Brain-behavior relationships in incidental learning of non-native phonetic categories

Sahil Luthra a,b,⁎, Pamela Fuhrmeister b, Peter J. Molfese c, Sara Guediche d, Sheila E. Blumstein e, Emily B. Myers a,b,f

a University of Connecticut, Department of Psychological Sciences, United States
b University of Connecticut, Department of Speech, Language and Hearing Sciences, United States
c National Institutes of Health, United States
d Basque Center on Cognition, Brain and Language, Spain
e Brown University, Department of Cognitive, Linguistic and Psychological Sciences, United States
f Haskins Laboratories, United States

ARTICLE INFO

Keywords:
Non-native phonetic perception
Implicit learning
Inferior frontal gyrus

ABSTRACT

Research has implicated the left inferior frontal gyrus (LIFG) in mapping acoustic-phonetic input to sound category representations, both in native speech perception and non-native phonetic category learning. At issue is whether this sensitivity reflects access to phonetic category information per se or to explicit category labels, the latter often being required by experimental procedures. The current study employed an incidental learning paradigm designed to increase sensitivity to a difficult non-native phonetic contrast without inducing explicit awareness of the categorical nature of the stimuli. Functional MRI scans revealed frontal sensitivity to phonetic category structure both before and after learning. Additionally, individuals who succeeded most on the learning task showed the largest increases in frontal recruitment after learning. Overall, results suggest that processing novel phonetic category information entails reliance on frontal brain regions, even in the absence of explicit category labels.

1. Introduction

Speech sounds have a complex internal structure, and in general, processing the fine-grained detail of these sounds relies on temporal brain regions such as the left superior temporal gyrus (LSTG; Desai, Liebenthal, Waldron, & Binder, 2008; Liebenthal, Binder, Spitzer, Possing, & Medler, 2005; Mesgarani, Cheung, Johnson, & Chang, 2014; Myers, 2007). These temporal areas show tuning that is specific and structured according to the acoustic details of one’s native language phonetic categories. However, a number of studies suggest that the perception of phonetic detail, even if largely supported by superior temporal cortex, is not entirely divorced from frontal brain regions. Individuals with Broca’s aphasia, for instance, have shown subtle deficits in phoneme discrimination, though they make fewer errors than individuals with posterior brain damage (Blumstein, Baker, & Goodglass, 1977). This notion has also been supported by functional neuroimaging studies of native language perception, with frontal brain regions implicated in different aspects of acoustic-phonetic processing (Lee, Turkeltaub, Granger, & Raizada, 2012; Myers, 2007; Rogers & Davis, 2017; Xie & Myers, 2018)). In particular, the left inferior frontal gyrus (LIFG) is sensitive to the proximity between an acoustic token and a phonetic category boundary (Myers, 2007; Myers, Blumstein, Walsh, & Eliassen, 2009) and responds to phonetic ambiguity in naturally-produced, continuous speech (Xie & Myers, 2018). While there are likely differentiable roles for frontal structures in the perception of speech, in general inferior frontal regions show evidence of abstraction away from low-level acoustic details in order to access category-level information about speech tokens (Chevillet, Jiang, Rauschecker, & Riesenhuber, 2013; Lee et al., 2012; Myers et al., 2009).

Further evidence for a role of frontal brain regions in speech perception comes from studies examining the acquisition of non-native phoneme categories. Non-native speech distinctions, especially those that are perceptually similar to existing native language categories, are very difficult to acquire in adulthood (Best & Tyler, 2007), with most adults falling short of native-like perceptual performance, even with targeted training (Golestani & Zatorre, 2009; Pruitt, Strange, Polka, &
Aguilar, 1990; Strange & Dittmann, 1984)). The extant research suggests that acquisition of new speech categories invokes processes in left frontal areas, among other neural systems. For instance, Golestani and Zatorre (2004) showed that newly-learned non-native stimuli activated the bilateral IFG (pars opercularis) and LSTG relative to a noise baseline, and Myers and Swan (2012) showed that an area of the left middle frontal gyrus (MFG) immediately adjacent to Broca’s area was sensitive to newly-acquired non-native category structure. One interpretation of these patterns is that non-native tokens activate emerging perceptual category information stored in the frontal lobe.

While several studies have shown frontal recruitment for non-native learning, evidence points to increased reliance on temporoparietal structures as listeners become more proficient (see Myers, 2014 for review). For instance, individual success in learning has been associated with reduced activation of LIFG (Golestani & Zatorre, 2004; Myers & Swan, 2012) and increased recruitment of temporoparietal regions such as the bilateral angular gyri (AG) (Golestani & Zatorre, 2004). These findings can be taken as evidence that listeners may initially recruit frontal regions to process non-native sounds but that as listeners develop better-elaborated representations of the novel phonetic categories, processing of these sounds may increasingly recruit temporal regions associated with sensory perception. Under such a view, the early reliance on frontal regions may reflect access to articulatory codes or abstract category-level representations that can be used to guide perception, or else may reflect high demands on phonological working memory (Callan, Jones, Callan, & Akahane-Yamada, 2004; Golestani & Zatorre, 2004; Myers, 2014).

The interpretation of the role of frontal areas for native as well as non-native speech perception is complicated because many studies examining phonetic learning have used explicit tasks during scanning, such as phoneme categorization (Callan et al., 2004; Golestani & Zatorre, 2004). What is not clear is whether category-relevant neural activation is driven by the metalinguistic demands of the task or by speech perception per se. Indeed, Hickok and Poeppel (2000, 2004) have argued that the involvement of frontal brain structures in perceiving acoustic-phonetic detail is limited to situations in which participants must explicitly attend to sub-lexical details of the stimulus, as is required in phoneme identification tasks.

Nonetheless, frontal recruitment for phonetic learning has been observed in the absence of an explicit task. In a study by Myers and Swan (2012), participants were exposed to a dental-retroflex-velar continuum (i.e., da-da-ga) and trained to categorize stimuli into two categories. Half of the participants learned that the category boundary was between the dental and retroflex tokens, and for the other half of the participants, the category boundary was between the retroflex and velar tokens. A short-interval habituation design (Zevin & McCandliss, 2005) was used during scanning: On every trial, participants heard a train of identical stimuli followed by a distinct stimulus, which either came from the same phonetic category as the preceding stimuli or came from the other category. Notably, participants were not asked to identify the category for the tokens they heard and instead only responded to occasional high-pitched catch trials. The bilateral MFG showed sensitivity to the learned category structure, suggesting a role for frontal regions in perceiving non-native phonemic distinctions even in the absence of an explicit identification task. However, it is important to note that the Myers and Swan (2012) study did use an explicit categorization task during training, so it is possible that participants were categorizing stimuli during the fMRI scan, despite not being required to do so.

