

# Effects of Category Learning on Neural Sensitivity to Non-native Phonetic Categories

Emily B. Myers<sup>1</sup> and Kristen Swan<sup>2</sup>

## Abstract

■ Categorical perception, an increased sensitivity to between-compared with within-category contrasts, is a stable property of native speech perception that emerges as language matures. Although recent research suggests that categorical responses to speech sounds can be found in left prefrontal as well as temporoparietal areas, it is unclear how the neural system develops heightened sensitivity to between-category contrasts. In the current study, two groups of adult participants were trained to categorize speech sounds taken from a dental/retroflex/velar continuum according to two different boundary locations. Behavioral results suggest that for successful learners, categorization training led to

increased discrimination accuracy for between-category contrasts with no concomitant increase for within-category contrasts. Neural responses to the learned category schemes were measured using a short-interval habituation design during fMRI scanning. Whereas both inferior frontal and temporal regions showed sensitivity to phonetic contrasts sampled from the continuum, only the bilateral middle frontal gyri exhibited a pattern consistent with encoding of the learned category scheme. Taken together, these results support a view in which top-down information about category membership may reshape perceptual sensitivities via attention or executive mechanisms in the frontal lobes. ■

## INTRODUCTION

Research on speech perception has proposed an intriguing duality: The existence of a perceptual system for processing speech that is both flexible and stable. On the one hand, there is evidence of apparent plasticity in accommodating acoustic variations due to lexical context (Ganong, 1980), speech rate (Miller, 1981), and speaker identity (Nygaard & Pisoni, 1998), while at the same time perceptual sensitivities to native language speech contrasts acquired in childhood appear to be permanently ingrained and, in fact, may interfere with the subsequent acquisition of non-native contrasts in adulthood (Best & McRoberts, 2003). Such a dichotomy is particularly observable in the categorical perception of native language speech. Adults have a robust tendency to perceive speech stimuli that fall into two different categories (e.g., “d” and “t”) as more distinctive than stimuli that fall into the same category (e.g., “d1” and “d2”) even when they are equally spaced on a physical continuum (Pisoni & Tash, 1974; Liberman, Delattre, & Cooper, 1958; Liberman, Harris, Hoffman, & Griffith, 1957), a phenomenon known as *categorical perception*. Although native language categories learned over the course of a lifetime likely have ingrained, stable representations, categorical perception, defined behaviorally as selective improvements in discrimination of between-category contrasts, can be induced in adults in the short-term via category training (Swan & Myers, under review;

Golestani & Zatorre, 2009). In the current study, participants were trained to categorize tokens along a non-native speech contrast to examine the neural systems underlying the contribution of top-down category-level information to emerging categorical perception.

Perception of non-native speech contrasts is notoriously difficult for adults, although with sufficient training, participants show gains in their ability to categorize sounds from non-native speech continua (Golestani & Zatorre, 2009; Flege, MacKay, & Meador, 1999; Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997). The emergence of discontinuous perception may be impacted by the manner in which categories are learned (Guenther, Husain, Cohen, & Shinn-Cunningham, 1999), how the sounds map onto native language categories or minimal pairs in the learning environment (Yeung & Werker, 2009; Best, McRoberts, & Goodell, 2001), or even the way attention is distributed among category-relevant features (Francis & Nusbaum, 2002; Francis, Baldwin, & Nusbaum, 2000). For example, second language learners often receive explicit instruction as to the nature of phonetic categories in the new language, which may be reinforced by a mapping between the new contrast and the orthography of the language. Previous behavioral work from our laboratory (Swan & Myers, under review) and others (Golestani & Zatorre, 2009; McCandliss, Fiez, Protopapas, Conway, & McClelland, 2002) suggests that participants who successfully learn to categorize items from a non-native speech continuum using labels also show increased ability to discriminate tokens that cross the category boundary, with no change

<sup>1</sup>University of Connecticut, <sup>2</sup>Birkbeck, University of London

in discrimination accuracy for tokens within a learned category. This asymmetry in the ability to discriminate between-category and within-category contrasts provides a behavioral measure of categorical perception.

Ultimately, category learning may impact between-category discrimination accuracy via at least two mutually compatible routes. First, category training may result in a warping of perceptual sensitivities to speech sounds (Guenther, Nieto-Castanon, Ghosh, & Tourville, 2004; Guenther et al., 1999; Guenther & Gjaja, 1996; Kuhl, 1991), effectively retuning neural sensitivity to tokens that cross a category boundary. An alternative hypothesis is that improvements in discrimination for between-category contrasts instead reflect the application of a decision threshold once the learned category boundary has been established. According to this view, the underlying perceptual space is not altered, but rather category learning is mediated by cognitive processes; for example, the distribution of attention toward meaningful differences in the speech signal and away from irrelevant distinctions (Francis & Nusbaum, 2002; Francis et al., 2000). Note that these two accounts are not entirely incompatible: namely, it may be the case that the redistribution of attention to relevant parts of the acoustic signal in the short-term leads to retuning of perceptual sensitivities to phonetic contrasts over time. It is assumed that behavioral sensitivity to within- and between-category contrasts has its root in neural sensitivity of category distinctions. However, it is unclear where in the neural processing stream category level information is encoded, and in particular, whether such coding suggests a retuning of perceptual space or rather the top-down imposition of a learned category scheme. Neuroimaging studies, in particular fMRI, offer a tool for exploring these mechanisms.

### Neural Systems Underlying Phonetic Category Learning

Perception of native language phonetic category continua has revealed a network of areas involving frontal (IFG, precentral) and posterior (STG, SMG) areas (Liebenthal et al., 2010; Myers, Blumstein, Walsh, & Eliassen, 2009; Blumstein, Myers, & Rissman, 2005; Binder, Liebenthal, Possing, Medler, & Ward, 2004; Guenther et al., 2004). Broadly speaking, these studies suggest a division of labor between temporal and frontal areas. The left superior temporal gyrus (STG) shows sensitivity to the mapping between the fine-grained acoustic details of the signal and phonetic category space and responds to acoustic variation both within and between phonetic categories (Myers et al., 2009; Myers, 2007). In contrast, frontal areas show patterns of activation, which reflect a greater reliance on the details of the category structure: the left inferior frontal gyrus (IFG) responds preferentially to tokens that are ambiguous and, as such, may be involved in resolving competition at the level of the phonetic category (Myers, 2007; Binder et al., 2004). An adjacent area, which spans

BA 44 and precentral gyrus, has been shown to exhibit a pattern of phonetic category invariance: namely, it responds to phonetic contrasts that cross the category boundary, but not to acoustically equivalent contrasts within a phonetic category (Myers et al., 2009). Taken together, these results are consistent with a model in which the acoustic-phonetic details of the speech signal are processed in the left (or perhaps bilateral) STG and phonetic categorization processes occur in the IFG and precentral gyrus.

Investigations of the neural systems activated in categorizing non-native speech sounds before and after training have examined native English speakers learning a non-native dental/retroflex contrast (Golestani & Zatorre, 2004), a lexical tone contrast (Wang, Sereno, Jongman, & Hirsch, 2003), and Japanese speakers learning a non-native r/l contrast (Callan et al., 2003). In all three studies, participants showed post-training increases in activation for phonetic categorization tasks in a network of areas associated with performance on native language phonetic categorization tasks (Myers, 2007; Blumstein et al., 2005), including the bilateral STG, left supramarginal gyrus (SMG), and sometimes left or bilateral IFG. However, these studies were not designed to consider the question of how categorization training in turn affects sensitivity to within- and between-category contrasts in the learned acoustic-phonetic space. Only a handful of studies have examined the relationship between a category structure imposed on a novel acoustic space and the brain's sensitivity to exemplars from this space, and these studies generally support the notion that regions in the STG show sensitivity to the learned category structure (Leech, Holt, Devlin, & Dick, 2009; Guenther et al., 2004). Given that the left STG has been implicated in the perception of fine-grained details of native-language phonetic categories (Chang et al., 2010; Liebenthal et al., 2010; Formisano, De Martino, Bonte, & Goebel, 2008; Myers, 2007; Liebenthal, Binder, Spitzer, Possing, & Medler, 2005), a finding that the STG are also modulated by a learned phonetic category scheme would suggest that category training induces changes in the perceptual representation of speech sounds.

