DIFFERENTIAL EFFECTS OF STIMULUS VARIABILITY AND LEARNERS' PRE-EXISTING PITCH PERCEPTION ABILITY IN LEXICAL TONE LEARNING BY NATIVE ENGLISH SPEAKERS

Jiyeon Lee[1], Tyler K. Perrachione[2,4], Tasha M. Dees[1], & Patrick C.M. Wong[1,3]

[1]The Roxelyn and Richard Pepper Department of Communication Sciences and Disorders, [2]Department of Linguistics, [3]Northwestern Institute for Neuroscience, [4]Cognitive Science Program
Northwestern University, Evanston, Illinois, U.S.A.
pwong@northwestern.edu

ABSTRACT

We examined the role of stimulus variability in learning non-native phonetic contrasts (suprasegmentals) for word identification by adults, considering whether all learners benefit from high-variability training. Forty-seven English-speakers were trained to use Mandarin lexical tones to identify 18 English pseudowords with stimuli produced by either four talkers (multi-talker training) or only one of the four talkers (single-talker training). Subjects' pre-training ability to identify the pitch patterns in a non-lexical context was also measured. Subjects with high pitch-identification ability learned more successfully than those with lower pitch-identification ability. Importantly, multi-talker training was beneficial only for learners with high pitch-identification ability, whereas learners with low pitch-identification ability benefited more from single-talker training. These findings provide support for phonetic-phonology continuity in adult sound-to-word learning and suggest that in order for high-variability training to be beneficial, learners first need to be able to successfully encode phonetic details.

Keywords: Talker variability, Second-language acquisition, Lexical tone, Suprasegmental learning

1. INTRODUCTION

One aspect of learning a new language involves learning non-native phonetic contrasts to identify word meaning. Adult learners can learn non-native phonetic contrasts after training, including both segmental [1,2] and suprasegmental contrasts [3]. Only a few studies, however, have shown that adult learners can learn to use these contrasts for word identification. Curtin et al. [4] trained English speakers to use a 3-way voicing contrast to identify 18 actual Thai words. Wong & Perrachione [5] showed that adults can learn novel suprasegmental contrasts for use in words. They trained English-speaking adults to use Mandarin lexical tones to identify 18 artificial words, and found that learning success was strongly associated with trainees' pre-training ability to identify pitch patterns in non-lexical contexts. Successful learners had higher pre-training pitch identification (pitch-ID) scores than less-successful learners

Training stimulus variability plays an important role in perceptual learning by encouraging robust category formation [2,6]. The efficacy of training with high-variability (e.g. contrasts produced by multiple talkers and occurring across various phonetic environments) is evident in generalization to new stimuli and new speakers [3,6] and retention measured by 3- or 6-month follow-up tests [3,8,9]. For instance, Lively et al. [6] trained Japanese listeners with English /r/ and /l/, using a two-alternative forced choice procedure. When listeners were trained with 5 talkers, the learning generalized to untrained stimuli and new talkers. However, this pattern was not observed in listeners who were trained with a single talker. Although the efficacy of multi-talker training is established in non-lexical perceptual learning for segmental contrasts, it is unknown whether it is efficacious in learning suprasegmental contrasts, and whether these effects extend to learning contrasts for lexical (higher-level) purposes. Finally, previous studies have not investigated whether and to what extent learners' pre-existing auditory ability interacts with the type of training received.

The purpose of the current study was to examine the efficacy of multi-talker training in Mandarin lexical-tone (suprasegmental) learning by native-English listeners and to examine the interaction between training type and learners' pre-training pitch identification ability.

2. METHODS

2.1. Subjects

A group of 47 native speakers of American English (age 18-28, M = 21.6 years) participated in this study. All were recruited from the university or local community. None reported any hearing or speech problems at the time of training. No subject had ever had any prior exposure to a tone language.

2.2. Stimuli

The training stimuli consisted of 18 English pseudowords manipulated with three pitch patterns, resembling Mandarin tone 1 (level), 2 (rising), and 4 (falling). Tone 3, the dipping tone, was not included because it has been shown to be the most confusable to second-language learners of Mandarin [10]. As shown in Table 1, there are six sets of words with three minimal pitch contrasts in each set, resulting in a total of 18 words.

