

# Lexical exposure to native language dialects can improve non-native phonetic discrimination

Annie J. Olmstead<sup>1</sup> · Navin Viswanathan<sup>1,2</sup>

Published online: 30 October 2017  
© Psychonomic Society, Inc. 2017

**Abstract** Nonnative phonetic learning is an area of great interest for language researchers, learners, and educators alike. In two studies, we examined whether nonnative phonetic discrimination of Hindi dental and retroflex stops can be improved by exposure to lexical items bearing the critical non-native stops. We extend the lexical retuning paradigm of Norris, McQueen, and Cutler (*Cognitive Psychology*, 47, 204–238, 2003) by having naive American English (AE)-speaking participants perform a pretest-training-posttest procedure. They performed an AXB discrimination task with the Hindi retroflex and dental stops before and after transcribing naturally produced words from an Indian English speaker that either contained these tokens or not. Only those participants who heard words with the critical nonnative phones improved in their posttest discrimination. This finding suggests that exposure to nonnative phones in native lexical contexts supports learning of difficult nonnative phonetic discrimination.

**Keywords** Speech perception · Perceptual learning

Successful speech perception requires perceptual learning of differences between phonemes in a given language at the expense of perceptual sensitivity to within phoneme differences. Consequently, learning nonnative phoneme contrasts can be especially difficult when the to-be-learned contrast requires differentiation of nonnative phones that are treated as

allophones in the learners' native language (Best, 1994; Best, McRoberts, & Goodell, 2001). In such a situation, learners must first learn to discriminate these phones and then to identify them correctly. A well-known example of such a situation is American English (AE) listeners learning to discriminate the Hindi stops, [ʈ] - [ɖ]. These stops differ in their places of articulation, retroflex versus dental, respectively, but are not phonemically contrasted in AE; instead, they exist as allophones of /t/ (Pruitt, Akahane-Yamada, & Strange, 1998). As such, this contrast presents an interesting case for studies of perceptual learning.

Previous studies have trained AE listeners on this contrast by repeatedly exposing listeners to sublexical examples and then testing for improvements in either discrimination or identification. For example, Pruitt, Jenkins, and Strange (2006) examined improvement in Hindi consonant identification after 1.5 hours of training in which AE listeners repeatedly heard these consonants in different contexts with feedback on consonant category membership. After this familiarization, participants showed modest improvement (10–16%) in identification and some evidence of generalization of learning to novel vowel contexts. In a different approach, Tees and Werker (1984) trained listeners in up to 300 trials in which participants listened to one of the consonants until they requested a change to the other consonant. This training did not lead to improvement in the discrimination test, which required participants to monitor a stream of sounds for a change from one phoneme to the other. Similar studies (Werker, Gilbert, Humphrey, & Tees, 1981; Werker & Logan, 1985; Werker & Tees, 1984) have shown that while English speakers can discriminate the Hindi consonants, this ability is task-dependent and does not seem to improve over trials. In all of these studies, researchers relied on training auditory discrimination through repeated exposure to the different phones without lexical context.

✉ Annie J. Olmstead  
anne.j.olmstead@ku.edu

<sup>1</sup> Department of Speech-Language-Hearing, University of Kansas, Lawrence, KS, USA

<sup>2</sup> Haskins Laboratories, New Haven, CT, USA

One strategy for phonetic learning that has not been explored with this contrast is the use of lexical knowledge. In a seminal study, Norris, McQueen, and Cutler (2003) showed that lexical knowledge may be used to retune stable native phonetic categories. The authors replaced /f/ and /s/ in Dutch words with an ambiguous token [ʔ] that could be heard as either fricative. They then presented either the f-replaced tokens or the s-replaced tokens to Dutch listeners in an auditory lexical decision task. Those listeners who heard f-replaced tokens were more likely to categorize the ambiguous token as [f] in a categorization task than those in the other group. This indicated that brief exposure to an ambiguous segment in a lexical context was sufficient to alter stable native phonetic categories. Several studies have since used this paradigm across languages and contexts successfully (see Kraljic & Samuel, 2005, 2006, 2007; for review).

