What Does the Right Hemisphere Know about Phoneme Categories?

Michael Wolmetz[1], **David Poeppel**[2], and **Brenda Rapp**[1]

[1]Johns Hopkins University

[2]New York University

Abstract

Innate auditory sensitivities and familiarity with the sounds of language give rise to clear influences of phonemic categories on adult perception of speech. With few exceptions, current models endorse highly left-hemisphere-lateralized mechanisms responsible for the influence of phonemic category on speech perception, based primarily on results from functional imaging and brain-lesion studies. Here we directly test the hypothesis that the right hemisphere does not engage in phonemic analysis. By using fMRI to identify cortical sites sensitive to phonemes in both word and pronounceable nonword contexts, we find evidence that right-hemisphere phonemic sensitivity is limited to a lexical context. We extend the interpretation of these fMRI results through the study of an individual with a left-hemisphere lesion who is right-hemisphere reliant for initial acoustic and phonetic analysis of speech. This individual's performance revealed that the right hemisphere alone was insufficient to allow for typical phonemic category effects but did support the processing of gradient phonetic information in lexical contexts. Taken together, these findings confirm previous claims that the right temporal cortex does not play a primary role in phoneme processing, but they also indicate that lexical context may modulate the involvement of a right hemisphere largely tuned for less abstract dimensions of the speech signal.

INTRODUCTION

The categorical perception of phonemes is a widely investigated aspect of the speech perception system. Early formulations of categorical perception proposed that the receptive language system collapses the continuous acoustic speech signal into the discrete phonemic categories of a language. This proposal was based on the finding that linguistically defined phonemes have psychophysical validity: listeners could discriminate acoustically slightly distinct speech sounds when—and only when—the listeners identified those speech sounds as coming from two distinct phonemic categories (Liberman, Harris, Hoffman, & Griffith, 1957). Subsequent work has shown the initial proposal of perfectly discrete speech perception to be underspecified. The degree to which segments are perceived categorically is influenced by numerous factors (Schouten, 2003). Also, subphonemic details that can aid in phoneme identification and lexical disambiguation and be used for speaker, dialect, or mood

identification are retained by speech decoding mechanisms (McMurray, Aslin, Tanenhaus, Spivey & Subik, 2008).

Although speech perception may be somewhat less than categorical, there is a clear categorical influence. The perceptual space is not isomorphous to physical space but warped, with regions of heightened and diminished sensitivities. The influence of phonemic categories can result in a "continuous physical dimension … perceived in a discontinuous manner," (Pastore, 1987, p. 41), such as the dimension of VOT—the time lag between the onset or initial release of an obstruent consonant and the subsequent vibration of the vocal fold. For example, in the range of VOTs between prevocalic /b/ and /p/, the ability to distinguish tokens with similar VOTs is not constant from minimal to maximal VOT but is lowest near the canonical VOTs for /b/ and /p/ and peaks somewhere between canonical /b/ and /p/, forming a phonemic category boundary.

There are (at least) four classes of explanation for the discontinuity in VOT perception and categorical influences more generally: (1) listeners pick up on real acoustic discontinuities in the signal, (2) nonlinear temporal filters are applied to the signal by early auditory mechanisms, (3) perception relies on contact with the relatively discrete articulatory representations or programs used to produce segments, and (4) well-learned and relatively stable phonemic labels (unrelated to motor representations) influence perception to different degrees as gradient sensory traces fade more or less rapidly, depending on task demands and listening context. The evidence for each of these explanations (reviewed in Rosen & Howell, 1987) suggests that the categorical influence on perception is due to an interaction of all four factors because any subset has limitations in accounting for the 50 years of related results.

Given these multiple cognitive mechanisms, it is unlikely that one brain region is the exclusive "seat" of phonemic processing. This is borne out by fMRI studies that have consistently associated several areas with the categorical influence on perception: left hemisphere (LH), middle and posterior STS, and peri-sulcal regions (Desai, Liebenthal, Waldron, & Binder, 2008; Myers & Blumstein, 2008; Joanisse, Zevin, & McCandliss, 2007; Myers, 2007; Blumstein, Myers, & Rissman, 2005; Dehaene-Lambertz, 2005; Liebenthal, Binder, Spitzer, Possing, & Medler, 2005); LH temporo-parietal regions including the TPJ; and parts of the supramarginal and angular gyri (Joanisse et al., 2007; Raizada & Poldrack, 2007; Blumstein et al., 2005) as well as bilateral frontal regions (Myers & Blumstein, 2008; Myers, 2007; Raizada & Poldrack, 2007; Blumstein et al., 2005; Dehaene-Lambertz, 2005).

On the basis of cumulative evidence, a model of the functional neuroanatomy of categorical influences on speech perception begins to take shape. This model tentatively includes left-lateralized primary auditory areas specialized for higher frequency acoustic/phonetic temporal filtering (Liegeois-Chauvel, de Graaf, Laguitton, & Chauvel, 1999; Steinschneider, Schroeder, Arezzo, & Vaughan, 1995), a left middle and posterior temporal lobe mechanism related to speech-specific phonemic analysis[1] (Andoh et al., 2006; Boatman & Miglioretti, 2005), a left temporo-parietal locus engaged in sound-to-articulation mapping (Hasson, Skipper, Nusbaum, & Small, 2007; Hickok & Poeppel, 2004, 2007), left frontal influences of articulatory representations (Meister, Wilson, Deblieck, Wu, & Iacoboni, 2007), and bilateral frontal regions related to task-influenced decision processes (Myers, 2007).

Conspicuously absent from this emerging model is the contribution of the right hemisphere (RH). Although some functional anatomical models explicitly highlight RH involvement (Hickok & Poeppel, 2007), a recent review of abstract representations used for speech perception dispenses with the RH entirely (Obleser & Eisner, 2009). In an attempt to evaluate the degree to which this exclusion is warranted, Figure 1 provides a summary of findings from LH and RH temporal and temporo-parietal regions as reported by 19 neuroimaging studies (17 fMRI and 2 PET) categorized by level of processing targeted in the research. The studies were selected on the basis of three criteria: (1) discussion of phonemic processing, (2) RH data collection, and (3) publication of tables of results (a summary of studies can be found in Supplementary Table 1). The basic pattern is illustrated by plotting reported activation peaks color coded by level of processing (colder colors indicating less abstract acoustic dimensions and warmer colors indicating more abstract phonemic dimensions). Sensitivity to spectral complexity and temporal cues (dark blue) is roughly equally distributed between the two hemispheres; as more abstract and less acoustic aspects of speech processing are investigated (warmer colors), the balance changes. All eight studies that perform contrasts related to phoneme-specific processing (red) reported LH temporal or temporo-parietal results, whereas only two reported RH results.