Whereas the left or bilateral superior temporal areas may play an important role in detecting differences within and between acquired sound categories as well as during native language acquisition, frontal regions, especially the left IFG, are likely to play a role in the perception of learned speech categories. With regard to mature native language categories, the left IFG has been implicated in segmenting phonemes from the surrounding phonetic context (Burton, Small, & Blumstein, 2000) and resolving competition between acoustically similar phonetic categories (Blumstein et al., 2005). Moreover, studies of non-native category learning suggest that inferior frontal regions are among the set of regions recruited in successful non-native sound categorization (Golestani, Molko, Dehaene, LeBihan, & Pallier, 2007; Callan, Jones, Callan, & Akahane-Yamada, 2004; Golestani & Zatorre, 2004; Callan

et al., 2003). These results are all consistent with a role for inferior frontal areas in categorizing the sounds of speech to map them to an established language category. Modulation of activity in the left IFG and/or left precentral gyrus associated with the learned category structure, in the absence of modulation of temporal activity, would suggest that discontinuous behavioral performance along an acoustic–phonetic continuum is due to cognitive, rather than perceptual factors.

The current study was designed to investigate the nature of neural sensitivity to non-native phonetic contrasts after categorization training and in particular to determine whether newly acquired sensitivities to non-native contrasts are encoded in temporal or frontal regions. Native English-speakers were trained to categorize items along a synthetic acoustic continuum that spanned three voiced stop categories, from a dental /*da*/, to a retroflex /*ɖa*/, to a velar /*ga*/ sound (see Methods for details). Although the endpoints approximate native English phonemes (/da/ and /ga/), the retroflex category is a non-native category. Participants were trained to categorize tokens from this continuum into categories “A” and “B” using one of two different boundaries, and the behavioral consequences of categorization training were measured with an AX discrimination task. Neural sensitivity to the between- and within-category contrasts was assessed using a short-interval habituation paradigm in fMRI (Zevin & McCandliss, 2005). Participants listened to trains of five syllables, with the final syllable either the same or different from the first four. Regions showing increased activation for “different” trials compared with “same” trials are thought to either signify a release from adaptation (Grill-Spector, Henson, & Martin, 2006; Grill-Spector & Malach, 2001) or an active change detection response to the contrast (Zevin, Yang, Skipper, & McCandliss, 2010). Importantly, while in the scanner, participants were not required to explicitly categorize stimuli or discriminate phonetic contrasts but instead monitored auditory stimuli for rare, high-pitched target stimuli. This implicit paradigm has the advantage of avoiding some of the additional, extralinguistic task demands that accompany explicit categorization and discrimination tasks.

To examine possible changes in behavioral and neural sensitivity due to category learning, the current study compared stimulus pairs that form a between-category contrast for one training group, but a within-category contrast for the other group. Consistent with previous work (Myers et al., 2009; Joanisse, Zevin, & McCandliss, 2007), it was predicted that between-category contrasts would show greater activation than within-category contrasts, as dictated by the learned category scheme. Given the role of the left, and perhaps also right, STG in processing fine-grained details of the speech stream, between-group differences in activation for the same speech stimuli in temporal areas would be taken as evidence that categorization training results in a reshaping of the auditory sensitivities to tokens from the learned categories. Alternatively, between-group differences in activation in left inferior

frontal regions would support the hypothesis that reshaped perceptual sensitivities reflect the application of a decision mechanism as tokens are mapped to the new category representation. Finally, sensitivity of both frontal and temporal regions to the learned category boundaries would suggest that changes in perceptual sensitivity are achieved via both reshaping of auditory sensitivities to learned speech contrasts as well as diversion of attention to the relevant parts of acoustic space.

## METHODS

### Participants

Twenty-nine adults (21 women, 8 men) between the ages of 18 and 45 years old were recruited from the Brown University community. Having completed a behavioral study of category training (Swan & Myers, under review), participants were invited to take part in a subsequent study using fMRI. All participants were native speakers of American English, were right-handed as confirmed by the Edinburgh Handedness Inventory (Oldfield, 1971), and reported no hearing or neurological deficits. Informed consent was obtained, and all participants were screened for ferromagnetic materials according to guidelines approved by the Human Subjects Committees of Brown University. Participants were compensated \$30 for the 2-hr session. Movement of more than 3 mm (the width of one voxel) was cause for exclusion of one participant from further analyses; thus, the total number of participants in the study was 28, with 14 (10 women, 4 men) in each group.

### Stimuli

Speech syllables used in both the behavioral and fMRI testing sessions were taken from a synthetic continuum that varied in the onset frequencies of the second and third formants and in the frequency of the burst, constituting a 9-point acoustic continuum ranging from a dental /*da*/ to a retroflex /*ɖa*/ to a velar /*ga*/ places of articulation (Stevens & Blumstein, 1975). Although Stevens and Blumstein (1975) report individual differences in placement of dental/ retroflex and retroflex/velar boundaries among native speakers of languages utilizing the dental/ retroflex/velar contrast, they found that, in general, the first three points can be classified as Dental sounds, the middle three points as Retroflex sounds, and the remaining three points as Velar sounds. For native speakers of English, the endpoints of the continuum comprise a near-native phonetic contrast, whereas the dental and retroflex points are more difficult for these speakers to discriminate (Polka, 1991; Werker & Tees, 1984). A high-pitched version of the 9-point continuum was created by shifting the entire pitch contour of a stimulus upward by 100 Hz using Praat (Boersma, 2001). These stimuli were used in the

*TARGET* trials, replacing one of the five tokens in 10 trials per run.

## Procedure

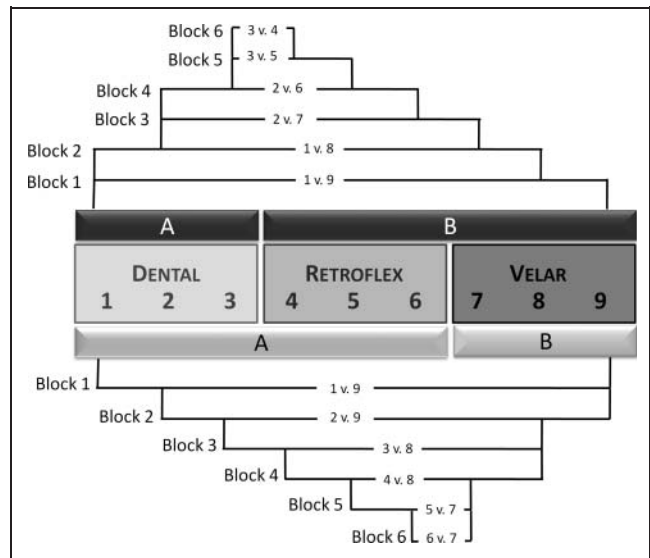
In the weeks before this experiment, participants had completed a single 45-min behavioral testing session, which included (1) an AX discrimination pretest, (2) six blocks of categorization training with feedback, (3) an identification task without feedback, and (4) a final AX discrimination posttest. The mean number of days between testing sessions was 60.11 ( $SE = 7.31$ , median = 39, range = 9–147) and did not differ between training groups,  $t(26) = 1.24$ ,  $p = .228$ . Because it was unknown how much the effects of training would decay between sessions, participants underwent a replication of the behavioral training upon arriving at the scanner. Participants were assigned to two training groups, Dental/Retroflex and Retroflex/Velar, just as in their first session; thus, their training was consistent with the category boundary with which they were already familiar. The replication included (1) the AX discrimination pretest, (2) categorization training with feedback, and (3) the identification task without feedback. 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 evaluated sensitivity to the dental/retroflex/velar contrasts before and after learning to categorize the tokens. Participants heard 60 pairs of syllables (separated by a 250-msec ISI) from the three category centers, Dental (point 2), Retroflex (point 5), and Velar (point 8). Thirty pairs were repeated (e.g., 2 vs. 2), and 30 pairs were contrasts (e.g., 5 vs. 8). They judged whether the stimuli sounded the same or different from one another by pushing the corresponding button.

## Categorization Training

Participants heard 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 training group categorized phonetic contrasts converging on the boundary between points 3 and 4, whereas the Retroflex/Velar training group categorized contrasts converging on the boundary between points 6 and 7, as 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 posttests. 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 feedback (i.e.,



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

unique sounds for a correct or incorrect response) was given immediately after each response.

## Identification

Following training, participants heard 10 tokens of each sound on the 9-point continuum, presented in a random order for a total of 90 trials. They identified each token as either “Category A” or “Category B” by pushing a corresponding button and no feedback was given.