Given that the difficulty of Hindi contrast discrimination partly stems from the stable structure of AE phonetic categories, we asked whether the lexical retuning paradigm can be utilized to promote the discrimination of the Hindi phones. To do so, we take advantage of a particular regularity that occurs in Indian English (IE): a dialect of English spoken on the Indian subcontinent. In IE, dental and retroflex voiceless stops are used systematically to instantiate the [t]-[θ] distinction of AE. Specifically, IE talkers use [t̪] in place of AE [t], such as in the word “tick” and [t̪] instead of [θ], such as in the word “thick” (Sailaja, 2009). This systematic use of the retroflex and dental consonants by IE speakers provides us with a unique opportunity to examine whether nonnative phonetic category learning can result from exposure to naturally produced (as opposed to synthetic or spliced, see Sjerps & McQueen, 2010, for discussion) words in the native language that contain the segments of interest. This design differs from Norris et al. (2003) in important ways. First, while lexical retuning tasks typically focus on a single, ambiguous token, we focus on two different phones that are perceived as the same phoneme (i.e., a one-category assimilation contrast; Best, 1994; Flege, 1995). The task for the listener is to retain the retroflex stop as English /t/ and differentiate it from the dental stop that can now be assimilated into their /θ/ category. Second, while lexical retuning tasks typically examine learning of variations that center on the ambiguous token itself (but see Reinisch, Weber, & Mitterer, 2013), we rely on natural dialectal variations in English pronunciation.

Xie, Theodore, and Myers (2017) applied the lexical retuning paradigm to examine phonetic category reorganization in English listeners exposed to Mandarin-accented English. In this accent, words ending in voiced alveolar stops (“seed”) are produced as unvoiced, aspirated ones (“seet<sup>h</sup>”). They demonstrated that listeners performing auditory lexical decision on naturally produced accented tokens, compared with controls, show changes in both category boundaries and internal category structure of /d/ and /t/. Critically, this

indicates that naturally produced accented versions of native speech can alter native phonetic categories through lexical retuning. Our study will take this methodology further by examining whether lexical exposure can support a basic level of phonetic retuning: being able to reliably discriminate two nonnative phones.

The differences between the IE and AE dialects go beyond the specific consonantal differences described. Because of both segmental and suprasegmental differences from AE, even relatively short words are produced differently in IE. Consequently, to ensure that listeners accept IE tokens as words, we used a transcription task (Bradlow & Bent, 2008) in the exposure phase to both implicitly indicate to participants that they are listening to words as well as obtain data on how the nonnative phones were heard. We did so because pilot studies using auditory lexical decision suggested that listeners had difficulty accurately judging isolated words produced in IE dialects (Olmstead & Viswanathan, 2016).

In this study, we investigate whether the systematic mapping of nonnative phones to native categories in IE dialects can be used to support AE listener’s learning of the Hindi contrast. In two experiments, we examined whether exposure to naturally produced words in IE containing the Hindi dental (“thick”) and retroflex consonants (“tick”) can help AE listeners learn to discriminate those consonants in a diphone context. We expect that participant’s discrimination of the nonnative contrasts will improve as a function of lexical exposure when words contain the target stops (Experiment 1) but not when they do not (Experiment 2).

## Experiment 1

We asked AE listeners to transcribe naturally produced IE words containing the target contrasts: dental versus retroflex stops. To determine whether this exposure helped participants make the distinction, we tested them on AXB diphone discrimination before and after the exposure task.

## Method

**Participants** Twenty-two (9 males, 13 females) University of Kansas students participated in our study. All were native English speakers who reported no proficiency in any other language, normal or corrected to normal vision, normal hearing, and no known speech or language deficits.

**Stimuli** Two female talkers recorded transcription and discrimination stimuli. We used two talkers so that we could examine whether any learning that occurred during the experiment resulted from perceptual adjustment to talker idiosyncracies or general adjustments that affected the perceptions of multiple talkers. Talker 1 was a 28-year-old native

Hindi speaker who reported fluency in Urdu, English, and familiarity with Kannada. She started speaking IE at age 4 years. At the time of recording, she had been in the United States for 2 months and estimated that 90–100% of her current language use was English. Talker 2 was a 30-year-old native Hindi speaker who was fluent in English (IE), which she began speaking at age 5 years. At the time of recording, she had lived in the United States for 8 months and reported that 90% of her current language use was English. Naturally produced words from these talkers were equated for average intensity. Importantly, unlike earlier studies (e.g., Norris et al., 2003), no other manipulations were performed.

