Neural Systems Underlying Perceptual Adjustment to Non-Standard Speech Tokens

Emily B. Myers[a,b,c,d,*] and **Laura M. Mesite**[c,d]

[a]University of Connecticut, Department of Speech, Language, and Hearing Sciences, 850 Bolton Road, Storrs, CT 06269

[b]University of Connecticut, Department of Psychology, 406 Babbidge Road, Storrs, CT 06269

[c]Brown University, Department of Cognitive, Linguistic, and Psychological Sciences, 190 Thayer Street, Providence, RI 02912

[d]Haskins Laboratories, 300 George Street #900, New Haven, CT 06511

Abstract

It has long been noted that listeners use top-down information from context to guide perception of speech sounds. A recent line of work employing a phenomenon termed 'perceptual learning for speech' shows that listeners use top-down information to not only resolve the identity of perceptually ambiguous speech sounds, but also to adjust perceptual boundaries in subsequent processing of speech from the same talker. Even so, the neural mechanisms that underlie this process are not well understood. Of particular interest is whether this type of adjustment comes about because of a retuning of sensitivities to phonetic category structure early in the neural processing stream or whether the boundary shift results from decision-related or attentional mechanisms further downstream. In the current study, neural activation was measured using fMRI as participants categorized speech sounds that were perceptually shifted as a result of exposure to these sounds in lexically-unambiguous contexts. Sensitivity to lexically-mediated shifts in phonetic categorization emerged in right hemisphere frontal and middle temporal regions, suggesting that the perceptual learning for speech phenomenon relies on the adjustment of perceptual criteria downstream from primary auditory cortex. By the end of the session, this same sensitivity was seen in left superior temporal areas, which suggests that a rapidly-adapting system may be accompanied by more slowly evolving shifts in regions of the brain related to phonetic processing.

*Corresponding Author: Emily Myers, Department of Speech, Language, and Hearing Sciences, University of Connecticut, 850 Bolton Road, Storrs, CT 06269, Emily.myers@uconn.edu, 860-486-2630.

1. Introduction

The speech we listen to is full of perceptual problems to be solved. Slips of the tongue produce imperfect or ambiguous speech tokens, environmental noise may occlude the speech signal, and different talkers, for reasons of idiolect or dialect, implement speech sounds in ways that consistently deviate from 'standard' pronunciations (e.g. Allen, Miller, & DeSteno, 2003). An important question in speech perception is how listeners resolve such indeterminacies in order to map the speech signal to meaning. Solving this problem requires a degree of functional plasticity in our speech perception system.

A line of research suggests that listeners can capitalize upon regularities in the speech signal in order to adjust phonetic category boundaries. One source of information that may help listeners resolve ambiguities in the speech signal is top-down context. Evidence suggests that listeners use the local linguistic context to adjust their responses to phonetic input, particularly when that input is ambiguous. For example, Ganong (1980) observed that lexical context shifted participants' phonetic category boundaries in a categorization task, such that an ambiguous token midway between/g/ and /k/ was likely to be perceived as a /g/ sound in the context of a gift—kift continuum, and as a /k/ sound in the context of a giss—kiss continuum. These types of boundary adjustments have been shown as a function of other lexical phenomena, such as neighborhood density and lexical frequency, and from semantic and syntactic expectation (Borsky, Tuller, & Shapiro, 1998; Connine, Titone, & Wang, 1993; Newman, Sawusch, & Luce, 1997). Taken together, these results suggest that disambiguating lexical, sentential, or semantic information in the local linguistic context plays a role in guiding the resolution of phonetic ambiguity.

At the same time, top-down information not only plays a role in shaping responses to local perceptual indeterminacies, but also in guiding longer-term adjustments to talker idiolect. A line of research in recent years has extended Ganong-type effects to investigate *persisting* effects of lexically-conditioned shifts in phonetic perception, termed 'perceptual learning in speech' (Norris, McQueen, & Cutler, 2003). In a series of related studies, researchers demonstrated that listeners who are exposed to multiple exemplars of an ambiguous phoneme blend (e.g. one between 's'-/s/ and 'sh'-/ʃ/) in an unambiguous lexical context (e.g. in the place of the 's' in 'Tennessee'), show subsequent shifts in the categorization function for items along an unbiased non-word to non-word continuum (e.g. /asi/ to /aʃi/) (see among others Clarke-Davidson, Luce, & Sawusch, 2008; Kraljic & Samuel, 2005, 2007; Norris et al., 2003).

One interpretation of these studies is that phonetic processing must be permeable to the effects of top-down information. Crucially, the phonetic processing system must adjust to top-down expectation both to resolve local perceptual indeterminacies (see the Ganong effect, above) but also to accommodate phonetic detail as it varies between talkers. However, this view has not gone unchallenged. In particular, it has been argued that lexically-guided phonetic category adjustments result not from an influence of top-down information on a phonetic level of processing, but instead from post-perceptual processing (McClelland, Mirman, & Holt, 2006, Norris et al., 2003).

Neuroimaging studies have informed this debate. Two investigations of the classic lexical effect phenomenon reported that activation in the left superior temporal gyrus (STG) was modulated by the top-down effects of lexical context (Gow, Segawa, Ahlfors, & Lin, 2008; Myers & Blumstein, 2008). The lexically-biased shift in the location of the phonetic category boundary was reflected by modulation of activity in the left STG as well as left inferior frontal areas (Myers & Blumstein, 2008). Data from a multi-modal imaging study (Gow et al., 2008) suggested that lexically-modulated shifts in perception may arise from the early interaction of information in the supramarginal gyrus (SMG) with fine-grained phonetic processing in the STG. The fact that lexically-biased shifts in perception are reflected in regions relatively early in the *neural* processing stream gives rise to the suggestion that lexical information contacts phonetic stages of processing quite early in the *language* processing stream

What is unclear, however, is how neural systems underlying phonetic processing respond when lexical information guides *longer-term* adjustment of the phonetic category boundary, as in the case of a listener adapting to a new talker's accent. This question differs from that investigated in Ganong effect studies in that rather than focusing on the integration of top-down information during phonetic processing (i.e. during lexical decision), the question is how this top-down information is stored and used for subsequent processing. When listeners use lexical information to adjust to a talker idiosyncrasy, either online or offline use of lexical information disambiguates the ambiguous sounds during the lexical decision task. As listeners are exposed to increasing numbers of these resolved ambiguities, shifts in categorization for speech sounds for that talker begin to emerge. At least two non-exclusive mechanisms for these shifts are possible. First, it may be the case that exposure to disambiguated non-standard speech variants over time will fundamentally re-tune the responsiveness of populations of neurons that respond to phonetic category identity and phonetic category structure in the superior temporal lobes (Chang et al., 2010; Desai, Liebenthal, Waldron, & Binder, 2008; Guenther, Nieto-Castanon, Ghosh, & Tourville, 2004; Myers, Blumstein, Walsh, & Eliassen, 2009). On the other hand, one might imagine that dynamically adjusting the perceptual tuning for non-standard sounds for each talker that we encounter might introduce instability into the phonetic processing system. Given that these non-standard variants deviate significantly from the expected phonetic category shape, it may instead be the case that the processing of such tokens is guided by decision-related or attentional mechanisms. Specifically, a new decision criterion may be generated for non-standard tokens such that they are processed according to the talker's novel speech pattern.

The goal of the current study was to investigate the neural systems that underlie lexically-mediated perceptual learning for speech using fMRI. While in the scanner, participants listened to speech from a talker in which an ambiguous phonetic token between /s/ and /ʃ/ substituted for either an /s/ sound (S Group) or a /ʃ/ sound (SH Group). Subsequent shifts in the location of the phonetic category boundary were assessed by examining categorization functions for stimuli sampled from a non-word to non-word, /asi/ to /aʃi/ continuum. Behavioral responses as well as activation to the ambiguous tokens were measured, and data from a separate control group who did not participate in scanning was used to estimate the average location of the 'unshifted' phonetic category boundary in the population.

Behaviorally, we predicted that there would be a difference in the location of the phonetic category boundary across groups, such that the S Group would show more 's' categorizations for tokens near the category boundary, and the SH Group would show the opposite pattern, thus replicating previous studies using this paradigm (e.g. Kraljic & Samuel, 2005).

Previous imaging work has demonstrated increased activation in both frontal and temporal areas for phonetic tokens which fall on the phonetic category boundary compared to those further from the boundary (Binder, Liebenthal, Possing, Medler, & Ward, 2004; Myers & Blumstein, 2008; Myers, 2007). This increase in activation for near-boundary tokens may reflect increased perceptual and executive resources necessary for categorization of ambiguous tokens. While many regions are likely to show greater activation to boundary value tokens, we predicted that only a subset would show sensitivity to the shifted boundary. In particular, behavioral group differences in the location of the phonetic category boundary should be accompanied by differences in activation between groups for these near-boundary tokens. To be more concrete: if the 40% /s/ token falls on the category boundary for the 'S Group' and the 50% /s/ token falls on the category boundary for the 'SH Group,' we expect to see greater activation for each group for exactly the token that is nearest to the boundary. Activation data were queried for this type of pattern by searching for a Group by Continuum Point interaction across brain regions. This approach resembles that of a previous investigation of the Ganong effect (Myers & Blumstein, 2008). The location of those regions which show a lexically-conditioned sensitivity to the category boundary are informative in that they allow us to make predictions regarding the locus of this perceptual effect.