Research on brain-lesioned individuals has also supported a left-lateralized network. Most relevant neuropsychological studies use the following logic: By testing individuals with LH lesions, one can observe the abilities of the RH. For example, Basso, Casati, and Vignolo (1977) identified the "position and extent of [the] boundary zone" separating /da/ from /ta/ in a VOT continuum for the following groups: LH-lesioned aphasics, LH-lesioned non-aphasics, and RH-damaged nonaphasics. RH-lesioned and LH-lesioned nonaphasic individuals produced typical categorical identification functions, but 74% of LH-lesioned aphasics had some degree of deficit, ranging from slight (identifiable but abnormally wide boundary zone), to severe (no identifiable boundary zone, but linearly correlated with VOT), to very severe (uncorrelated with VOT). Similarly, Blumstein, Baker, and Goodglass (1977) assessed the phonemic categorization deficits of aphasics, showing an overall tendency toward poor identification in Wernicke's aphasics but seemingly spared discrimination. These data appear consistent with an inability of the RH to support the typical phonemic influence on perception.

Given the functional imaging and neuropsychological data, the degree to which anesthesia (Wada) and (epilepsy-related) hemi-decortication findings suggest substantial RH involvement is surprising. For example, Boatman's studies show that neither complete LH anesthetization (Boatman et al., 1998) nor complete removal (Boatman et al., 1999) impaired auditory discrimination of monosyllabic minimal pairs. Specifically, LH anesthetization resulted in a complete absence of auditory comprehension, object naming and contralateral limb strength, indicating that the LH was truly anesthetized and solely responsible for higher level language functions. Nonetheless, the participant not only

---

[1]To be clear, by phonemic analysis, we refer to representations and processes that utilize stored sublexical distributional information about phonemic categories. Here, a phonemic category is thought of as probability mass across various feature values. We have opted to use the term phonemic but do not intend to distinguish between phonemes, allophones, or bundles of features in the current report. By acoustic/phonetic, we refer to less abstract processes or representations of a finer grain size (acoustic dimensions or individual features) unbiased by category information.

discriminated between different word minimal pairs but could recognize that two acoustically different yet phonemically identical words were the same, indicating intact phonemic processes supported by the RH (Boatman et al., 1998). This capacity was similarly shown in six children who each received a left hemidecorticectomy. Following surgery, minimal pair discrimination was uniformly intact (Boatman et al., 1999). (The possibility of cortical reorganization or abnormal development in these participants complicates the generalizability of these findings.)

Although the RH can support discrimination, this may not be so when tokens are presented in noise (Boatman, Vining, Freeman, & Carson, 2003; Zaidel, 1978). Zaidel (1978) suggested an account according to which the RH relies primarily on continuous acoustic-dependent representations easily hindered by noise, whereas LH phonetic feature extraction and abstract phoneme representations are more robust to noise.

With the current report, we aim to further test the hypothesis that the RH does not engage in phonemic analysis. This will help resolve the discrepancies between, on the one hand, the rarity of RH BOLD response reported during phonemic processing and absence of categorization deficits in some LH-lesioned individuals and, on the other, the apparent capacities of the RH to execute phonemic processing as evident from decortication and anesthetization. Although similar tasks were used across these methodologies, one key difference was whether the target phonemes were presented within words or nonwords. The data suggesting an RH inability to support typical phonemic processing in perception used nonwords, whereas those data demonstrating the phonemic capacities of the RH used words.

In the first study presented here, we investigated whether this lexical distinction was important for resolving the issue of RH phonemic processing by using f MRI to identify cortical sites sensitive to phonemes in both word and pronounceable nonword contexts. We next present a case study of an individual DMN, who has a left temporoparietal lesion. This case is particularly well suited for extending the interpretation of the fMRI results as DMN's lesion affected phonemically sensitive areas identified by fMRI in our first study. This allowed us to assess the integrity of speech perception when regions typically involved in phonemic processing are damaged. We are able to show with fMRI and magneto-encephalography (MEG) that DMN's LH is essentially deafferented for auditory input, rendering him RH reliant for early cortical processing of speech. As a result, we can attribute disparities between DMN's performance and that of typical listeners to disparities between the hemispheres, in the degree to which phonemic categories influence speech perception.

The earlier neuropsychological studies discussed above (Basso et al., 1977; Blumstein et al., 1977) had similar goals, but with several key methodological differences. All previous testing was done with nonwords, preventing investigation of phonemic perception in word contexts. Second, previous testing used synthesized stimuli—but there are marked differences in the perception of synthetic and natural speech among aphasic individuals (Gow & Caplan, 1996; Huntress, Lee, Creaghead, Wheeler, & Braverman, 1990). In contrast, we use naturally produced word and nonword stimuli. Lastly, we use functional

imaging to examine the extent of structural and functional LH damage such that inferences about speech processing by the RH can be made with greater certainty.

STUDY 1: FUNCTIONAL LOCALIZATION OF CORTEX SENSITIVE TO PHONEMIC CATEGORY

Methods

We used a neural adaptation paradigm (Kourtzi & Grill-Spector, 2005) to identify cortex sensitive to phonemic category in unimpaired listeners (Zevin & McCandliss, 2005). To reveal neural populations selectively responsive to phonemic category, we first presented the same stimulus multiple times to habituate a given phonemic category and then presented a stimulus that differed from the habituating stimulus either acoustically and phonemically or only acoustically. In this way, we identified phonemic category-sensitive regions as those exhibiting phoneme-specific dishabituation: a larger rebound response when both the phonemic category and the acoustics changed, as compared with the response when the acoustics alone changed.

Participants—Eight healthy volunteers (six women, mean age = 23.4 years, range = 19–27 years, right-handed, no reported hearing deficits, and native fluency in American English) took part in informed consent, MRI safety screening, task practice, anatomical and functional MRI scans, and received compensation totaling $25/hour of participation. The institutional review boards of the Johns Hopkins Medical Institutions or University approved all studies reported.

Behavioral Procedure—During fMRI, participants heard different trains of four consecutive syllables delivered in periods of scanner silence. Participants reported by button press with the right index finger when they perceived the fourth syllable as acoustically identical to the first three and with the right middle finger when they perceived it as acoustically different. A practice paradigm administered pre-scanning familiarized participants with the notion of acoustic difference rather than the default conception of phonemic difference.

Syllable trains were manipulated on two dimensions: dissimilarity and lexicality In the dissimilarity dimension (Figure 2B), the onset of the fourth syllable was (1) acoustically and phonemically different from the first three, forming a between-category trial, (2) equally acoustically different as in the between-category trial, but phonemically identical to first three syllables, forming a within-category trial, or (3) acoustically identical to the preceding syllables, forming a repetition trial. Each of these three trial types had 48 trials per participant. In the lexicality dimension, on half of trials (72) the syllables formed words, on the other half syllables had acoustically identical onsets and nuclei to the words but codas that resulted in pronounceable nonwords. Responses were obtained and analyzed for discrimination sensitivity and response time by condition.

Stimuli—Four syllable pairs differing in onset voicing (beach/peach, best/pest, goat/coat, and dent/tent) and matched pronounceable nonwords (beesh/peesh, besk/pesk, gobe/cobe,

and deg/teg) were recorded at 22,050 Hz using male native English speakers. VOT continua were created from each nonword pair using standard cross-splicing procedures (McMurray & Aslin, 2005): Sound was deleted from the onset of the voiced member of the pair; an equivalent duration of sound from the onset of the unvoiced member of the pair was spliced onto the onset of the truncated voiced nonword. The procedure was repeated for eight splice points, 4–7 msec apart, occurring at zero-amplitude points to avoid discontinuities in the manufactured sounds. In this way, four nonword continua were created with nine steps in each. Word codas were spliced into nonword tokens to create four acoustically matched word continua. Four-token trains were created by concatenating the same token from a continuum three times followed by a token two steps away in the continuum and inserting 50 msec of silence between each token in the train (Figure 2B).