Immediately after replicating these behavioral tasks, participants completed a 1-hr fMRI session. During the scanning session, participants alternated between two tasks: a high-pitch detection task during the functional runs and a shortened category training practice during the anatomical scan and between functional runs. Participants listened to speech sounds over headphones adjusted to ensure comfortable listening over the noise of the scanner, and they responded by pushing buttons on an MRI-compatible button box. They were instructed to keep their eyes closed and refrain from movement, especially of the head and upper body.

A short-interval habituation paradigm was used during the functional scans, with each trial containing five 266-ms tokens separated by 50-ms ISIs (see Figure 2). In *SAME* trials, a token from one of the three category centers, Dental (point 2), Retroflex (point 5), and Velar (point 8), was repeated five times, whereas in *DIFFERENT* trials, one token was repeated four times and followed by a contrasting fifth token. There were three functional runs, each containing 55 trials: 30 *DIFFERENT* trials (10 of each contrast), 15 *SAME* trials (5 of each) and 10 high-pitched *TARGET* trials.



### High-pitch Detection

Although participants heard *SAME* and *DIFFERENT* stimulus trains, their task was not to discriminate between the phonetic tokens. Rather, they were instructed to listen for rare *TARGET* trials, in which one of the five syllables was replaced by a high-pitched syllable, and to push a button whenever they detected one. *TARGET* trials, as well as trials in which participants erroneously detected a high-pitched target, were included in a separate stimulus regressor and were not analyzed further. Participants were highly accurate at detecting these stimuli, with only five errors across all participants.

### Category Training Practice

To be sure that effects of training were maintained in the scanner, participants practiced categorizing sounds with feedback during acquisition of the anatomical scan and between functional runs. Participants completed a single 60-trial block with just 10 trials for each progressively narrower step of training. No functional data were acquired during these blocks. Finally, immediately after scanning, participants completed one more 60-trial block of practice category training followed by the AX discrimination posttest.

### Image Acquisition and Analysis

A 3T Siemens Trio scanner was used to perform anatomical and functional scans. High-resolution 3-D anatomical images were acquired with a T1-weighted magnetization prepared rapid acquisition gradient-echo sequence (repetition time = 1900 msec, echo time = 4.15 msec, inversion time = 1100 msec, 1 mm<sup>3</sup> isotropic voxel size, 256 × 256 matrix). Functional images were acquired in ascending, interleaved order with an EPI sequence (15 slices, 5 mm thick, 3 mm<sup>2</sup> axial in-plane resolution, 64 × 64 matrix, 192 mm<sup>3</sup> field of view, flip angle = 90°). The functional slab was aligned parallel to the sylvian fissure, with

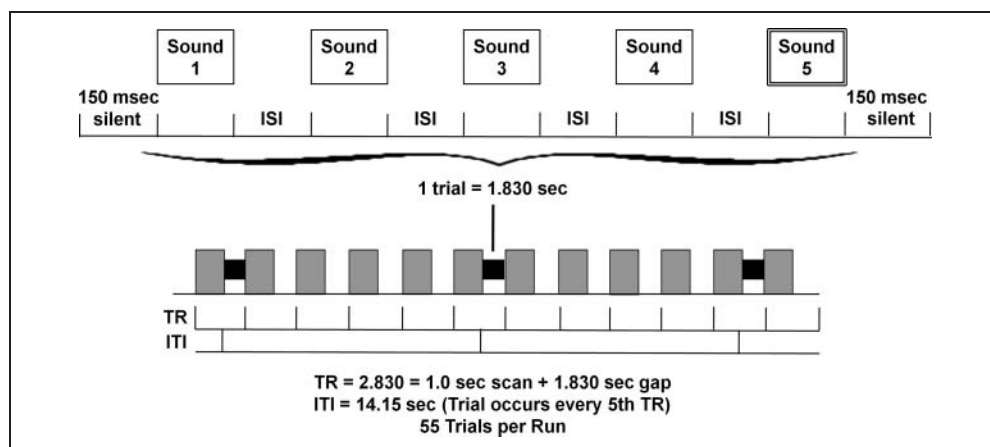
coverage from the middle temporal gyrus (MTG) caudally to the middle frontal gyrus (MFG) dorsally, including coverage of the STG, MTG, SMG, AG, IFG, and portions of the MFG adjacent to the IFG. Each functional volume was acquired with a 1-sec acquisition time, followed by 1.830 sec of silence during which auditory stimuli were presented, for a total repetition time of 2.830 sec. Trials occurred every fifth volume acquisition (see Figure 2), and two additional volumes were added to the beginning of each run to accommodate saturation effects, resulting in a total of 277 volumes collected per run.

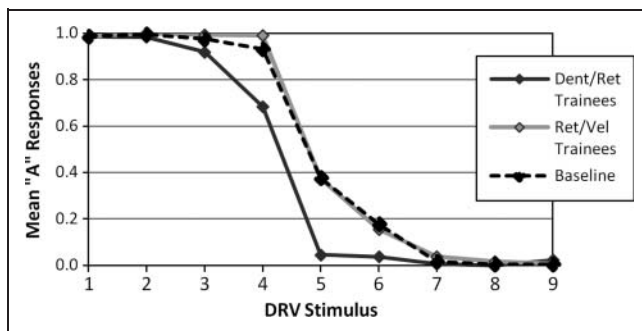
Images were analyzed using AFNI (Cox, 1996). Preprocessing of images included motion correction using a six-parameter rigid body transform coregistration with the anatomical data set (Cox & Jesmanowicz, 1999), normalization to Talairach space (Talairach & Tournoux, 1988), and spatial smoothing with a 6-mm Gaussian kernel. Masks were created using each participant's anatomical data to eliminate activated voxels located outside the brain. Individual masks were used to generate a group mask, which included only those voxels imaged in at least 25 of 28 participants' functional data sets.

Time series vectors were generated for each stimulus type: *DIFFERENT* trials (Dental vs. Retroflex, Retroflex vs. Velar, and Dental vs. Velar), *SAME* trials (Dental vs. Dental, Retroflex vs. Retroflex, and Velar vs. Velar) and *TARGET* trials. These vectors contained the start time of each stimulus train and were convolved with a gamma-variate function. Each participant's preprocessed functional data were submitted to a regression analysis with 13 regressors, namely the seven condition vectors and the six parameters of the motion correction process, which were included as nuisance variables. This 3dDeconvolve analysis returned by-voxel fit coefficients for each condition for each participant.

Fit coefficients for each condition were entered into a mixed-factor ANOVA with Condition as a fixed factor and Subject as a random factor. The contrasts of interest (Dental vs. Retroflex and Retroflex vs. Velar) were compared both between and within groups, and main effects

**Figure 2.** Each trial included the five 266-msec tokens separated by 50-msec ISI with 150-msec of silence at each end, for a total trial length of 1.830 sec. Trials occurred every fifth silent gap between 1.0 sec EPI scans.





**Figure 3.** Percentage of “A” responses to the nine stimulus tokens on the identification task without feedback.

and interactions were also investigated. In addition, the effect of adaptation (*DIFFERENT–SAME*) was considered overall and for each contrast type. The ANOVA results were masked with a small volume corrected group mask, which included anatomically defined regions typically implicated in language processing: bilateral IFG, MFG, STG, and MTG, left SMG, TTG, and AG. To correct for multiple comparisons, Monte Carlo simulations were run on the small-volume corrected group mask, producing a corrected threshold of  $p < .05$ , and statistical maps were thresholded to include only clusters of 63 contiguous voxels at a voxel-level threshold of  $p < .05$ . Mean fit coefficients were then extracted from each participant’s data and submitted to paired and two-sample  $t$  tests.

## RESULTS

### Behavioral Results

#### Categorization Training

Because categorization training became more difficult over the six blocks as the phonetic contrasts narrowed, all participants were required to reach a criterion for inclusion of at least four blocks completed before falling below 60% accuracy. The mean number of blocks completed above criterion was 5.29 ( $SE = 0.12$ ) with no significant difference between groups ( $t(26) = 0.57, p < .576$ ).