Two questions are addressable using this data. First, we ask whether the same neural systems implicated for processing typical native speech sounds, namely the bilateral STG and left Inferior Frontal Gyrus (IFG), are likewise implicated for this behavioral phenomenon. Second, we ask how perceptual regions in the STG/STS and executive regions in the IFG may work in concert to produce shifts in phonetic category decisions. Namely, if neural evidence of category boundary adjustment is seen in the STG/STS, this may suggest that feedback from the lexicon to phonetic category structure results in a subtle alteration of perceptual sensitivities. On the other hand, if this kind of interaction is seen only frontal regions, this may be taken as evidence that adjustments in phonetic boundary are instead driven by shifts in decision criteria or attention.

Finally, because this study employs an alternating exposure-test design, it is possible that sensitivity to the shifted category boundary might emerge over time as participants hear increasing numbers of 'altered' tokens in each successive block of the LD task. To investigate the time course of emerging neural sensitivity to the shifted boundary, we searched for regions that showed a significant interaction between time, group, and condition. Of particular interest in the current study is whether increasing exposure results in a shift in the neural systems handling the perceptual adjustment.

While activation differences across groups for the phonetic categorization task are of the greatest theoretical interest, differences in activation for phonetically altered vs. unaltered words in the LD task might reflect the engagement of additional neural resources for

processing perceptually indeterminate speech sounds (e.g. "per?onal"). Given that both groups of participants encountered a mix of altered and unaltered words, no group difference in activation during the LD task is predicted.

2. Methods and Materials

2.1. Stimulus selection

We based the items for the lexical decision task on stimuli from Appendix A of Kraljic and Samuel (2005). These consisted of 20 critical /s/ words, 20 critical /ʃ/ words, 60 filler words, and 100 filler non-words. The critical words were two to five syllables in length and contained either /s/ or /ʃ/ in a syllable-initial position towards the end of the word. None of the critical words contained any instance of /z/ or /ʒ/ nor did the filler words contain any instance of /s/, /ʃ/, /z/, or /ʒ/. The two sets of critical words were matched with each other and the filler words on number of syllables and frequency (see Kraljic & Samuel, 2005 for details]. Kraljic and Samuel (2005) created the non-words by changing one phoneme per syllable for each of the 60 filler words from the study. They created an additional 40 non-words in the same manner using real words that were not included in the study. In order to match for biphone probability between real words and non-words, we substituted ten of the non-words; we calculated biphone probabilities using the Irvine Phonotactic Online Dictionary (Vaden, Halpin, & Hickok, 2009). We replaced one of Krajic and Samuel's critical /ʃ/ words, *negotiate*, with *brushing*.

2.2. Stimulus construction

We recorded all stimuli digitally using a female native speaker of American English. For the critical /s/ and /ʃ/ words, the speaker produced a second version of each word in which she replaced the critical phoneme that normally occurs in the word (either /s/ or /ʃ/) by the other phoneme (e.g. 'dinosaur' and 'dinoshaur'). We trimmed all of the items to the borders of the word using BLISS, (Mertus, 2000) and we normalized RMS energy to 70 db for all LD items using PRAAT (Boersma & Weenink, 2013).

We extracted the /s/ and /ʃ/ portions for both versions of each critical word at zero-crossings at the onset of the frication noise, and cut the longer fricative to match the length of the shorter one. We blended the fricatives in five different proportions from 30% /s/ -- 70% /ʃ/, to 70% /s/ -- 30% /ʃ/ by using waveform averaging in PRAAT. Based on experimenter judgment, we determined that the 50% blend would provide the most ambiguous sounding fricative for each critical word. We then inserted each of these unique 50% blends into the /s/-frame (e.g., *Arkansas, ambison*) of twenty /s/ and twenty /ʃ/ words. Finally, we created a continuum by blending the fricatives extracted from /asi/ and /aʃi/ together at several different proportions from 20% /s/ to 80% /s/ and inserting each of the resulting blends into the /asi/ frame. This resulted in a seven point continuum ranging in equal steps from /asi/ to /aʃi/. Because time limitations during scanning prevented the collection of a full sample of neural data on items from the full continuum, we conducted a behavioral pilot to determine which points along the continuum were likely to be susceptible to shifts in categorization following the LD exposure. On this basis, we selected four points along the continuum that constituted an unambiguous /asi/ exemplar (70%s), an unambiguous /aʃi/

exemplar (20%s), and two ambiguous exemplars (40%s and 50%s) that had been previously shown to lie near the phonetic category boundary (50% /s/ categorization point) for both the S and SH Groups, respectively. We used these four tokens in a phonetic categorization (PC) task (see 'Behavioral Design and Procedure' below).

2.3. Participants

We recruited twenty-eight participants from Brown University and its surrounding community to participate in the fMRI experiment and paid them for their participation. We eliminated four participants from the analysis because they scored less than 75% correct on one or more condition in the LD task (see Behavioral Design and Procedure, below) leaving a total of 24 participants (13 females). Participants ranged from 18 to 40 years of age (mean age = 26 years) and were all right-handed as indicated by a modified version of the Oldfield Handedness Inventory (Oldfield, 1971). All participants reported that they were native speakers of American English with normal hearing and no known history of neurological disease.

We recruited fifteen participants as behavioral controls (see 'Baseline' Control Group, below). All control subjects were native speakers of American English between the ages of 18 and 45 years of age who reported no known hearing deficits. All participants gave written informed consent in accordance with the guidelines established and approved by Brown University's Human Subjects Committees and we screened MRI participants for MRI compatibility

2.4. Behavioral Design and Procedure: Pretest in the lab for fMRI participants

We randomly assigned participants in the experimental group to either the ambiguous /ʃ/ group (SH:12 participants, 8 female) or the ambiguous /s/ group (S:12 participants, 5 female). In order to confirm that participants were able to perform the LD and PC tasks, we pre-trained both groups of participants in the lab before the MRI session. While the behavioral procedure resembled that of Kraljic & Samuel (2005), we chose an alternating task design, such that 5 LD runs and 5 PC runs alternated with a LD run always preceding each PC run. We chose this alternating design in order to maximize the probability that shifts in perception would be maintained over the course of the experiment and would not be cancelled out by experience during the PC task. Moreover, this alternating design allowed us to look for changing sensitivity as participants were exposed to increasing numbers of 'altered' words. We presented all ten blocks in a fixed, alternating order. We trained participants in the lab first, and they then returned one to two weeks later to repeat the training during the MRI portion of the study. Previous work has shown that shifts in category boundary emerge quickly and persist at least 24 hours after exposure to the LD task (Eisner & McQueen, 2006). As such, we expected a certain amount of carry-over of learning from the in-lab pre-test to the MRI session.

2.4.1. Lexical Decision Task—During the LD task, we asked participants to indicate whether they heard a word or non-word by pressing the corresponding button as quickly and as accurately as possible with either their right index finger or middle finger. Stimuli consisted of words in which an ambiguous /s/-/ ʃ / phoneme was embedded in critical /s/-

words (S Group) or in the critical /ʃ/-words (SH Group) while the phonemes in other critical, filler, and non-words remained unaltered. Each LD run consisted of 4 /s/-words (ambiguous for the S Group, unaltered for the SH Group), 4 /ʃ/-words (unaltered for the S Group, ambiguous for the SH Group), 12 filler words, and 20 non-words for a total of 40 trials per LD run.

2.4.2. Phonetic Categorization Task—Each PC run consisted of eight exemplars of each of the four continuum points taken from the /asi/ to /aʃi/ continuum. During the PC task, participants were asked to indicate whether they heard /asi/ or /aʃi/ by pressing the corresponding button in a similar manner.

2.5. 'Baseline' Control Group

Fifteen control subjects participated in a behavioral version of the fMRI experiment in which no ambiguous stimuli were presented. The purpose of this control study was to determine the average phonetic category boundary between /s/ and /ʃ/ in the absence of any altered tokens. The design and procedure were identical to that of the behavioral pretest except that subjects heard only unaltered versions of the critical words, and that the trials for all subjects were presented in the same pseudorandom order that they were presented in at the scanner. Participants were tested one at a time in a soundproof booth and did not receive any feedback as to the accuracy of their responses.

2.6. fMRI Design and Procedure

Participants in the experimental groups performed a replication of the experimental procedure from the behavioral pretest in the MRI scanner. Importantly, participants were assigned to the same exposure group (e.g. 'SH group') during pretest and scanning. All responses were registered using an MRI-compatible button box placed under each subject's right hand (Mag Design and Engineering, Redwood City, CA), and button mapping was counterbalanced across participants. As in the behavioral pilot, the fMRI experiment consisted of ten separate runs presented in a fixed order (alternating LD and PC) with trials within the runs presented in a fixed, pseudorandom order. Stimuli were delivered over air-conduction headphones (Avotec Inc) that were custom-designed to fit inside the ear canal and provide approximately 30 dB of ambient sound attenuation. Stimuli were evenly assigned to SOAs of 3, 6, and 9 seconds. Accuracy data and reaction time (RT) data were collected with RTs measured from the onset of the stimulus.