Indeed, the vast majority of studies examining the perception of non-native phonemes have used training tasks in which participants are explicitly taught a category label that corresponds to each stimulus. This explicit information about category identity may reinforce the early, frontally-mediated stages of non-native phonetic learning (Myers, 2014). That is, the frontal activation associated with non-native phonetic learning may specifically reflect a mapping between stimuli and category labels, rather than reflecting (bottom-up) sensitivity to the underlying acoustic-phonetic category structure. As such, a more stringent test of a role for frontal regions in non-native phonetic learning would require the use of implicit paradigms during both the training and fMRI portions of the study, such that participants do not have labels for the categories being learned and therefore cannot categorize the stimuli, even implicitly.

In recent years, researchers have increasingly utilized implicit paradigms to train participants on novel categories. For instance, Leech, Holt, Devlin, and Dick (2009) examined the neural underpinnings of implicit auditory learning using complex non-speech stimuli. Over the course of several training sessions, participants played a video game where auditory cues were diagnostic of whether an upcoming visual exemplar was a member of one category (aliens to be captured) or another (aliens to be shot). Pre- and post-training fMRI sessions utilized an implicit oddball detection task, meaning that neither behavioral training nor the scanner task entailed explicit categorization. Results showed that better auditory learning was associated with increased reliance on STS post-training. More recently, Lim, Fiez, and Holt (2019) measured BOLD activity while participants played this incidental learning video game in the MRI scanner. The authors manipulated whether the non-speech auditory exemplars were organized into linearly separable categories (structured categories) or not (unstructured categories). Critically, the time course of activation in the basal ganglia – and more specifically, in the striatum – differed between structured and unstructured categories, consistent with a proposed role for the striatum in acquiring new behaviorally-relevant sound categories (Lim, Fiez, & Holt, 2014; Yi, Maddox, Mumford, & Chandrasekaran, 2016). While the authors focused their discussion on the striatum, this same pattern was also observed in a number of additional regions including the bilateral IFG. Further, striatal activity was positively correlated with changes in behavior and functionally connected to superior temporal sulcus. Taken together, such results suggest the involvement of a coordinated network of frontal, striatal, and temporal areas in auditory category learning, at least for non-speech sounds.

In general, incidental or implicit learning paradigms can yield successful non-native learning (Gabay & Holt, 2015; Lim & Holt, 2011), showing that consistent associations between category information and behaviorally relevant stimulus properties can increase sensitivity to novel sound distinctions. Vlahou, Protopapas, and Seitz (2012) used an incidental training paradigm to examine learning of two different sound categories. Native speakers of Greek heard two pairs of speech sounds (four sounds total) on every trial and were asked to identify whether tokens within the first pair or second pair differed in volume. Unbeknownst to subjects, one pair always consisted of two Hindi dental sounds while the other consisted of two Hindi retroflex sounds. Critically, the volume difference emerged only within the retroflex pair (i.e., the correct response always corresponded to the retroflex category). To ensure the task was appropriately challenging, the size of the volume difference within the retroflex pair was set adaptively, such that the task got harder (i.e., the volume difference got smaller) if participants succeeded on easier levels. Following training, subjects’ discrimination and identification abilities were tested explicitly. Vlahou and colleagues found that participants who completed the incidental learning task performed as well as or better than a group who received explicit training on the speech sounds, and both groups performed better than a group of naïve listeners. Thus, even though the incidental learning task itself did not require learning of the non-native phonemic contrast, the consistent temporal yoking of category-level information (the phonetic category difference) with a behaviorally relevant dimension (the volume difference) resulted in learning, consistent with other similarly structured studies of incidental learning (Seitz & Watanabe, 2005).

The aim of the current study is to examine the neural systems underlying the learning of a non-native phonetic category distinction using an incidental speech sound learning paradigm, specifically testing whether frontal regions are involved in non-native phonetic category
learning in the absence of explicit category labels. In Experiment 1, we leveraged the incidental learning paradigm used by Vlahou et al. (2012) to promote non-native learning of the Hindi dental-retroflex contrast. Functional activation was measured with fMRI both before and after three days of incidental learning, allowing us to examine whether frontal brain regions are recruited for processing phonetic detail when participants are not explicitly aware that they are being exposed to two novel speech sound categories. In Experiment 2, we examined the extent to which behavioral gains over the course of the incidental learning sessions depend on consistent associations between the phonetic category structure and the task-relevant changes in volume.

2. Experiment 1

In Experiment 1, we collected fMRI data to measure changes in brain activity that occur after three days of an incidental learning task designed to induce sensitivity to a non-native phonetic category difference. Crucially, participants were not informed of the categorical structure of the stimuli until after all scanning was completed, at which point their sensitivity to the non-native phonetic category structure was assessed explicitly.

2.1. Methods

2.1.1. Participants

Twenty participants who self-identified as right-handed were recruited from the University of Connecticut campus community. All participants were native speakers of American English with no history of speech, language, hearing or neurological impairments. All participants received monetary compensation for participating. Due to experimenter error, two participants received inconsistent condition assignments during the incidental learning task and were therefore excluded from analyses, resulting in a final n of 18 (age: mean = 22.6, SD = 2.2; 14 female). All procedures were approved by the University of Connecticut Institutional Review Board, and all subjects provided informed consent prior to participating.

2.1.2. Stimuli

Voiced dental (/d̪ɑ/) and retroflex (/ɖɑ/) stops had been recorded for a previous study (Earle & Myers, 2015) by a native speaker of Hindi. For each type of stop, five unique productions were edited to the onset of the burst, equated for length and scaled to 70 dB SPL mean amplitude. For each token, volume variants were created in Praat (Boersma & Weenink, 2017) by scaling the tokens to 70.5 dB, 71 dB, 72 dB, 73 dB and 74 dB. Additionally, a high-pitched variant was created for each token by increasing the F0 contour by 100 Hz in Praat.