Post-training identification task responses to each of the nine stimulus tokens are plotted alongside a baseline function provided by a group of 14 untrained native English speakers, shown in Figure 3. Responses were then converted to  $z$  scores, and the point at which each response function crossed the 0.50 mark was calculated as the participant’s category boundary. The two groups demonstrated very different functions following training, particularly on the middle tokens 4–6, with members of the Retroflex/Velar training group exhibiting a mean category boundary of 5.396 ( $SE = 0.119$ ) and members of the Dental/Retroflex training group exhibiting a mean category boundary of 4.027 ( $SE = 0.132$ ). A univariate ANOVA comparing the baseline and training groups’ boundaries

showed a significant main effect of Group,  $F(2, 41) = 20.11, p < .001$ . Post hoc tests (Bonferroni-corrected at  $p < .05$ , yielding a functional threshold of  $p < .017$ ) revealed that, whereas the Dental/Retroflex trainees’ category boundary was significantly different from both the baseline ( $t(26) = 3.97, p < .001$ ) and the Retroflex/Velar trainees’ boundaries, the Retroflex/Velar trainees’ boundary did not differ from baseline ( $t(26) = 1.67, p = .110$ ). This result suggests that categorization training had a strong impact on how participants identified tokens from the continuum.

#### Discrimination

$d'$  scores were computed for all three contrasts in the AX discrimination task (displayed in Table 1) for Pretest 1 and Post-test 1, completed in the first behavioral testing session, and for Pretest 2 and Posttest 2, completed before and after the 1-hr scanning session. Of particular interest is the mean difference in  $d'$  sensitivity from pretest to posttest, shown in Figures 4 and 5, as these reflect the effects of categorization training between the two categorization training groups.

These  $d'$  scores were submitted to two separate three-way repeated-measures ANOVAs. The first ANOVA was conducted using  $d'$  scores from the first behavioral testing session to examine the effects of the original categorization training manipulation, and it included factors of Group (Dental/Retroflex and Retroflex/Velar), Contrast (Dental vs. Retroflex and Retroflex vs. Velar only), and Effect of Training (Pretest 1 and Posttest 1). This ANOVA revealed a main effect of Contrast,  $F(1, 26) = 8.26, p < .008$ , and no other significant main effects. The ANOVA also showed a significant interaction of Contrast  $\times$  Training,  $F(1, 26) = 10.77, p < .003$ , an interaction of Group  $\times$  Training approaching significance,  $F(1, 26) = 2.47, p < .128$ , and a significant three-way interaction of Group  $\times$  Contrast  $\times$  Training,  $F(1, 26) = 6.68, p < .016$ .

The results of the first ANOVA were examined for simple effects with a two-way repeated-measures ANOVA for the Dental vs. Retroflex contrast, which revealed a significant main effect of Training,  $F(1, 26) = 10.81, p < .003$ , as well as an interaction of Group  $\times$  Training,  $F(1, 26) = 12.56, p < .001$ . The two-way repeated measures ANOVA for the Retroflex vs. Velar contrast showed a significant main effect of Training,  $F(1, 26) = 4.14, p < .052$ , but no interaction with Group. These simple effects indicated that whereas the Dental/Retroflex training group improved considerably in discriminating the Dental vs. Retroflex contrast, the Retroflex/Velar group did not increase or decrease their performance on either contrast.

The second three-way repeated-measures ANOVA was conducted using  $d'$  scores from the first pretest and the final post-fMRI posttest to consider the overall effects of training across the two sessions. It included factors of Group (Dental/Retroflex and Retroflex/Velar), Contrast (Dental vs. Retroflex and Retroflex vs. Velar only), and Effect of Training (Pretest 1 and Post-test 2). This ANOVA

**Table 1.** AX Discrimination Results (Mean  $d'$  Scores)

	Prior Behavioral Testing Session		Change: Post 1 – Pre 1	fMRI Scanning Session		Overall Change: Post 2 – Pre 1
	Pretest 1	Post-test 1		Pretest 2	Post-test 2	
<i>Dental/Retroflex Trainees</i>						
Dental vs. Retroflex	1.471	2.930	1.459	2.781	3.202	1.731
SE	0.292	0.298	0.331	0.374	0.372	0.480
Retroflex vs. Velar	1.001	0.160	–0.841	0.756	1.035	0.034
SE	0.275	0.267	0.325	0.289	0.190	0.375
Dental vs. Velar	3.281	4.004	0.723	4.499	4.196	0.916
SE	0.362	0.187	0.259	0.104	0.172	0.441
<i>Retroflex/Velar Trainees</i>						
Dental vs. Retroflex	1.978	1.918	–0.060	2.005	2.533	0.554
SE	0.437	0.433	0.267	0.372	0.302	0.420
Retroflex vs. Velar	1.408	1.074	–0.334	1.471	1.667	0.260
SE	0.446	0.452	0.477	0.428	0.338	0.306
Dental vs. Velar	3.750	4.419	0.669	4.497	4.574	0.824
SE	0.314	0.106	0.358	0.082	0.079	0.341

revealed a significant main effect of Training,  $F(1, 26) = 12.18, p < .002$ , a significant main effect of Contrast,  $F(1, 26) = 8.37, p < .008$ , and an interaction of Contrast  $\times$  Training,  $F(1, 26) = 5.39, p < .028$ . A three-way interaction of Group  $\times$  Contrast  $\times$  Training approached significance,  $F(1, 26) = 2.67, p < .114$ .

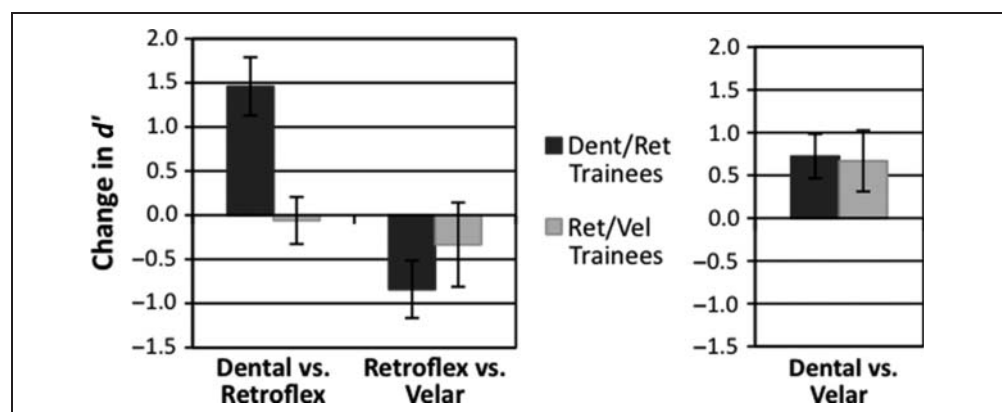
The second ANOVA was also examined for simple effects with a two-way repeated-measures ANOVA for the Dental vs. Retroflex contrast, which revealed a significant main effect of Training,  $F(1, 26) = 12.85, p < .001$ , and a near-significant interaction of Group  $\times$  Training,  $F(1, 26) = 3.407, p < .076$ . A two-way repeated-measures ANOVA for the Retroflex vs. Velar contrast did not show any main

effects or interactions. These simple effects again indicated that the Dental/Retroflex training group's substantial improvement in discriminating the Dental vs. Retroflex contrast after training was driving the three-way interaction.

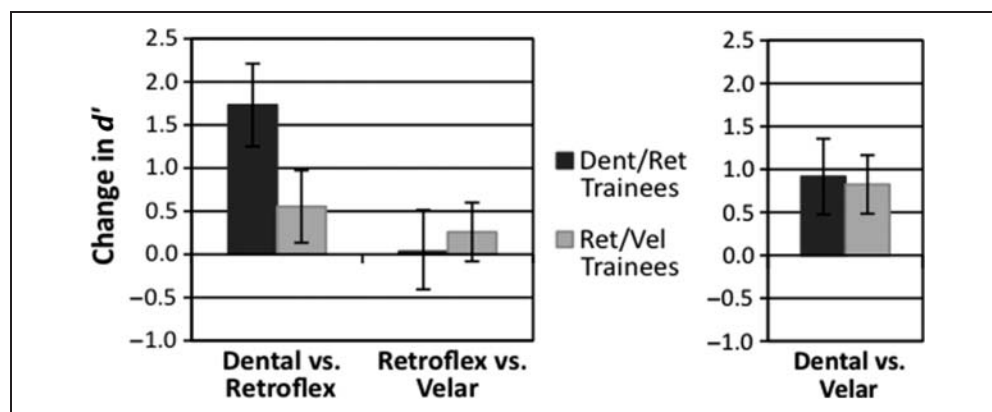
### Imaging Results

Comparison of *DIFFERENT* trials to *SAME* trials across groups showed clusters in left and right IFG, pictured in Figure 6, that responded differentially to contrasting syllables. This effect was significantly stronger among members of the Dental/Retroflex training group (in the left IFG,  $t(1, 79) = 3.009, p < .003$  and in the right IFG,

**Figure 4.** Mean change in  $d'$  scores on AX discrimination task due to first categorization training (difference from Pretest 1 to Post-test 1 in first behavioral testing session). Left: Contrasts of interest included in ANOVA. Right: Near-native contrast shown for comparison.