**Discrimination** Each talker also produced the Hindi unvoiced retroflex stop and the Hindi unvoiced dental stop in three different vowel contexts (/a/, /e/, /i/). Multiple vowel contexts were chosen to vary coarticulatory information and assess generalizability of learning. The talkers read the Hindi script for each diphone, and the two best tokens, based on clarity of articulation and recording, per vowel context were used. Pilot testing by two other Hindi listeners showed ceiling performance on categorization of these stops.

**Transcription** Four types of words were used: (i) /t-/θ/ minimal pairs (e.g., tin-thin), (ii) /t/ words (e.g., taste), (iii) /θ/ words (e.g., think), and (iv) fillers that did not contain these stops. Other than containing the target consonants, the phonetic structure of the stimuli was unrestricted.

Each talker read the word lists in a sound attenuated booth. Speech was recorded using a Shure S58 stand microphone placed on the desk in front of the talker using Audacity at a sampling rate of 44.1 kHz.

**Procedure** Participants provided informed consent and basic demographic and linguistic background information. They performed preexposure AXB discrimination, followed by word transcription, and the postexposure AXB discrimination.

**AXB tasks** The pre-test and post-test discrimination tasks were identical. Participants sat in a quiet room and listened to stimuli over Sennheiser HD 558 headphones run through a Behringer, HA400 microamplifier at 70 dB. For each trial, participants heard three diphones, each separated by 100 ms of silence and indicated whether the first two sounds or the last two sounds were the same. All three diphones were different tokens with two of the tokens drawn from the same category. Each combination (RRD, RDD, DRR, DDR) was presented four times in each talker's voice for a total of 48 trials. Trial order was randomized. Each test took about 5 minutes.

**Transcription task** After the pretest, participants transcribed the words of either Talker 1 or Talker 2. They heard each word once and typed their responses. They were told that the talker

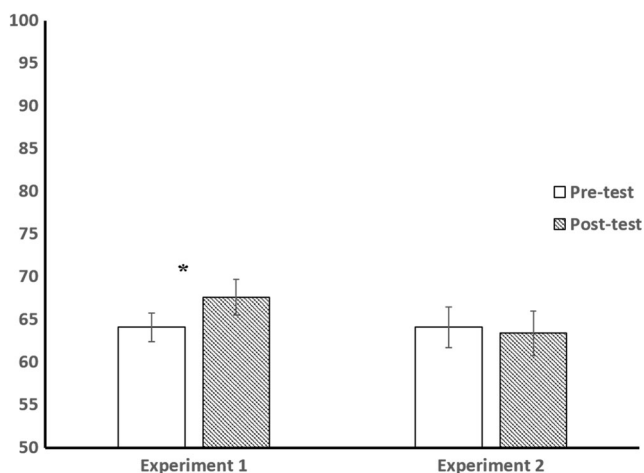
spoke a different English dialect and were asked to type the word they thought the talker said. After each response the next word played after a 1,000 ms. delay. Words were presented in two blocks. In the first block, 27 /t/ words, 17 /θ/ words and 26 filler words were presented. In the second block, 26 /t-/θ/ minimal pairs (52 words overall) and 25 fillers (different from those presented in the first block) were presented. Half of the participants heard the words spoken by Talker 1 and half heard Talker 2 (this factor is referred to as training talker). The transcription task took about 15 minutes.

## Results and discussion

**Discrimination** Overall discrimination accuracy was calculated by averaging accuracy for each subject across AXB triplets by session (pre-test, post-test), discrimination talker (one producing the tokens to be discriminated), training talker, and vowel.