For the training stimuli, four native speakers of American English (2 female, 2 male) produced the 18 pseudowords with a high pitch. For generalization test stimuli, the same words were produced by another four speakers (2 female, 2 male). The recorded words were then resynthesized to include variants consisting of the three different pitch patterns, following Wong & Perrachione [5]. Pitch patterns were interpolated linearly through the voiced portion of each word, using the Pitch-Synchronous Overlap and Add (PSOLA) method implemented in the software Praat. The reliability of the pitch patterns was identified by 5 native Mandarin speakers, with identification accuracy of 97%.

Table 1: Training stimuli in IPA (numbers indicate tones; words meanings are in quotations)

[pʰɛʃ1] 'glass'	[dɹi1] 'arm'	[nɛɹ1] 'boat'	[vɛs1] 'hat'	[nʌk1] 'brush'	[fjut1] 'shoe'
[pʰɛʃ2] 'pencil'	[dɹi2] 'phone'	[nɛɹ2] 'potato'	[vɛs2] 'tape'	[nʌk2] 'tissue'	[fjut2] 'book'
[pʰɛʃ4] 'table'	[dɹi4] 'cow'	[nɛɹ4] 'dog'	[vɛs4] 'piano'	[nʌk4] 'bus'	[fjut4] 'knife'

2.3. Non-lexical pitch identification test

Prior to lexical training, subjects' ability to identify pitch patterns was tested in a non-lexical context (pitch-ID test), adopted from Wong & Perrachione [5]. Subjects listened to the three (level, rising, falling) tone contours over cardinal vowels produced by 4 talkers (2 female, 2 male). Two different arrows representing pitch contours appeared on the computer screen, and subjects pressed a button to match each aurally presented pitch pattern with its representative arrow. Subjects were familiarized with the contours and task before the test, and accuracy was determined by correct responses out of 180 trials. In case a subject performed at or below chance level (50% or lower), the subject was re-tested with the same pitch-ID test to ensure the reliability of the score.

2.4. Procedure

Subjects were randomly assigned to single- vs. multi-talker training groups. Within each training group, subjects were further divided into high vs. low pitch-ID groups, based on their scores in the pitch-ID test. High pitch-ID was defined by 70% accuracy and above, and scores below 70% defined the low pitch-ID group. Similarly, Wong & Perrachione [5] found 74% pitch-ID accuracy was the criterion for successful learning.

The subgroups were matched for their mean pitch-ID score between training types. In both multi- and single-talker training, the mean was 77% for high pitch-ID groups and 58% for low pitch-ID groups.

2.4.1. Training

Subjects were trained to identify word meanings depicted by drawings, similar to Wong & Perrachione [5]. Word meanings assigned to the stimuli represent high frequency English nouns [11]. Each training session consisted of a practice phase and a daily word identification (word-ID) test. During the practice phase, similar to Curtin et al. [4], the 18 words were divided into 6 groups of 3 words to facilitate learning. Each group contained all three lexical tones, as minimal triads (e.g., [nʌk1], [nʌk2], and [nʌk4]). Each group of words was presented aurally, simultaneously with the drawings. Subjects were then quizzed to match the heard words with the three drawings, with feedback provided. At the end of each training session, the subjects were given the daily word-ID test. Subjects heard each word and were asked to identify its meaning by selecting the appropriate drawing out of 18 choices.

For multi-talker training, subjects were trained and tested with words produced by all four talkers resulting in a total of 72 tokens (18 words x 4 talkers). For single-talker training, the subjects

were trained with only one of the four talkers. Thus, each word was repeated four times, resulting in the same number of 72 tokens (18 words x 4 times). All 18 words were learned within one session. A training session, including both training phase and word-ID test, lasted about 30 minutes. All subjects received 8 days of training. They were trained at least 5 days per week, with no longer than a two-day interval between sessions.

2.4.2. Generalization test

Subjects participated in the generalization test the day after training was terminated. For all subjects, the stimuli consisted of the same words produced by the four untrained talkers. Stimuli were blocked by talkers. Each block consisted of the stimuli from one talker repeated four times. Subjects were asked to identify each word by selecting the corresponding drawing out of 18 possible choices with no feedback given, similar to the daily word-ID test described above.