**Figure 5.** Mean change in  $d'$  scores on AX discrimination task due to overall categorization training (difference from Pretest 1 in first behavioral testing session to Post-test 2 following scanning session). Left: Contrasts of interest included in ANOVA. Right: Near-native contrast shown for comparison.



$t(1, 82) = 2.051, p < .043$ ) than for the Retroflex/Velar training group ( $p > .342$  for both regions).

The mixed-factor ANOVA was conducted to look for main effects, interactions, and between group and within group comparisons, all summarized in Table 2. The Group  $\times$  Contrast type interaction revealed significant clusters of activation in primarily in the left and right MFG, extending into the pars triangularis of the IFG on each side: on the left,  $F(1, 26) = 11.16, p < .003$ , and on the right,  $F(1, 26) = 9.20, p < .005$ . The results suggest that these regions respond differently depending on the nature of the categorization training experienced (see Figure 7). The clusters from the Group  $\times$  Contrast interaction are paralleled by significant clusters in the same regions for the Within-Group comparisons, indicating that the MFGs respond less actively for the trained contrast than for the untrained contrast. Paired  $t$  tests showed a significant difference between contrasts in the right MFG for members of the Retroflex/Velar training group,  $t(13) = 3.086, p < .009$ , and in the left MFG for members of the Dental/Retroflex training group,  $t(13) = -3.167, p < .007$ .

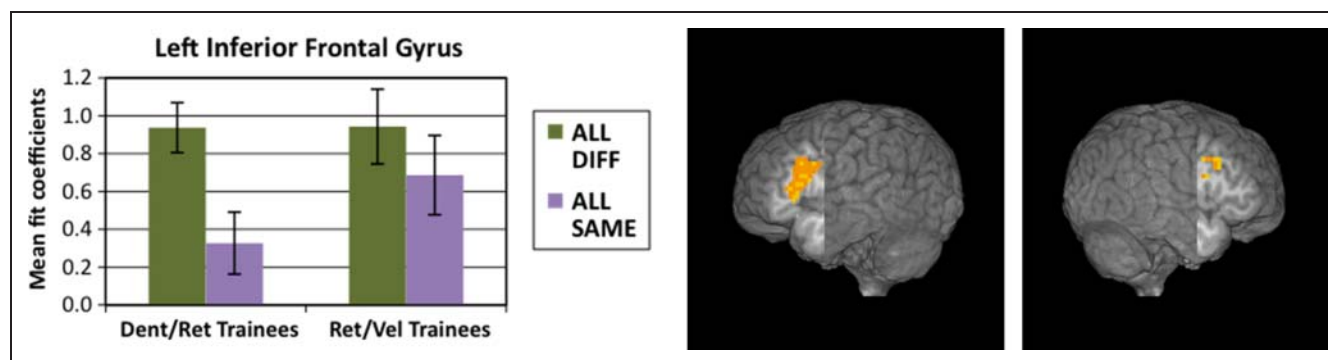
Between-groups comparisons revealed two clusters in frontal regions for which the Retroflex/Velar training group showed greater signal change in response to the

Dental vs. Retroflex contrast than the Dental/Retroflex training group: in the right IFG,  $t(26) = 4.174, p < .001$ , and in the left MFG,  $t(26) = 3.954, p < .001$ . However, no significant differences between groups were apparent in response to the Retroflex vs. Velar contrast.

Results of the regression analysis, reported in Table 3 and pictured in Figure 8, indicated that a change in discrimination sensitivity (Posttest 2 – Pretest 1  $d'$  score) was negatively correlated with signal change in two regions, right STG/MTG,  $r(84) = -0.309$ , and left IFG,  $r(84) = -0.334$ . This latter cluster was immediately adjacent to the MFG cluster revealed in the interaction analysis and overlapped this cluster by five voxels. The negative correlation suggests that participants who improved at discriminating the contrasts showed less activation than those who did not improve.

## DISCUSSION

A central question in the study of speech perception is how categorical perception of speech emerges. The results of this study suggest that explicit categorization training can produce increased discrimination sensitivity to between-category contrasts, as well as differential neural sensitivity to the learned category scheme. Specifically, one



**Figure 6.** Main effect of adaptation: DIFFERENT trials  $>$  SAME trials. Left inferior (213 voxels, cut:  $x = -46, y = 0$ ) and right middle (63 voxels, cut:  $x = 47, y = 0$ ) frontal gyri. Bar graphs display mean fit coefficients within the displayed functional clusters.



**Table 2.** Significant Clusters, Thresholded at a Corrected Threshold of  $p < .05$ , Voxel-wise  $p < .05$ , Clusters of 63 Contiguous Voxels

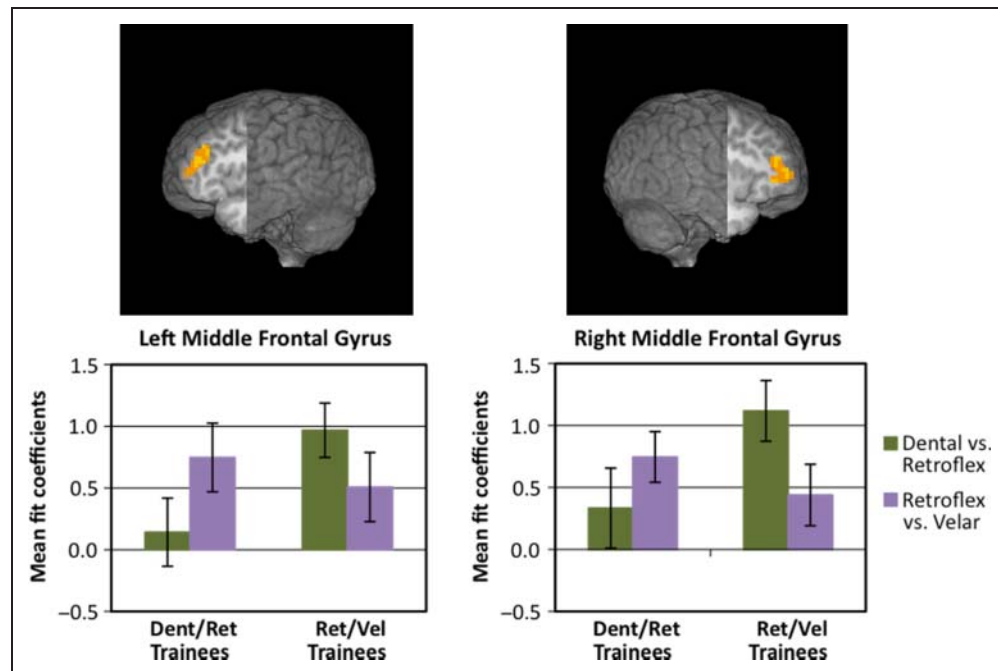
Area	Number of Activated Voxels	Maximum Intensity Coordinates (T-T)			Maximum <i>t</i> Value
		<i>x</i>	<i>y</i>	<i>z</i>	
<i>Main Effect of Adaptation [All DIFFERENT trials &gt; All SAME trials]</i>					
Left IFG and left BA 9	213	-46	14	26	3.581
Right MFG	63	47	17	29	2.921
Left MTG and left BA 21 <sup>a</sup>	51	-61	-37	2	2.153
Left MTG and left BA 19 <sup>a</sup>	46	-37	-79	23	-3.255
<i>Main Effect of Adaptation: [Dent vs. Vel DIFFERENT &gt; Dent SAME + Vel SAME]</i>					
Left IFG and left BA 47	94	-46	17	2	2.442
Right MTG	81	41	-61	23	2.450
Left IFG and left BA 9 <sup>a</sup>	59	-43	11	26	3.037
Right MFG <sup>a</sup>	57	38	8	32	3.331
Left MTG and left BA 19 <sup>a</sup>	43	-37	-79	23	-2.829
<i>Main Effect of Group [Ret/Vel Trainees &gt; Dent/Ret Trainees]</i>					
Right IFG	120	38	20	-7	2.914
<i>Interaction Group × Contrast Type</i>					
Left MFG	116	-40	35	26	8.500
Right MFG and right BA 10	78	38	41	17	9.060
<i>Between Group Effects: Dent vs. Ret [Ret/Vel Trainees &gt; Dent/Ret Trainees]</i>					
Right IFG	146	44	41	2	3.810
Left MFG	80	-37	38	26	2.964
Right IFG <sup>a</sup>	52	50	5	20	3.908
<i>Between Group Effects: Ret vs. Vel [Dent/Ret Trainees &gt; Ret/Vel Trainees]</i>					
No significant clusters emerged.					
<i>Within Group Effects: Ret/Vel Trainers [Dent vs. Ret &gt; Ret vs. Vel]</i>					
Right MFG	67	38	38	17	2.980
<i>Within Group Effects: Dent/Ret Trainers [Dent vs. Ret &lt; Ret vs. Vel]</i>					
Left MFG and left BA 10 <sup>a</sup>	53	-40	41	20	-2.715
Right MTG <sup>a</sup>	40	53	-40	-1	-3.833