**Transcription** Transcriptions were scored for overall accuracy (whether the entire word was correctly reported) and /t-/θ/ distinction accuracy (whether the target segment was correctly reported). Homophones (e.g., typing “TACKS” when the sound was “TAX”) and obvious typos (e.g., “THEIF” for “THIEF”) were marked correct. All other errors were marked incorrect.

**Discrimination results** For each subject, discrimination accuracy was calculated per talker for each vowel in each session. Logit transformed accuracies were submitted to a 2 session (pre-test, post-test)  $\times$  2 discrimination talker (Talker 1, Talker 2)  $\times$  3 vowel (/a/, /e/, /i/)  $\times$  2 training talker (Talker 1, Talker 2) mixed ANOVA with training talker as the between subjects factor. The analysis revealed a main effect of Session,  $F(1,20) = 6.523$ ,  $p < 0.05$ ,  $\eta_p^2 = 0.235$ . Examination of the means indicates that accuracy in the post-test (67.6%) was better than in the pre-test (64.1% correct; Fig. 1). This indicates that participants' exposure to nonnative tokens in lexical contexts in the transcription task produced phonetic learning. In addition, the analysis showed main effects of both discrimination talker,  $F(1,20) = 9.688$ ,  $p < 0.01$ ,  $\eta_p^2 = 0.325$  and vowel,  $F(2,40) = 6.532$ ,  $p < 0.01$ ,  $\eta_p^2 = 0.239$ . Examination of the means for discrimination talker shows that, across sessions, listeners performed worse in discriminating Talker 1's productions (62.9%) compared with Talker 2's (68.8%). For vowels, means show that performance was worst with an /a/ context (60.4%), followed by /i/ (66.8%) and /e/ (70.9%). Unplanned, Bonferroni corrected comparisons indicated that /a/ and /e/ were different from one another ( $p < 0.01$ ), but /i/ was not different from either. There was no main effect of training talker,  $F(1,20) = 1.84$ ,  $p = 0.19$ , indicating that it did not matter which talker's words were used in the transcription task. The three-way session  $\times$  discrimination talker  $\times$  training



**Fig. 1** Discrimination accuracy for the pre-test in white and post-test in shaded bars for Experiments 1 and 2. Pre-test and post-test differ in Experiment 1 but not in Experiment 2. Error bars denote standard error of the mean. The star indicates a significant difference

talker interaction that may have indicated talker-specific training effects did not approach significance ( $F < 1$ ). However, there was a marginal interaction of session and discrimination talker,  $F(1,20) = 4.034$ ,  $p = 0.07$ ,  $\eta_p^2 = 0.148$ . No other interactions approached significance.

**Transcription results** Transcription accuracy was logit transformed before being submitted to a 4 word type (minimal pair, t-word, th-word, filler)  $\times$  2 talker (talker 1, talker 2) mixed ANOVA. This analysis revealed a main effect of word type,  $F(3,60) = 264.598$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.89$ . Means are presented in Table 1. All pairwise comparisons of error rates were different, except between t-words and th-words. Specifically, filler words were most accurate, minimal pairs were least accurate and t- and th- words were intermediate. There also was a main effect of talker,  $F(1,20) = 5.475$ ,  $p < 0.05$ ,  $\eta_p^2 = 0.215$ , indicating that accuracy on talker 1s words (63.2%) was worse than accuracy on talker 2s words (68.7%). Word-type and talker did not interact.

To assess whether the transcription errors were primarily due to predicted t-th confusion, logit transformed t-th accuracies were submitted to a 3 word type (minimal pair, t word, th word)  $\times$  2 talker (talker 1, talker 2) mixed ANOVA. The analysis revealed main effects of word type,  $F(2,40) = 109.747$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.846$  and talker,  $F(1,20) = 12.671$ ,  $p < 0.01$ ,

$\eta_p^2 = 0.388$ . Pairwise comparisons revealed that t-words had the highest accuracy, followed by th-words, and minimal pairs had the worst accuracy (Table 1). An analysis of the difference between t-accuracy and th-accuracy within the minimal pairs also showed that participants were more accurate at transcribing t-words, as expected. Participants were more accurate in transcribing the t's and th's of Talker 2 (70%) than Talker 1 (60.4%). A word type  $\times$  talker interaction,  $F(2,40) = 7.689$ ,  $p < 0.01$ ,  $\eta_p^2 = 0.278$ , also reached significance. However, the pattern of word-type differences was the same for both Talkers indicating more pronounced differences for Talker 2. We predicted that a listener's ability to accurately hear the "t" or "th" in lexical context would correlate with their scores in the post-test diphone discrimination or their improvement from pre-test to post-test; however, the relationships were not strong ( $r = 0.21$  and  $r = 0.11$ , respectively).

