Article

second language research

Second Language Research
29(4) 391–411

DOI: 10.1177/0267658313491763
slr.sagepub.com
SAGE

Category labels induce boundary-dependent perceptual warping in learned speech categories

Kristen Swan
Birkbeck, University of London, UK

Emily Myers
University of Connecticut, USA

Abstract

Adults tend to perceive speech sounds from their native language as members of distinct and stable categories; however, they fail to perceive differences between many non-native speech sounds without a great deal of training. The present study investigates the effects of categorization training on adults' ability to discriminate non-native phonetic contrasts. It was hypothesized that only individuals who successfully learned the appropriate categories would show selective improvements in discriminating between-category contrasts. Participants were trained to categorize progressively narrow phonetic contrasts across one of two non-native boundaries, with discrimination pre- and post-tests completed to measure the effects of training on participants' perceptual sensitivity. Results suggest that changes in adults' ability to discriminate a non-native contrast depend on their successful learning of the relevant category structure. Furthermore, post-training identification functions show that changes in perceptual categories specifically correspond to their relative placement of the category boundary. Taken together, these results indicate that learning to assign category labels to a non-native speech continuum is sufficient to induce discontinuous perception of between- versus within-category contrasts.

Keywords

categorization, second language acquisition, speech perception

I Introduction

Categorical perception – the tendency to perceive stimuli that are spaced equally on a physical continuum as more or less distinctive depending on their category membership

Corresponding author:
Emily Myers, University of Connecticut, 850 Bolton Road Unit 1085, Storrs, CT 06269, USA.
Email: emily.myers@uconn.edu

– is especially characteristic of adults' native-language speech perception. Adults tend to discriminate syllables across a category boundary with much greater accuracy and faster reaction time than syllables within a given category (Liberman et al., 1957; Pisoni and Tash, 1974). Though the sources of these perceptual discontinuities are not entirely known, it is likely that limitations of the human perceptual system affect the relative discrimination of between and within-category contrasts (Sharma et al., 2000), with certain native-language phonetic contrasts falling in regions of acoustic space that the human auditory system can easily perceive. Moreover, properties of the acoustic signal itself may provide leverage for the perception of phonetic contrasts. Stevens' Quantal Theory posits that invariant properties of the acoustic input lead to a perceptual advantage for stimuli that fall into certain acoustic categories (see Stevens, 1972). While properties of the acoustic stimuli and of the human auditory system may provide an initial advantage for perception of certain acoustic contrasts, these bottom-up factors cannot explain category-specific changes in perceptual sensitivity that result from learning. In particular, given that phonetic boundaries along the same acoustic continuum differ between languages (Lisker and Abramson, 1964), the operational category boundary must be a learned property of the maturing language system.

What remains unknown is precisely how this discontinuous perception around phonetic categories arises as individuals acquire new phonetic categories. Some suggest that structural properties of the input, namely the statistical distribution of tokens in acoustic space, may help infants and adults alike attune to regions that underline category-level information. This presumably leads to perceptual warping around phonetic category centers and bootstraps the infant or adult learner into recognizing distinctions that correspond to a meaningful speech sound category (Hayes-Harb, 2007; Kuhl et al., 1992; Maye et al., 2002). While statistical information may be present in the input, at the same time, learners are also exposed to top-down information, such as category labels, phoneme–grapheme correspondences, and minimal pairs, all of which may influence the way attention is distributed among category-relevant acoustic features (Francis and Nusbaum, 2002; Yeung and Werker, 2009). For example, second language learners often receive explicit instruction about the nature of speech sound categories in the new language, and those sounds often (although not always) map onto distinct graphemes. As such, it is possible that the top-down imposition of explicit category knowledge provides a sufficient condition for the warping of perceptual space along a phonetic continuum.

Studies of adults' non-native phonetic category learning have demonstrated the inherent difficulty of acquiring sensitivity to the relevant distinctions of a new language; but with sufficient training, adults can learn to categorize non-native contrasts and, given the 'right' training conditions, may also generalize to new tokens and speakers (Bradlow and Pisoni, 1999; Lively et al., 1993; Logan et al., 1991). However, learning the location of a category boundary does not entail that perceptual space has been warped around these categories, as seen in mature phonetic perception. Fewer studies have investigated the effect of categorization training on subsequent discrimination of sound contrasts, and those that have offer an incomplete picture. Guenther et al. (1999) trained participants to categorize tokens from two novel non-speech categories, and while their results showed decreases in discrimination of within-category contrasts, the concurrent effect of training on between-category discrimination was not evaluated. When McCandliss et al. (2002)

trained Japanese participants to categorize the English /l/–/r/ contrast – a distinction that is difficult for Japanese listeners to perceive – they observed better categorization of both the training continuum ('load'–'road') and an untrained continuum ('lock'–'rock'). However, the resulting peak in discrimination sensitivity was measured near the center of the /l/–/r/ continuum, and the relationship between this peak and the location of the category boundary was not explicitly assessed. Similar results were reported by Golestani and Zatorre (2009), who trained native speakers of English to categorize tokens from a synthetic voiced stop continuum comprising a dental versus retroflex place of articulation contrast. As in McCandliss et al. (2002), categorization training resulted in a peak in discrimination sensitivity near the center of the phonetic continuum, but the peak was limited to the subset of participants who showed successful learning of the non-native contrast, as measured by a positive identification slope after training.

While this research suggests that successful learning is accompanied by changes in perceptual sensitivity, it remains unclear whether these changes result from a heightened acuity to properties across the phonetic continuum due to prolonged input, or whether they are specific to participants' learning and localization of a new phonetic category boundary. Since all of the studies mentioned above examined learning of a symmetric phonetic category distribution (i.e. one in which the trained boundary was at the center of the acoustic distribution), it is not possible to deduce whether the observed discrimination peaks are specific to the learned categories or rather the result of increased perceptual sensitivity near the center of the training continuum. Further, it is unclear how category-relevant training will concurrently affect sensitivity to tokens within and between the corresponding phonetic categories, and to what extent the relative placement or difficulty of the phonetic contrast itself will impact individuals' ability to acquire such sensitivity. Previous investigations have focused on a single phonetic contrast and a single training group and, as such, have lacked the necessary comparisons to address these questions.

The current study investigates the relationship between explicit category training and perceptual sensitivity in the phonetic domain, considering how the learning of a particular category scheme influences participants' discrimination of non-native contrasts. Across two experiments, participants were exposed to a nine-point continuum of synthetic speech sounds, comprising dental, retroflex and velar CV syllables (Stevens and Blumstein, 1975). This three-way phonetic distinction, found in languages such as Hindi and Malayalam, is not native to American English, and retroflex tokens have been found to be confusable with alveolar or dental tokens for native English-speaking listeners (Polka, 1991; Tees and Werker, 1984). In Experiment 1, two groups of participants were trained to categorize these tokens according to either the dental/retroflex or retroflex/velar boundaries using a two-alternative forced-choice task with feedback and a perceptual fading regimen. Participants completed an AX discrimination test before and after training to evaluate the effects of category learning on perceptual sensitivity to non-native contrasts. In Experiment 2, two new groups of participants listened passively to the same distributions of tokens used during training in Experiment 1. The same AX discrimination tests were administered before and after listening, in order to control for the possibility that mere exposure, as opposed to explicit training, could be sufficient to elicit category-specific changes in perceptual sensitivity.