One set of within-category trains from the voiced end of the continuum, one set from the voiceless end, and one set of between-category trains were selected from each continuum on the basis of pilot testing for categorical discrimination and identification. Each set consisted of an AAAX and XXXA train. VOT differences between members of within-category (mean = 12.8 msec) and between-category (mean = 13.5) trains were not significant, indicating an equivalent acoustic difference. In addition to four functional runs of between- and within-category discrimination, one final run of continuum end-point discrimination (e.g., peesh—50 msec/peesh—50 msec/peesh—50 msec/beesh—5 msec) was collected for piloting a future study, but behavioral data were analyzed and reported here. Each run contained 36 task trials and 12 silent trials in which no auditory stimulus was presented.

fMRI Protocol—To avoid scanner noise masking the stimuli, a sparse sampling event-related design (Hall et al., 1999) was used. Each acquisition followed a period during which the scanner was silent. Stimuli were delivered in those periods of scanner silence (Figure 2A, illustrating the timing for one trial with four acquisitions over the course of 14.8 sec). Each stimulus train (lasting 1.95 sec) was delivered at a comfortable listening level via noise-attenuating electrostatic earphones 400 msec after volume acquisition offset, and 150 msec was left between stimulus offset and volume acquisition onset. Anatomical scans occurred before functional runs to partially habituate participants to scanner noise.

Images were acquired on a Phillips 3-T magnet with an eight-channel SENSE coil. Four functional runs were collected per participant, with 193 T2*-weighted volumes per run using an echo-planar pulse sequence for BOLD imaging: volume acquisition time = 1200 msec, repetition time = 3700 msec, intertrial interval = 14800 msec, echo time = 30 msec, and flip angle = 30°. Volumes contained twenty-four, 80×80, 3-mm-thick transverse slices (3-mm isotropic voxels), with a prescription of the middle slice aligned to the Sylvian fissure. Three additional volumes were discarded at the start of each run to permit T2* signal levels to stabilize. Stimuli sequences were pseudorandomized, and 12 trials/run were silent for baseline collection and a varied distribution of stimulus onset asynchronies.

Statistical Analysis—Preprocessing and statistical analysis was performed using SPM2 statistical parametric mapping software (Friston, Frith, Frackowiak, & Turner, 1995) and AFNI (Cox, 1996). Preprocessing consisted of slice-timing correction, affine motion correction, high-pass filtering (cutoff = 128 sec), spatial smoothing using a Gaussian filter of

6-mm FWHM, and co-registration with individual anatomical images as well as normalization to standard Montreal Neurological Institute space. Time-series analyses were performed using a general linear model constructed to evaluate rebound from neural adaptation.

Adaptation regressors and rebound (or dishabituation) regressors were defined for nonsilence trials. Adaptation regressors were time locked to syllable train onsets (Figure 2A, black vertical bar), and rebound regressors were time locked to onsets of dishabituation syllables—the fourth stimulus in the train (Figure 2A, gray vertical bar). Volumes were acquired 600 msec post-dishabituation onset and then every 3700 msec until the next trial. Design matrices contained eight regressors of interest: (1) adaptation regressor for speech trials time locked to trial onset, (2) baseline regressor time locked to the beginning of silent trials, (3–5) rebound regressors for between-category, within-category, and repetition trials of word, and (6–8) nonword syllables. Six motion and four run nuisance regressors were included. Random effects beta weights were estimated from these predictors convolved with a canonical hemodynamic response function (HRF).

Planned comparisons included the following contrasts: (1) speech onsets as compared with silence, (2) main effect of phoneme-specific dishabituation (between-category rebound compared with within-category rebound, collapsing across lexical status), (3) phoneme-specific dishabituation specifically for word-embedded phonemes (between-category rebound compared with within-category rebound for words), (4) phoneme-specific dishabituation specifically for non-word-embedded phonemes, and (5) interaction of phoneme-specific dishabituation and lexicality testing the difference between items 3 and 4 (between-category rebound compared with within-category rebound for words compared with nonwords).

Group t tests for each comparison were calculated, and multiple comparison correction was carried out through AFNI's AlphaSim Monte Carlo estimation of a false detection of $\alpha < .05$, yielding a voxel threshold of $p < .001$ ($t = 5.21$) for a cluster threshold of seventy-seven $2 \times 2 \times 2$-mm voxels. A relaxed threshold of $\alpha < .1$ was also tested for each of the contrasts and reported where relevant.

Results

Behavioral Results—Means and standard deviations for sensitivities (d') and response times (RT) for the in-scanner discrimination task are shown in Table 1. The d' score represents acoustic discrimination ability within or across the categorical boundary. A two-way repeated measures ANOVA (Dishabituation Stimulus $\times$ Lexicality) on d' scores yielded a significant main effect of within versus across boundary dishabituation stimulus, $F(1, 6) = 47.39$, $p < .001$, and no significant main effect of lexicality or interaction. The effect of dishabituation stimulus was consistent across participants. A similar ANOVA on response times yielded no significant main effects of dishabituation stimulus or lexicality but a trend toward a significant interaction, $F(1, 6) = 4.6867$, $p = .074$.

The finding of greater discriminability for between versus within category differences that was comparable for words and nonwords indicates that participants perceived perceptual

differences for between-category versus within-category despite an imperfect listening environment and normal individual variability in category boundary. In addition, the trend for interaction between dishabituation stimulus and lexicality in RT, despite the acoustically identical onset material, suggests that discrimination of word-embedded phonemes may differ from non-word-embedded phonemes. Specifically, the result indicates that lexical context provides greater ease in detecting between-category differences but a greater challenge for detecting within-category differences, whereas the nonword environment appears to be more neutral.

fMRI Results—The contrasts tested were (1) task-related response, (2) a main effect of phoneme-specific dishabituation, (3) phoneme-specific dishabituation for words, (4) phoneme-specific dishabituation for nonwords, and (5) the interaction of phoneme dishabituation and lexicality (Table 2).

<u>**Speech-related response:**</u> Large regions of frontal, temporal, parietal, occipital, and cerebellar cortex were bilaterally responsive during speech habituation as compared with silence. Heschl's gyrus (HG) was significantly more active bilaterally in each individual subject during speech versus silence.

<u>**Main effect of phoneme-specific dishabituation:**</u> Evidence of phonemic dishabituation above and beyond acoustic dishabituation (between category > within category) was found only in the left posterior middle temporal gyrus (MTG)/STS. Trends were explored at a relaxed threshold ($\alpha < .1$), revealing several additional areas, but no trends were found in the RH outside of the cerebellum. The main effect of phonemic dishabituation collapses across word- and non-word-embedded phonemes. The next set of contrasts examined dishabituation in lexical and nonword contexts separately.