These results indicated that lexical exposure in the transcription task improved listeners' ability to discriminate the dental-retroflex contrast. Diphone discrimination was better in the post-test than in the pre-test. Furthermore, this effect was not restricted to the exposed talker, i.e., there was no three-way interaction of session, discrimination talker, and training talker indicating that the pattern of improvement between the pretest and the posttest was the same regardless of who the talker was in the transcription task. However, discrimination performance did not improve uniformly. Instead, examination of the marginal interaction between session and discrimination talker revealed that there was no difference in accuracy on the two talker's tokens (talker 1 = 64.6%, talker 2 = 63.6%) in the pretest, but there was a difference in the posttest (talker 1 = 61.2%, talker 2 = 71.4%). In short, participant's improvement on discrimination was driven by discrimination of talker 2's productions irrespective of which talker they were trained on. Participants also performed more accurately in the transcription task when they were transcribing talker 2's words.

This pattern of results raises two broad questions. The first relates to differences between Talker 1 and 2. Specifically, why are participants able to show improvement more for Talker 2 than Talker 1? Transcription results suggest that participants may be better at discerning differences in Talker 2's productions perhaps indicating that it is easier for AE listeners to tune into overall (such talker differences are not uncommon; see Pruitt et al., 2006). However, discrimination for Talker 1's tokens and Talker 2's tokens were the same in

**Table 1** Percentage of correct transcriptions for different word types in Experiment 1

Speaker	Minimal pairs (tick-thick)		T-words (tall)		Th-words (thermos)		Fillers (boil)	
	1	2	1	2	1	2	1	2
Overall accuracy	33.8	43.3	66.7	74.5	65.8	66.3	86.7	90.2
T-Th accuracy	42.7	51.9	72.7	87.8	65.7	61.7		

the pretest, so any benefit that AE listeners have for Talker 2 seems to require at least some exposure or learning. The second relates to why performance improved in the post-test. There are two explanations beyond our hypothesis. First, perhaps the transcription task did not support the improvement in discrimination and the improvement was simply a function of practice in the discrimination task itself. Second, perhaps the transcription task merely provided information to the listeners regarding the general characteristics of the talker's accent and this information was sufficient to lead to improvement in discrimination. According to both alternative explanations, discrimination accuracy should improve whether or not the IE lexical items presented in the transcription task contain the target phones. In Experiment 2, we test these hypotheses.

## Experiment 2

We had another set of participants perform the same tasks as in Experiment 1, but critically, we changed the English words heard during the transcription task to only filler items, i.e., those that do not contain either of the critical segments. If discrimination still improves, then we can conclude that exposure to the specific segments is not necessary for learning to occur.

## Method

**Participants** Twenty-two (7 males, 15 females) University of Kansas students participated in our study. All were native English speakers, only three of whom reported very low proficiency in Spanish. All participants had normal or corrected-to-normal vision, normal hearing, and no known speech or language deficits.

**Stimuli** Stimuli used in this experiment were the same as those used in Experiment 1 except that only fillers (145) were used for the transcription task.