Consistent with previous results (Golestani and Zatorre, 2009; McCandliss et al., 2002), we hypothesize that explicit categorization training will produce discontinuities in perception of stimuli taken from a non-native speech continuum. Moreover, we hypothesize that the emergence of categorical perception, that is, better discrimination of between-category compared to within-category contrasts, will be linked to successful category learning. We expect this dependency to persist across variations in both individuals' categorization performance and in phonetic contrast difficulty. Specifically, we predict that only participants who successfully acquire a new category scheme – as measured by a shift in their phonetic category boundary post-training – will show a concomitant increase in discrimination sensitivity for contrasts that cross the learned category boundary. Likewise, we expect phonetic contrasts that are easier to learn will elicit a greater change in discrimination sensitivity.

II Experiment 1: Categorization training

1 Methods and materials

a Participants. Sixty-four adults (18 males, 18–45 years old) were recruited from the Brown University community to participate in this experiment. All were monolingual native speakers of American English and reported no hearing deficits. Informed consent was obtained according to the guidelines approved by the Human Subjects Committees of Brown University, and participants were compensated for participation.

Seven participants were eliminated for failing to discriminate the endpoints of the continuum at pre-test, three for failing to reach the categorization training criterion, and two for equipment failure. The resulting sample of 52 adults (13 males) was randomly assigned to two experimental groups of 26 each.

b Stimuli. Speech syllables were taken from a synthetic nine-point continuum ranging from a dental /d̪a/, to retroflex /ɖa/, to velar /ga/ places of articulation (for stimulus details, see Stevens and Blumstein, 1975 and Table 1). These stimuli were chosen because of the detailed data provided by Stevens and Blumstein (1975) in their seminal work on both the acoustic properties and the cross-linguistic perception of the synthetic continuum. The three-way contrast differs primarily in the frequency of the burst at the onset of the syllable and the trajectory of the transition of the third formant from onset to steady-state frequency. While Stevens and Blumstein (1975) found individual differences in the placement of boundaries between categories among native speakers of languages containing this phonemic contrast, the first two points were classified as dental sounds with above 90% consistency, the third token fell near the category boundary, the middle three to four points were categorized as retroflex sounds, and the remaining two to three points as velar sounds. For native English listeners, the endpoints (dental vs. velar) are a near-native phonetic contrast, while the dental vs. retroflex contrast has been reported as difficult to discriminate (Polka, 1991; Tees and Werker, 1984). Although no published data to our knowledge refers to the ability of English listeners to discriminate velar and retroflex tokens, several studies (Polka, 1991; Tees and Werker, 1984) suggest that retroflex and dental contrasts are both assimilated to the native-language alveolar

Table 1. Stimulus details.

Phonetic	Stimulus	Burst frequency	F1	F2	F3
Dental	1	4500	(440) 655	(1650) 1185	(3080) 2885
Dental	**2**	**4200**	**(440) 655**	**(1650) 1185**	**(2913) 2885**
Dental	3	3900	(440) 655	(1650) 1185	(2746) 2885
Retroflex	4	3600	(440) 655	(1650) 1185	(2580) 2885
Retroflex	**5**	**3300**	**(440) 655**	**(1650) 1185**	**(2413) 2885**
Retroflex	6	2900	(440) 655	(1650) 1185	(2246) 2885
Velar	7	2600	(440) 655	(1650) 1185	(2080) 2885
Velar	**8**	**2300**	**(440) 655**	**(1650) 1185**	**(2080) 2885**
Velar	9	2100	(440) 655	(1650) 1185	(2080) 2885

Notes. All frequency values are given in Hz. Parenthetical values indicate the onset frequency of a formant when it differs from the steady state frequency. F4 and F5 were synthesized with steady-state values of 3600 Hz and 4500 Hz, respectively, with no transition. All stimuli were used in categorization training. Bolded stimuli are those that were used in discrimination tests.

category. If this is the case, then the retroflex vs. velar contrast should be relatively easy to discriminate, as it is thought to resemble the perception of a native-language alveolar vs. velar contrast. However, given that the retroflex tokens are situated more closely to the velar tokens in acoustic space, the discrimination of this contrast might nonetheless be more difficult than the dental vs. retroflex contrast.

A baseline identification function for the nine speech sounds was obtained from 14 native-English-speaking adults (6 males). These 'baseline' participants listened to 10 tokens of each point on the continuum, for a total of 90 randomized trials, and identified each token as either 'DA' or 'GA' by pushing a corresponding button. The resulting mean category boundary was located at 4.98, or nearly at stimulus 5, suggesting that without training or exposure to these tokens, native English listeners are inclined to place the category boundary essentially in the middle of the phonetic continuum.

c Procedure. In a single 45-minute testing session, participants completed the following tasks: (1) a discrimination pre-test, (2) six blocks of categorization training, (3) an identification task, and (4) a discrimination post-test. Each task was immediately followed by the next, with no deliberate gaps between training and testing. Participants were seated in a sound-proof booth and communicated via an intercom to an experimenter in the neighboring room. Participants listened to speech sounds over headphones and indicated their responses by pushing the appropriate buttons on the button box as quickly and accurately as possible.

Discrimination: The pre- and post-tests consisted of an AX discrimination task, which was used to establish participants' sensitivity to tokens from the nine-point continuum. Participants heard 60 pairs of syllables (separated by a 250 ms inter-stimulus interval) from the three category centers, Dental (point 2), Retroflex (point 5), and Velar (point 8). Thirty pairs were identical stimuli (e.g. 2 vs. 2), and 30 pairs were contrasts (e.g. 5 vs. 8). Each participant judged whether the stimuli sounded the same or different from one another by pushing a corresponding button. Responses to the AX discrimination pre-test