<u>**Phoneme-specific dishabituation and lexicality:**</u> The analysis revealed that phonemic dishabituation was conditioned by the context in which the onset phoneme occurred, despite the context being irrelevant for the task being performed. Relatively few areas showed phoneme-specific dishabituation in nonword contexts: a left MTG cluster similar to that reported in the previous contrast (Figure 3A), a hippocampal cluster, and an RH cluster at the temporal pole, bordering on the parahippocampal gyrus. A very different picture emerged for word-embedded phonemic dishabituation, where many areas registered phonemic change beyond acoustic change, as shown in Table 2: an LH posterior temporal response common to word and non-word-embedded phonemes as well as an LH angular gyrus and an RH posterior temporal response not found for nonwords (Figure 3B).

A direct comparison—in the form of the interaction between dishabituation (between category, within category) and lexicality (word, nonword) revealed that although no areas were significant for greater phonemic dishabituation for nonwords as compared with words, several areas showed a larger dishabituation response to words as compared with nonwords in the LH—a precentral gyrus area superior to Broca's area and overlapping with the inferior frontal junction, an inferior parietal area encompassing parts of the angular gyrus, supramarginal gyrus, and extending into the TPJ, and the left anterior cingulate gyrus—and in the RH—the posterior aspect of right MTG (Figure 3C). These results suggest a

qualitative difference in the neurotopography of phonemic sensitivity during word and nonword processing.

Study 1 Summary

This study identified areas showing an influence of phonemic category on BOLD response, using natural speech stimuli in an experimental paradigm that allowed for a direct comparison of word and nonword phonemic context. The results reveal a highly left-lateralized pattern of phonemic sensitivity, with only the posterior left MTG/STS showing phonemic sensitivity in both word and nonword contexts. This area is slightly more posterior to areas already shown to be sensitive to phonemic processing (Desai et al., 2008; Myers, 2007; Dehaene-Lambertz, 2005; Liebenthal et al., 2005). It should be noted that although there are categorical influences on BOLD responses from this area, gradient responses have also been detected using recovery from neural adaptation (Myers, Blumstein, Walsh, & Eliassen, 2009), and differences in response may be related to specific task demands and acoustics of stimuli. In addition, the study also reports a number of LH areas that showed phonemic sensitivity conditioned by lexical/nonlexical context. In particular, the left anterior cingulate, angular gyrus, and precentral gyrus/inferior frontal junction were more sensitive to word-embedded phonemes than non-word-embedded phonemes. Although it is difficult to separate these effects from those of general lexical processing, these results suggest that the influence of phonemic category during speech perception is affected by lexical context.

Importantly, with regard to the RH, the study provides evidence of RH phonemic sensitivity that is limited to a lexical context. Specifically, posterior right MTG was significantly more responsive for between versus within-category dishabituation trials for words but not nonwords.

As reviewed in the Introduction, evidence of RH phonemic processing is scarce; nonetheless, our findings are consistent with those from Myers (2007), one of the few studies also attempting to compare word-embedded and non-word-embedded phonemic sensitivity (although in a different experimental paradigm). Myers and Blumstein reported an interaction between the lexicon and the phonemic category in the posterior temporal lobes bilaterally, the IPL bilaterally (as well as the left anterior cingulate, precentral, and middle frontal gyri and the precuneus). Right temporal involvement was also found by Gow, Segawa, Ahlfors, and Lin (2008) when combining structural MRI with MEG and ERP data collected while participants identified members of word–nonword continua. These researchers reported that the left and right pSTG (and left AG) receive both bottom–up and top–down input throughout token identification.

The nature of these RH processes, whether the observed activations are the result of bottom–up or top–down computations and whether and how they are modulated by context is less clear. Several factors complicate the interpretation of the relevant data, including lexical-semantic processing, attentional mechanisms, and limited power. Although these results suggest that phonemic processing is affected by lexical context, it is difficult to separate effects of phonemic influence from those of general lexical processing. A competing interpretation is that the interaction between phoneme-specific dishabituation and lexicality

resulted from dishabituation of the lexical items themselves and all the phonology and semantics that come with them. Under this view, activation differences in areas responding differentially to word and non-word-embedded phonemes, such as the posterior right MTG, result from lexical and/or semantic processing. Indeed, auditorily presented words have been found to evoke different BOLD responses than pronounceable nonwords in this region (Kotz, Cappa, von Cramon, & Friederici, 2002; Newman & Twieg, 2001). Another competing interpretation of activation differences reported here is that they reflect attentional and decision-related modulation that occurs in response to differing degrees of difficulty between conditions or coincides with amodal detection of change. As compared with a passive design, the use of a discrimination task may increase the risk of attentional confounds, but even without an explicit task, listeners may deploy attention in ways that correlate with the experimental conditions. Finally, we must interpret the absence of an RH phonemic response for nonwords with caution as detection of weak or variable BOLD responses is always a concern. For these reason, Study 2 is presented in an effort to mitigate some of the ambiguities of f MRI interpretation and further investigate the relationship between phonemic processing, hemisphere, and context.

STUDY 2: CASE STUDY OF SPEECH PERCEPTION IN AN APHASIC LISTENER WITH A TEMPOROPARIETAL LESION

Methods

The methods used include auditory and language screenings to characterize the language profile, MRI to characterize the lesion locus, fMRI and MEG to examine the functional locus of the lesion with regard to speech perception and comprehension, and several identification and discrimination paradigms to test the consequences of the lesion for speech perception. Control participants were also tested.

Case Description and Control Participants—DMN, a native English speaker with a college education, was studied when 68 years old, 1 year after a cerebrovascular accident. At the time of the cerebrovascular accident, he was hospitalized after experiencing global aphasia and right-lateralized buccofacial and limb apraxia. The severity of the aphasia lessened over time (rehabilitation was also delivered), but some language functions remained impaired. At the time of testing, DMN was healthy, living independently, had moderate conversational comprehension and production, but had auditory comprehension difficulty in situations lacking semantic and syntactic cues. He was right-handed according to the Edinburgh Handedness questionnaire (Oldfield, 1971); however, he recalled childhood ambidextrous tendencies that were discouraged.

Preliminary screenings indicated severe difficulty in word recognition and comprehension specific to the auditory modality (7th percentile on the auditory administration of the Peabody Picture Vocabulary Test; Dunn & Dunn, 1997) that contrasted sharply with his excellent word comprehension in the written modality (75th percentile on the written administration of the Peabody Picture Vocabulary Test; the two administrations used different forms of the test). The dissociation between intact written word comprehension and impaired auditory comprehension was documented in a number of additional tasks.