2.6.1. Magnetic Resonance Imaging—Functional and anatomical brain images were collected using a 3T Siemens Trio scanner (Siemens Medical Systems, Erlangen, Germany). For anatomical coregistration, high-resolution 3D T1-weighted anatomical images were acquired using a multi-echo MPRAGE sequence (time repetition [TR] = 2530 ms; time echoes [TE] = 1.64 ms, 3.5 ms, 5.36 ms, 7.22 ms; time for inversion [TI] = 1100 ms; 1-mm^3 isotropic voxels; 256×256 matrix) and reconstructed into 176 slices. Functional images were acquired using a 32-channel head coil. EPI data was acquired using a clustered acquisition design, such that stimuli were presented during one second of silence that followed each two-second volume acquisition (effective TR: 3 seconds). Slices were acquired in an ascending interleaved order and consisted of 27 2mm-thick echo planar (EPI)

axial slices with a 2mm isotropic in-plane resolution (TR = 3000 ms, TE = 32 ms, flip angle = 90 degrees 256 × 256 matrix, FOV = 256mm^3). Functional slabs were angled to cover perisylvian regions including the superior temporal gyri and superior temporal sulci as well as inferior frontal gyri extending to inferior portions of the middle frontal gyri. We told participants to keep their eyes closed during the tasks and reminded them to remain as still as possible.

2.7. Behavioral Data Analysis

We scored LD data for accuracy such that we considered responses to altered and unaltered words to be correct if they were labeled as words. We removed participants who scored < 75% accurate on any of the four conditions from the analysis. For the PC data, we plotted the identification functions for each subject, and calculated s/ʃ boundary values by converting percent /ʃ/-scores for each continuum point to z-scores, selecting the critical range from the minimum to maximum z-score values, submitting the selected z-scores to a linear regression that output slope and y-intercept values, calculating the x-intercept (which corresponds to the category boundary), and averaging the boundary value for each PC run in order to arrive at a single boundary value for each subject. We ran a one-way analysis of variance (ANOVA) to compare boundary values between groups, as well as a repeated measures ANOVA including group (S, SH, and baseline control) as a between-subjects factor and continuum point (20%s, 40%s, 50%s and 70%s) as a within-subjects factor to compare categorization on all points between groups.

2.8. MRI Data Analysis

2.8.1. Image Preprocessing—We used AFNI (Cox, 1996) to analyze the imaging data. EPI images were transformed from oblique to cardinal orientation, motion-corrected using a six-parameter rigid-body transform, aligned to the subject's own anatomical dataset, transformed to Talairach space (Talairach & Tournoux, 1988) using an automated affine registration algorithm (@auto_tlrc, AFNI), and spatially smoothed with a 4-mm Gaussian kernel. Further analysis was restricted to those voxels in which signal was recorded for at least 20 out of the 24 subjects. We created individual time series files for each of the 4 LD conditions (/ʃ/-words, /s/-words, fillers, non-words) for each subject and removed LD errors [see Behavioral Data Analysis, above] from their respective conditions and included them as a separate vector (Errors). We also created time series files for each of the 4 PC conditions (20%s, 40%s, 50%s, 70%s). For a secondary analysis, we modeled separate stimulus vectors for each condition for each run, which allowed us to include Time as a factor. Each stimulus start time was convolved with a stereotypic gamma hemodynamic function (Cohen, 1997), which resulted in a modeled hemodynamic time course for each of the nine stimulus conditions.

2.8.2. Statistical Analysis—In order to validate phonetic categorization results with respect to previous imaging studies on the perception of phonetic category structure, we performed a planned comparison between Boundary and Endpoint tokens. To this end, we entered percent signal change data for continuum points into a mixed-factor ANOVA with group as a fixed factor (S, SH), condition as a fixed factor (4 PC conditions), and subjects as a random factor. Planned comparisons included the contrast between Boundary value (40%

and 50% /s/) and Endpoint (20% and 70% /s/) tokens in the PC task. A separate t-test was performed to contrast Altered and Unaltered words in the LD task.

Behavioral results (see below) suggested that, as predicted, shifts in perception were most evident at the two near-boundary tokens (40%s and 50%s) while the endpoint tokens were treated similarly by both groups. Moreover, given the alternating design of the current study, these near-boundary would be those most likely to show changes in activation over successive experimental runs. To this end, we performed a 2×2×5 mixed-factor ANOVA including within-subjects factors of Condition (40%s vs. 50%s) and Time (each of 5 successive PC runs), and a between-subject factor of Group (S vs. SH Group), with subjects entered as a random factor.

We masked ANOVA outputs with a small volume-corrected group mask that included anatomically defined regions selected from the San Antonio Talairach Daemon (Lancaster et al., 2000) that are typically implicated in language processing, including both left and right homologues: middle frontal gyrus (MFG), inferior frontal gyrus (IFG), the insula, STG, middle temporal gyrus (MTG), left supramarginal gyrus (SMG), Heschl's gyrus, and angular gyrus (AG). Cluster-level correction for multiple comparisons was determined by Monte Carlo simulations using AFNI's 3dClustSim program. Two approaches were used to estimate smoothness values entered into the simulation. Using a less conservative approach (Ward, 2000), the 4mm smoothing kernel was entered as a smoothness estimate. In addition, a more conservative approach was applied in which smoothing in each dimension was estimated using the residuals image from the GLM. Ten thousand iterations using each estimate were performed using the small-volume corrected group mask. Statistical masks were thresholded at a voxel level p<0.05, and clusters which survive correction for multiple comparisons at p<0.05 using the less conservative estimation (106 contiguous voxels) are displayed. In cluster tables, below, a single asterisk indicates clusters that survive the more lenient smoothing estimation (*, 106 contiguous voxels) and double asterisks mark clusters that survive the more conservative estimation (**, 166 contiguous voxels).

3. Results

3.1. Phonetic Categorization

3.1.1. Behavioral Results—A repeated measures ANOVA on all continuum points for all three groups revealed a main effect of condition ($F_{3,108} = 201.59$, MSE = 290.33, $p < 0.001$, $\eta^2 = 0.820$), reflecting the fact that tokens at the /s/ end of the continuum were more likely to be categorized as /s/, irrespective of group [see Figure 1A]. The ANOVA also revealed an interaction between group and condition ($F_{6,108} = 4.113$, $p < 0.001$, $\eta^2 = 0.033$), and a main effect of group ($F_{2,36} = 3.927$, MSE = 975.10, $p < .029$, $\eta^2 = 0.179$). Post-hoc between-groups tests revealed that the S Group made significantly more /s/ responses during PC than the SH Group overall (p < 0.03), while the control group did not differ significantly from either the S or the SH Group.

The current study employed an alternating LD/PC design, which has not been utilized in previous studies of this type. Given that we spaced the LD information over the course of the experiment, it is possible that successive PC runs might show changes in the location of

the phonetic category boundary over time as participants were exposed to accumulating information in the lexical decision task. We computed the location of the phonetic category boundary for each experimental run separately. We submitted these boundary values to a 3×5 ANOVA with Group (Control, S, and SH) as a between-subjects factor and Time (5 successive PC Runs) as the within-subjects factor [Figure 1C]. This analysis revealed a main effect of Group that approached significance ($F_{2,38}$=3.179, p=0.053, η^2=0.143), but no significant main effect of Time ($F_{4,152}$=1.173, p=0.325), nor any Group*Time interaction ($F_{8,152}$=1.692, p=0.105). The boundary value for the S Group fell closer to the 40%s continuum point, while the boundary value for the SH Group fell closer to the 50%s continuum point [see Figure 1A, 1C]. Post-hoc tests revealed that while the S Group (M = 37.54, SD = 13.26) had a significantly different boundary value than the SH Group (M = 47.60, SD = 9.41; significant at $p < 0.05$, Bonferroni corrected), the control group (M = 41.089, SD = 9.05) did not differ significantly from the other two groups.

We collected reaction time (RT) data for all PC decisions. Of note, the control group showed much faster RTs as compared to the fMRI participants. This is likely due to the substantial differences in environment for these two tasks—namely fMRI participants are lying on their backs and cannot see the button box. In order to determine whether RT data would also support the finding that the S and SH Groups have different category boundaries, we performed a repeated measures ANOVA comparing RTs to the two boundary continuum points and the S and SH Groups. We found no main effect of group, but there was a significant main effect of continuum point ($F_{3,66} = 30.177$, MSE = 157,168, $p < 0.001$, η^2=0.511), and a significant interaction between continuum point and group ($F_{3,66} = 6.884$, MSE = 35,853, $p < 0.001$, η^2=0.117) such that both groups had slower RTs to the stimuli that fell closest to their category boundary. Specifically, the S Group's slower RT response was to the 40%S continuum point, and the SH Group had a slower RT to the 50%s continuum point [see Figure 1B], which suggests that LD training shifted the groups' s/ʃ boundary. This result is consistent with many other studies demonstrating increased RTs in phonetic categorization tasks as stimuli approach the category boundary (Connine & Clifton, 1987; Miller, 1994; Myers, 2007).