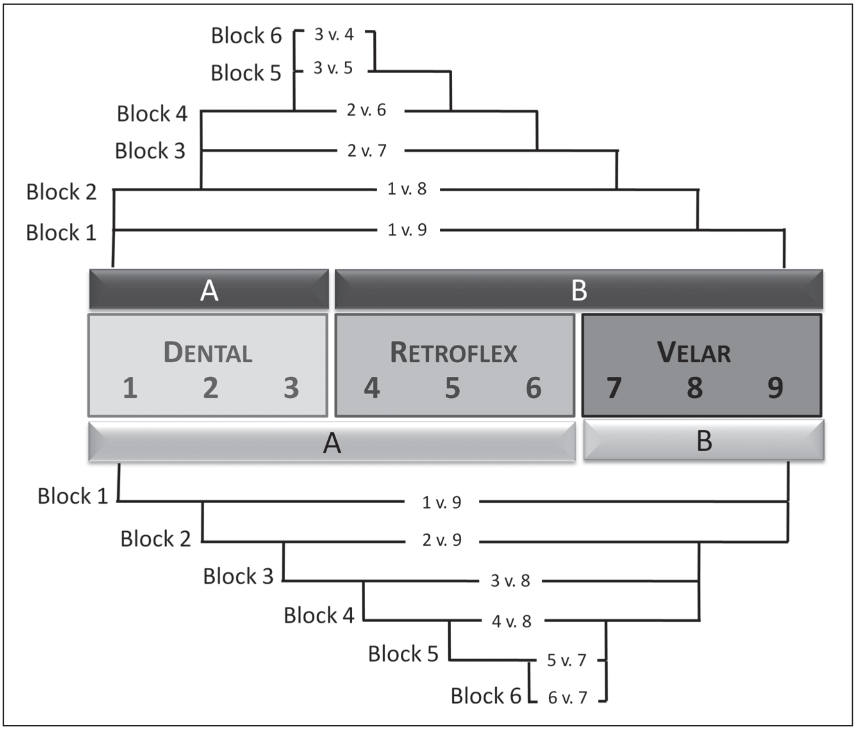

Figure 1. Schematic of perceptual fading technique.
Notes. Nine-point continuum shown in center, appropriate category labels denoted with 'A' and 'B'. Dental/Retroflex group shown with upper pattern and Retroflex/Velar group with lower pattern.

and post-test were analysed using *d′*, a sensitivity statistic used in signal detection theory (MacMillan and Creelman, 2005);[1] *d′* takes into account both correct discriminations ('Hits') and incorrect discriminations of identical tokens ('False Alarms'), and is computed as follows: $d' = Z_{score}(\text{Hits}) - Z_{score}(\text{False Alarms})$.

Category training: Participants were randomly assigned to two groups, Dental/Retroflex and Retroflex/Velar, according to the placement of their to-be-learned category boundary. They listened to six blocks of 40 single-token trials, beginning with the endpoints of the continuum and stepping inward on each subsequent block to present progressively narrower phonetic contrasts. The Dental/Retroflex Trainees categorized phonetic contrasts converging on the boundary between points 3 and 4, such that points 1–3 made up one category and points 4–9 made up the other category, whereas the Retroflex/Velar Trainees categorized phonetic contrasts converging on the boundary between points 6 and 7, such that points 1–6 constituted one category and points 7–9 constituted the other category. This training progression is illustrated in Figure 1. Critically, neither group was exposed within a given block to the stimulus contrasts that were tested in the discrimination pre- and post-tests. Participants were familiarized with the labels 'Category A' and 'Category B', and after each token was played, they pressed a button to indicate in which category the sound belonged. Auditory

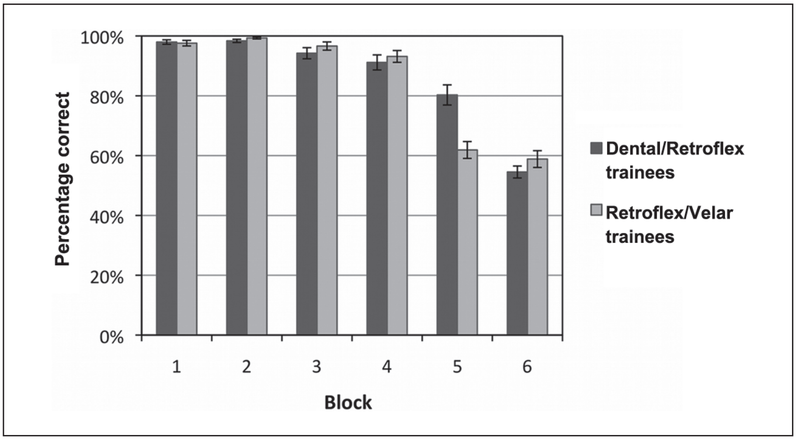

Figure 2. Percentage of correct responses during categorization over six progressively more difficult training blocks in Experiment 1.

feedback (i.e. unique sounds for a correct or incorrect response) was given immediately after each response.

Identification: Following training, participants identified stimuli as either 'Category A' or 'Category B' by pressing the corresponding button. Ten tokens of each sound on the nine-point continuum were presented in a random order for a total of 90 trials, and no feedback was given.

2 Results

a Categorization training results. Categorization training became more difficult over the six blocks as phonetic contrasts narrowed; thus, for inclusion the study, participants had to complete at least four out of six blocks before falling below chance, which was operationalized as 60% accuracy. There were no significant differences between training groups in overall accuracy ($M = 85.33\%$, $SE = 0.78\%$; $t(50) = 0.93$, $p < 0.356$) or in the number of categorization blocks completed above this criterion ($M = 4.94$, $SE = 0.10$; $t(43.2) = 0.55$, $p < 0.585$), which suggests that groups achieved similar success during the training task (see Figure 2).

Responses to each of the nine stimulus tokens on the identification task following training were plotted alongside the function provided by 14 untrained baseline participants, shown in Figure 3. Responses were converted to z scores, and the point at which each participant responded at chance (i.e. a z-score of 0) was calculated as the participant's category boundary. Following training, the groups demonstrated different placements of the category boundary: the Retroflex/Velar Trainees had a mean boundary of 5.17 ($SE = 0.14$), whereas the Dental/Retroflex Trainees had a mean boundary of 3.93 ($SE = 0.15$), closer to the dental end of the continuum. The baseline category boundary, provided by untrained participants, was 4.98 ($SE = 0.12$).

A univariate ANOVA comparing the baseline and experimental groups' post-training boundaries showed a significant main effect of Group ($F(2,77) = 24.08$, $\eta^2_{partial} = .385$,

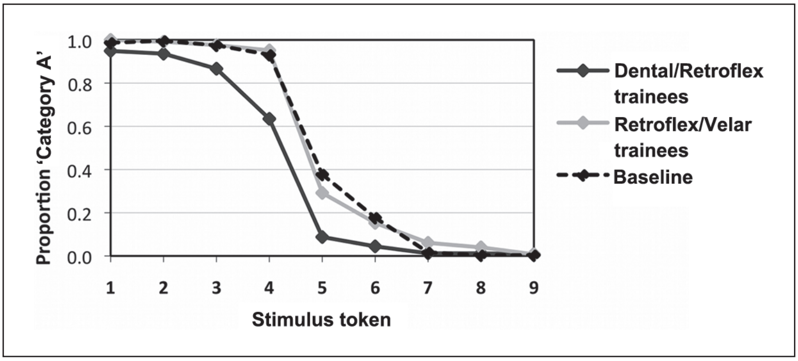

Figure 3. Percentage of 'Category A' responses to the nine stimulus tokens on the identification task without feedback in Experiment 1.