2.1.3. Procedure

The experiment took place over four consecutive days; the full schedule of tasks is displayed in Fig. 1A, and the procedure for each task is described in detail below. On the first day, subjects participated in an initial fMRI scan and then completed their first session of the incidental learning task. On days 2 and 3, participants completed additional sessions of the incidental learning task. On day 4, participants completed another fMRI scan, after which they were informed that they had been exposed to two speech sounds from Hindi. At this point, a behavioral posttest was conducted to explicitly assess identification and discrimination for the two categories.

2.1.3.1. fMRI sessions

Scanning took place on a 3-T Siemens Prisma using a 64-channel head coil. Anatomical images were acquired sagittally using a T1-weighted magnetization-prepared rapid acquisition gradient echo (MP-RAGE) sequence (TR = 2300 ms, TE = 2.98 ms, FOV = 256 mm, flip angle = 9 degrees, voxel size = 1 mm × 1 mm × 1 mm). Diffusion-weighted images were acquired during the first scan session and magnetic resonance spectroscopy data were collected during the second session; those data are not presented here. Functional images were acquired with a T2*-weighted EPI sequence using a slow event-related design (TR = 4.0 s [2.1 s scan with a 1.9 sec delay], TE = 25 ms, FOV = 192 mm, flip angle = 90 degrees, slice thickness = 2.5 mm, in-plane resolution = 2 mm × 2 mm). Thirty-six slices were acquired per TR, and slices were acquired in an interleaved, ascending fashion. A sparse sampling method was used to ensure that auditory information was presented during silent intervals between scans (Edmister, Talavage, Ledden, & Weisskoff, 1999; Fig. 1C), and trials were presented during the silent gap every 3 TRs, yielding a stimulus onset asynchrony of 12 s. In each scan session, participants completed five functional runs of 36 trials each; 110 volumes were acquired per run, and each run lasted approximately 7.5 min.

A short-interval habituation paradigm was used for the in-scanner task. On each trial, participants heard four stimuli in quick succession with a 50-ms ISI (Fig. 1B). The first three tokens on a given trial were always repetitions of a single token. On within-category trials, the fourth token was a different production of the same syllable, spoken by the same speaker. On between-category trials, the fourth token was from the other phonetic category. Because subjects did not have to make judgments about the category membership of the stimuli, the design provided an implicit measure of neural responses to phonetic category information (between-category versus within-category) (Zevin & McCandliss, 2005). There were 80 between-category and 80 within-category trials in each scan session. Each session also included 20 attentional catch trials; on these trials, a high-pitched token replaced one stimulus in the train. Participants were instructed to press a button whenever they heard a high-pitched stimulus. Two versions of the scanner task were created, with subjects receiving one version during their first scan and the other during the second (with order counterbalanced). Participants also completed a set of 15 practice trials before each session.

2.1.3.2. Incidental learning sessions

The incidental learning task was programmed in MATLAB (The Mathworks Inc., Natick, MA, USA) using Psychtoolbox (Brainard & Vision, 1997). Each day’s session took approximately 20 min to complete and consisted of 200 trials, presented in four blocks of 50 trials. On each trial, participants heard two pairs of tokens and were asked to identify whether the tokens in the first pair or the tokens in the second pair differed in volume. Unbeknownst to the subject, one pair was composed of two dental tokens and the other was composed of two retroflex tokens (Fig. 1D).

Critically, the volume difference was always associated with the same non-native phoneme (counterbalanced across participants). The volume difference could be either 4 dB, 3 dB, 2 dB, 1 dB or 0.5 dB, with the precise difference determined by a 3-down-1-up adaptive staircase (Fig. 1E), such that if a subject correctly identified the target pair on three consecutive trials, the volume difference was reduced by one step. The volume difference was made larger after any incorrect response (e.g., it was raised to a 4-dB difference if the incorrect response occurred on a trial with a 3-dB difference). In this way, task difficulty was modulated by the subject’s performance on the task. The rationale for this design is that because the task-relevant dimension (the volume difference) was consistently associated with phonetic information, it would enhance learning of the phonetic category information, particularly as the volume difference became smaller.

The quieter token (always 70 dB) was presented first for half the target pairs, and the louder token was presented first for the remaining half. The amplitude of non-target stimuli was consistent with the subject’s place on the adaptive staircase. For instance, if a subject was on a 3 dB trial, the non-target stimuli were either both at 73 dB or both at 70 dB. Within each pair, the same dental or retroflex token was always presented, even if the amplitude differed; across all trials, all dental and all retroflex tokens were presented. There were an equal number of retroflex-first trials as dental-first trials, with order held constant across subjects. Tokens within a pair were separated by a mean ISI of 250 ms (SE = 3 ms), and pairs of tokens were separated by a mean ISI of 500 ms.
The full procedure took place over four days. In both Experiment 1 and Experiment 2, participants completed three days of incidental behavioral training and received explicit behavioral assessments (an identification task and an AX discrimination task) on the fourth day. Participants in Experiment 1 also completed MRI sessions (shown in dashed boxes) on Days 1 and 4.

A short-interval habituation task was used during scanning. On within-category trials, participants first heard three identical stimuli and then an acoustically distinct token from the same phonetic category. On between-category trials, participants heard three identical stimuli and then a stimulus from the contrastive phonetic category. Participants did not make overt responses to either within-category or between-category trials; instead, they only responded to occasional attentional catch trials, in which one of the four tokens in the stimulus train was presented at a higher pitch (raised by 100 Hz). A sparse sampling design was used for the scanner task, whereby stimuli were presented in silent gaps that fell between scans. To induce sensitivity to the phonetic category distinction, the incidental learning task used training trials in which participants heard four speech sounds and had to decide whether the first pair or second pair contained a volume difference. Unbeknownst to participants, one pair consisted of two dental tokens and the other of two retroflex tokens. In Experiment 1, the volume difference was consistently associated with one phonetic category, and Experiment 2 tested whether learning depended on this consistent association. The second half of each learning session also included some probe trials, in which neither pair contained a volume difference; of interest was whether participants would make their selection in line with previous trials. On the fourth day of the experiment, participants completed ID and discrimination posttests, as shown. The incidental learning sessions used an adaptive staircase structure. Participants needed to respond correctly to three consecutive trials before they moved to the next difficulty level (smaller volume difference), and an incorrect response moved them back to the previous difficulty level (larger volume difference). A sample participant’s trajectory is shown here; a threshold value is computed as the average of the inflection points on the staircase within a block of 50 trials. Note that it should be very challenging to detect very small volume differences based on volume information alone, so it behooves participants to leverage the consistent association between phonetic category information and volume information to succeed on the volume task. As such, we assume that lower thresholds on the volume task reflect incidental learning of the phonetic category distinction.
(SE = 3 ms).
To assess whether participants were relying on phonetic cues, ten probe trials were scattered throughout the second half of each session (Fig. 1D). For these trials, neither pair contained a volume difference; of interest was whether subjects would consistently choose the pair type on which they were being trained or whether they would be at chance in their responses. Unlike the results reported by Vlahou et al. (2012), no significant differences in probe trial performance were found, so detailed results are not reported here.