3.1.2. Imaging Results

3.1.2.1. Sensitivity to the phonetic category boundary across groups: We entered all continuum points (20% s, 40%s, 50%s, and 70%s) into a mixed-factor ANOVA in order to explore the differences in processing of the two near-boundary stimuli (40%s and 50%s) versus the two endpoint stimuli (20%s and 70%s) across groups. The output of the ANOVA indicated that while regions in the bilateral IFG, and the left TTG were significantly more activated for boundary points (40% and 50%) than endpoints (20% and 70%) across groups, clusters in the bilateral MTG and left SMG were significantly more activated for endpoints than boundary points [see Table 1. Figure 2]. Clusters in both the right and left IFG were centered in the pars triangularis and extended into the pars opercularis and insula. Both clusters in the bilateral temporal lobes were centered in the MTG with only the cluster in the left MTG extending into the left SMG. Mean percent signal change data were extracted from each functional cluster and are displayed in Figure 2. When we entered the mean percent signal change extracted from these functional clusters into a 2×2 repeated measures

ANOVA comparing boundary points and group, only the right MTG cluster indicated a significant interaction ($F_{1,22} = 5.85$, MSE = 0.007, $p = 0.02$). This cluster overlaps with the right MTG cluster revealed in the Group × Condition interaction analysis, below.

3.1.2.2. Neural sensitivity to shifted phonetic category boundaries: Because shifts in the location of the category boundary most affected the two continuum points near that boundary (40%s and 50%s), we performed a targeted analysis to investigate differences in activation for these two points as a function of Group and Time. Below we report results of a 2×2×5 ANOVA including factors of Group (S vs. SH), Condition (40%s and 50%s) and Time (five successive PC runs). Results of all main effects and interactions are reported in Table 2, below. Of primary interest for the current discussion are those interactions that involve both Group and Condition, as these analyses reflect brain regions that show differential sensitivity to phonetic tokens as a function of the type of exposure that the participants received during the LD task.

Sensitivity to the shifted category boundary, irrespective of time: Based on previous work (Blumstein, Myers, & Rissman, 2005; Myers & Blumstein, 2008; Myers, 2007), we predicted that participants would show greater activation for tokens that fell near the phonetic category boundary (40%s for S group, 50%s for SH Group-see Figure 1). Clusters in the right MFG and IFG, and STG/MTG revealed a Group × Condition interaction [see Table 2, Figure 3]. The pattern of activation in the frontal cluster showed greater activation for tokens that fell on each group's shifted phonetic category boundary. The pattern in the right MTG cluster was reversed, but also showed deactivation overall (see Discussion). In general, both clusters show a pattern consistent with the allocation of more neural resources for the token which falls nearest the to the category boundary after exposure.

Changes in sensitivity to the shifted category boundary over time: While regions that showed a significant Group × Condition interaction were centered in the right hemisphere, the regions identified in the Group × Condition × Time interaction were instead centered in the left hemisphere. In particular a cluster in the left IFG and two clusters in the left STG, one in the anterior portions of the left STG and extending into the MTG, and the other in the posterior margins of the STG, all showed such an interaction (Table 2, Figure 4). While there were differences in the pattern of activation, particularly in early runs, all three clusters shared a pattern of activation in which boundary value tokens (40%s for the S group, 50%s for the SH Group) showed increasing divergence from non-boundary tokens by the end of scanning (Figure 4). This suggests that these regions reflect an increasing differentiation of neural sensitivity to the shifted category boundary over increasing exposure to ambiguous tokens within the LD task.

Main Effects of Group and Condition: While in principle there is no reason to expect a main effect of group, behavioral data showed that compared to behavioral controls, only one group (the SH Group) showed a significant shift in category boundary compared to a behavioral Control group. A plausible interpretation of this pattern is that the SH Group successfully engaged lexical-level information that resulted in a boundary shift whereas the S Group did not, which might be reflected in group-level differences in activation. The

output of the 2×2×5 mixed-factor ANOVA revealed a main effect of group in functional clusters within the bilateral STG, right AG, and right IFG (see Table 2). Overall, the S Group had significantly greater activation in response to these boundary points than the SH Group in all four functional clusters. While we interpret this result with caution, this may reflect greater ease of processing in the SH group when talker-specific information can successfully be brought to bear.

We observed a main effect of continuum point in a cluster centered in the left insula. Subjects had more activation for the 40%s continuum point than the 50%s continuum point in this functional region [See Table 2]. No other clusters emerged for this contrast.

3.2. Lexical Decision

3.2.1. Behavioral Results—Table 2 summarizes the accuracy data for all conditions among all three groups. A 2×2 repeated-measures ANOVA comparing LD accuracy for Word Type (Altered vs. Unaltered) and the two fMRI groups (S Group and SH Group) was performed in order to examine whether the presence of an altered phoneme significantly affected accuracy in the lexical decision task. Specifically, /s/-containing words were 'Altered' in the S Group whereas /ʃ/-containing words were 'Altered' for the SH Group. No main effect of Word Type (Altered vs. Unaltered) nor any main effect of Group was observed, although there was a significant Word Type by Group interaction ($F_{1,22}$=8.267, MSE=39.22, p=0.009, η^2=0.273). Tests of simple effects yielded only one significant comparison, namely a significant difference for Unaltered words between the S and SH groups (t(22)=3.190, p<0.004), significant when corrected for multiple comparisons. Importantly, post-hoc tests revealed that the S and SH Groups did not differ significantly in accuracy for Altered vs. Unaltered words, suggesting that altered stimuli were reliably perceived as words.

3.2.2. Imaging Results—While no significant differences in accuracy emerged in between altered and unaltered words in the lexical decision task, it is nonetheless possible that the presence of an altered token might trigger increased lexical-level or phonetic-level processing as listeners resolve the perceptual ambiguity. For this reason, we explored activation differences between Altered and Unaltered tokens using a t-test (see Table 4). In general, greater activation for Unaltered vs. Altered words was seen in the left AG as well as in the right IFG.

4. Discussion

Phonetic category processing is considered to be relatively fixed in adulthood. Certainly, the observation that adults have difficulty learning new non-native phonetic contrasts (e.g. Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Flege, MacKay, & Meador, 1999) suggests that there must be some entrenchment of native language phonetic category representations. Nonetheless, a significant body of work suggests that listeners are able to use contextual information of many varieties to recalibrate speech processes (see Samuel & Kraljic, 2009 for review). These findings can be captured under the umbrella term of 'perceptual learning for speech,' and together encompass varied phenomena such as non-native speech sound learning (e.g. Bradlow et al., 1997), adaptation to accented, synthesized

or degraded speech (e.g. Bradlow & Bent, 2008; Fenn, Nusbaum, & Margoliash, 2003) as well as the use of visual or linguistic context to resolve perceptual indeterminacies and guide later processing (e.g. Bertelson, Vroomen, & de Gelder, 2003; Norris et al., 2003). In general, these investigations provide evidence that disambiguating information, whether delivered explicitly, as in the form of corrective feedback, or implicitly, as in the use of linguistic information to resolve an ambiguity, results in persisting changes in the perception of speech stimuli. Despite this general similarity, there is no reason to suspect that all varieties of perceptual learning are attributable to the same mechanism. Significant differences exist in terms of the level at which contextual information is provided (e.g. at the level of lexical form, or at the speech sound level) and at the level at which these effects are seen (e.g. a global increase in intelligibility, or a shift in the phonetic category boundary).

In the current study, we investigate one variety of perceptual learning, in which the use of top-down lexical information results in longer-term, talker-specific shifts in the location of the phonetic category boundary. Indeed, contextually-mediated shifts in the categorization of speech sounds could account for more general changes in intelligibility, as in the case of accent learning or processing degraded speech, although this has not yet been thoroughly tested. The behavioral literature on this phenomenon suggests that the new mapping may be talker-specific (Kraljic & Samuel, 2007) and can generalize to new instances and new syllable positions within a word (Jesse & McQueen, 2011; Norris et al., 2003). Taken together, this suggests that the listener must adopt some level of abstraction in the mapping between acoustic token and phonetic category, and that listeners learn to associate the non-standard token with a particular talker rather than adjusting boundaries more generally. In general, this effect is thought to be attributable to shifts in processing at the phonological level (Samuel & Kraljic, 2009).

This investigation addressed two questions. First, we asked whether the neural systems engaged for perceptual learning for speech are the same as those implicated in processing typical native-language speech sounds. Second, we sought to determine whether behavioral shifts in the location of the phonetic category boundary were primarily attributable to shifts in activation within perceptually-sensitive brain areas such as the STG—that is, retuning of perceptual sensitivities—or instead (or additionally) mediated by shifts in decision criteria or attention guided by processes in the frontal lobe. Imaging data can be informative in the sense that neural data does not always parallel behavior. In general, results support a system that initially relies on lexical-level or decision-related systems in the right hemisphere but shows emerging sensitivity in left temporal areas over repeated exposure.

Neural systems for processing ambiguous speech

In order to adjust a phonetic category boundary, the neural system must detect the presence of a non-standard acoustic variant, that is, the ambiguous token between /s/ and /ʃ/ which is embedded in a lexical context. Previous work suggests that regions of the STG posterior to Heschl's gyrus are sensitive to the fine-grained details of native language phonetic categories, and moreover respond specifically to the 'goodness-of-fit' of a token to its phonetic category (Myers, 2007). These regions have been shown to be sensitive to acoustic changes that fall both within and between phonetic categories (Guenther et al., 2004; Myers

et al., 2009). At the same time, frontal areas have been found to be sensitive to competition between phonetic categories, with greater activation for tokens that fall near the category boundary (Myers & Blumstein, 2008; Myers, 2007). In the current study, greater activation for tokens that fell near the category boundary than for those at the endpoints was found in both the right and left IFG (Boundary vs. Endpoint comparison) and in the left TTG, replicating previous results. Differences in deactivation emerged in middle temporal areas, bilaterally, here too showing greater deactivation for tokens close to the boundary compared to those at the endpoints. The explanation of patterns of deactivation in fMRI studies has been the subject of some debate (e.g. Buckner, Andrews-Hanna, & Schacter, 2008; Gilbert, Bird, Frith, & Burgess, 2012). One potential explanation is that deactivation results from disengagement from 'default network' areas that operate while participants are not engaged in the task. Differences in activation here could reflect the greater attentional 'pull' that Boundary stimuli exert compared to the easy Endpoint stimuli. Because the MTG responds to meaningful contrasts in stimulus condition, we believe this region reflects functionally-relevant activation for the processing of speech sounds.