MSE = 11.7, $p < 0.001$). Post-hoc tests (Bonferroni-corrected at a level of $p < 0.05$, which yielded a functional threshold of $p < 0.017$) revealed that while the Dental/Retroflex Trainees' category boundary was significantly different from both the baseline ($t(52) = 5.602$, $p < 0.001$) and the Retroflex/Velar Trainees' boundaries ($t(50) = 6.154$, $p < 0.001$), the Retroflex/Velar Trainees' boundary did not differ from baseline ($t(52) = 1.016$, $p = 0.314$).

b Discrimination results. Responses to the AX discrimination pre-test and post-test were converted to *d′* scores for the three contrasting pairs (Dental vs. Retroflex, Retroflex vs. Velar, and Dental vs. Velar), displayed in Table 2. Of particular interest is the mean change in *d′* from pre-test to post-test, displayed in Figure 4, as it reflects any differences in the effect of categorization training between the two training groups. It was predicted that, for each group, contrasts crossing the learned category boundary would show increases in discrimination sensitivity (larger *d′*), and that contrasts within the learned category would show either a decrease in sensitivity or no change.

A three-way repeated measures ANOVA, including factors of Group (Dental/Retroflex and Retroflex/Velar training groups), Contrast (Dental vs. Retroflex and Retroflex vs. Velar discrimination pairs), and Training (Pre/Post-test), revealed significant main effects of Training ($F(1,50) = 5.38$, $\eta^2_{partial} = 0.097$ $p < 0.025$), and Contrast ($F(1,50) = 25.35$, $\eta^2_{partial} = 0.336$ $p < 0.001$), but no significant main effect of Group. Moreover, there was a significant Group × Training interaction ($F(1,50) = 5.64$, $\eta^2_{partial} = 0.101$ $p < 0.021$), a significant Contrast × Training interaction ($F(1,50) = 16.49$, $\eta^2_{partial} = 0.248$, $p < 0.001$), and a significant three-way interaction of Group × Training × Contrast type ($F(1,50) = 7.10$, $\eta^2_{partial} = 0.124$ $p < 0.010$), reflecting the finding that groups treated the Dental vs. Retroflex and Retroflex vs. Velar contrasts differently depending on how they had been trained to categorize these tokens. This interaction was examined for simple effects within each group using *t*-tests to compare pre-test and post-test *d′* scores for each contrast. These *t*-tests revealed a significant increase in sensitivity to the Dental vs. Retroflex contrast in the Dental/Retroflex group, $t(50) = 4.398$, $p < 0.001$. No other simple effects were found ($p > 0.231$), suggesting that this particular instance of learning was driving the three-way interaction.[2]

Table 2. Experiment 1: AX discrimination results: Mean *d′* score (standard error in parentheses).

	Dental/retroflex training group			Retroflex/velar training group		
	Pre-test	Post-test	Change [post-test–pre-test]	Pre-test	Post-test	Change [post-test–pre-test]
Dental vs. retroflex	1.422 (0.260)	2.916 (0.219)	1.494 (0.248)	1.938 (0.291)	2.119 (0.286)	0.180 (0.206)
Retroflex vs. velar	0.771 (0.184)	0.458 (0.181)	–0.313 (0.238)	1.094 (0.278)	0.899 (0.291)	–0.195 (0.333)
Dental vs. velar	2.915 (0.287)	4.080 (0.121)	1.165 (0.279)	3.535 (0.239)	4.314 (0.098)	0.778 (0.237)

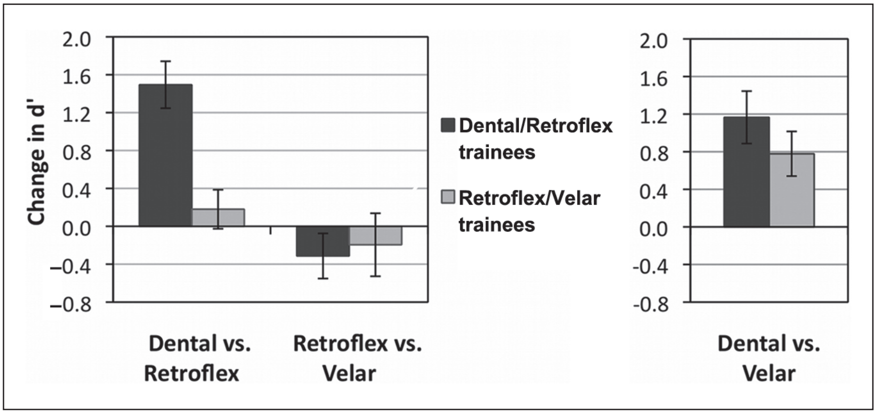

Figure 4. Mean change in *d'* scores on AX discrimination task due to categorization training [post-test]–[pre-test] in Experiment 1. Left: Contrasts of interest included in ANOVA. Right: Near-native contrast shown for comparison.

The Dental vs. Velar contrast, a native English contrast that spans six points on the continuum and should be relatively easy to distinguish, was examined separately in a repeated measures ANOVA with factors of Group and Training (Pre/Post-test). A significant main effect of Training, $F(1,50) = 28.159$, $p < 0.001$, and a marginal effect of Group, $F(1,50) = 3.776$, $p < 0.058$, indicated improvement from pre-test to post-test in both groups and slightly higher sensitivity in the Retroflex/Velar group at both pre-test and post-test. No interaction emerged, suggesting that the two groups improved equally on this wider contrast as a result of categorization training.

c Correlations between category learning and discrimination sensitivity. Regression analyses were performed separately for each group to examine the relationship between persistent category learning and changes in perceptual sensitivity. Each regression included post-training category boundary location and changes in sensitivity (d') to the Dental vs. Retroflex and Retroflex vs. Velar contrasts as regressors. Results of the regression analyses revealed a significant positive correlation between category boundary and change in sensitivity to the Retroflex vs. Velar contrast within the Retroflex/Velar training group, $r = 0.389$, $p < 0.049$. This finding, illustrated in Figure 5, suggests that, although the Retroflex vs. Velar contrast may have been more difficult to learn, participants who successfully learned this boundary showed concomitant improvements on discrimination of the between-category contrast. However, no relationship was found between category boundary location and change in sensitivity to the Dental vs. Retroflex contrast within the Retroflex/Velar group alone. Moreover, no significant correlations between category boundary and contrast sensitivity were found within the Dental/Retroflex group for either contrast, $p > 0.492$.

Results of Experiment 1 suggest that explicit training with category labels can lead to significant category-specific changes in perceptual sensitivity. However, as participants in Experiment 1 were both listening to and categorizing the distributions of speech

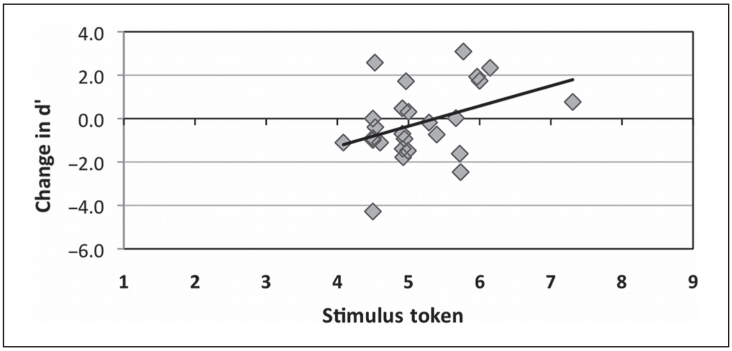

Figure 5. Correlation between post-training category boundary and perceptual sensitivity (change in *d′*) to the Retroflex vs. Velar contrast within the Retroflex/Velar training group in Experiment 1.
Notes. $r = 0.389$, $p < 0.049$.

tokens, it is impossible to determine whether these category-specific changes were elicited by training with the corresponding labels or simply through exposure to the contrasts. To control for this possibility, Experiment 2 was conducted with a new group of participants, who listened to the same distributions of tokens as in Experiment 1 without categorizing or assigning labels. If exposure alone were sufficient to produce discontinuities in perceptual sensitivity, then one would expect to observe similar changes in discrimination from pre-test to post-test as seen in Experiment 1.