2.1.3.3. Behavioral posttest. All participants completed a behavioral posttest to assess their sensitivity to the dental-retroflex contrast; token amplitude was not manipulated during this posttest. The posttest began with an initial familiarization portion, during which participants heard each of the 10 tokens (5 from each category) paired with a category label (D1 or D2). Participants then completed a singleton identification task in which they heard one token at a time and were asked to categorize the stimulus as belonging to D1 or D2. Each participant received ten trials for each condition, resulting in 100 total trials; the same random order was used for all participants. Finally, participants completed a pair discrimination task in which they heard two tokens and were asked whether they came from the same category or different categories. There were 100 total discrimination trials, allowing each possible trial combination to be presented once, and the same random order was used for all participants. No feedback was given during the identification or discrimination trials.

2.1.4. fMRI analyses

Neuroimaging data were analyzed in AFNI (Cox, 1996). Because we obtained oblique data for our functional runs, EPI data were first rotated to cardinal orientation to match the coordinates of the anatomical data. Preprocessing was done separately for each run. In particular, an afni.proc.py script was used to register functional volumes to the first volume of each run, to align EPI data to the anatomical data, and to align functional data with AFNI's Colin27 template in Talairach and Tournoux (1988) space; these transformations were done in a single warp to minimize interpolation. Data were smoothed using a 4-mm full-width half-maximum Gaussian kernel and scaled to a mean of 100 for each run to represent percent signal change.

To conduct univariate analyses, idealized hemodynamic response functions (HRFs) for each condition (between-category, within-category, attentional catch trials) were created by convolving stimulus onset times with a gamma function; trial onsets were measured from the start of the first token in the stimulus train. For each subject, pre-processed BOLD data were submitted to a deconvolution analysis with the condition HRFs (between, within, and catch trials) and six rigid-body motion parameters as regressors; false alarm trials (in which participants had made a button response to a trial without a pitch change) were censored (i.e., those rows were removed from the regression matrix). Using the AFNI program 3dMVM (Chen, Adleman, Saad, Leibenluft, & Cox, 2014), beta coefficients were submitted to two group-level ANCOVAs with Session (Pre/Post) and Phonetic Category (Between/Within) as categorical factors. Individual Threshold scores from the final day of training were used as a continuous covariate in one ANCOVA (as these scores are assumed to reflect the degree of learning in the incidental learning task), and another ANCOVA looked for relationships with individual post-test identification performance.

A group mask containing only voxels that were imaged in all the participants was applied at the ANCOVA stage. Subsequently, a small volume correction was applied, limiting analyses to regions known to be involved in language processing (bilateral IFG, MFG, insula, supramarginal gyr, AG, STG, middle temporal gyr, and transverse temporal gyr) as well as the striatum, a subcortical region thought to be involved in incidental learning of auditory categories (e.g., Lim et al., 2019). These regions were defined using the Talairach and Tournoux (1988) atlas built into AFNI; the set of voxels considered is shown in Fig. 3A. To correct for multiple comparisons, we first estimated the noise smoothness for each subject by applying the 3dFWHMx command to the residual times series; notably, we used a mixed autocorrelation function as suggested by Cox, Chen, Glen, Reynolds, and Taylor (2017) in order to address concerns about Type I error raised by Eklund, Nichols, and Knutsson (2016). Estimated smoothness values were averaged across subjects, and these averages were used in Monte Carlo simulations to estimate the likelihood of noise-only clusters. Simulations indicated that a cluster size of 218 voxels was needed at a voxel-level significance of p < 0.05 to yield cluster-level threshold of α < 0.05.

2.2. Results

2.2.1. Behavioral results

To examine the trajectory of learning during the incidental learning sessions, we calculated a dB threshold measure for each block of 50 trials by taking the average of the inflection points (dB level at a change in direction) on the adaptive staircase within that block (Fig. 1E). If there were no inflection points within a block (i.e., if the participant never advanced beyond the initial 4 dB difficulty level), the modal difficulty level was taken as the threshold. Conceptually, this threshold measure estimates the smallest volume difference at which participants can reliably make a correct response. The threshold scores for each block are visualized in Fig. 2A.2 We then used a linear mixed effects model to estimate how threshold levels changed over time, implementing the model in R using the mixed function of the “afex” package (Singmann, Bolker, Westfall, & Aust, 2018). This model included a fixed factor of Block (mean-centered) nested within a fixed factor of Session (mean-centered); we also included random intercepts for each subject. 3 A likelihood ratio test yielded only a significant main effect of Session, χ²(2) = 39.74, p < 0.0001, suggesting that participants in Experiment 1 improved at the volume detection task from day to day.

Next, we analyzed participants’ ability to identify the correct phonetic category for each speech sound during the behavioral posttest. To account for participants who may have been able to distinguish the speech sounds but confused the labels in the categorization phase, response labels were reversed if a subject’s mean accuracy was at or below 0.41; this affected 3 participants. This criterion was selected because the binomial probability of obtaining a score of 0.41 or lower by chance was less than 5%. Performance on the identification task is visualized in Fig. 2B. Participants in Experiment 1 had a mean accuracy of 0.67 (SE: 0.03). A one-sample t-test indicated that this was significantly above chance, t(17) = 24.89, p < 0.001.