Audiological testing revealed some minor hearing loss at high frequencies not expected to have a marked effect on speech perception, but DMN was severely impaired on tasks requiring intact phoneme recognition (phoneme monitoring, minimal pair discrimination, syllable repetition, segment counting) as well as word repetition (committing both formal and semantic paraphasias and unable to respond more than half of the time). He demonstrated a range of performance on tasks involving perception of fine-scale acoustic timing. Although DMN could consistently detect gaps in noise within normal thresholds (< 5 msec), similar tasks such as gap detection in tones and click fusion produced inconsistent thresholds on different administrations, ranging from normal (4–6 msec) to abnormal (30–40 msec). In contrast, his pitch perception, ability to discriminate pure tones and frequency-modulated sweeps, and ability to name complex environmental sounds (Marcell, Borella, Greene, Kerr, & Rogers, 2000) were within normal range. The specific combination of spared naming of environmental sounds in the context of poor word repetition, potential auditory temporal deficits, and left superior temporal damage is consistent with what is referred to as *word deafness*, deep dysphasia, or auditory agnosia (Poeppel, 2001; Griffiths, Rees, & Green, 1999; Simons & Ralph, 1999). However, the vast majority of the reported cases involve bilateral lesions.

Testing took place over eight sessions, up to 2 hours each. DMN received $10/hour for behavioral testing and $25/hour for imaging participation. Nine age and education matched (three women, mean age = 66.9 years, range = 62–70 years) right-handed, native English-speaking control subjects participated in behavioral testing and were compensated $10/hour. Controls reported no history of neurological disorders and had normal to slightly impaired hearing. Five younger controls also participated in MEG.

Experiment 1: Anatomical and Functional MRI—MRI and fMRI data were collected using the same methods and stimuli reported in Study 1. However, it was not appropriate to perform the same analyses of the functional data as were carried out in Study 1 to identify cortex sensitive to phonemic categorization because DMN's behavioral responses were grossly abnormal. Therefore, the fMRI data were used to contrast the neural response to speech versus silence to examine the integrity of auditory input to the two hemispheres and, in this way, evaluate the extent of deafferentation of the LH. Nonetheless, DMN's behavioral discriminations to the dishabituation task were analyzed and compared with controls. All statistical comparisons between DMN and control groups were performed using the Crawford and Garthwaite (2007) Bayesian point estimate.

Experiment 2: Magneto-encephalography—To investigate DMN's neurophysiological responses to speech, MEG measurements were collected on a 160-channel whole-head axial gradiometer MEG system (KIT, Japan) at the University of Maryland. Data were collected with a sampling rate of 500 Hz during passive listening to tones (100 Hz, 250 Hz, 1 kHz, and 4 kHz) and to bisyllabic words and nonwords (45 words, 45 nonwords, mean length of 613 and 610 msec, respectively). Noise reduction was achieved using adaptive noise suppression (Ahmar & Simon, 2005). Waveforms were baseline corrected (100 msec prestimulus onset) and low-pass filtered (30 Hz). For each participant (DMN and five controls), for each response component, and for each

hemisphere, the channels most representative of the source and sink distribution were selected for quantitative analysis. Here, we report the peak root mean squared (RMS) amplitudes, latencies at peak, and surface topographies for the M100 response to tones as well as the M350 response to auditorily presented words and nonwords. The M100 was selected as the first RMS peak after 70 msec showing a plausible M100 distribution (see, e.g., Salajegheh et al., 2004) and the M350 as the first RMS peak after 300 msec that showed the M350 distribution (following Fiorentino & Poeppel, 2007; Pylkkanen, Stringfellow, & Marantz, 2002).

Experiment 3: Perception of Phonemic Category — Identification and Discrimination—To assess the degree to which phonemic categories influenced DMN's speech perception, we considered DMN's performance on the in-scanner discrimination task as well as VOT identification tasks that included tokens from a range of word and nonword VOT continua. For the identification tasks, stimuli consisted of VOT continua similar to those described in the fMRI methods: three word and three matched nonword continua described earlier made up the primary stimulus set (beach/peach, dent/tent, goat/coat and beesh/peesh, deg/teg, gobe/cobe). An additional three word (bale/pale, base/pace, game/came) and nonword continua (baysh/paysh, bem/pem, gice/kice) were piloted to confirm a highly categorical response from typical listeners and then used with DMN.

In VOT identification tasks, participants were familiarized with the continuum end points then given a two-alternative forced-choice identification task with members of that continuum. For example, for testing of the beach/peach continuum, each participant heard over headphones (Sony MDR V-15) calibrated to a volume comfortable for the listener (approximately 72–79 dB) a member of the beach/peach continuum and saw the choices "beach/b" and "peach/p" on the left- and right-hand side of the screen, respectively. The participant was instructed to press the 1 key if "beach" was the percept or 0 if "peach" was the percept. This was repeated for nine presentations of each member of the nine-step continuum, for a total of 81 randomized trials per continuum. Each trial began 500 msec after the response from the previous trial. The procedure was repeated for each continuum in a random order. For each continuum, we obtained the number of times each token was identified as a particular type as well as response times. Stimuli were presented, and identifications and response times were recorded using Alvin presentation software (Hillenbrand & Gayvert, 2005).

DMN was evaluated with the same procedure but received more extensive familiarization with the continua end points and two testing sessions of each continuum on separate days because of his greater response variability. DMN received one additional identification paradigm in which members of one word continuum (beach/peach) and nonword continuum (beesh/peesh) were mixed, but choices for identification were limited to "beach/b" and "peach/p." Here, instructions were purposefully vague and did not mention the inclusion of nonwords. As a result, DMN was under the impression he would only hear word stimuli. The goal of this final set of identifications was to help determine the role of stimulus-driven and knowledge-driven influences on lexical effects of phonemic processing.

We adopted a curve-fitting approach to assess the structure of phonemic category representations. Typical VOT identification strongly influenced by phonemic category produces data sharply shifting from near exclusive identification of one category to near exclusive identification of the other. This type of sigmoid response is well fit by logistic functions but poorly fit by linear functions (McMurray & Spivey, 2000). In contrast, gradient or continuous identification yields data that are similarly fit by linear and logistic functions because logistic functions can approximate linear functions, but not the reverse. To evaluate the degree of phonemic category influence on speech perception, both linear and logistic regressions were performed, yielding a linear best-fit line and deviance residual as well as a logistic best-fit curve and deviance residual (calculated for proportional data) for each continuum, for each participant. The difference between linear and logistic deviance yielded a metric of categoricity such that high values indicated high categoricity whereas low deviance values indicated a lack of categoricity. Pearson correlations between VOT and percent unvoiced token identified were also calculated for each continuum and tested for significance. Statistics were calculated using MATLAB 7.0.

Results

Experiment 1: MRI/fMRI—DMN's MRI showed an LH lesion affecting gray and white matter integrity in HG, the middle and posterior aspects of the MTG and STG, extending caudally into the TPJ and rostrally to the rolandic operculum. When normalized, DMN's lesion overlapped with the pMTG/STS area shown in Study 1 to be selectively responsive to phonemic category in normal participants (Figure 4) as well as with posterior temporal findings for phonemic sensitivity in the literature (Desai et al., 2008; Joanisse et al., 2007; Myers, 2007; Blumstein et al., 2005; Dehaene-Lambertz, 2005; Liebenthal et al., 2005).

The degree to which DMN's LH early auditory cortex was still functional was unclear from structural MRI because some highly misshapen gray matter remained in the HG. However, a total absence of white matter in adjacent regions indicated possible deafferentation of LH from auditory input. fMRI data allowed us to address this issue. As mentioned above, bilateral activation of HG was observed in each of our normal participants for speech relative to silence. DMN, however, showed a total lack of left HG response for the contrast of speech > silence, even at liberal thresholds ($\alpha < .1$), at the same time displaying the typical right HG activation for this contrast, even at strict thresholds (Bonferonni corrected $P < .05$).