III Experiment 2: Exposure only

1 Methods and materials

a Participants. Twenty-eight adults (12 males, 18–45 years old) were recruited from the Brown University community, in the same manner detailed in Experiment 1. Informed consent was obtained, and participants were compensated for participation.

b Stimuli. The speech syllables used in Experiment 2 were identical to those presented in Experiment 1, comprising a synthetic nine-point continuum from a dental /d̪a/, to retroflex /ɖa/, to velar /ga/ places of articulation (see above; for stimulus details, see also Stevens and Blumstein, 1975).

c Procedure. In a single 30-minute testing session without breaks, participants completed the following tasks: (1) a discrimination pre-test, (2) six blocks of exposure, and (3) a discrimination post-test. Participants were seated in a sound-proof booth and listened to speech sounds over headphones. They indicated their responses by pushing the appropriate buttons on the button box as quickly and accurately as possible.

d Discrimination. The pre- and post-tests consisted of the same AX discrimination task used in Experiment 1 to establish participants' sensitivity before and after exposure to tokens from the nine-point continuum. Participants heard 60 pairs of syllables from the three category centers: Dental (point 2), Retroflex (point 5), and Velar (point 8). They judged whether the stimuli sounded the same or different from one another by pushing the corresponding button.

c Exposure. Participants were randomly assigned to two groups, Dental/Retroflex and Retroflex/Velar, and they listened to the same six blocks of 40 single-token trials that participants had categorized in Experiment 1. That is, the Dental/Retroflex group heard phonetic contrasts converging on the boundary between points 3 and 4, while the Retroflex/Velar group heard contrasts converging on the boundary between points 6 and 7. As in Experiment 1, participants were presented with progressively narrow phonetic contrasts on each subsequent block, stepping inward from the endpoints of the continuum toward each group's assigned boundary, as illustrated in Figure 1. Importantly, participants were not introduced to any category labels nor trained to group the sounds into categories; instead, participants simply listened and responded by pushing a button when they heard each token.

2 Results

a Discrimination results. Responses to the AX discrimination pre-test and post-test were converted to d' scores for the three contrasting pairs (Dental vs. Retroflex, Retroflex vs. Velar, and Dental vs. Velar), displayed in Table 3. Of particular interest is the mean change in d' from pre-test to post-test, displayed in Figure 6, as it reflects any differences in the effect of exposure between the two exposure groups.

A three-way repeated measures ANOVA, including factors of Group (Dental/Retroflex and Retroflex/Velar), Contrast (Dental vs. Retroflex and Retroflex vs. Velar only), and Exposure (Pre/Post-test) revealed a significant main effect of Exposure ($F(1,26) = 4.18$, $\eta^2_{\text{partial}} = 0.138$, $p < 0.051$) and a significant main effect of Contrast ($F(1,26) = 7.57$, $\eta^2_{partial} = 0.226$, $p < 0.011$), suggesting that participants improved in discrimination from pre-test to post-test and were more sensitive to one contrast (the Dental vs. Retroflex contrast) than the other. In addition, a significant main effect of Group was apparent ($F(1,26) = 5.80$, $\eta^2_{partial} = 0.640$, $p < 0.023$), suggesting that higher discrimination scores were exhibited by one group compared to the other. Importantly, no significant interactions emerged, which indicated that the two groups had comparable changes in performance as a function of exposure to the speech sounds.

These results were compared to Experiment 1 in a 2(Experiment) × 2(Group) × 2(Contrast) × 2(Pre/Post-test) repeated measures ANOVA. Results showed significant main effects of Group, $F(1,76) = 6.14$, $p < 0.015$, of Contrast, $F(1,76) = 27.12$, $p < 0.001$, and of Pre/Post-test, $F(1,76) = 9.45$, $p < 0.003$. A significant Contrast × Pre/Post-test interaction emerged, $F(1,76) = 5.03$, $p < 0.028$, due to a greater increase in sensitivity to the Dental vs. Retroflex contrast than to the Retroflex vs. Velar contrast. A significant three-way interaction of Experiment × Contrast × Pre/Post-test also emerged, $F(1,76) = 5.74$, $p < 0.019$, driven by a greater increase in sensitivity to the Dental vs. Retroflex contrast in Experiment 1 (Training) than in Experiment 2 (Exposure). Results showed no main effect of Experiment or interaction of Experiment × Contrast, likely due to the

Table 3. Experiment 2 AX Discrimination Results: Mean *d*' score (standard error in parentheses).

	Dental/retroflex exposure group			Retroflex/velar exposure group		
	Pre-test	Post-test	Change [post-test–pre-test]	Pre-test	Post-test	Change [post-test–pre-test]
Dental vs. retroflex	1.414 (0.418)	1.824 (0.441)	0.410 (0.376)	2.263 (0.288)	2.586 (0.424)	0.323 (0.396)
Retroflex vs. velar	0.658 (0.292)	1.011 (0.317)	0.354 (0.337)	1.009 (0.298)	1.460 (0.405)	0.451 (0.443)
Dental vs. velar	2.891 (0.485)	3.519 (0.376)	0.628 (0.482)	3.406 (0.374)	4.550 (0.070)	1.144 (0.368)

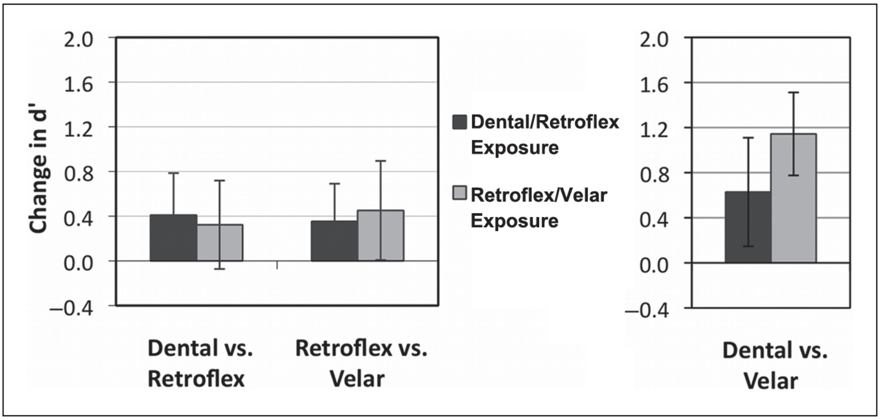

Figure 6. Mean change in *d′* scores on AX discrimination task due to exposure only [post-test]–[pre-test] in Experiment 2. Left: Contrasts of interest included in ANOVA. Right: Near-native contrast shown for comparison.

slight increases in sensitivity with exposure in Experiment 2 balancing out the increases in sensitivity in one training group and not the other in Experiment 1. Further, no significant four-way interaction emerged, suggesting that Group did not play a strong role across experiments in the differential changes in sensitivity to the two contrasts. However, as the individual analyses showed, participants in Experiment 1 were differentially sensitive to the Dental/Retroflex and Retroflex/Velar contrasts depending on their training group, while participants in Experiment 2 did not show differential changes in sensitivity depending on exposure group.