We next considered participants’ explicit discrimination of the non-native categories, as assessed on the behavioral posttest. To account for...
potential effects of response bias, percent accuracy scores were con-
verted to $d'$ scores (Macmillan & Creelman, 2004). Discrimination data
displayed in Fig. 2C. Participants in Experiment 1 had an averaged‘
of 0.56 (SE: 0.10), significantly above what would be expected by
chance, $t(17)=5.44$, $p < 0.001$.

Finally, we examined the relationship between subjects' perfor-
mance on the various behavioral tasks. Correlation tests revealed that
participants who performed well on the identification task also per-
formed well on the discrimination task, $r=0.72$, $t(16)=4.21$,
$p < 0.001$, and that participants who succeeded on the volume task (as
measured by lower mean threshold scores on the third day) did better
on the discrimination posttest, $r = -0.47$, $t(16) = -2.13$, $p = 0.05$.
There was no significant correlation between success on the volume
task and performance on the identification task, though the relationship
was in the expected direction, $r = -0.35$, $t(16) = -1.50$, $p = 0.15$.

2.2.2. Univariate fMRI results

A univariate analysis examined potential effects of Session (pre-
training vs post-training) and Phonetic Category (between-category trials
vs within-category trials). To examine potential differences in activation
due to individual differences in learning, mean Threshold scores from the final day of incidental learning were included as a continuous
covariate. Results are summarized in Table 1 and visualized in Fig. 3;
the full mask of voxels considered in analyses is shown in Fig. 3A.

The MFG and IFG bilaterally were less active after training (main
effect of Session, Fig. 3B). We also observed that even though particip-
ants did not have explicit labels for the two categories when com-
pleting the scanner task, frontal regions – specifically, the LIFG and
LMFG – were sensitive to phonetic category-level information, showing
relatively more activation for between-category trials than within-
category trials even before training (main effect of Phonetic Category,
Fig. 3C). Finally, we observed an interaction between Session and
Threshold, whereby participants with the lowest final thresholds (i.e.,
those who succeeded most in the learning task, as measured on the
third session) showed relatively stronger recruitment of left IFG/MFG
after learning compared to before; by contrast, those who performed
worst on the incidental learning task showed reduced activation of LIFG
on the second scan compared to the first (Fig. 3D).

In a parallel analysis, we considered effects of Session and Phonetic
Category, using accuracy on the ID task as a continuous covariate.
Results are summarized in Table 2 and visualized in Fig. 4. As expected,
comparable results were obtained for the main effects of Session and of
Phonetic Category. This analysis also revealed a cluster in left tempor-
oparietal cortex that was sensitive to overall accuracy on the identifi-
cation task; however, this effect was driven by three outlier partici-
pants, and so we do not consider it further (Fig. 4A). Finally, this
analysis identified several clusters showing an interaction between
Session and ID, such that participants who were most accurate on the
identification task showed the most pronounced increases in recruit-
ment of bilateral IFG and left insula after training as compared to before
(Fig. 4B).

2.3. Discussion

2.3.1. Frontal contributions to novel perceptual category learning

The current results support other findings that frontal structures –
particularly in the left inferior frontal gyrus and middle frontal gyrus –
detect phonetic changes in passive, oddball-type paradigms (Myers &
Swan, 2012; Myers et al., 2009). Of interest, this sensitivity to the ca-
tegory status of the tokens was present across sessions, as participants
showed differential activation in response to a change in phonological category (for between-category dental-retroflex trials compared to within-category trials). Previous studies showing category-level sensitivity in the left frontal lobe have used well-established native language phonological categories (Chevillet et al., 2013; Lee et al., 2012; Myers et al., 2009) or explicitly-trained non-native categories (Myers & Swan, 2012). This frontal sensitivity to phonetic category changes has been interpreted as reflecting access to information about the category status of tokens at a level abstracted from the acoustic input. Yet in the current study, the dental and retroflex stops are presumably both heard as variations on the alveolar /d/ for English-speaking listeners and should therefore activate the same phonological category, at least prior to incidental learning (e.g., the perceptual assimilation model, Best & Tyler, 2007). Put differently, prior to any learning, between-category tokens did not yet have any category status (implicit or otherwise) to map on to. Nonetheless, dental and retroflex tokens do differ from one another acoustically, and the acoustic differences encountered in between-category trials (i.e., the difference between a retroflex token and a dental token) are necessarily greater than the acoustic differences encountered in within-category trials (e.g., the difference between two dental tokens). The frontal response to phonetic category change may thus reflect a passive detection of auditory change (see Zevin, Yang, Skipper, & McCandliss, 2010), rather than necessarily activation of established phonological categories. We suggest that this response can still be modulated by experience, since lifetime exposure to native sounds or laboratory training on non-native sounds may change the salience of certain relevant (between-category) phonetic dimensions (see also Holt & Lotto, 2006). That is, frontal regions may be recruited in response to any auditory change, though the specific degree of recruitment may depend on the relevance of changes along that particular auditory dimension, which may in turn vary among individuals depending on their

### Table 1
Results of analysis considering Session, Phonetic Category and Threshold. Coordinates and F-value correspond to peak activation in cluster. Approximate Brodmann areas are given in parentheses.