A recent report on postinfarct hemodynamics suggested that perilesional BOLD responses may go undetected by models applying canonical HRF convolutions (Bonakdarpour, Parrish, & Thompson, 2007). To examine this possibility, we directly measured DMN's BOLD response and carried out an ROI analysis measuring the first eigenvariate for spherical ROIs of 10-mm radius centered at the anatomically defined HG bilaterally. Because the left HG was less defined than the intact right HG, several ROIs for the left HG were tested, with the ROI producing the largest signal reported. ROI analysis showed a right HRF for DMN similar to controls and a slow, largely attenuated left HRF (Figure 5). A one-way AN OVA on the time course of the left HG ROI was not significant for a main effect of time bin, $F(4, 710) = 1.92, p = .105$, whereas the RH ROI was highly significant, $F(4, 710)$

= 52.43, *p* < .0001, suggesting that the LH was either devoid of bottom–up input or was receiving extremely abnormal input. These findings were further investigated using MEG during passive listening to tones and speech.

Experiment 2: MEG—We report the latencies, amplitudes, and surface topographies of the M100 response to passively presented tones and the M350 response to passively presented words and nonwords for DMN and five control participants. Each control participant showed typical surface distributions (Figure 6) and evidence of bilateral auditory-evoked M100 responses to 1 kHz tones (Figure 7A, LH latency/amplitude: μ = 97.2 msec/116 fT, range = 88–119 msec/69–157 fT; LH latency/amplitude: μ = 101 msec/116 fT, range = 89–115 msec/60–177 fT).

DMN showed typical amplitudes in the RH but markedly low amplitudes across left channels. For the 125-Hz tone, no M100 could be detected in DMN's LH channels, whereas typical topography and amplitude was detected for the RH source across tone frequency (125 Hz—104 fT, 250 Hz— 147 fT, 1 kHz—165 fT, and 4 kHz—85 fT). For the higher frequency tones, nonzero M100 components were identified for DMN on the basis of topography and timing, but the drastically attenuated responses were well below the control range (250 Hz—46 fT, 1 kHz—48 fT, and 4 kHz—39 fT), and time series data were not convincingly dipolar. This supports the fMRI finding that the LH is either lacking in auditory input or that the input is grossly abnormal.

A robust LH M350 in response to passively heard words and nonwords was found for each of the controls (Figure 7B, words latency/amplitude: μ = 359 msec/86 fT, range = 324–408 msec/68–109 fT; nonwords latency/amplitude: μ = 331 msec/73 fT, range = 310–370 msec/57–89 fT), with similar LH surface distributions to each participant's LH M100 (Figure 6). Interestingly, DMN showed a normal range M350 response to words (latency/amplitude = 372 msec/78.4 fT) but differed from controls in two aspects: (1) he had no identifiable M350 response to nonwords, and (2) whereas each control showed similar LH M100 and M350 topography, each individual's M350 response to both words and nonwords was reduced in amplitude relative to their M100 response (range = 1–85 fT greater for M100). In contrast, DMN showed markedly larger M350 amplitude relative to the attenuated M100 (+30.2 fT).

Experiment 3: Perception of Phonemic Category — Identification and Discrimination—DMN's discrimination of tokens from word and nonword continua (in the task presented for f MRI scanning) revealed that relative to controls, DMN was significantly less sensitive to between-category differences for words (*p* < .005) and nonwords (*p* < .05) but was within normal range for sensitivity to within-category and end-point differences (Figure 8). Although DMN was numerically better at word-embedded phoneme discrimination (*d′* = .41) than non-word-embedded phoneme discrimination (*d′* = −.15), this difference was not statistically reliable. DMN was within normal RT range for end-point word trials but significantly slower to respond to end-point nonword (*p* < .05), between-category word (*p* < .001), between-category nonword (*p* < .05), within-category word (*p* < .05), within-category nonword (*p* < .05) trials.

With regard to VOT identifications, we found that for each continuum, for each control participant, linear fits produced greater deviance than did logistic fits (Figure 9B and D). Accordingly, their identification functions were significantly better fit by logistic as compared with linear functions (Wilcoxon sign-rank test on deviance differences, $P < .05$ for each control), demonstrating the standard influence of phonemic category on identification for phonemes in words or nonword onsets. DMN showed two departures from this typical response.

First, DMN did not show the characteristic advantage for logistic function fit. Instead, DMN's word-embedded phoneme identifications were gradient: His responses were equally well fit by logistic and linear functions (Wilcoxon signed rank test for difference, $P > .68$), and each function had a strong linear correlation (Pearson correlation coefficients ranging from .63 to .98, p values ranging from .001 to .07). The difference in deviances between logistic and linear fits for DMN was significantly less than that of the controls ($p < .05$) in each of the word and nonword continua. The same results were observed with the additional word continua completed only by DMN. Second, DMN showed a marked difference in identification when the onset was nonword embedded (compared with word-embedded). Although DMN's word-embedded phoneme identification was gradient (significantly linearly correlated with VOT but not better fit by a logistic), there was no correlation between VOT and DMN's non-word-embedded phoneme identification (and logistic and linear fits were equally poor).

In a final identification paradigm, word and nonword tokens (from beach/peach and beesh/ peesh continua) were mixed and presented in random order; DMN was not told of the inclusion of nonwords, and response choices were limited to word responses ("beach/b," "peach/p"). When asked, DMN did not report anything odd about the paradigm nor did he think any nonwords were present. Although onset phonemes were acoustically identical in the word and nonword continua and DMN had no conscious awareness of the inclusion of nonwords, performance differed for the word and nonword continua (Figure 10), indicating stimulus-driven differences in identification. Although both word and nonword identifications were gradient (similarly well fit by linear and logistic best-fit functions and significantly correlated with VOT), word-embedded phoneme identification produced a 36% steeper slope than non-word-embedded phoneme identification, demonstrating differences in the use of acoustic cues across word and nonword contexts. In addition, this task also provided some indication of knowledge-driven or top-down influence on identification: When DMN thought he was hearing and responding to words (as in this task), nonword identification was correlated with VOT ($r = .83$, $p < .01$), but when DMN thought he was hearing and responding to nonwords (as he was in the other VOT identification tasks), there was no correlation between nonword identification and VOT ($r = 2$, $p < .6$).

Study 2 Summary

Functional imaging (fMRI and MEG) revealed that DMN was RH reliant for, at a minimum, the initial acoustic and phonetic analyses of auditory input, rendering him an informative case study for understanding the speech processing capacities of the RH. Specifically, both his BOLD responses to speech from the left HG as well as his M100 auditory-evoked

responses (AER) to tones were absent or weak. The source of the auditory-evoked M100 has been localized to the superior temporal area bilaterally and reflects acoustic processing particularly important for speech (Gage, Poeppel, Roberts, & Hickok, 1998).