**Procedure** The procedure was identical to that of Experiment 1.

## Results and discussion

**Discrimination** As in Experiment 1, means were computed for session, talker, and vowel. These values were logit transformed and analyzed using a 2 (session: pre-test, post-test)  $\times$  2 (discrimination talker: talker 1, talker 2)  $\times$  3 (vowel context: /a/, /e/, /i/)  $\times$  2 (training talker: talker 1, talker 2) mixed ANOVA. The analysis revealed that, unlike in Experiment 1, the main effect of session was not significant ( $F < 1$ ), indicating that pretest performance ( $M = 64.1\%$ ) did not differ from post-test performance ( $M = 63.4\%$ ; Fig. 1). However, similar to the pattern in Experiment 1, there was a main effect of

talker,  $F(1,20) = 5.216$ ,  $p < 0.05$ ,  $\eta_p^2 = 0.207$ , indicating that performance on the two talkers' diphones differed. The means reveal that overall participants discriminated Talker 1's ( $M = 62.5\%$ ) tokens less accurately than Talker 2's ( $M = 66.1\%$ ). This pattern, however, did not change as a function of session ( $F = 1$ ) or training talker ( $F < 1$ ). Unlike in Experiment 1, there was no effect of vowel,  $F(2,40) = 2.301$ ,  $p = 0.11$ ,  $\eta_p^2 = 0.103$ .

**Transcription** Only fillers were used in the transcription task. Logit transformed accuracies were submitted to an independent samples  $t$  test. This analysis revealed a difference in accuracy between Talker 1's words and Talker 2's,  $t(20) = -5.395$ ,  $p < 0.001$ . Examination of the means revealed that, as in Experiment 1, participants were again more accurate in transcribing Talker 2's words ( $M = 90.9\%$ ) than in transcribing Talker 1's words ( $M = 82.5\%$ ).

Experiment 2 results show that exposure to lexical items without the critical phones is insufficient to improve listener's discrimination. These data do not support the hypotheses that 1) discrimination improvement is a function of practice in the discrimination task or 2) that exposure to the general characteristics of an IE accent is sufficient for learning the dental-retroflex distinction.

## General discussion

In two experiments, we examined whether exposure to IE lexical items containing Hindi retroflex and dental consonants can help AE listeners learn the distinction between them. We showed that such exposure does lead to improvement in diphone discrimination (Experiment 1), but that the words must contain these segments for there to be learning (Experiment 2). These findings are consistent with a lexical retuning perspective and make new contributions to the literature. First, we showed that naturally produced lexical items can be used to aid phonetic learning. Importantly, in conjunction with Xie et al. (2017), this shows that the lexically supported perceptual learning is a likely and available mechanism for perceptual learning outside of the laboratory. Secondly, we have shown that even short exposure to lexical stimuli in a dialectical variant of listeners' L1 can support perceptual learning of a difficult nonnative contrast. These findings are especially important given that the dental-retroflex distinction has required intensive, long-term training to achieve modest gains in performance on multiple types of tasks. However, it is difficult to directly compare studies, because the current study uses an AXB task, whereas previous studies have used identification tasks (Pruitt et al.), AX discrimination tasks (Werker & Tees, 1984), and change detections (Tees & Werker, 1984). While participants did not achieve native-like perceptual discrimination, the ease and effectiveness of our training procedure may make it useful and informative to laboratory

researchers. Moreover, this is the first time that lexical retuning has been demonstrated to be useful in learning difficult nonnative, single category contrasts. In particular, such training often is used in situations in which existing categories are tuned to new information. We have demonstrated that the same type of exposure, i.e., native-language lexical information, can be used to help learn nonnative contrasts.

However, our pattern of findings also raises questions. First, our data do not indicate a simple relationship between lexical exposure and phonetic discrimination. Performance in the transcription task did not correlate with performance in the discrimination task, and overall transcription performance was not exceptionally good, especially when it came to distinguishing the retroflex from the dental to determine the correct word. While listeners were better at doing this when one phoneme or the other was strongly suggested by the lexical context (e.g., in “thermos”), they were still far from perfect. Interestingly, the errors in the three critical conditions could not be fully explained by participants hearing /t/ for both the dental and retroflex consonants, as would be expected. Instead, both consonants were sometimes erroneously heard as voiced (tie being transcribed as “die”) or at a completely different place of articulation (thugs being transcribed as “pugs,” “jugs,” or “hugs”). Perhaps most surprising is that both consonants were sometimes heard as “th.” Although these types of errors represent a small proportion of the overall errors, this suggests that the exposure phase of our study did not always provide participants with a clear and systematic mapping of the produced phone to the perceived phoneme. What is driving discrimination improvement, then, is unclear. However, participants benefit from the exposure to these phones in lexical contexts indicating that consistent mapping of the phones to particular categories in their transcription responses is not critical to success in low-level perceptual tasks like discrimination. Follow-up research is required to assess the relationships between improvements in transcription, discrimination, and identification.