<table>
<thead>
<tr>
<th>Anatomical region</th>
<th>Maximum intensity coordinates (x y z)</th>
<th>Number of activated voxels</th>
<th>F value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Session (Post-Pre)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Left inferior frontal gyrus (BA 45)/Left insula (BA 13)</td>
<td>−37 17 8</td>
<td>257</td>
<td>56.31</td>
</tr>
<tr>
<td>2. Right middle frontal gyrus (BA 9)</td>
<td>45 27 34</td>
<td>239</td>
<td>25.80</td>
</tr>
<tr>
<td><strong>Phonetic Category (Between-Within)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Left inferior frontal gyrus (BA 47)</td>
<td>−43 39 −12</td>
<td>356</td>
<td>45.82</td>
</tr>
<tr>
<td>2. Left middle frontal gyrus (BA 9)</td>
<td>−45 11 24</td>
<td>226</td>
<td>12.65</td>
</tr>
<tr>
<td><strong>Session × Phonetic Category</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No significant clusters</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Threshold</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No significant clusters</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Session × Threshold</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Left inferior frontal gyrus (BA 45)/Left middle frontal gyrus (BA 46)</td>
<td>−43 33 8</td>
<td>430</td>
<td>31.81</td>
</tr>
<tr>
<td>No significant clusters</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Phonetic Category × Threshold</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No significant clusters</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Session × Phonetic Category × Threshold</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No significant clusters</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 3. Results of fMRI analysis considering effects of Session, Phonetic Category and Threshold (our behavioral measure of performance during incidental learning). While a volumetric approach was used for statistical analyses, results are visualized on the anatomical surface of a single subject. FreeSurfer (Fischl, 2012) was used for surface reconstruction, and SUMA (Saad & Reynolds, 2012) was used to map volume-based statistical maps to the surface reconstruction. (A) Analyses were limited to a set of cortical regions known to be involved in language processing and the striatum, which has been implicated in the learning of auditory categories. (B) Frontal regions bilaterally showed greater activation on the first scan session than the second. (C) Between-category trials elicited greater activation in left frontal regions than did within-category trials. (D) An interaction between Session, Phonetic Category and Threshold emerged in left inferior frontal gyrus. This interaction is visualized in the associated scatterplot, with each data point indexing an individual subject; a trend line shows the general relationship. Threshold scores are plotted on the x-axis, while the y-axis indicates the change in activation from the first scan to the second one.
individual language experiences.

We also observed a general reduction in frontal recruitment from the first scan session to the second one. However, the fact that there is often considerable variability in behavioral success makes it challenging to interpret group-level changes in activation over time without also accounting for behavioral performance. Indeed, recruitment of frontal brain regions has been shown to differ as a function of learning, with better learners tending to rely less on frontal brain regions over time (Golestani & Zatorre, 2004; Myers & Swan, 2012). We thus examined how frontal involvement was modulated by subjects’ out-of-scanner behavioral performance. Critically, we observed increased recruitment of left frontal structures in those participants who consistently reached the hardest difficulty levels on the task used in training (the Session × Threshold interaction). The training paradigm was structured so that phonetic information (i.e., whether the pair was dental or retroflex) served as a redundant cue to the volume difference, so we infer that successful performance at the hardest levels of the volume task could result from implicit detection of the phonetic category differences. One interpretation of this finding is that listeners who capitalized on the phonological structure of the training developed emerging category sensitivity in frontal regions, consistent with results from more explicit training paradigms (Myers & Swan, 2012). Greater categorization success at post-test was associated with a similar pattern. Namely, we observed increased recruitment of bilateral frontal structures in the participants who performed best on the explicit identification task that took place after the final scan session (the Session × ID interactions). Taken together, the results suggest that participants who succeeded most on the behavioral tasks were also those who showed the largest increases in reliance on frontal brain structures.

These findings can be explained in the context of reverse hierarchy theory (Ahissar, Nahum, Nelken, & Hochstein, 2009). This theory proposes that rapid perception is based primarily on higher-level representations and that changes in perceptual encoding of fine-grained low-level detail emerge only over time. For speech sound learning, the anatomical correlate of this hypothesis is that sensitivity to novel phonetic category distinctions is predicted to emerge first in more domain-general, non-sensory neural systems (i.e., frontal systems), and only over time does experience retune perceptually-sensitive regions (i.e., temporal regions) (see also Myers, 2014; Reetzke, Xie, Llanos, & Chandrasekaran, 2018). Notably, we did not observe sensitivity to phonetic category structure in temporal regions. In light of previous work showing that temporal recruitment is tied to participants’ degree of behavioral success (Leech et al., 2009), it may be the case that participants in the current study did not show temporal recruitment because they did not progress past relatively early stages of learning; this notion is also supported by the generally weak identification and discrimination abilities of our participants (Section 2.3.3).

### 2.3.2. Potential striatal contributions to category learning

Recent work has posited a role for the basal ganglia, and more specifically the striatum, in the acquisition of novel auditory categories ((Lim et al., 2014, 2019; Yi et al., 2016)). In particular, activity of the prefrontal cortex and the anterior dorsal striatum (the head of the caudate) have been linked to the use of a reflective, rule-based system for category learning, whereas engagement of the posterior dorsal striatum (the tail and body of the caudate as well as the putamen) has been linked to the use of a more procedural reflexive system, the latter which has been argued to be better suited to learning speech sound categories ((Chandrasekaran, Yi, & Maddox, 2014; Lim, Fiez, & Holt, 2014; Yi, Maddox, Mumford, & Chandrasekaran, 2016)). In this way, the striatum has been theorized to support the coordination between frontal and temporal regions in auditory category learning ((Lim et al., 2014)).

Notably, no significant effects were observed in the striatum in the present study. We suspect that this is attributable to the fact that incidental learning sessions were conducted outside the scanner, in contrast to studies where the learning task and scanning occur concurrently (e.g. Lim et al., 2019). Other studies have suggested that the activity in specific sub-regions of the striatum may depend on trial-by-trial performance feedback (((Lim, Fiez, & Holt, 2014; Tricomi, Delgado, McCandliss, McClelland, & Fiez, 2006)), which was not provided in the present study.

---

### Table 2

Results of analysis considering Session, Phonetic Category and ID Accuracy. Coordinates and F-value correspond to peak activation in cluster. Approximate Brodmann areas are given in parentheses.

<table>
<thead>
<tr>
<th>Anatomical region</th>
<th>Maximum intensity coordinates</th>
<th>Number of activated voxels</th>
<th>F-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Session (Post-Pre)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Left middle frontal gyrus (BA 46)</td>
<td>$x = -31$</td>
<td>$y = 35$</td>
<td>$z = 30$</td>
</tr>
<tr>
<td>2. Left inferior frontal gyrus (BA 45)/Left insula (BA 13)</td>
<td>$x = -37$</td>
<td>$y = 17$</td>
<td>$z = 8$</td>
</tr>
<tr>
<td>3. Right middle frontal gyrus (BA 9)</td>
<td>$x = 43$</td>
<td>$y = 27$</td>
<td>$z = 34$</td>
</tr>
<tr>
<td><strong>Phonetic Category</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Left inferior frontal gyrus (BA 47)</td>
<td>$x = -43$</td>
<td>$y = 39$</td>
<td>$z = -12$</td>
</tr>
<tr>
<td>2. Left middle temporal gyrus (BA 21)</td>
<td>$x = -55$</td>
<td>$y = -13$</td>
<td>$z = -14$</td>
</tr>
<tr>
<td>3. Left middle frontal gyrus (BA 9)</td>
<td>$x = -51$</td>
<td>$y = 13$</td>
<td>$z = 44$</td>
</tr>
<tr>
<td><strong>ID</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Left superior temporal gyrus (BA 42)/Left supramarginal gyrus (BA 40)/Left transverse temporal gyrus (BA 41)</td>
<td>$x = -55$</td>
<td>$y = -41$</td>
<td>$z = 22$</td>
</tr>
<tr>
<td><strong>Session × Phonetic Category</strong></td>
<td>No significant clusters</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Phonetic Category × ID</strong></td>
<td>No significant clusters</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Session × Phonetic Category × ID</strong></td>
<td>No significant clusters</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
2.3.3. Behavioral success in incidental learning