While not the focus of our analysis, we did explore differences in processing for words in the LD task that contained an ambiguous phoneme ('Altered') stimuli compared to those that did not ('Unaltered'). Despite the fact that the same type of phonetic ambiguity was present in the Altered stimuli, no regions showed greater activation for Altered compared to Unaltered words. Two interpretations are available for this finding. First, it may be that the LD task did not engage phonetic processing to the same degree as the PC task. Second, it may be that top-down lexical information for the Altered stimuli was powerful enough that the ambiguous phoneme did not pose a difficulty for neural processing.

Rapid sensitivity in the right hemisphere to lexically-mediated boundary shifts

Behavioral results of the current study replicate the perceptual learning in speech phenomenon (Clarke-Davidson et al., 2008; Kraljic & Samuel, 2005; Norris et al., 2003) with participants showing a significant shift in the location of the phonetic category boundary (and a concomitant shift in the peak of the RT function) such that they were more likely to categorize ambiguous tokens in a way that was consistent with the input during familiarization. Given that the change in boundary emerged quickly, and was stable over successive runs, regions which show a time-invariant response to the shifted category boundary can be said to underlie the rapid adaptation to non-standard speech tokens.

Modulation of activity as a function of top-down factors has been seen at multiple levels in the neural processing stream (Gow et al., 2008; Guediche, Salvata, & Blumstein, 2013; Myers & Blumstein, 2008). When local lexical or sentential context constrains the interpretation of an ambiguous sound, modulation of activity has been especially observable in temporal cortex (Gow et al., 2008; Guediche et al., 2013). The use of lexical information to guide perception may be attributable to access to lexical-level information in the SMG which in turn impacts phonetic processing in the STG (Gow et al., 2008). In a more general sense, a frontal-to-temporal processing route may be engaged as listeners apply top-down knowledge to difficult or ambiguous speech (see Sohoglu, Peelle, Carlyon, & Davis, 2012). In the case of the lexical effect studies, the linguistic information that allows the listener to

resolve the identity of the ambiguous speech sound immediately follows that sound—that is, the remainder of the word (e.g. 'iss' in the case of '[g/k]iss"), or the preceding sentence context provides the disambiguating information. In the current paradigm, listeners must use lexical information to update information about the way individual talkers produce this phonetic contrast. This means that the identity of the ambiguous token cannot be calculated on-the-fly based on the local context, but instead must be resolved based on the information that has built up in memory regarding the idiosyncrasies of the specific talker.

We observed sensitivity to the learned mapping between acoustic exemplar and phonetic category in the right MFG/IFG and in the right posterior STG, extending into MTG. This was manifested as an interaction between group (S vs. SH training group) and token on the phonetic continuum (40/60 vs. 50/50 blend, see Figure 3). In general, the token that fell closest to the phonetic category boundary for a given group showed the greatest activation (frontal clusters) or the greatest magnitude of deactivation (posterior cluster).

Unlike previous studies examining contextually mediated perceptual shifts in phonetic categorization, modulation of activation as a function of group was not observed in either Heschl's gyrus or the STG, but rather in the MTG and IFG/MFG. While the roles of these frontal and temporal areas likely support different functions, the fact that regions beyond the STG are activated suggests that this type of learning in fact does not result in a retuning of early acoustic-phonetic sensitivities, but rather operates at a level of abstraction from the phonetic input.

The MTG, particularly on the left side, has been linked to access to lexical form or phonology (e.g. Baldo, Arévalo, Patterson, & Dronkers, 2013; Binder & Price, 2001; Gow, 2012; Minicucci, Guediche, & Blumstein, 2013; Zhuang, Tyler, Randall, Stamatakis, & Marslen-Wilson, 2012). While the MTG region revealed in our analysis is on the right, rather than the left, taken together this suggests that this process may invoke lexical-level or phonological-level processing.

Most studies of phonetic categorization have shown modulation of activity in frontal areas such that there is more activation for tokens near the phonetic category boundary, on either the left side or bilaterally (Myers & Swan, 2012; Myers, 2007). It is important to note that in the present study, tokens that fell on the category boundary were similarly more difficult to process, as measured by reaction time. As such, we cannot rule out the possibility that increases in activation, in particular in frontal areas, may reflect greater engagement of these regions as stimuli become more difficult to process. Specifically, engagement of these frontal areas may reflect a more domain-general role for inferior frontal areas in resolving competition at multiple levels of language processing (Kan & Thompson-Schill, 2004; Righi, Blumstein, Mertus, & Worden, 2010) or in access to category-level phonetic representations (Chevillet, Jiang, Rauschecker, & Riesenhuber, 2013; Myers et al., 2009; Myers & Swan, 2012)

Perhaps surprisingly, and contrary to other studies, sensitivity to the shifted category boundary was only seen in the right, rather than left hemisphere. We speculate that recruitment of right hemisphere structures is driven by the type of information that impels

listeners to shift their phonetic category boundary. This paradigm forces listeners to adjust to the way a particular *talker* pronounces a given phoneme. Both imaging studies (Belin & Zatorre, 2003; von Kriegstein, Eger, Kleinschmidt, & Giraud, 2003), and patient studies (Van Lancker & Canter, 1982) have implicated right hemisphere structures for talker identification. Given this, processing non-standard phonetic tokens in line with the talker expectation may likewise rely on right hemisphere structures.

Dynamics of sensitivity to lexically-mediated perceptual shifts

In the current study, behavioral evidence of a boundary shift was evident by the second block of the phonetic categorization task, suggesting either that carry-over of learning persisted from the first round of training, or that the few ambiguous tokens present in the first block of the lexical decision task were sufficient to create this shift. Despite this stability, a search for areas which showed group-wise differences in sensitivity to the category boundary over time revealed a network of left-hemisphere regions that are similar to those typically implicated studies of phonetic processing. Of interest, in the left IFG and two STG clusters, greater activation was seen for tokens that fell on the learned boundary in both groups, but only at later blocks of the experiment. This finding is consistent with other studies that have investigated within-experiment perceptual learning at a word level. In particular, Adank and Devlin (2010) report results of a study in which listeners adapted to time-compressed speech within a scanning session. Behavioral improvement on these sentences was accompanied by modulation in activation in left premotor as well as bilateral STG. A similar network has been found to correlate with trial-by-trial comprehension of difficult noise-vocoded sentences (Erb, Henry, Eisner, & Obleser, 2013). In these latter two studies, this activation pattern could be taken as evidence of engagement of a general language system as unintelligible speech became more intelligible via learning.

Our results add to this interpretation and suggest that sufficient exposure to contextual information may result in transfer of decision-related or lexical-level information to more bottom-up or perceptual processes in the temporal lobes. If this is the case, we may propose that listeners have at least two mechanisms at their disposal in using top-down information for speech processing. When contextual information is available to solve perceptually ambiguous speech, listeners may engage a fast-adapting system that immediately adjusts on the fly to resolve the identity of the phoneme. However, when this contextual information is applied many times, listeners may learn the association between the auditory input and the ultimate phonemic percept. At this point, modulation of perceptual codes in the superior temporal gyrus is evident. What is unclear is the nature of this learning. In the present study, as in other behavioral studies of this type, participants heard multiple tokens of each point on the continuum. Given this, it is possible that the activation pattern revealed in the STG reflects episodic learning for precisely these stimuli rather than a shift in criteria at a more abstract level. In general, this type of model is similar to one proposed for learning the categorical structure of non-native speech sounds (Myers & Swan, 2012) in which early sensitivity to phonetic category structure is seen only in executive regions, and only after sufficient training emerges in the STG.

4.1. Conclusion

The behavioral literature suggests that listeners are able to use many sources of information to inform phonetic decisions, including lexical status, semantic expectation, and, in the context of the current study, talker information. While all of these phenomena suggest that the neural system handling phonetic information must be at least somewhat permeable to top-down information, the present data suggest that different mechanisms may give rise to adaptation to different sources of information. In particular, neural systems handling adjustment to a particular talker's idiosyncratic pronunciations appear to take advantage of right hemisphere systems which may be driven by the linkage of the acoustic phonetic information to a talker representation. A complete model of the neural systems underlying speech perception must take into account not only the systems which process stable aspects of speech perception, but also describe the influence of other sources of information on phonetic processing.