<sup>a</sup>Cluster is subthreshold.

group (Dental/Retroflex training group) showed improvements in behavioral discrimination performance for tokens that crossed the learned category boundary (Dental vs. Retroflex) but no improvement for those within a single learned category (Retroflex vs. Velar). This same group

also showed a significant shift in categorization function compared with individuals who had not undergone training, whereas the Retroflex/Velar group showed no such shift. Of interest, previous work from our laboratory using the same stimuli and presentation order (Swan & Myers,

**Figure 7.** Group  $\times$  Condition interaction. Left (116 voxels, cut:  $x = -40, y = 0$ ) and right (78 voxels, cut:  $x = 38, y = 0$ ) MFGs extending into pars triangularis of IFGs. Bar graphs display mean fit coefficients within the displayed functional clusters.



under review) showed that mere exposure to the tokens was not sufficient to produce category-specific changes in discrimination. In this study, discrimination asymmetries that resulted from categorization training were accompanied by differences in activation almost exclusively in left and right frontal areas, specifically in the left IFG and left and right MFG. Surprisingly, no interaction between Group and Contrast was found in temporal areas, which have often been associated with the perception of fine-grained acoustic details of speech. The implications of these results for our understanding of the neural systems underlying speech perception are discussed below.

The effect of categorization training on behavioral discrimination differed significantly between groups. Behavioral data showed that, despite similar success during categorization training, only the Dental/Retroflex training group demonstrated a persisting shift in the phonetic category boundary and subsequently transferred their category learning to a discrimination paradigm (cf. also Swan & Myers, under review).<sup>1</sup> These behavioral results lead to

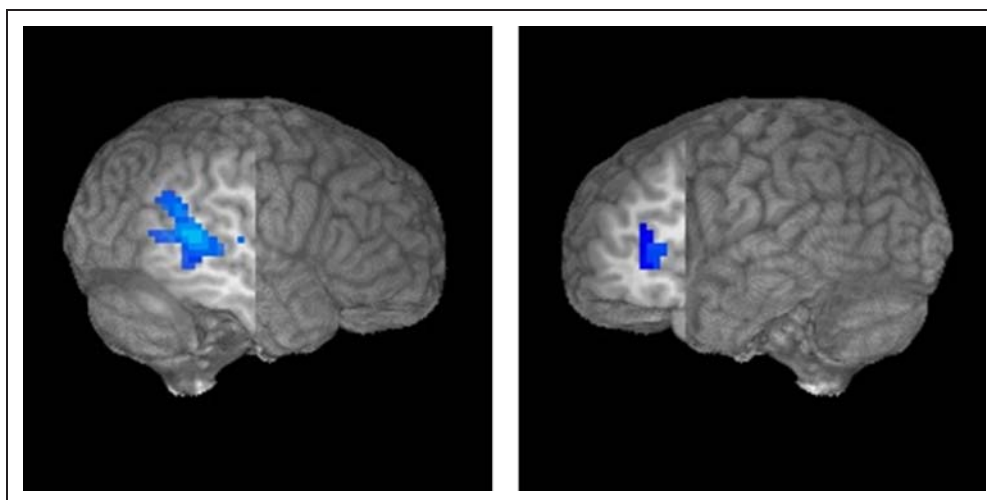
the suggestion that only the Dental/Retroflex group showed a reorganization of perceptual space around the learned categories. Nonetheless, fMRI data revealed an interaction between group and contrast in left and right MFG, with significantly greater activation for the learned within-category contrast compared with the learned between-category contrast in both groups. This finding is consistent with the view that sensitivity to category-level information may precede the reshaping of sensitivities along the acoustic–phonetic continuum, as measured by discrimination (Liebenthal et al., 2010). Evidence from ERP paradigms also suggests that neural correlates of phonetic category learning may be found before behavioral sensitivity to the same contrast (Tremblay, Kraus, & McGee, 1998).

Both groups showed greater activation for “different” trials than “same” trials in the left STG as well as left IFG, replicating previous studies using this paradigm (Myers et al., 2009). However, the relationship between activation patterns on between- and within-category trials was

**Table 3.** Regression with Behavioral Scores, Significant Clusters Thresholded at a Corrected Threshold of  $p < .05$ , Voxel-wise  $p < .05$ , Clusters of 63 Contiguous Voxels

Area	Number of Activated Voxels	Maximum Intensity Coordinates (T-T)			Maximum $t$ Value
		$x$	$y$	$z$	
<i>Regression: Change in <math>d'</math> Perceptual Sensitivity with % Signal Change</i>					
Right STG/MTG	237	56	-40	11	-3.194
Left IFG	77	-37	29	5	-3.586

**Figure 8.** Negative correlation for change in  $d'$  score. Right STG/MTG (237 voxels, cut:  $x = 52, y = -15$ ) and left IFG (77 voxels, cut:  $x = -36, y = 15$ ).



unexpected. Previous research from our group (Myers et al., 2009) and others (Zevin et al., 2010; Joanisse et al., 2007; Zevin & McCandliss, 2005) has shown greater activation in adaptation/habituation designs for more perceptible stimulus contrasts. In the current study, participants showed the expected behavioral pattern: between-category contrasts were easier to perceive than within-category contrasts after category training, yet greater activation was shown for the difficult within-category contrast in both groups. A possible explanation accounting for this difference is that the results reflect a cognitive set that participants enter when they are tasked with performing an active categorization task immediately before scanning. That is, although participants were not required to attend to category-level information during scanning, they may nonetheless be implicitly categorizing stimuli, consistent with the demands of the behavioral task performed outside the scanner. If participants are implicitly assigning category labels to stimuli, the results may be expected to reflect active detection of category distinctions akin to a discrimination task (Desai, Liebenthal, Waldron, & Binder, 2008; Hutchison, Blumstein, & Myers, 2008; Liebenthal et al., 2005). Of interest, Desai and colleagues (2008) report activation of the left and right MFG in a study that used an ABX discrimination task to measure participants' sensitivity to acoustic distinctions in nonspeech sine-wave sounds that can be perceived either as nonspeech sounds or with phonetic content. Changes in activation in both the left and right MFG were larger for a continuum that participants ultimately perceived as speech than for a set of sounds that could not be mapped to speech categories, suggesting that the MFG bilaterally are sensitive to category-level information about speech tokens. Of interest, a separate region in the left STG was correlated with increases in the sharpness of participants' categorization function. Given that these authors did not investigate correlations between brain and behavior in the MFG, it is unknown whether frontal activation shows a similar relationship to changes in behavior.

The left MFG has also been shown to discriminate between naturally produced and synthetic speech (Benson et al., 2001). Similarly, in a meta-analysis of 23 studies of speech perception, Turkeltaub and Branch Coslett (2010) report a small focus in the posterior left MFG, which was related to activation for speech stimuli compared with low-level control stimuli. Although greater activation in the MFG has been observed while attending rather than ignoring auditory, and particularly speech, stimuli (Sabri et al., 2008), the MFG are also modulated by nonlinguistic input. The bilateral MFG have been implicated in access to category-level information in abstract visual patterns (Vogels, Sary, Dupont, & Orban, 2002), and sensitivity to learned visual object categories has been shown in a homologous region of lateral pFC in nonhuman primates (Freedman, Riesenhuber, Poggio, & Miller, 2001). This evidence suggests that the recruitment of the bilateral MFG may reflect a domain-general resource for categorization processes.