Another point of interest is the differences in listener’s accuracy at judging productions of Talkers 1 and 2. Specifically, in both Experiment 1 and 2, participants were better at transcribing tokens from Talker 2 than tokens from Talker 1. Additionally, the post-test improvement seen in Experiment 1 was really driven by improvement on Talker 2’s tokens, whereas in Experiment 2, participants were slightly more accurate overall in discriminating Talker 2’s tokens. Past studies have found similar differences (Pruitt et al., 2006), and what leads to these differences is not clear.

## Conclusions

In two experiments, we showed that exposure to nonnative phones in dialectical variants of a listener’s native language

can help the listener learn to discriminate those phones. This shows that lexical information can support phonetic learning even in the very early stages of simply learning to discriminate phones. Future work should focus on how stable this learning is, what information is made available to listeners in the dialectical variants, and whether the improved discrimination performance supports further learning of categorization and production.

**Acknowledgments** NV was partially supported by NICHD Grant # NICHD P01 HD001004-96 awarded to Haskins Laboratories.

## References

- Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. *The development of speech perception: The transition from speech sounds to spoken words*, 167(224), 233–277.
- Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener’s native phonological system. *The Journal of the Acoustical Society of America*, 109(2), 775–794.
- Bradlow, A.R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, 106(2), 707–729.
- 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*. Timonium, MD: York Press.
- Kraljic, T. & Samuel, A.G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology*, 51, 141–178.
- Kraljic, T. & Samuel, A.G. (2006). Generalization in perceptual learning for speech. *Psychonomic Bulletin & Review*, 13, 262–268.
- Kraljic, T. & Samuel, A.G. (2007). Perceptual adjustments to multiple speakers. *Journal of Memory and Language*, 56, 1–15.
- Norris, D., McQueen, J.M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, 47, 204 – 238.
- Olmstead, A. J. & Viswanathan, N. (2016, May). Training non-native contrast discrimination using dialectical variants. Poster presented at the International Meeting of the Psychonomic Society.
- Pruitt, J. S., Akahane-Yamada, R., & Strange, W. (1998). Perceptual assimilation of Hindi dental and retroflex stop-consonants by native English and Japanese speakers. *The Journal of the Acoustical Society of America*, 103(5), 3091–3091.
- Pruitt, J.S., Jenkins, J.J., & Strange, W. (2006). Training the perception of Hindi dental and retroflex stops by native speakers of American English and Japanese. *Journal of the Acoustical Society of America*, 119, 1684 – 1696.
- Reinisch, E., Weber, A., & Mitterer, H. (2013). Listeners retune phoneme categories across languages. *Journal of Experimental Psychology: Human Perception and Performance*, 39 (1), 75 –86.
- Sailaja, Pingali (2009). Indian English. Edinburgh University Press: Edinburgh.
- Sjerps, M. J. & McQueen, J.M. (2010). The bounds of flexibility in speech. *Journal of Experimental Psychology: Human Perception and Performance*, 36 (1), 195 – 211.
- Tees, R.C. & Werker, J.F. (1984). Perceptual flexibility: Maintenance or recovery of the ability to discriminate non-native speech sounds. *Canadian Journal of Psychology*, 38, 579 – 590.
- Werker, J.F., Gilbert, J.H.V., Humphrey, K., & Tees, R.C. (1981). Developmental aspects of cross-language speech perception. *Child Development*, 52, 349 – 355.

- Werker, J.F. & Logan, J.S. (1985). Cross-language evidence for three factors in speech perception. *Perception & Psychophysics*, 37, 35 – 44.
- Werker, J.F. & Tees, R.C. (1984). Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America*, 75, 1866 – 1877.
- Xie, X., Theodore, R. M., & Myers, E. B. (2017). More than a boundary shift: Perceptual adaptation to foreign-accented speech reshapes the internal structure of phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance*, 43(1), 206.