Acknowledgments

Works Cited

Adank P, Devlin JT. On-line plasticity in spoken sentence comprehension: Adapting to time-compressed speech. NeuroImage. 2010; 49(1):1124–1132. [PubMed: 19632341]

Allen JS, Miller JL, DeSteno D. Individual talker differences in voice-onset-time. The Journal of the Acoustical Society of America. 2003; 113(1):544–552. [PubMed: 12558290]

Baldo JV, Arévalo A, Patterson JP, Dronkers NF. Grey and white matter correlates of picture naming: evidence from a voxel-based lesion analysis of the Boston Naming Test. Cortex. 2013; 49(3):658–667. [PubMed: 22482693]

Belin P, Zatorre RJ. Adaptation to speaker's voice in right anterior temporal lobe. 2003; 14(16):2105–2109.

Bertelson P, Vroomen J, de Gelder B. Visual Recalibration of Auditory Speech Identification. Psychological Science. 2003; 14(6):592–597. [PubMed: 14629691]

Binder JR, Liebenthal E, Possing ET, Medler DA, Ward BD. Neural correlates of sensory and decision processes in auditory object identification. Nature Neuroscience. 2004; 7(3):295–301.

Binder, JR.; Price, C. Functional Neuroimaging of Language. In: Cabeza, R.; Kingstone, A., editors. Handbook of Functional Neuroimaging of Cognition. Cambridge: MIT Press; 2001. p. 187-251.

Blumstein SE, Myers EB, Rissman J. The perception of voice onset time: an fMRI investigation of phonetic category structure. Journal of Cognitive Neuroscience. 2005; 17(9):1353–1366. [PubMed: 16197689]

Boersma P, Weenink D. Praat: doing phonetics by computer(Version 5.3.57). 2013 Retrieved from http://www.praat.org/.

Borsky S, Tuller B, Shapiro LP. "How to milk a coat:" The effects of semantic and acoustic information on phoneme categorization. Journal of the Acoustical Society of America. 1998; 103(5, Pt 1):2670–2676. [PubMed: 9604360]

Bradlow AR, Bent T. Perceptual adaptation to non-native speech. Cognition. 2008; 106(2):707–729. doi:S0010-0277(07)00112-6 [pii] 10.1016/j.cognition.2007.04.005. [PubMed: 17532315]

Bradlow AR, Pisoni DB, Akahane-Yamada R, Tohkura Y. Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. The Journal of the Acoustical Society of America. 1997; 101(4):2299–2310. [PubMed: 9104031]

Buckner RL, Andrews-Hanna JR, Schacter DL. The brain's default network: anatomy, function, and relevance to disease. Annals of the New York Academy of Sciences. 2008; 1124:1–38. [PubMed: 18400922]

Chang EF, Rieger JW, Johnson K, Berger MS, Barbaro NM, Knight RT. Categorical speech representation in human superior temporal gyrus. Nature Neuroscience. 2010; 13(11):1428–1432.

Chevillet MA, Jiang X, Rauschecker JP, Riesenhuber M. Automatic phoneme category selectivity in the dorsal auditory stream. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience. 2013; 33(12):5208–5215. [PubMed: 23516286]

Clarke-Davidson CM, Luce PA, Sawusch JR. Does perceptual learning in speech reflect changes in phonetic category representation or decision bias? Perception & Psychophysics. 2008; 70(4):604–618. [PubMed: 18556922]

Cohen MS. Parametric analysis of fMRI data using linear systems methods. Neuroimage. 1997; 6(2):93–103. [PubMed: 9299383]

Connine CM, Clifton C. Interactive use of lexical information in speech perception. Journal of Experimental Psychology: Human Perception and Performance. 1987; 13(2):291–299. [PubMed: 2953858]

Connine CM, Titone D, Wang J. Auditory word recognition: extrinsic and intrinsic effects of word frequency. Journal of Experimental Psychology: Learning Memory and Cognition. 1993; 19(1):81–94.

Cox RW. AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. Computers and Biomedical Research. 1996; 29:162–173. [PubMed: 8812068]

Desai R, Liebenthal E, Waldron E, Binder JR. Left posterior temporal regions are sensitive to auditory categorization. Journal of Cognitive Neuroscience. 2008; 20(7):1174–1188. [PubMed: 18284339]

Eisner F, McQueen JM. Perceptual learning in speech: Stability over time. The Journal of the Acoustical Society of America. 2006; 119(4):1950. [PubMed: 16642808]

Erb J, Henry MJ, Eisner F, Obleser J. The brain dynamics of rapid perceptual adaptation to adverse listening conditions. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience. 2013; 33(26):10688–10697. [PubMed: 23804092]

Fenn KM, Nusbaum HC, Margoliash D. Consolidation during sleep of perceptual learning of spoken language. Nature. 2003; 425(6958):614–616. [PubMed: 14534586]

Flege JE, MacKay IR, Meador D. Native Italian speakers' perception and production of English vowels. J Acoust Soc Am. 1999; 106(5):2973–2987. [PubMed: 10573909]

Ganong WF. Phonetic categorization in auditory word perception. Journal of Experimental Psychology: Human Perception and Performance. 1980; 6(1):110–125. [PubMed: 6444985]

Gilbert SJ, Bird G, Frith CD, Burgess PW. Does "task difficulty" explain "task-induced deactivation?". Frontiers in Psychology. 2012; 3:125. [PubMed: 22539930]

Gow DW Jr. The cortical organization of lexical knowledge: a dual lexicon model of spoken language processing. Brain and Language. 2012; 121(3):273–288. [PubMed: 22498237]

Gow DW, Segawa JA, Ahlfors SP, Lin F-H. Lexical influences on speech perception: a Granger causality analysis of MEG and EEG source estimates. NeuroImage. 2008; 43(3):614–623. [PubMed: 18703146]

Guediche S, Salvata C, Blumstein SE. Temporal cortex reflects effects of sentence context on phonetic processing. Journal of Cognitive Neuroscience. 2013; 25(5):706–718. [PubMed: 23281778]

Guenther FH, Nieto-Castanon A, Ghosh SS, Tourville JA. Representation of sound categories in auditory cortical maps. Journal of Speech Language and Hearing Research. 2004; 47(1):46–57.

Jesse A, McQueen JM. Positional effects in the lexical retuning of speech perception. Psychonomic Bulletin & Review. 2011; 18(5):943–950. [PubMed: 21735330]

Kan IP, Thompson-Schill SL. Selection from perceptual and conceptual representations. Cogn Affect Behav Neurosci. 2004; 4(4):466–482. [PubMed: 15849891]

Kraljic T, Samuel AG. Perceptual learning for speech: Is there a return to normal? Cognitive Psychology. 2005; 51(2):141–178. [PubMed: 16095588]

Kraljic T, Samuel AG. Perceptual adjustments to multiple speakers. Journal of Memory and Language. 2007; 56(1):1–15.

Lancaster JL, Woldorff MG, Parsons LM, Liotti M, Freitas CS, Rainey L, Fox PT. Automated Talairach atlas labels for functional brain mapping. Hum Brain Mapp. 2000; 10(3):120–131. [PubMed: 10912591]

McClelland JL, Mirman D, Holt LL. Are there interactive processes in speech perception? Trends in cognitive sciences. 2006; 10(8):363–369. [PubMed: 16843037]

Mertus, J. Brown Lab Interactive Speech System. Providence, RI: Brown University; 2000.

Miller JL. On the internal structure of phonetic categories: a progress report. Cognition. 1994; 50(1–3): 271–285. [PubMed: 8039364]

Minicucci D, Guediche S, Blumstein SE. An fMRI examination of the effects of acoustic-phonetic and lexical competition on access to the lexical-semantic network. Neuropsychologia. 2013; 51(10): 1980–1988. [PubMed: 23816958]

Myers EB. Dissociable effects of phonetic competition and category typicality in a phonetic categorization task: An fMRI investigation. Neuropsychologia. 2007; 45(7):1463–1473. [PubMed: 17178420]

Myers EB, Blumstein SE. The neural bases of the lexical effect: An fMRI investigation. Cerebral Cortex. 2008; 18(2):278. [PubMed: 17504782]

Myers EB, Blumstein SE, Walsh E, Eliassen J. Inferior frontal regions underlie the perception of phonetic category invariance. Psychological Science. 2009; 20(7):895–903. doi:PSCI2380 [pii] 10.1111/j.1467-9280.2009.02380.x. [PubMed: 19515116]

Myers EB, Swan K. Effects of category learning on neural sensitivity to non-native phonetic categories. Journal of Cognitive Neuroscience. 2012; 24(8):1695–1708. [PubMed: 22621261]

Newman RS, Sawusch JR, Luce PA. Lexical neighborhood effects in phonetic processing. Journal of Experimental Psychology: Human Perception and Performance. 1997; 23(3):873–889. [PubMed: 9180048]

Norris D, McQueen JM, Cutler A. Perceptual learning in speech. Cognitive Psychology. 2003; 47(2): 204–238. [PubMed: 12948518]

Oldfield RC. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia. 1971; 9(1):97–113. [PubMed: 5146491]

Righi G, Blumstein SE, Mertus J, Worden MS. Neural systems underlying lexical competition: an eye tracking and fMRI study. Journal of Cognitive Neuroscience. 2010; 22(2):213–224. [PubMed: 19301991]

Samuel AG, Kraljic T. Perceptual learning for speech. Attention, Perception, & Psychophysics. 2009; 71(6):1207.

Sohoglu E, Peelle JE, Carlyon RP, Davis MH. Predictive top-down integration of prior knowledge during speech perception. Journal of Neuroscience. 2012; 32(25):8443–8453. [PubMed: 22723684]

Talairach, J.; Tournoux, P. Co-planar stereotaxic atlas of the human brain. Stuttgart: Thieme; 1988.