Sensitivity to the Dental vs. Velar contrast, a native English between-category contrast, was examined separately. A repeated measures ANOVA, including factors of Group and Pre/Post-Test, showed a significant main effect of Pre/Post-Test, $F(1,26) = 8.538$, $p < 0.007$, indicating improvement from pre-test to post-test in both groups. No main effect of Group or interaction emerged, suggesting that both groups improved equally on this contrast as a result of exposure. This result was compared to Experiment 1 in a three-way ANOVA, including factors of Experiment (Training or Exposure), Group (Dental/Retroflex or Retroflex/Velar), and Pre/Post-test. Results revealed a main effect of Pre/Post-test, $F(1,76) = 30.974$, $p < 0.001$, and a main effect of Group, $F(1,76) = 8.014$, $p < 0.006$; however, no main effect of Experiment and no interactions emerged. Such findings indicate that changes in participants' sensitivities to the Dental vs. Velar contrast due to categorization training were not significantly different from changes in sensitivities due to exposure; all types of experience led to equal improvement on the Dental vs. Velar contrast.

3 *Discussion*

This study provides evidence that training with category labels affects perceptual sensitivity to non-native between- and within-category speech contrasts in adults. In

Experiment 1, explicit categorization training led to between-group differences in participants' discrimination of the non-native sounds as revealed by a Group × Contrast interaction. However, in Experiment 2, passive listening to the same frequency and distribution of speech tokens did not lead to group differences in sensitivity for between-category and within-category contrasts. As the statistical distribution of tokens heard was kept constant, top-down knowledge of a category structure must have induced changes in participants' sensitivities in a way that exposure alone could not.

In Experiment 1, the Dental/Retroflex training group evinced a significantly different category boundary placement following training compared to that of the baseline group, while the Retroflex/Velar training group showed no such difference. Furthermore, the Dental/Retroflex training group showed improved discrimination of the Dental vs. Retroflex contrast and worse discrimination of the Retroflex vs. Velar contrast, while the Retroflex/Velar group exhibited little change in sensitivity to either contrast. While at the group-level the Retroflex/Velar group showed no evidence of a boundary shift, within this group, individuals who showed a greater boundary shift also showed greater increases in discrimination sensitivity to the Retroflex vs. Velar contrast. While these findings are consistent with results reported by McCandliss et al. (2002) and Golestani and Zatorre (2009), what is novel about the present results is that changes in perceptual sensitivity were linked both to learning of a particular phonetic contrast, rather than the continuum at large, and to the success of the individual in acquiring a new category boundary. Implications of these findings are discussed below.

a Contrast difficulty and individual variation in category learning. Phonetic contrasts differ in terms of the ease with which they are acquired in infancy (Best et al., 1988) as well as in adult second language learning (e.g. Polka, 1991). In the current study, the Dental vs. Retroflex and Retroflex vs. Velar contrasts were not learned with the same proficiency. Although both training groups completed categorization training with identical success, only one group (Dental/Retroflex) showed an impact of this training on their perceptual structure at the group-level. Several explanations may account for the relative difficulty of acquiring a given phonetic contrast. Some attribute the inconsistencies of non-native contrast learning to inherent competition between the non-native category structure and the adult's robust native language categories, proposing that successful learning depends heavily on how these structures are integrated (Perceptual Assimilation Model: Best et al., 1988, 2001; Speech Learning Model: Flege, 1988; 1991; for a review, see Kingston, 2003). Broadly speaking, these models posit that non-native contrasts that map onto the same perceptual native language category will be more difficult to discriminate because differences in the non-native phonetic scheme are opposed by the equivalent native language category membership. Since both dental and retroflex sounds have previously been found to assimilate to the /d/ category for English listeners (Tees and Werker, 1984), while Retroflex and Velar sounds should assimilate to two different English categories (/d/ and /g/, respectively), the current results are not predicted by assimilation to native language categories.

A simpler explanation for the asymmetries in learning success of the two training groups may exist; namely, the retroflex and velar exemplars may, at least for this stimulus set, be closer together in acoustic and perceptual space (see Best and Tyler, 2007).

Differences in acoustic proximity have been found to influence phonetic category learning in adulthood (Polka, 1991). However, acoustic differences between the stimuli used for the discrimination task (points 2, 5, and 8) do not demonstrate a sizable asymmetry between the dental vs. retroflex and retroflex vs. velar contrasts.[3] At pre-test, all participant groups showed poorer discrimination of the Retroflex vs. Velar contrast compared to the Dental vs. Retroflex contrast, suggesting that the tokens were unevenly distributed in participants' naïve perceptual space. Moreover, in the data from native speakers of languages containing the three-way phonemic contrast, only points 8 and 9 were reliably identified as velar tokens at above-chance levels (Stevens and Blumstein, 1975); thus the velar tokens selected for this study may not be extremely strong examples of velars for our English-speaking listeners.

It has been demonstrated that individuals vary in their ability to perceive and learn non-native speech contrasts (e.g. Golestani and Zatorre 2009) and even in their ability to perceive differences among native speech contrasts (Surprenant and Watson, 2001). Although we did not explicitly assess differences in these factors, it is likely that both low-level (e.g. perceptual acuity) as well as higher-level resources (e.g. auditory memory, selective auditory attention) play a role (see Crowder, 1982; Francis and Nusbaum, 2002). In the current study, individuals varied within and across training groups in their placement of the phonetic category boundary post-training and in changes in discrimination sensitivity from pre- to post-test. Even in the Retroflex/Velar group, which failed as a whole to show shifts in phonetic category boundary or changes in discrimination sensitivity, there were individuals who were able to learn the trained contrast. The correlation between category boundary placement and change in sensitivity to the between-category contrast among individuals in this group suggests that the re-warping of perceptual space that appears to accompany category learning is dependent on the individual's ability to learn to attend to the trained details of the novel phonetic continuum. At the same time, no correlation between boundary placement and perceptual sensitivity (change in d') was observed within the Dental/Retroflex group alone. The failure to find this relationship may reflect the homogeneity within this group: nearly all participants showed boundary shifts in the predicted direction, and nearly all improved on the Dental/Retroflex contrast from pre- to post-test. Taken together, these results suggest that individual variation and differences in contrast difficulty influence the transfer of category-level information to perceptual space.