3. RESULTS

3.1. Training results

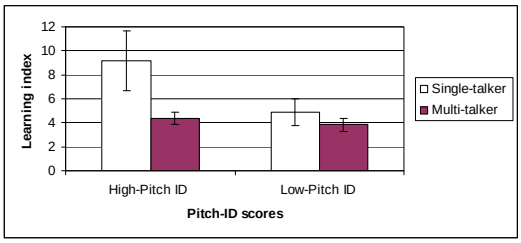
All data reported were Box-Cox transformed to improve normality. Daily word-ID scores from the first and eighth session of training from all subjects were entered into a repeated measures ANOVA, which revealed significant training improvement [$F(1,43) = 567.153$, $p < 0.001$]. On average, subjects improved by 56%.

3.2. Learning index

As an indicator for the amount of learning over the training, a 'learning index' (the difference in daily word-ID score between the last and first session devided by the first-session score) was calculated for each subject (Figure 1). A 2 x 2 (training type x pitch ID group) ANOVA revealed no main effect of training type [$F(1,43) = 0.001$, $p = .984$]. However, there was a main effect of pitch-ID scores, suggesting that the learners with high pitch-ID scores learned more (M = 6.78) than those with low pitch-ID scores (M = 4.36) [$F(1,43) = 5.675$, $p = 0.022$].

We did not find a significant interaction between training-type and learners' pitch-ID [$F(1,43) = 1.431$, $p = .238$]. However, it is worth mentioning that t-tests on the transformed learning index data did not show reliable differences between training types among learners with high Pitch-ID scores (t=-0.811, p = 0.426) and among learners with low pitch-ID scores (t=0.886, p = 0.386). In addition, there was a significant correlation between learning index and generalization scores, suggesting that the amount of generalization is a function of the amount of learning (Pearson's $r = .393$, $p = .006$).

Figure 1. Learning index (difference in the scores of daily word-ID tests between last and first session, divided by the first-session score).

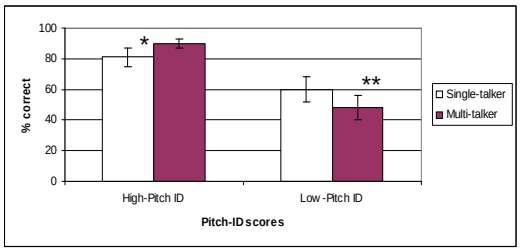


3.3 Generalization results

Results from the generalization test are shown in Figure 2. A 2 x 2 (training type x pitch-ID group) ANOVA indicated no reliable main effect of training type [$F(1,43) = 0.812$, $p = 0.08$]. But there was a main effect of pitch-ID group [$F(1,43) = 3.430$, $p < 0.001$], indicating learners with high pitch-ID showed greater generalization (M = 85%) than those with low pitch-ID scores (M = 53%). Post-hoc analysis showed that, within the multi-talker group, learners with high pitch-ID scores showed better generalization (90%) than learners with low pitch-ID scores (48%) [t = 4.870, $p < 0.001$]. Within single-talker training, the learners with high pitch-ID scores (81%) showed better generalization than those with low pitch-ID scores (60%) [t = 3.763, $p = 0.001$].

Importantly, we found a significant interaction between training type and the learners' pitch-ID scores [$F(1,43) = 10.893$, $p = 0.002$]. Post-hoc comparisons showed that among learners with high pitch-ID scores, multi-talker training resulted in better generalization (M = 90%) than single-talker training (M = 81%) [t = -2.321, $p = 0.015$]. On the other hand, among learners with low pitch-ID scores, single-talker training yielded better generalization (60%) than multi-talker training (48%) [t = 2.996, p = 0. 003].

Figure 2. Generalization results * p = 0.015, ** p = 0.003



4. DISCUSSION

In both learning index and generalization results, we found that learners with high pitch-ID ability performed better than those with low pitch-ID ability, regardless of types of training stimuli. This finding is consistent with Wong & Perrachione [5], suggesting that the learners' pre-training pitch-ID ability plays an important role in learning new words using non-native suprasegmental contrasts.