Vaden KI, Halpin HR, Hickok G. Irvine Phonotactic Online Dictionary. Version 2.0. 2009 [Data file]. Available from Http://www.iphod.com.

Van Lancker DR, Canter GJ. Impairment of voice and face recognition in patients with hemispheric damage. Brain and Cognition. 1982; 1(2):185–195. [PubMed: 6927560]

Von Kriegstein K, Eger E, Kleinschmidt A, Giraud AL. Modulation of neural responses to speech by directing attention to voices or verbal content. Brain Research. Cognitive Brain Research. 2003; 17(1):48–55. [PubMed: 12763191]

Ward, BD. AFNI 3dDeconvolve Documentation. Medical College of Wisconsin; 2000. Simultaneous inference for fMRI data.

Zhuang, J.; Tyler, LK.; Randall, B.; Stamatakis, EA.; Marslen-Wilson, WD. Cerebral Cortex (New York, N.Y.: 1991). 2012. Optimally Efficient Neural Systems for Processing Spoken Language.

Highlights

- We investigate the neural systems involved in the perception of nonstandard speech tokens

- Left and right-hemisphere areas respond to boundary value sounds compared to endpoint sounds

- Only right hemisphere areas are sensitive shifts in category boundary that come from previous experience with a talker

- By the end of the session, left hemisphere regions are also sensitive to the shift
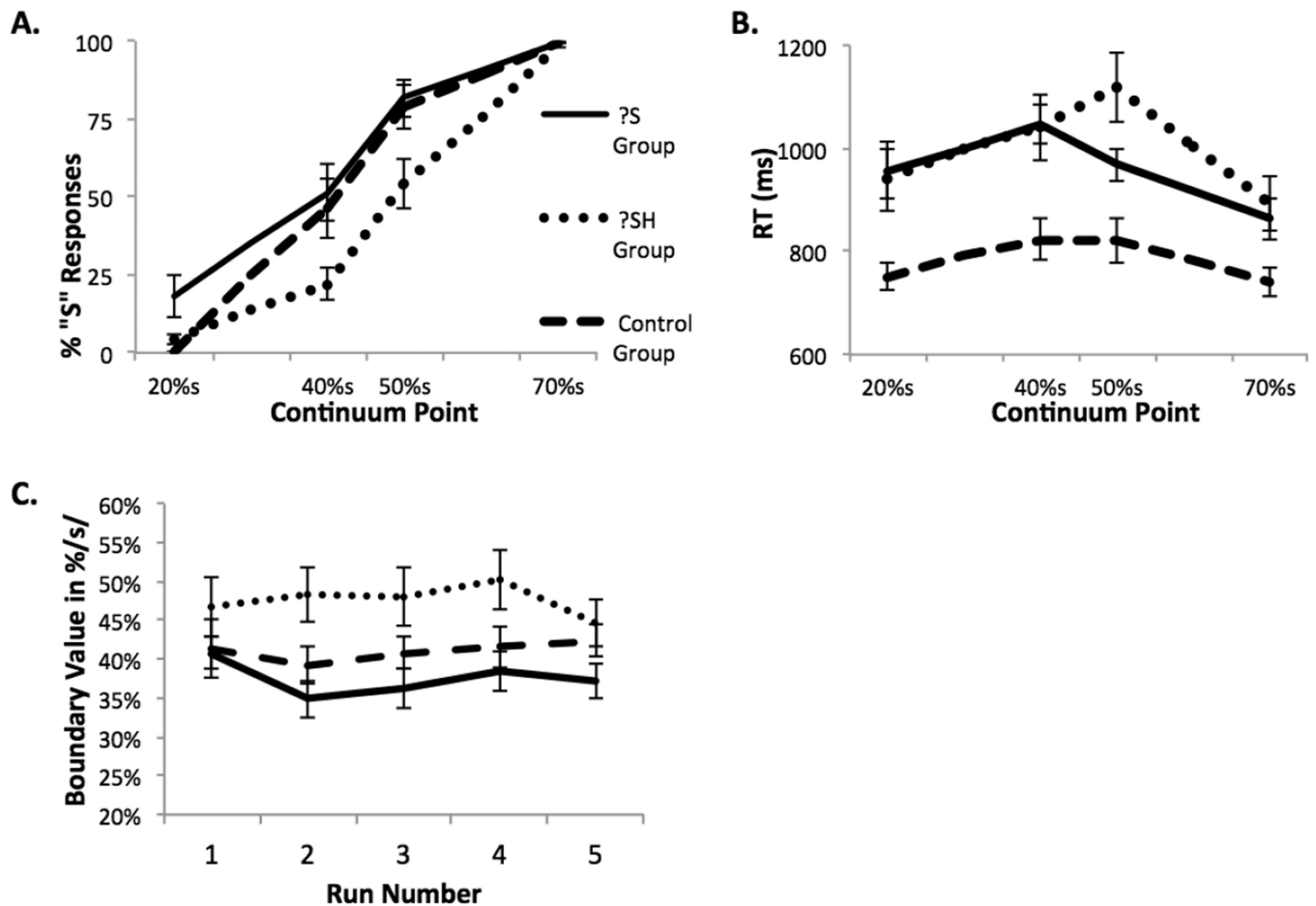
A.

B.

C.

Figure 1.
(A) Mean percent /s/ responses across all runs of the phonetic categorization task. (B) Reaction time for the same decisions. (C) Location of the phonetic category boundary over successive phonetic categorization runs. Lexical decision and phonetic categorization tasks alternate five times over the course of the experiment. All error bars indicate standard error of the mean.
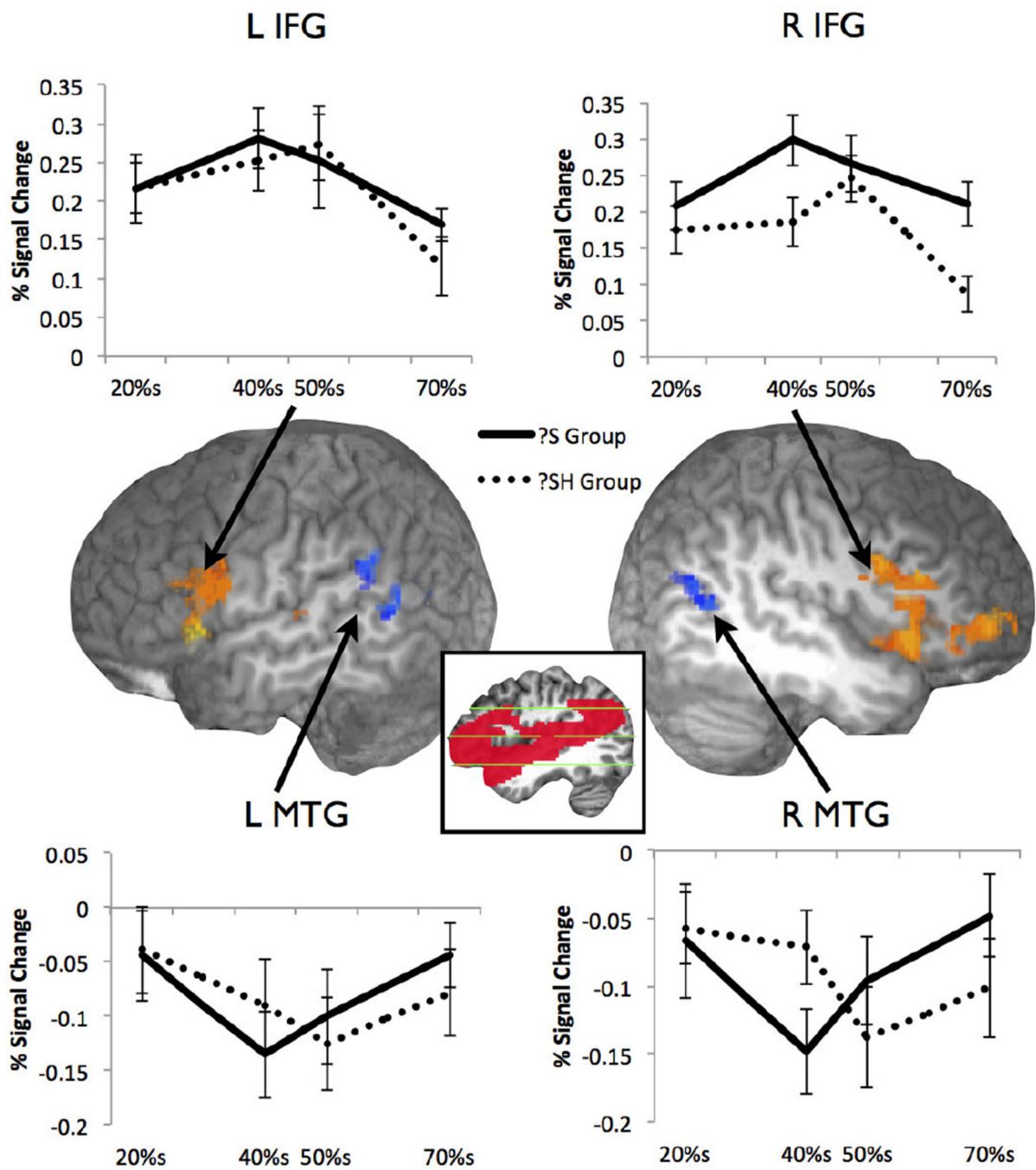
Figure 2.
Boundary vs. Endpoint stimuli, see Table 1 (LH plane cut x = −56, RH plane cut x = 45).
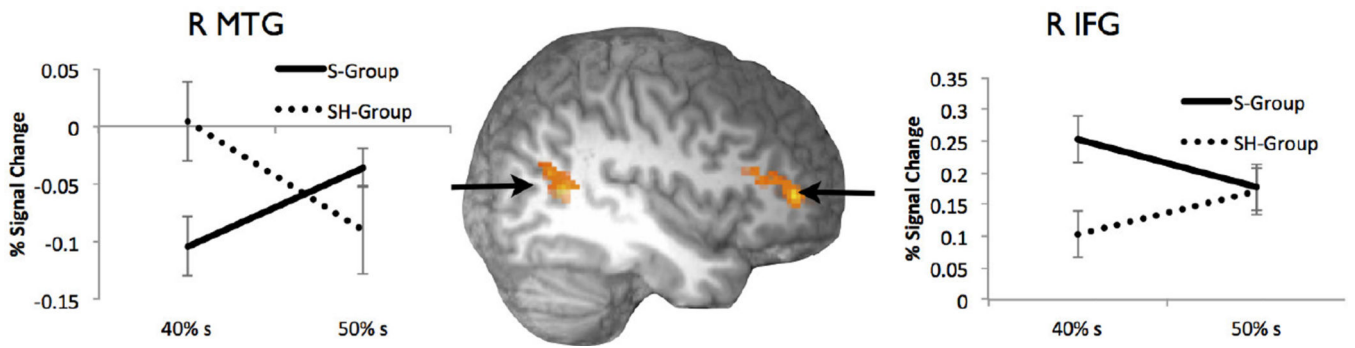Inset shows extent of small volume mask (see Methods).