While the present results offer substantial evidence of 'acquired distinctiveness' – namely, selective improvement on discrimination performance on between-category contrasts – there was little evidence of 'acquired similarity,' or decreases in discrimination accuracy for tokens which fall within the same category (Liberman et al., 1957). Nonetheless, it should be noted that in both experiments, both groups showed substantial improvements in performance on the Dental vs. Velar contrast. While it may seem odd that participants should be less than at ceiling for a contrast that is essentially a native language contrast, it has been observed that synthetic stimuli are sometimes hard for individuals to decode (Logan et al., 1989; Pisoni and Koen, 1982). Participants were not given practice trials and may have needed time to acclimate to the stimuli, the acoustic environment, and the task demands. All three contrasts were randomized during pre- and post-test, but when both groups were explicitly trained on dental and velar tokens in the first two blocks of categorization training, their responses were highly accurate, suggesting that the endpoints of the continuum were easy to map to category labels. Since identical improvements in performance on this contrast were observed in both participant groups, they likely reflect a general practice effect as participants became more familiar with the stimuli and task. As such, the failure to find any significant changes from pre- to post-test in the Retroflex/Velar group and in the Dental/Retroflex group for the Retroflex vs. Velar contrast may in fact reflect a general trend towards acquired similarity (cf. Guenther et al., 1999) which is concurrently offset by a more general practice effect.

b Category training and the role of attention in perceptual warping. Whatever the underlying reason for the asymmetries reflected in the post-training results, the presence of between-group differences supports the hypothesis that re-warping of perceptual space, as measured by discrimination sensitivity, is contingent on persistent changes in the category structure, and not on successful categorization training, *per se* (cf. Guenther et al., 1999). Moreover, these results suggest that language learning paradigms using categorization may, for successful learners and successfully-learned contrasts, produce discrimination patterns that begin to resemble those of native-language categories. The emergence of discontinuous patterns of sensitivity to within- and between-category contrasts may confer a processing advantage by distributing attention only to acoustic details that are functionally relevant.

Cognitive models of the role of selective attention in categorical perception (cf. Goldstone, 1994; Nosofsky, 1986), referred to as attention-to-dimension (A2D) models, posit that experience with a specific language shifts attention toward dimensions of the acoustic space that signal meaningful (phonetic) differences between stimuli and away from dimensions that are irrelevant. Such shifts in attention lead to dimensional warping, the apparent stretching of the dimension to which attention is focused and shrinking of the dimension from which attention is withdrawn (Francis and Nusbaum, 2002), thereby offering an attentional account for the classic findings of acquired distinctiveness and acquired similarity in phonetic category learning (Liberman et al., 1957; Pisoni and Tash, 1974).

An A2D model can account for a broad range of experimental results obtained in phonetic categorization studies and has implications for first and second language learning. For example, some contrasts may be easier to learn when their differences occur on a dimension that is already attended to in the listener's native structure (Francis and Nusbaum, 2002). In the current study, while one group's training converged toward a particular boundary, the other group trained away from that boundary (and toward their own to-be-learned boundary). Acquired knowledge of these category structures may have led participants to selectively attend to different properties of the stimuli, resulting in between-group differences in post-training contrast sensitivity despite having practiced with the same phonetic continuum. In this way, categorization training may have the effect of amplifying attention to regions of acoustic space that have functional significance for distinguishing between categories. Minimal pairs may represent an example of a natural functional category distinction in language, guiding attention to regions in acoustic space which signal meaningful words; for instance, when infants heard distinct tokens from a non-native contrast paired with two different objects, they discriminated these tokens better than infants who heard those tokens randomly assigned to two objects (Yeung and Werker, 2009). Converging evidence that the phonetic learning process may rely on higher-level (e.g. attentional and executive) processes rather than low-level, perceptual processes comes from a recent study from our lab investigating the neural consequences of phonetic category learning using fMRI (Myers and Swan, 2012). In this study, two groups of participants underwent training on dental, retroflex, and velar contrasts as described in the current study. Neural sensitivity to the learned phonetic category contrasts was found in inferior frontal regions which have been associated with attention and executive function, rather than in superior temporal regions which have been more closely linked to perception of the fine-grained aspects of the phonetic category (Myers, 2007). Models of speech sound acquisition that take into account cognitive/top-down factors may have more explanatory power than those that rely on bottom-up cues from the input alone.

IV Conclusions

While recent efforts to explain the existence of discontinuities in perception have focused on the influence of statistical cues in the input (Hayes-Harb, 2007; Maye et al., 2002; McMurray et al., 2009) we demonstrate an apparent causal link between successful category learning and perceptual sensitivity (Yeung and Werker, 2009). This phenomenon is not unique to speech perception. For example, when Goldstone (1994) trained participants to categorize squares along dimensions of size and brightness, he found increases in sensitivity to local distinctions along the trained dimension and decreases in sensitivity along the untrained dimension. Moreover, color is categorized differently among people of different language backgrounds, and people tend to perceive only those color distinctions for which there are categories in their native language (Kay and Kempton, 1984; Özgen and Davies, 2002; Roberson et al., 2005). Finally, the categorical nature of face perception, which allows a changing set of facial features to be consistently identified as the same person, seems to depend heavily on familiarization with the identities of the faces and can be induced with training with labels (Angeli et al., 2008; Kikutani et al., 2008). The current study provides consistent evidence that non-native phonetic category learning may resemble other domains, as successful category acquisition is a sufficient condition for the warping of perceptual space that accompanies mature phonetic categories.

Acknowledgement

The authors thank Sheila Blumstein for assistance with the stimuli.

Declaration of Conflicting Interest

The author declares that there is no conflict of interest.

Funding

This work was supported by Grant R03 DC009495 from the National Institutes of Health (NIH), National Institute on Deafness and Other Communication Disorders (NIDCD). The content is the responsibility of the authors and does not necessarily represent official views of NIH or NIDCD.

Notes

1. β and $\beta_{normalized}$ were used to measure response bias for each contrast, and calculated as follows: $\beta = -0.5(Z_{\text{score}}(\text{Hits})+Z_{\text{score}}(\text{False Alarms}))$ and $\beta_{normalized} = \beta/d'$ (see Zarate et al., 2012). Bias scores were analysed with repeated measures ANOVA; however, as no effects or interactions emerged, suggesting no differences in response bias, these analyses are not reported or discussed.
2. Although acquired distinctiveness (or better discrimination of between-category tokens) can be accompanied by acquired similarity (or worse discrimination of within-category tokens), statistical analyses show no significant negative changes in d-prime for either contrast in either training group. The 'change in d-prime' measure, as displayed in tables and figures, includes both pre-test and post-test sources of error, so it is possible for a apparent change in sensitivity to be non-significant when both are taken into account in a statistical test (e.g. the change of –0.313 in the Dental/Retroflex training group).
3. Specifically, the difference in the frequency of the burst is 900 Hz for Dental vs. Retroflex, and 1,000 Hz for Retroflex vs. Velar, and the difference in the onset frequency of F3 is 500 Hz for Dental vs. Retroflex and 333 Hz for Retroflex vs. Velar. This latter difference seems to indicate an asymmetry in the same direction as the behavioral results; however, transforming these values into the Mel scale (Stevens et al., 1937) to more closely correspond with human pitch discrimination acuity reveals more similar values (the difference in onset frequency of F3 is 168 Mel for Dental vs. Retroflex, 127 Mel for Retroflex vs. Velar).