Behavioral performance during the training task suggests that participants may have begun to learn the relevant dimensions for non-native category discrimination as they reached progressively harder difficulty levels (lower dB thresholds) on the volume change task. However, behavioral gains did not consistently generalize to behavioral success on the posttests, where there was considerable variability in subjects’ behavioral performance. Indeed, many subjects performed at near-chance levels on the identification and discrimination tasks. The relatively inconsistent posttest performance of participants in Experiment 1 may be partly attributable to fatigue, as these participants completed the posttest assessment immediately after spending an hour in the MRI scanner. Furthermore, the unstructured exposure to the sounds that subjects encountered during the in-scanner sessions may have attenuated the overall amount of non-native phonetic learning in this group (Fuhrmeister & Myers, 2017). We suggest that future work examines how frontal recruitment relates to the ultimate level of behavioral success on non-native phonetic learning tasks, given previous work showing reduced reliance on frontal regions in better learners (in contrast to the current findings; Golestani & Zatorre, 2004; Myers & Swan, 2012) as well as theoretical accounts positing relatively greater reliance on temporoparietal regions as individuals’ perceptual performance improves (e.g., Myers, 2014).

3. Experiment 2

While Experiment 1 supports a role for frontal brain regions in the development of sensitivity to non-native phonetic category structure, it is unclear how much of this is attributable to learning per se. The incidental learning paradigm used in Experiment 1 was adapted from a study conducted by Vlahou et al. (2012), who demonstrated that subjects who had completed incidental learning sessions were more sensitive to phonetic category structure than a group of naïve participants. Learning observed in the incidentally trained participants could be
attributable to the structure of the incidental learning task, as Vlahou et al. suggested. However, it is also possible that the results reflect the fact that incidentally trained participants had more exposure to the stimuli than did the naïve group of listeners. That is, it is unclear whether behavioral gains brought about by this incidental learning task are merely a result of exposure to the stimuli, or whether a consistent pairing of the volume discrimination with one of the sound categories is necessary for learning. Experiment 2 examines this issue directly.

3.1. Experiment 2: Methods

3.1.1. Participants

Sixty adults (38 female) were recruited from the University of Connecticut. All subjects were monolingual native speakers of American English with no history of neurological, speech, hearing or language impairments. All participants received course credit or monetary compensation for their participation, and all provided informed consent prior to participating.

Participants were assigned to one of three groups (20 subjects per group). One group of participants completed the incidental learning task used in Experiment 1, in which phonetic category information was consistently associated with the task-relevant volume change. A second group completed the same protocol, but for these participants, the phonetic category information was inconsistently associated with the volume difference. Finally, a third group completed only the posttest assessments without any prior exposure to the stimuli.

3.1.2. Stimuli

The same stimuli were used as in Experiment 1 apart from the pitch-shifted tokens, since participants in Experiment 2 did not complete the scanner task.

3.1.3. Procedure

Participants in the consistent association and inconsistent association groups participated in three lab training sessions and completed a behavioral posttest on the fourth session, with each session occurring on consecutive days. Subjects in the no exposure group completed only the behavioral posttest and were used as a baseline against which to compare the other two groups.

Participants in the consistent association group completed the same protocol as in Experiment 1, apart from completing the fMRI sessions. For participants in the inconsistent association group, the protocol was identical with one key difference: The volume difference occurred within the reflexbox for half of a subject’s trials and within the dental pair for the other half. In this way, participants in the consistent association and inconsistent association groups both received equal exposure to the auditory tokens over the three incidental learning sessions, and both groups engaged in a challenging volume-discrimination task. However, individuals in the inconsistent association group were not able to take advantage of a systematic association between phonetic cues and volume to succeed on the training task. As such, the comparison between the consistent association and inconsistent association groups allows us to evaluate the extent to which consistent associations between the phonetic category distinction and the volume difference support incidental learning.

3.2. Experiment 2: Results

3.2.1. Incidental learning sessions

The threshold data from Experiment 2 are displayed in Fig. 2A. Following our approach for analyzing threshold data in Experiment 1, a linear mixed effects model was used to assess group differences in threshold over time, allowing us to model both the fixed effects of interest and random variation between subjects. In particular, we modeled fixed effects of Group (consistent association, inconsistent association) as well as fixed effects of Block (mean-centered) nested within Session (mean-centered). As before, we also modeled random intercepts for each subject. A likelihood ratio test revealed a main effect of Session, \( \chi^2(2) = 21.71, p < 0.0001 \), and a Group by Session interaction, \( \chi^2(2) = 9.23, p = 0.002 \). No other main effects or interactions were significant.

To unpack this Group by Session interaction, we ran separate mixed effects models for each group, dropping the fixed effect of Group but otherwise keeping the same model structure as in the omnibus model. For the inconsistent association group, a likelihood ratio test revealed no significant effects. In contrast, we obtained a main effect of Session for the consistent association group, \( \chi^2(2) = 28.46, p < 0.0001 \). The effect of Session indicates that participants in the consistent association group improved in their performance on the volume detection task from day to day, suggesting that these participants may have been implicitly relying on the phonetic category information in order to perform the task. By contrast, the inconsistent association group did not receive a systematic pairing between the volume difference and the phonetic category distinction and therefore could not rely on this information to improve performance on the volume detection task.