Contrary to our hypotheses, although the IFG responded more to "Different" than "Same" trials, there was no interaction centered in the IFG between contrast and group, suggesting that this region responded similarly regardless of how participants had been trained. Most previous studies showing IFG sensitivity to phonetic category-level information have used active categorization tasks (Callan et al., 2004; Golestani & Zatorre, 2004), and activation in the IFG for categorization tasks has been linked to difficulty in phonetic tasks (Blumstein et al., 2005; Binder et al., 2004). Moreover, some have suggested that the contribution of Broca's area to speech perception is limited to situations in which participants are required to actively inspect the acoustic signal (Hickok & Poeppel, 2004). It is possible that the implicit nature of the current design did not require participants to inspect the signal and thus did not differently engage the IFG as a function of the trained category membership.

Most remarkably, no evidence of reorganization of cortical sensitivity as a function of training was seen in temporal

areas. Although the left STG did show sensitivity to phonetic differences, with a sub-threshold (50-voxel) cluster exhibiting greater activation for phonetic change trials than repeated trials, the two different training groups showed no differences in sensitivity in temporal areas to the learned within- and between-category contrasts. This may seem surprising given that some previous studies (Liebenthal et al., 2010; Desai et al., 2008; Guenther et al., 2004) have shown engagement of temporal areas related to categorization training. The relatively rapid learning that our participants displayed may result from an ability to deploy attention appropriately to the distinguishing aspects of the acoustic waveform rather than a fundamental reorganization of perceptual space (cf. also Alain, Campeanu, & Tremblay, 2010). Accordingly, the presence of category labels may modulate attention and momentarily highlight stimulus features containing meaningful distinctions (cf. Landau, Dessalegn, & Goldberg, 2010) or may reinforce a progressively more consistent decision threshold but does not alter the underlying perceptual space or its representation in the temporal lobes in the short term. The suggested role of attention in learning to perceive new speech contrasts is not unique to this study. “Attention to dimension” models of speech perception (Francis & Nusbaum, 2002; Francis et al., 2000) propose that shifts in attention to relevant parts of acoustic phonetic space may be crucial not only for learning new speech sounds, but for attending to distinctions between sounds in one’s native language repertoire.

It is possible that the activation observed in the MFGs does not relate to learned sensitivities to category information, but rather results from attentional biases that arise during training. For instance, it may be the case that as participants learn to assign different category labels to tokens along the acoustic phonetic continuum, attention is directed to between-category contrasts in a way that is unrelated to or independent of the demonstrated changes in behavioral sensitivity to the same contrasts. This account might explain the finding that, although only one group showed significant changes in their behavioral performance, there was a full cross-over interaction in the bilateral MFG, with significant differences in activation for both groups.

We argue that the engagement of regions involved in attention and categorization need not be seen as epiphenomenal. Categorization training in this study resulted in shifts in discrimination accuracy that resemble those seen for native-language phonetic category structure. The engagement of frontal regions associated with a learned phonetic category structure may be the first step in encoding more long-term sensitivities to speech sounds. The fact that sensitivity to category membership was observed in the MFG rather than the temporal lobes suggests that this change in apparent perceptual sensitivity results from the influence of an enhanced distinctiveness from mapping two tokens to two distinct abstract categories, rather than due to the changes in the perceptual encoding of these stimuli per se. One hypothesis is that over the course

of learning, top-down feedback signals mediate the tuning of acoustic-phonetic processors in the temporal lobes, which over time may result in long-term changes in tuning properties of temporal areas. Results of the regression analysis support this view: less activation in inferior frontal regions was seen for individuals who showed the greatest improvements from pretest to posttest. Such a model would account for data showing that incidental auditory category learning over many sessions, or even over a lifetime of second-language learning (Liebenthal et al., 2010; Raizada, Tsao, Liu, & Kuhl, 2010; Leech et al., 2009) produce changes in encoding of between-category contrasts in temporal areas.

Results of this study highlight the similarities between category learning for speech and category learning in other modalities. Across a variety of domains, the use of category labels seems to influence how items from physical continua are perceived. Increased discrimination of between-category distinctions for which there are established labels has been shown in perception of color, size and brightness, facial identity and emotional valence (Kikutani, Roberson, & Hanley, 2008; Özgen & Davies, 2002; Kay & Kempton, 1984). In the current study, exposure to category-level information produced increases in behavioral sensitivity to between-category contrasts. Parallels also exist at the level of neural function: primate studies of learned animal categories (Seger & Miller, 2010; Freedman et al., 2001) and face identity categories (Rotshtein, Henson, Treves, Driver, & Dolan, 2005), implicate inferior frontal regions in the processing of category membership. As such, it seems plausible that changes in behavioral sensitivity arising from category learning in non-native speech acquisition may be driven by domain-general executive processes, operating in the frontal lobes.

## Conclusion

Speech scientists are forced to reconcile evidence of both plasticity and durability in the speech perceptual system. This contrast is readily apparent in the discontinuous perception of speech categories, which although robust, nonetheless adapts flexibly to changes in the language environment. We suggest that such tensions are resolved in the neural system through parallel processes; in this case combining stable long-term sensitivities encoded in the temporal lobes with flexible, context-dependent decision processes in the frontal lobes.

## Acknowledgments

This work was funded by R03 DC009495 to Emily Myers. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIDCD or the NIH. The authors thank Sheila Blumstein for providing stimuli and Laura Mesite for assistance with data analysis.

Reprint requests should be sent to Emily B. Myers, Department of Communication Sciences, University of Connecticut, 850 Bolton Rd., Storrs, CT 06269, or via e-mail: emily.myers@uconn.edu.



## Note

1. The failure of the Retroflex/Velar group to show persisting effects of categorization training on subsequent categorization and discrimination tasks remains somewhat of a mystery. The most likely explanation is that the retroflex and velar tokens are closer to one another in acoustic space, which has been shown to affect the relative discriminability of non-native contrasts (e.g., Polka, 1991). This was reflected by dramatically lower  $d'$  scores at pretest for both groups on the Retroflex/Velar contrast compared with the Dental/Retroflex contrast (Retroflex/Velar contrast: mean  $d' = 0.9325$ , Dental/Retroflex contrast: mean  $d' = 3.225$ ).