Figure 3.
Interaction between Group (S Group or SH Group) and near-boundary continuum point (40% /s/ and 50% /s/). All clusters active at a cluster-corrected threshold of p<0.05. Figure displays a RH cut at x=42.
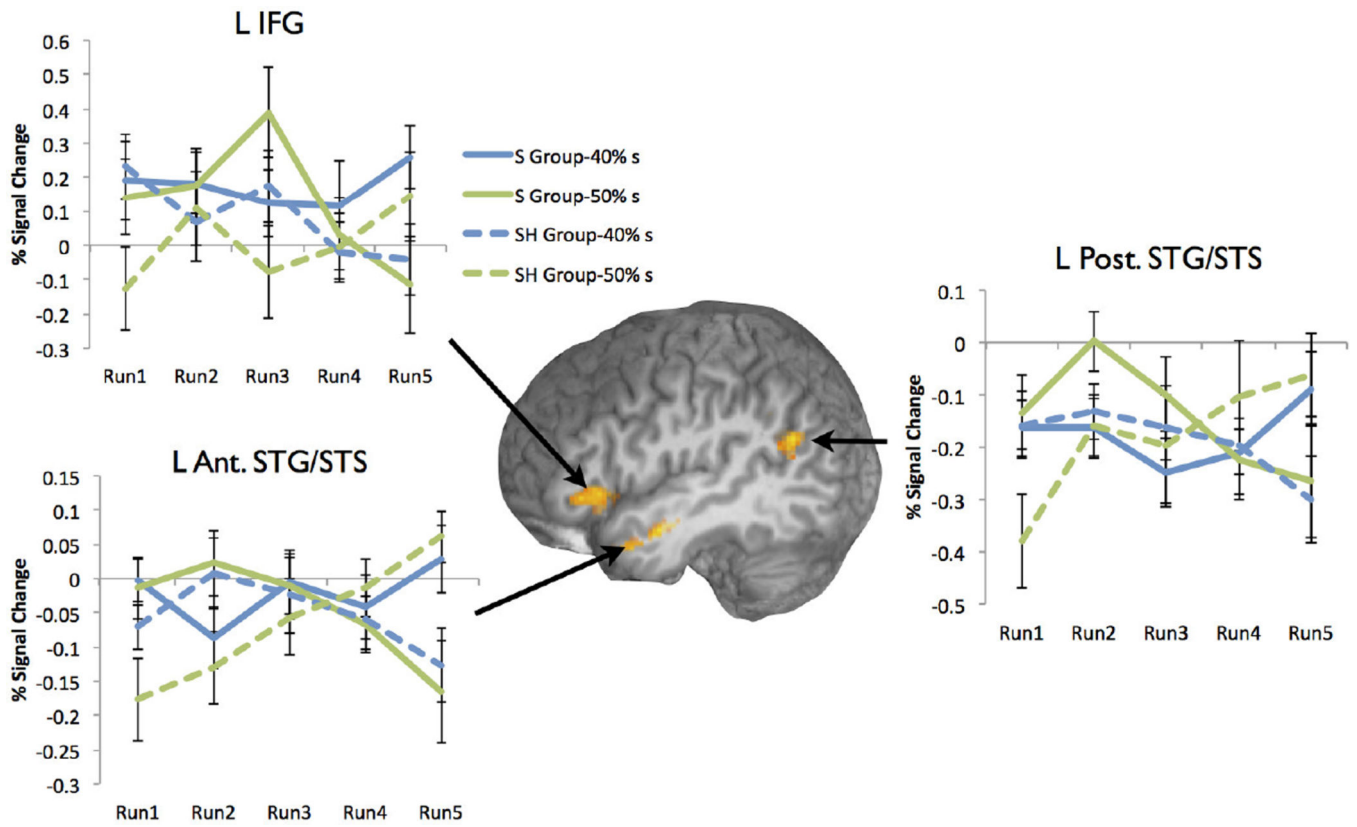
Figure 4.
Interaction between Group (S Group or SH Group), near-boundary continuum point (40% /s/ and 50% /s/), and Time (5 PC runs). S Group data is indicated by solid line, SH Group by a dashed line. Blue lines indicate activation for the 40% s token and green lines indicate activation for 50% s tokens.

Table 1

Boundary vs. Endpoints Comparison. Mixed-factor ANOVA output (percent signal change data) within the small volume-corrected group mask.

Area	Cluster size in voxels	Maximum Intensity Coordinates			Maximum *t* value
		x	**y**	**z**	
Boundaries vs. Endpoints					
Endpoints > Boundaries					
Left MTG**	281	−47	−59	20	2.92
Left SMG*	159	−59	−51	18	2.12
Right MTG*	135	45	−61	12	2.56
Boundaries > Endpoints					
Right IFG and insula**	1250	31	23	10	3.84
Left IFG and insula**	1084	−51	17	2	3.62
Right IFG**	271	49	29	−4	2.64
Left Heschl's gyrus*	135	−63	−19	12	2.41

Table 2

Boundary Values, 2×2×5 (Group, Condition, and Time) mixed-factor ANOVA output (percent signal change data) with small volume-corrected group mask.

Area	Cluster size in voxels	Maximum Intensity Coordinates			Maximum *F* value
		x	**y**	**z**	
Group × Condition × Time Interaction					
Left IFG (pars Orbitalis, pars Triangularis)**	197	−41	21	0	6.669
Left post STG, left post MTG, left SMG*	132	−47	−61	26	5.755
Left ant STG, left ant MTG*	125	−51	−7	−12	6.495
Group × Condition Interaction					
Right post MTG, post STG**	260	47	−59	10	20.942
Right MFG, IFG (pars Triangularis)**	225	39	45	10	26.531
Group × Time Interaction					
Left STG**	172	−43	−43	12	5.37
Right MTG, STG*	108	35	−73	22	5.532
Main Effect of Condition: 40% /s/ > 50% /s/					
Left Insula, STG**	223	−45	−13	4	31.102
Main Effect of Group: S Group > SH Group					
Left STG, MTG*	154	−63	−45	14	13.734
Right IFG (pars Orbitalis, pars Triangularis)*	135	47	−33	−2	10.215
Main Effect of Time					
Right STG, STS, MTG**	293	59	−25	2	7.338
Left STG, MTG**	258	−47	−33	14	7.379
Right IFG, MFG*	142	39	5	26	4.236

Table 3

Mean accuracy in the lexical decision task, standard deviations in parentheses. For the S Group, /s/ words contained the ambiguous /s/--/ʃ/ blend, for the SH Group /ʃ/-words contained the ambiguous blend, and in the Control group, neither set of words

	S Group	**SH Group**	**Control Group**
/s/-words	89.5% (7.6%)	88.3% (6.4%)	99.7% (1.3%)
/ʃ/-words	94.7% (5.0%)	93.7% (4.8%)	100.0% (0.0%)
Filler words	90.1% (6.7%)	88.1% (7.2%)	97.2% (2.2%)
Non-words	93.4% (4.9%)	93.5% (4.8%)	95.9% (4.6%)

Table 4

Activation results for the Lexical Decision task. A planned comparison was performed between Altered and Unaltered words.

Area	Cluster size in voxels	Maximum Intensity Coordinates (T-T)			Maximum *t* value
		x	y	z	
Altered *vs* Unaltered Words					
Unaltered > Altered					
Left AG**	212	−47	−71	32	2.36
Right insula**	194	45	−1	10	3.44