3.2.2. Behavioral posttests

Data from the identification task are visualized in Fig. 2B; note that we applied the same label correction as in Experiment 1. Participants who received a consistent association between the volume difference and phonetic category information had a mean accuracy of 0.69 (SE: 0.01), those who received an inconsistent association had a mean accuracy of 0.67 (SE: 0.01), and those who had received no exposure prior to testing had a mean accuracy of 0.62 (SE: 0.01). Label-corrected trial-by-trial data were submitted to a logistic mixed effects model using the glmer function of the “lme4” package (Bates, Maechler, Bolker, & Walker, 2015) to test for group differences. The fixed factor of Group (consistent association, inconsistent association, and no exposure) was dummy-coded with the consistent association set as the reference level; the model also included random by-subject intercepts. The model found that participants who received a consistent association between phonetic category information and volume differences performed marginally better than participants who received no exposure to the stimuli prior to the behavioral posttest, \( \beta = -0.33, SE = 0.19, z = -1.81, p = 0.07 \). Further, the model found no significant difference between the performance of participants who received consistent associations and participants who received inconsistent associations, \( \beta = -0.12, SE = 0.18, z = -0.63, p = 0.53 \).

Discrimination data are displayed in Fig. 2C. Participants in the consistent association group had a mean \( \text{d'} = 0.48 \) (SE: 0.08), those in the inconsistent association group had a mean \( \text{d'} = 0.60 \) (SE: 0.07), and those in the no exposure group had a mean \( \text{d'} = 0.41 \) (SE: 0.09). To assess the baseline ability of naïve participants to discriminate the two categories, we conducted a one-sample t test on the \( \text{d'} \) scores of the no exposure group; results indicated that participants were able to discriminate the two categories at above-chance levels without any training, \( t(19) = 4.44, p < 0.001 \). We then submitted \( \text{d'} \) scores from all three groups to a linear regression with Group (consistent association, inconsistent association, and no exposure) as a between-subjects factor. As before, the factor of Group was dummy-coded with the consistent association group used as a reference level. There were no significant differences in discrimination ability between participants who received consistent associations during incidental learning and those who received

---

4 Here, we benefited from an additional advantage of mixed effects models, which is in how they handle missing data (Baayen, Davidson, & Bates, 2008). Due to a programming error on one of the testing computers, some data collected in the first few days were not saved for the first few subjects. This resulted in the loss of 8 sessions’ worth of data from the inconsistent association group, distributed across 6 participants (13.3% of the all the inconsistent association data).
inconsistent associations, $\beta = 0.11$, SE = 0.12, $z = 0.94$, $p = 0.351$. Furthermore, there was no significant difference in the discrimination abilities of subjects who received consistent associations and those who received no exposure to the stimuli prior to the behavioral assessments, $\beta = -0.08$, SE = 0.12, $z = -0.66$, $p = 0.510$.

Finally, we examined potential correlations between the behavioral measures in each group. For the consistent association group, there was a significant positive correlation between performance on the identification task and performance on the discrimination task, $r = 0.51$, $t(18) = 2.54$, $p = 0.02$. There was also a significant correlation in the expected direction between performance on the volume task (measured by mean thresholds on the third session) and performance on the discrimination task, $r = -0.54$, $t(18) = -2.74$, $p = 0.01$, but no significant correlation between performance on the volume task and performance on the identification task, $r = -0.04$, $t(18) = -0.17$, $p = 0.87$. For the inconsistent association group, there was a significant correlation between discrimination and identification scores, $r = 0.50$, $t(18) = 2.46$, $p = 0.02$. However, there was no significant correlation between performance on the volume task and performance on the discrimination task, $r = -0.13$, $t(16) = -0.53$, $p = 0.60$, and no significant correlation between performance on the volume task and performance on the identification task, $r = 0.03$, $t(16) = 0.11$, $p = 0.92$.

3.3. Experiment 2: Discussion

Experiment 1 leveraged an incidental learning task used by Vlahou et al. (2012) to induce sensitivity to a non-native phonetic category distinction. Vlahou et al. found that participants who had completed this incidental learning task later showed better discrimination and identification of these non-native speech sounds than did a group of naïve participants. In Experiment 2, we showed weak evidence in support of their finding, as participants who completed an incidental learning task performed marginally better than naïve participants on an identification task. Vlahou et al. found that participants who had completed category-level codes (Myers et al., 2009) or articulatory codes (e.g., Wilson, Saygin, Sereno, & Iacoboni, 2004). However, the frontal activation in the current study cannot be attributed to these factors, since participants completed passive tasks during training and scanning (only completing the explicit phonetic categorization tasks after their final fMRI session); furthermore, participants had neither knowledge of how to produce the dental and retroflex tokens nor any knowledge or their differing category status. A necessary caveat in interpreting these results is that because the incidental paradigm did not result in strong perceptual performance at the group level, participants may not have developed clear dental and retroflex categories. As such, the evolution of the frontal response to categorical differences, and the degree to which the processing burden begins to include temporoparietal areas, may differ substantially when learners acquire sounds in a more elaborated, naturalistic fashion. We suggest that future investigations consider using increasingly naturalistic paradigms to differentiate the core neural systems involved in phonetic category acquisition from task-specific effects.

Author statement

All the authors declare that we had no conflict of interest with respect to the theoretical questions or results of this research. All experiments were conducted following ethical guidelines put forth by the University of Connecticut’s Institutional Review Board. The authors are all in agreement of the content of this manuscript, which has not been published elsewhere and is not under consideration at any other journal.

Statement of significance

Research on the acquisition of non-native speech sound categories has suggested an important role for frontal areas such as the left inferior frontal gyrus. Here, we investigate whether such frontal recruitment is a consequence of experimental procedures (e.g., task demands to map non-native sounds to explicit category labels).
Acknowledgments

This research was supported by NIH grant R01 DC013064 to EBM and NIH NIDCD Grant R01 DC066220 to SEB. The authors thank F. Sayako Earle for assistance with stimulus development; members of the Language and Brain lab for help with data collection and their feedback throughout the project; Elisa Medeiros for assistance with collection of fMRI data; Paul Taylor for assistance with neuroimaging analyses; and attendees of the 2016 Meeting of the Psychonomic Society and the 2017 Meeting of the Society for Neurobiology of Language for helpful feedback on this project. We also extend thanks to two anonymous reviewers for helpful feedback on a previous version of this manuscript.

References