Given the cumulative evidence for phonemic category sensitivity in the left posterior temporal lobe, DMN's poor performance with tokens from VOT continua may not be entirely unexpected. What is particularly noteworthy is that although DMN could differentiate between VOT end-point minimal pairs, he did not show the typical phonemic category influence on discrimination or identification, producing instead gradient identification functions only for word- and not non-word-embedded phonemes. His performance, therefore, reveals not only the RH hemisphere capacity for gradient subphonemic processing but also that this type of processing interacts with the lexical status of the acoustic stimulus. The topics of subphonemic RH processing and mechanisms underlying lexical effects in phoneme perception will be taken up in the General Discussion section.

GENERAL DISCUSSION

On the basis of converging evidence from different populations (neurologically intact and lesioned individuals) and methods, we examined the claim that "what underlies the left (hemisphere) dominance for speech consonants in the temporal lobes is their categorical perception" (Liebenthal et al., 2005).

We found that BOLD response in the superior posterior left temporal lobe showed selective sensitivity to phonemic status. Supporting the critical role of the LH in phoneme perception was the performance of DMN, an individual with a lesion affecting this region, whose performance revealed that the RH alone was insufficient for typical phonemic category effects. With regard to the RH's role, we found that it was not only active in normal listeners but that, when functioning in the context of a deafferented LH, it allowed the processing of gradient phonetic information, at least when provided with lexical support.

Highly relevant to DMN's gradient perception are the findings of Desai et al. (2008), who asked neurologically intact participants to identify and discriminate tokens from phonetic continua made from sine-wave speech before and after familiarization with sine-wave speech. This paradigm allows a comparison of performance with the same acoustic stimuli with and without the contributions of learned phoneme categories or lexical knowledge. Before familiarization, participants responded similarly to DMN, producing gradient identification functions well correlated with acoustic dimensions. After familiarization, they produced categorical identification and increasingly recruited the left posterior STS and STG, further linking this region, lesioned in DMN, to the influence of phonemic category.

Subphonemic Processing and the RH

DMN's gradient RH responsivity to VOT continua is consistent with other evidence that, despite the lack of typical phonemic category influence in the RH, it is capable of a good deal of the subphonemic processing needed for successful decoding. For example, although dichotic listening paradigms classically show a right ear advantage for speech, indicating an

LH bias for processing, a left ear advantage for VOT perception has been demonstrated (Cohen, 1981). Complementing these behavioral findings, Molfese (1980) and Simos et al. (2000) reported exclusively RH AER to VOT and tone onset time perception across several paradigms, indicating an RH facility for representing some phonetic features. Within-category acoustic-phonetic representations have been shown in the BOLD responses from the right STG (Myers et al., 2009) and subdural electrodes on the surface of the right STG (Steinschneider, Volkov, Noh, Garell, & Howard, 1999), and information associated with the second and third formants of syllables has been detected in right primary auditory cortex (Raizada, Tsao, Liu, & Kuhl, 2009). Given the ability of right temporal cortex to process phonetic cues, why should phonemic sensitivity be partitioned along hemispheric lines?

There are several plausible explanations for greater LH phonemic sensitivity, all relating to the way phonemic category information is learned. In one account, average temporal window size over which acoustic information is integrated or sampled is longer in the RH (Poeppel, 2003), which is likely to be deleterious to category formation. There are other similar proposals in the literature about this lateralization phenomenon (e.g., Zatorre, 1997), but the data reported here do not differentiate between them. Another account relies on general principles of learning: phonemic sensitivity in perception can develop from sensitivity to statistical structure in the environment (Salminen, Tiitinen, & May, 2009) combined with competitive processing resembling lateral inhibition (Wilson, Wolmetz, & Smolensky, 2008). It may be that RH cortex involved in decoding is simply less plastic than LH counterparts or that cytoarchitectonics or signaling is such that processing between assemblies is less competitive. A final factor may be the connectivity of this area with regions involved in production: LH phonemic representations may be more directly mapped on to relatively discrete left-lateralized motor plans than RH graded acoustic representations. In this way, laterality for phonemic representations would be a consequence of laterality for motor output.

The lack of phonemic influence on RH decoding may contribute to the perceptual impairments caused by LH neural damage but should not be considered a weakness for the typical listener. In some circumstances, within-category information remains on-line past the utterance (McMurray et al., 2008) and can aid in lexical disambiguation, syllabification, speaker-specific tuning, and other speech functions beyond the phoneme. It could be that the same sets of cells simply cannot simultaneously allow for a phonemic influence and encode and maintain the subphonemic details.

Lexical Support for Phoneme Perception

With lexical support, DMN showed evidence of gradient acoustic/phonetic processing but little or no evidence of phonemic influence. Without lexical support, DMN was essentially deaf to acoustic/phonetic detail. What's more, nonword processing seems to have broken down relatively early, as he showed a typical M350 response to words but no identifiable M350 to nonwords. These characteristics are similar to those of an individual studied by Caplan and Utman (1994) who, after a left peri-sylvian lesion, could only discriminate voiced from voiceless phonemes when they appeared in words. These deficits indicate that (1) the posterior superior left temporal lobe and temporo-parietal cortex, in addition to

mediating phonemic category influence, are crucial for some aspects of nonword perception and (2) in cases of impairment, lexical support facilitates the use of acoustic/phonetic detail in speech perception.

Lexical support could come in the form of feedback or lexical anchoring. In the lexical anchoring account, gradient sublexical representations were sufficient to contact the lexicon, and DMN's behavioral responses were based on subsequent lexical activation; as a result, his word responses were linearly related to VOT. Nonwords either did not contact word representation or did not do so to the same degree as words, so they showed no evidence of structured perception. In the feedback account, lexical support staves off decay of sublexical information (Norris, McQueen, & Cutler, 2000; or referred to as "resonance boost" by Grossberg, Boardman, & Cohen, 1997). In the impaired system, without phonemic organization of the perceptual space, fragile sublexical information decays rapidly, but with lexical support, sublexical acoustic/phonetic information decays slowly enough to permit gradient responses. In other words, without lexical support, transient acoustic/phonetic information does not persist long enough to impact behavioral responses.

In the case of DMN, both feedback and lexical anchoring seem at play. Feedback is indicated by DMN's increased ability to process phonetic cues when he perceived them as being word embedded (even when they were not). Lexical anchoring may account for the persistent word advantage. Imaging data presented here suggest that the posterior right temporal lobe has a role in this lexical mediation of phonetic perception. Further support for this interpretation comes from DMN's apparently intact M350 to auditorily presented words —a component with a similar dipole source and surface distribution to the M100 thought to reflect activation of lexical candidates (Pylkkanen et al., 2002). These results indicate that some regions of DMN's left temporal lobe, although inactive during early LH prelexical processes, participated in later lexical processing, presumably via RH input. Nonetheless, the details of RH-LH interaction remain the topic of future research.

Conclusions

We show that the right posterior temporal lobe is active during phoneme discrimination in normal listeners and contributes to speech perception in the context of a deafferentated LH. The RH was shown to be capable of gradient response, faithful to the acoustic properties of the speech input, when lexical support was available. The picture that emerges is one in which, in normal listeners, fast speech comprehension is achieved with RH acoustic/phonetic representations of speech working in concert with LH mechanisms more sensitive to phonemic category.

