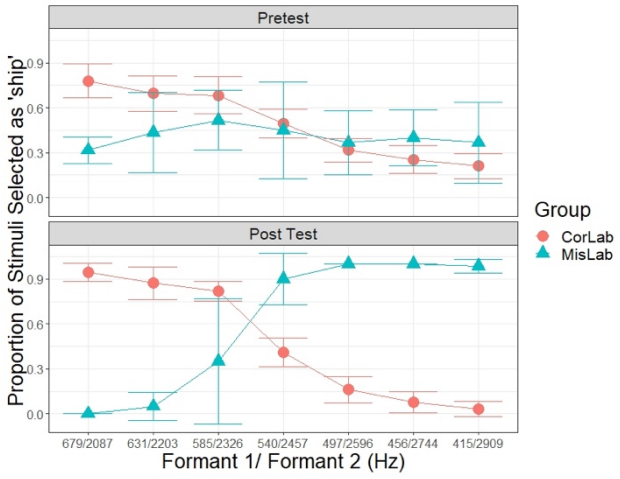

Figure 2: Results on pre-test (top panel) and post-test (bottom panel) for the group who mislabeled the vowels (MisLab) and the group who correctly labeled the vowels (CorLab) along the spectral continuum.

The use of the temporal cue (i.e., vowel duration) of this group before and after training is compared with that of the native English listeners in Figure 3 below. We can see that the participants in this group were using the temporal information in order to categorize the vowels before training, whereas the English listeners mostly ignored changes in vowel duration.

To evaluate whether the change in the use of vowel duration was significant over time, a two-way repeated-measures ANOVA was performed with the pret-test and post-test data of the Japanese group. As the data were not spherical (Duration: W = 0.24, p = .003; Time x Duration: W = 0.19, p < .001), the Greenhouse-Geisser correction was used. There was a significant effect of change in duration with a large effect size (F (3, 42) = 47.5, GGe = 0.53, p < .001, $\eta_p^2 = 0.57$) but no significant effect of Time alone (pre-test vs. post-test) (F (1, 14) = 0.216, p = 0.65, $\eta_p^2 = 0.003$). The effect of interest, however, is the Time x Duration interaction which was significant with a medium effect size (F (3, 42) = 22.9, GGe = 0.55, p < .001, $\eta_p^2 = 0.28$), meaning that the correct labeling (CorLab) group significantly altered their use of duration from pre-test to post-test.

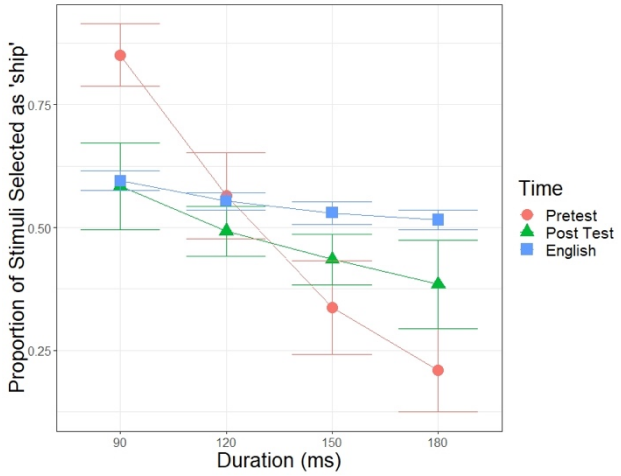

Figure 3: Results of the 28 test tokens across all vowel duration values for the CorLab group before (red circles) and after (green triangles) training compared with the values for the native English listeners (blue squares).

The English group's performance on the test was then compared to the post-test performance of the correctly labeling Japanese group with a mixed effect ANOVA, to see if the use of vowel duration by the Japanese group approximated the native norm. Again, the data were not spherical (Duration: W = 0.34, p < .001; Group x Duration: W = 0.34, p < .001) so the Greenhouse-Geisser correction was used. There was a significant effect of change in duration with a medium effect size (F (3, 159) = 22.8, GGe = 0.57, p < .001, $\eta_p^2 = 0.23$) and a significant effect of group, but with a small effect size (F (1, 53) = 26.0, p < .001, $\eta_p^2 = 0.13$). The effect of interest, however, is the Group x Duration interaction which was not significant (F (3, 159) = 4.10, GGe = 0.57, p = 0.024, $\eta_p^2 = 0.05$), meaning that the correct labeling group performed similarly to English speakers in their use of vowel duration after training.

Next, the use of spectral cues (i.e., formant frequencies) by the correctly labeling group before and after training was compared with that of native English listeners in Figure 4 below. While the correct labeling group could use spectral information before training, the shape of the slope (approximating linearity) suggests that these participants did not use formant information categorically yet. After training, however, their use of the spectral cues was more categorical and started to resemble the native speakers' performance.