References

Angeli A, Davidoff J and Valentine T (2008) Face familiarity, distinctiveness and categorical perception. *Quarterly Journal of Experimental Psychology* 61: 690–707.

Best CT and Tyler MD (2007) Non-native and second-language speech perception: Commonalities and complimentarities. In: Bohn O-S and Munro MJ (eds) *Language experience in second language speech learning: In honor of James Emil Flege*. Philadelphia, PA: John Benjamins, 13–34.

Best CT, McRoberts GW and Goodell E (2001) Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America* 109: 775–94.

Best CT, McRoberts GW and Sithole NM (1988) Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance* 14: 345–60.

Bradlow AR and Pisoni DB (1999) Recognition of spoken words by native and non-native listeners: talker-, listener-, and item-related factors. *Journal of the Acoustical Society of America* 106: 2074–85.

Crowder RG (1982) Decay of auditory memory in vowel discrimination. *Journal of Experimental Psychology. Learning, Memory, and Cognition* 8: 153–62.

Flege JE (1988) The production and perception of foreign language speech sounds. In: Winitz H (ed.) *Human communication and its disorders: A review, 2*. Westport, CT: Ablex, 224–401.

Flege JE (1991) Perception and production: The relevance of phonetic input to L2 phonological learning. In: Ferguson C and Huebner T (eds) *Crosscurrents in second language acquisition and linguistic theories*. Philadelphia, PA: John Benjamins, 249–89.

Francis AL and Nusbaum HC (2002) Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance* 28: 349–66.

Goldstone R (1994) Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General* 123: 178–200.

Golestani N and Zatorre RJ (2009) Individual differences in the acquisition of second language phonology. *Brain and Language* 109: 55–67.

Guenther FH, Husain FT, Cohen MA and Shinn-Cunningham BG (1999) Effects of categorization and discrimination training on auditory perceptual space. *Journal of the Acoustical Society of America* 106: 2900–12.

Hayes-Harb R (2007) Lexical and statistical evidence in the acquisition of second language phonemes. *Second Language Research* 23: 65–97.

Kay P and Kempton W (1984) What is the Sapir–Whorf hypothesis? *American Anthropologist* 86: 65–79.

Kikutani M, Roberson D and Hanley JR (2008) What's in the name? Categorical perception for unfamiliar faces can occur through labeling. *Psychonomic Bulletin and Review* 15: 787–94.

Kingston J (2003) Learning foreign vowels. *Language and Speech* 46: 295–349.

Kuhl PK, Williams KA, Lacerda F, Stevens KN and Lindblom B (1992) Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* 255: 606–08.

Liberman AM, Harris KS, Hoffman HS and Griffith BC (1957) The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology* 54: 358–68.

Lisker L and Abramson A (1964) A cross-language study of voicing in initial stops: acoustical measurements. *Word* 20: 384–422.

Lively SE, Logan JS and Pisoni DB (1993) Training Japanese listeners to identify English /r/ and /l/ II: The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America* 94: 1242–55.

Logan JS, Greene BG and Pisoni DB (1989) Segmental intelligibility of synthetic speech produced by rule. *The Journal of the Acoustical Society of America* 86: 566–81.

Logan JS, Lively SE and Pisoni DB (1991) Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America* 89: 874–86.

MacMillan N and Creelman C (2005) *Detection theory: A user's guide*. 2nd edition. Mahwah, NJ: Lawrence Erlbaum.

Maye J, Werker JF and Gerken LA (2002) Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition* 82: B101–B111.

McCandliss BD, Fiez JA, Protopapas A, Conway M and McClelland JL (2002) Success and failure in teaching the [r]–[l] contrast to Japanese adults: Tests of a Hebbian model of plasticity and stabilization in spoken language perception. *Cognitive and Affective Behavioral Neuroscience* 2: 89–108.

McMurray B, Aslin RN and Toscano JC (2009) Statistical learning of phonetic categories: insights from a computational approach. *Developmental Science* 12: 369–78.

Myers EB (2007) Dissociable effects of phonetic competition and category typicality in a phonetic categorization task: An fMRI investigation. *Neuropsychologia* 45: 1463–73.

Myers EB and Swan KA (2012) Effects of category learning on neural sensitivity to non-native phonetic categories. *Journal of Cognitive Neuroscience* 24: 1695–1708.

Nosofsky RM (1986) Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General* 115: 39–57.

Özgen E and Davies IRL (2002) Acquisition of categorical color perception: A perceptual learning approach to the linguistic relativity hypothesis. *Journal of Experimental Psychology: General* 131: 477–93.

Pisoni DB and Koen E (1982) Some comparisons of intelligibility of synthetic and natural speech at different speech-to-noise ratios. *The Journal of the Acoustical Society of America* 71: S94-S94.

Pisoni DB and Tash J (1974) Reaction times to comparisons within and across phonetic categories. *Perception and Psychophysics* 15: 285–90.

Polka L (1991) Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions. *Journal of the Acoustical Society of America* 89: 2961–77.

Roberson D, Davidoff J, Davies IRL and Shapiro LR (2005) Color categories: Evidence for the cultural relativity hypothesis. *Cognitive Psychology* 50: 378–411.

Sharma A, Marsh CM and Dorman MF (2000) Relationship between N1 evoked potential morphology and the perception of voicing. *Journal of the Acoustical Society of America* 108: 3030–35.

Stevens KN and Blumstein SE (1975) Quantal aspects of consonant production and perception: A study of retroflex stop consonants. *Journal of Phonetics* 3: 215–33.

Stevens SS, Volkman J and Newman EB (1937) A scale for the measurement of the psychological magnitude pitch. *Journal of the Acoustical Society of America* 8: 185–90.

Surprenant AM and Watson CS (2001) Individual differences in the processing of speech and non-speech sounds by normal-hearing listeners. *The Journal of the Acoustical Society of America* 110: 2085–95.

Tees RC and Werker JF (1984) Perceptual flexibility: Maintenance or recovery of the ability to discriminate non-native speech sounds. *Canadian Journal of Psychology* 38: 579–90.

Yeung HH and Werker JF (2009) Learning words' sounds before learning how words sound: 9-month-olds use distinct objects as cues to categorize speech information. *Cognition* 113: 234–43.

Zarate JM, Ritson CR and Poeppel D (2012) Pitch-interval discrimination and musical expertise: Is the semitone a perceptual boundary? *Journal of the Acoustical Society of America* 132: 984–93.