## REFERENCES

- Alain, C., Campeanu, S., & Tremblay, K. (2010). Changes in sensory evoked responses coincide with rapid improvement in speech identification performance. *Journal of Cognitive Neuroscience*, *22*, 392–403.
- Benson, R. R., Whalen, D. H., Richardson, M., Swainson, B., Clark, V. P., Lai, S., et al. (2001). Parametrically dissociating speech and nonspeech perception in the brain using fMRI. *Brain and Language*, *78*, 364–396.
- Best, C. C., & McRoberts, G. W. (2003). Infant perception of non-native consonant contrasts that adults assimilate in different ways. *Language and Speech*, *46*, 183–216.
- 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. *Journal of the Acoustical Society of America*, *109*, 775–794.
- Binder, J. R., Liebenthal, E., Possing, E. T., Medler, D. A., & Ward, B. D. (2004). Neural correlates of sensory and decision processes in auditory object identification. *Nature Neuroscience*, *7*, 295–301.
- Blumstein, S. E., Myers, E. B., & Rissman, J. (2005). The perception of voice onset time: An fMRI investigation of phonetic category structure. *Journal of Cognitive Neuroscience*, *17*, 1353–1366.
- Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glott International*, *5*, 341–345.
- Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America*, *101*, 2299–2310.
- Burton, M. W., Small, S. L., & Blumstein, S. E. (2000). The role of segmentation in phonological processing: An fMRI investigation. *Journal of Cognitive Neuroscience*, *12*, 679–690.
- Callan, D. E., Jones, J. A., Callan, A. M., & Akahane-Yamada, R. (2004). Phonetic perceptual identification by native- and second-language speakers differentially activates brain regions involved with acoustic phonetic processing and those involved with articulatory-auditory/orosensory internal models. *Neuroimage*, *22*, 1182–1194.
- Callan, D. E., Tajima, K., Callan, A. M., Kubo, R., Masaki, S., & Akahane-Yamada, R. (2003). Learning-induced neural plasticity associated with improved identification performance after training of a difficult second-language phonetic contrast. *Neuroimage*, *19*, 113–124.
- Chang, E. F., Rieger, J. W., Johnson, K., Berger, M. S., Barbaro, N. M., & Knight, R. T. (2010). Categorical speech representation in human superior temporal gyrus. *Nature Neuroscience*, *13*, 1428–1432.
- Cox, R. W. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, *29*, 162–173.
- Cox, R. W., & Jesmanowicz, A. (1999). Real-time 3D image registration for functional MRI. *Magnetic Resonance in Medicine*, *42*, 1014–1018.
- Desai, R., Liebenthal, E., Waldron, E., & Binder, J. R. (2008). Left posterior temporal regions are sensitive to auditory categorization. *Journal of Cognitive Neuroscience*, *20*, 1174–1188.
- Flege, J. E., MacKay, I. R., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *Journal of the Acoustical Society of America*, *106*, 2973.
- Formisano, E., De Martino, F., Bonte, M., & Goebel, R. (2008). "Who" is saying "what"? Brain-based decoding of human voice and speech. *Science (New York, N.Y.)*, *322*, 970–973.
- Francis, A. L., Baldwin, K., & Nusbaum, H. C. (2000). Effects of training on attention to acoustic cues. *Perception & Psychophysics*, *62*, 1668–1680.
- Francis, A. L., & Nusbaum, H. C. (2002). Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 349–366.
- Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. *Science*, *291*, 312–316.
- Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, *6*, 110–125.
- Golestani, N., Molko, N., Dehaene, S., LeBihan, D., & Pallier, C. (2007). Brain structure predicts the learning of foreign speech sounds. *Cerebral Cortex*, *17*, 575–582.
- Golestani, N., & Zatorre, R. J. (2004). Learning new sounds of speech: Reallocation of neural substrates. *Neuroimage*, *21*, 494–506.
- Golestani, N., & Zatorre, R. J. (2009). Individual differences in the acquisition of second language phonology. *Brain and Language*, *109*, 55–67.
- Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: Neural models of stimulus-specific effects. *Trends in Cognitive Sciences*, *10*, 14–23.
- Grill-Spector, K., & Malach, R. (2001). fMR-adaptation: A tool for studying the functional properties of human cortical neurons. *Acta Psychologica (Amsterdam)*, *107*, 293–321.
- Guenther, F. H., & Gjaja, M. N. (1996). The perceptual magnet effect as an emergent property of neural map formation. *Journal of the Acoustical Society of America*, *100*, 1111–1121.
- Guenther, F. H., Husain, F. T., Cohen, M. A., & Shinn-Cunningham, B. G. (1999). Effects of categorization and discrimination training on auditory perceptual space. *Journal of the Acoustical Society of America*, *106*, 2900–2912.
- Guenther, F. H., Nieto-Castanon, A., Ghosh, S. S., & Tourville, J. A. (2004). Representation of sound categories in auditory cortical maps. *Journal of Speech Language and Hearing Research*, *47*, 46–57.
- Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition*, *92*, 67–99.
- Hutchison, E. R., Blumstein, S. E., & Myers, E. B. (2008). An event-related fMRI investigation of voice-onset time discrimination. *Neuroimage*, *40*, 342–352.
- Joanisse, M. F., Zevin, J. D., & McCandliss, B. D. (2007). Brain mechanisms implicated in the preattentive categorization of speech sounds revealed using fMRI

- and a short-interval habituation trial paradigm. *Cerebral Cortex*, 17, 2084–2093.
- Kay, P., & Kempton, W. (1984). What is the Sapir-Whorf hypothesis? *American Anthropologist*, 86, 65–79.
- Kikutani, M., Roberson, D., & Hanley, J. (2008). What's in the name? Categorical perception for unfamiliar faces can occur through labeling. *Psychonomic Bulletin & Review*, 15, 787–794.
- Kuhl, P. K. (1991). Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics*, 50, 93–107.
- Landau, B., Dessalegn, B., & Goldberg, A. (2010). Language and space: Momentary interactions. In P. Chilton & V. Evans (Eds.), *Language, cognition and space: The state of the art and new directions. Advances in Cognitive Linguistics Series* (pp. 51–78). London: Equinox Publishing.
- Leech, R., Holt, L. L., Devlin, J. T., & Dick, F. (2009). Expertise with artificial nonspeech sounds recruits speech-sensitive cortical regions. *Journal of Neuroscience*, 29, 5234–5239.
- Lieberman, A. M., Delattre, P. C., & Cooper, F. S. (1958). Some cues for the distinction between voiceless and voiced stops in initial position. *Language and Speech*, 1, 153–157.
- Lieberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, 54, 358–368.
- Liebenthal, E., Binder, J. R., Spitzer, S. M., Posing, E. T., & Medler, D. A. (2005). Neural substrates of phonemic perception. *Cerebral Cortex*, 15, 1621–1631.
- Liebenthal, E., Desai, R., Ellingson, M. M., Ramachandran, B., Desai, A., & Binder, J. R. (2010). Specialization along the left superior temporal sulcus for auditory categorization. *Cerebral Cortex*, 20, 2958–2970.
- McCandliss, B. D., Fiez, J. A., Protopapas, A., Conway, M., & McClelland, J. L. (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, Affective, & Behavioral Neuroscience*, 2, 89.
- Miller, J. L. (1981). Some effects of speaking rate on phonetic perception. *Phonetica*, 38, 159–180.
- Myers, E. B. (2007). Dissociable effects of phonetic competition and category typicality in a phonetic categorization task: An fMRI investigation. *Neuropsychologia*, 45, 1463–1473.
- Myers, E. B., Blumstein, S. E., Walsh, E., & Eliassen, J. (2009). Inferior frontal regions underlie the perception of phonetic category invariance. *Psychological Science*, 20, 895–903.
- Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception and Psychophysics*, 60, 355–376.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, 9, 97–113.
- Özgen, E., & Davies, I. R. L. (2002). Acquisition of categorical color perception: A perceptual learning approach to the linguistic relativity hypothesis. *Journal of Experimental Psychology: General*, 131, 477–493.
- Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception and Psychophysics*, 15, 289–290.
- Polka, L. (1991). Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions. *Journal of the Acoustical Society of America*, 89, 2961–2977.
- Raizada, R. D. S., Tsao, F.-M., Liu, H.-M., & Kuhl, P. K. (2010). Quantifying the adequacy of neural representations for a cross-language phonetic discrimination task: Prediction of individual differences. *Cerebral Cortex*, 20, 1–12.
- Rotshtein, P., Henson, R. N. A., Treves, A., Driver, J., & Dolan, R. J. (2005). Morphing Marilyn into Maggie dissociates physical and identity face representations in the brain. *Nature Neuroscience*, 8, 107–113.
- Sabri, M., Binder, J. R., Desai, R., Medler, D. A., Leitl, M. D., & Liebenthal, E. (2008). Attentional and linguistic interactions in speech perception. *Neuroimage*, 39, 1444–1456.
- Seger, C. A., & Miller, E. K. (2010). Category learning in the brain. *Annual Review of Neuroscience*, 33, 203–219.
- Stevens, K., & Blumstein, S. (1975). Quantal aspects of consonant production and perception: A study of retroflex stop consonants. *Journal of Phonetics*, 3, 215–233.
- Swan, K., & Myers, E. (under review). Category labels induce boundary-dependent perceptual warping in learned speech categories.
- Talairach, J., & Tournoux, P. (1988). *A co-planar stereotaxic atlas of a human brain*. Stuttgart: Thieme.
- Tremblay, K., Kraus, N., & McGee, T. (1998). The time course of auditory perceptual learning: Neurophysiological changes during speech-sound training. *NeuroReport*, 9, 3557–3560.
- Turkeltaub, P. E., & Branch Coslett, H. (2010). Localization of sublexical speech perception components. *Brain and Language*, 114, 1–15.
- Vogels, R., Sary, G., Dupont, P., & Orban, G. A. (2002). Human brain regions involved in visual categorization. *Neuroimage*, 16, 401–414.
- Wang, Y., Sereno, J. A., Jongman, A., & Hirsch, J. (2003). fMRI evidence for cortical modification during learning of Mandarin lexical tone. *Journal of Cognitive Neuroscience*, 15, 1019–1027.
- 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, 1866–1878.
- Yeung, H. H., & Werker, J. F. (2009). Learning words' sounds before learning how words sound: 9-Month-olds use distinct objects as cues to categorize speech information. *Cognition*, 113, 234–243.
- Zevin, J. D., & McCandliss, B. D. (2005). Dishabituation of the BOLD response to speech sounds. *Behavioral and Brain Functions*, 1, 4.
- Zevin, J. D., Yang, J., Skipper, J. I., & McCandliss, B. D. (2010). Domain general change detection accounts for “dishabituation” effects in temporal-parietal regions in functional magnetic resonance imaging studies of speech perception. *Journal of Neuroscience*, 30, 1110.