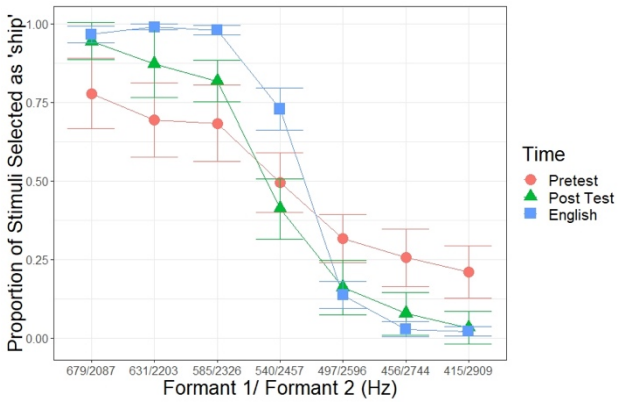

Figure 4: Results of the 28 test tokens across all formant values for the CorLab group before (red circles) and after (green triangles) training compared with the values for the native English listeners (blue squares).

To evaluate whether the change in the use of spectral information was significant over time, a two-way repeated-measures ANOVA was performed with the data of the correct labeling group. Once again, the data were not spherical (Formants: W = 0.003, p < .001; Time x Formants: W = 0.07, p = .049) so the Greenhouse-Geisser correction was used. There was a significant effect of change in formants with a very large effect size (F (6, 84) = 107, GGe = 0.30, p < .001, $\eta_p^2 = 0.77$) but no significant effect of Time alone (pre-test vs. post-test) (F (1, 14) = 0.22, p = 0.65, $\eta_p^2 = 0.003$). The effect of interest, however, is the Time x Formants interaction which was significant with a medium effect size (F (6, 84) = 12.0, GGe = 0.51, p < .001, $\eta_p^2 = 0.19$), meaning that the correct labeling group significantly altered their use of formant frequencies from pre-test to post-test.

The English group's test performance was then compared to that of the correct labeling group's post-test performance to see if the use of spectral cues by the correct labeling group approximated the native speakers' performance. Again, the data were not spherical (Formants: W = 0.03, p < .001; Group x Formants: W = 0.03, p < .001) so the Greenhouse-Geisser correction was used. There was a significant effect of Formants with a very large effect size (F (6, 312) = 604, GGe = 0.57, p < .001, $\eta_p^2 = 0.91$) and a significant effect of Group, but with a very small effect size (F (1, 52) = 26.7, p < .001, $\eta_p^2 = 0.08$). The effect of interest is the Group x Formants interaction which was significant with a medium effect size (F (6, 312) = 14.3, GGe = 0.57, p < 0.001, $\eta_p^2 = 0.19$), meaning that the correct labeling group performed differently from English speakers in their use of spectral information after training. What appears to be the case is that the Japanese learners set their categorical boundary between the English high front lax vowel and the tense vowel farther away from the center of the /i/ vowel than the monolingual English speakers. However, this is to be expected. The phonetic category dissimilation hypothesis, described in Flege's [22] speech learning model (SLM), posits that this phenomenon is a bilingual strategy in order to keep the L1 and L2 phonetic systems distinct. Importantly, the same phenomenon was observed when using an identification task [4], and hence, should not be considered a shortcoming of the discrimination task.

Moreover, going back to the results of the mislabeling group (N = 5) as presented in Figure 2 above, it is clear that these participants could use formant information after training, although they did not do so before training. Hence, although they may associate the vowels with the incorrect letters when assessed with an identification task, they appear to have forged a new vowel category within the vowel space. In that sense, discrimination training was successful even for the mislabeling group.

5. Conclusion

The results of this study indicate that the discrimination training task may be effective to help L2 learners become more sensitive to spectral information for vowel categorization. The current results, however, indicate caution needs to be taken when using this task for vowel training, as some learners demonstrate mislabeling issues, especially among those learners who are unaware of the proper association between sounds and letters at training onset. Caution may also be in order for interpretation of previous studies comparing the use of an identification training task with a discrimination training task, since these studies suggest that the identification task is superior to the discrimination task for training vowels. The current study demonstrated that while the discrimination task does not teach them the proper phoneme-grapheme association, it is still an efficient method for helping L2 listeners create new vowel categories within the spectral space. Hence, the discrimination training procedure could potentially be useful, for instance, with populations who are not literate in the target

language—such as Japanese beginner learners of Russian or young Japanese children learning English—as long as instructions about the association between sounds and letters are provided at an appropriate time.

6. Acknowledgments

We would like to warmly thank the numerous research assistants and participants without whom this research would not have been possible. Thanks too to Jim Tanaka. This research was also possible thanks to a grant-in-aid to scientific research by the Japan Society for the Promotion of Science, number 16K02915, granted to Izabelle Grenon.

7. References

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] Wang, X., Munro, M. J. 2004. Computer-based training for learning English vowel contrasts. *System, 32*, 539-552.

[7] Wang, Y., Spence, M. M., Jongman, A., Sereno, J. A. 1999. Training American listeners to perceive Mandarin tones. *JASA, 106* (6), 3649-3658.

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

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

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

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

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

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

[14] Grenon, I. 2012. The bi-level input processing model of first and second language perception (Doctoral dissertation, The University of Victoria, Canada). *Dissertation Abstracts International, 72* (8), 285.

[15] 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.

[16] 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.

[17] 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.

[18] Nishi, K., Kewley-Port, D. 2007. Training Japanese listeners to perceive American English vowels: Influence of training sets. *Journal of Speech, Language, and Hearing Research, 50*, 1496-1509.

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

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

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

[22] 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). Timonium, MD: York Press.