Interestingly, unlike findings from Lively et al. [6], our preliminary data suggest that not all learners benefit from multi-talker training. Multi-talker training was more efficacious than single-talker training only for learners whose pre-training pitch-ID scores were relatively high. For subjects with high pitch-ID scores, learning generalized to a greater degree when they were trained with multiple talkers as compared to a single talker (Fig. 2), despite the lack of a reliable difference in their learning indices (Fig.1). However, when learners had low pre-existing pitch-ID ability, single-talker training yielded greater generalization than multi-talker training (Fig. 2). The significant correlation between learners' learning indices and generalization scores possibly suggests that the learners' ability to generalize is attributable to the amount of learning.

As suggested in infant word learning [12], phonetic categories need to be established before the phonetic details are used phonologically, i.e., to contrast word meanings. Our results provide preliminary evidence that, in this process, pre-existing differences in the learners' ability to perceive relevant phonetic details (here, pitch patterns) plays an important role, affecting not only learning success but also training type efficacy. When the learners had higher sensitivity to pitch patterns, high-stimulus variability facilitated acquisition of phonetic categories. In contrast, low-variability stimuli facilitated acquisition of categories when the learners had low sensitivity to pitch patterns.

5. CONCLUSIONS

We found that learners' pre-existing pitch-ID ability in non-lexical contexts is associated with successfully learning to use pitch contrasts to identify words. Pre-existing pitch-ID ability also interacted with the efficacy of single- vs. multi-talker training types. High-variability training was beneficial only for learners with high pitch-ID ability, whereas low-variability training was more beneficial for learners with low pitch-ID ability.

6. REFERENCES

[1] Pisoni, D., Aslin, R., Perey, A., & Hennessy, B.. 1982. Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *J. Exp. Psychol. Human*, 8, 297-314.

[2] Jamieson, D. Morosan, D. 1989. Training non-native speech contrasts in adults: Acquisition of the English /θ/-/ð/ by francophones. *Percept. Psychophys.* 40. 205-215.

[3] Wang, Y., Spence, M., Jongman, A., Sereno, J. 1999. Training American listeners to perceive Mandarin tones. *J. Acoust. Soc.Am*, 106, 3649-3657.

[4] Curtin, S., Goad, H., Pater, J. 1998 Phonological transfer and levels of representation: the perceptual acquisition of Thai voice and aspiration by English and French speakers. *Second Lang. Res.*, 14. 389-405

[5] Wong, P., Perrachione, T. 2007. Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics*. in press.

[6] Lively, S., Logan, J., Pisoni, D.1993. Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *J. Acoust. Soc. Am*, 94. 1242-1255.

[7] Logan, I., Lively, S., Pisoni, D. 1991. Training Japanese listeners to identify English /r/ and /1/: A first report/ *J. Acoust. Soc. Am*. 89, 874--886.

[8] Lively, S., Pisoni, D., Yamada, R., Tokura, Y., Yamada, Y. 1994. Training Japanese listeners to identify English /r/ and /l/ III. Long-term retention of new phonetic categories, *J. Acoust. Soc.Am.* 96, 2076-208.

[9] Bradlow, A., Akahane-Yamada, R., Pisoni, D., Tokura, Y. 1997. Training Japanese listeners to identify English /r/ and /l/: long-term retention of learning in perception and production. *Percept. Psychophys.*, 61, 977-85

[10] Kiriloff, C. 1969. On the auditory perception of tones in Mandarin. *Phonetica*, 20:63-67, 1969.

[11] Raymer, A., Maher, L., Greenwald, M., Morris, M., Rothi, L. Heilman, K. 1990. *The Florida Semantics Battery*. Unpublished test.

[12] Werker, J., Fennell, C., Corcoran, K., Stager, C. 2002. Infants' ability to learn phonetically similar words: effects of age and vocabulary size. *Infancy*, 3, 1-30.

This work is supported by the National Institutes of Health (U.S.A.) grants HD051827 & DC007468 awarded to Patrick Wong who is the corresponding author.