Acknowledgments

REFERENCES

Ahmar, NE.; Simon, JZ. MEG adaptive noise suppression using fast IMS; Paper presented at the International IEEE EMBS Conference on Neural Engineering; March 16–19, 2005; Washington, D.C. 2005.

Andoh J, Artiges E, Pallier C, Riviere D, Mangin JF, Cachia A, et al. Modulation of language areas with functional MR image-guided magnetic stimulation. Neuroimage. 2006; 29:619–627. [PubMed: 16168674]

Basso A, Casati C, Vignolo LA. Phonemic identification defect in aphasia. Cortex. 1977; 13:85–95. [PubMed: 844311]

Blumstein SE, Baker E, Goodglass H. Phonological factors in auditory comprehension in aphasia. Neuropsychologia. 1977; 15:19–30. [PubMed: 831150]

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

Boatman D, Freeman J, Vining E, Pulsifer M, Miglioretti D, Minahan R, et al. Language recovery after left hemispherectomy in children with late-onset seizures. Annals of Neurology. 1999; 46:579–586. [PubMed: 10514094]

Boatman D, Hart J Jr, Lesser RP, Honeycutt N, Anderson NB, Miglioretti D, et al. Right hemisphere speech perception revealed by amobarbital injection and electrical interference. Neurology. 1998; 51:458–464. [PubMed: 9710019]

Boatman D, Vining EP, Freeman J, Carson B. Auditory processing studied prospectively in two hemidecorticectomy patients. Journal of Child Neurology. 2003; 18:228–232. [PubMed: 12731649]

Boatman DF, Miglioretti DL. Cortical sites critical for speech discrimination in normal and impaired listeners. Journal of Neuroscience. 2005; 25:5475–5480. [PubMed: 15944375]

Bonakdarpour B, Parrish TB, Thompson CK. Hemodynamic response function in patients with stroke-induced aphasia: Implications for fMRI data analysis. Neuroimage. 2007; 36:322–331. [PubMed: 17467297]

Caplan D, Utman JA. Selective acoustic phonetic impairment and lexical access in an aphasic patient. Journal of the Acoustical Society of America. 1994; 95:512–517. [PubMed: 8120262]

Cohen, H. PhD thesis. Montreal: Concordia University; 1981. Hemispheric contributions to the perceptual representation of speech sounds.

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

Crawford JR, Garthwaite PH. Comparison of a single case to a control or normative sample in neuropsychology: Development of a Bayesian approach. Cognitive Neuropsychology. 2007; 24:343–372. [PubMed: 18416496]

Dehaene-Lambertz G. Neural correlates of switching from auditory to speech perception. Neuroimage. 2005; 24:21–33. [PubMed: 15588593]

Desai R, Liebenthal E, Waldron E, Binder JR. Left posterior temporal regions are sensitive to auditory categorization. Journal of Cognitive Neuroscience. 2008; 20:1–15. [PubMed: 17919082]

Dunn, LM.; Dunn, DM. Peabody Picture Vocabulary Test. Circle Pines, MN: American Guidance Service; 1997.

Fiorentino R, Poeppel D. Compound words and structure in the lexicon. Language and Cognitive Processes. 2007; 22:953–1000.

Friston KJ, Frith CD, Frackowiak RS, Turner R. Characterizing dynamic brain responses with fMRI: A multivariate approach. Neuroimage. 1995; 2:166–172. [PubMed: 9343599]

Gage N, Poeppel D, Roberts TP, Hickok G. Auditory evoked m100 reflects onset acoustics of speech sounds. Brain Research. 1998; 814:236–239. [PubMed: 9838140]

Gow DW Jr, Caplan D. An examination of impaired acoustic-phonetic processing in aphasia. Brain and Language. 1996; 52:386–407. [PubMed: 8811970]

Gow DW Jr, Segawa JA, Ahlfors SP, Lin FH. Lexical influences on speech perception: A granger causality analysis of MEG and EEG source estimates. Neuroimage. 2008; 43:614–623. [PubMed: 18703146]

Griffiths T, Rees A, Green G. Disorders of human complex sound processing. Neurocase. 1999; 5:365–378.

Grossberg S, Boardman I, Cohen M. Neural dynamics of variable-rate speech categorization. Journal of Experimental Psychology: Human Perception and Performance. 1997; 23:481–503. [PubMed: 9104006]

Hall DA, Haggard MP, Akeroyd MA, Palmer AR, Summerfield AQ, Elliott MR, et al. "Sparse" temporal sampling in auditory fMRI. Human Brain Mapping. 1999; 7:213–223. [PubMed: 10194620]

Hasson U, Skipper JI, Nusbaum HC, Small SL. Abstract coding of audiovisual speech: Beyond sensory representation. Neuron. 2007; 56:1116–1126. [PubMed: 18093531]

Hickok G, Poeppel D. Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. Cognition. 2004; 92:67–99. [PubMed: 15037127]

Hickok G, Poeppel D. The cortical organization of speech processing. Nature Reviews Neuroscience. 2007; 8:393–402.

Hillenbrand JM, Gayvert RT. Open source software for experiment design and control. Journal of Speech, Language, and Hearing Research. 2005; 48:45–60.

Huntress LM, Lee L, Creaghead NA, Wheeler DD, Braverman KM. Aphasic subjects' comprehension of synthetic and natural speech. Journal of Speech and Hearing Disorders. 1990; 55:21–27. [PubMed: 2137187]

Joanisse MF, Zevin JD, McCandliss BD. Brain mechanisms implicated in the preattentive categorization of speech sounds revealed using fMRI and a short-interval habituation trial paradigm. Cerebral Cortex. 2007; 17:2084–2093. [PubMed: 17138597]

Kotz SA, Cappa SF, von Cramon DY, Friederici AD. Modulation of the lexical-semantic network by auditory semantic priming: An event-related functional MRI study. Neuroimage. 2002; 17:1761–1772. [PubMed: 12498750]

Kourtzi, Z.; Grill-Spector, K. Fmri adaptation: A tool for studying visual representations in the primate brain. In: Rhodes, G.; Clifford, C., editors. Fitting the mind into the world: Adaptation and after effects in high-level vision, advances in visual cognition. Oxford, UK: Oxford University Press; 2005.

Liberman AM, Harris KS, Hoffman HS, Griffith BC. The discrimination of speech sounds within and across phoneme boundaries. Journal of Experimental Psychology. 1957; 54:358–368. [PubMed: 13481283]

Liebenthal E, Binder JR, Spitzer SM, Possing ET, Medler DA. Neural substrates of phonemic perception. Cerebral Cortex. 2005; 15:1621–1631. [PubMed: 15703256]

Liegeois-Chauvel C, de Graaf JB, Laguitton V, Chauvel P. Specialization of left auditory cortex for speech perception in man depends on temporal coding. Cerebral Cortex. 1999; 9:484–496. [PubMed: 10450893]

Marcell MM, Borella D, Greene M, Kerr E, Rogers S. Confrontation naming of environmental sounds. Journal of Clinical and Experimental Neuropsychology. 2000; 22:830–864. [PubMed: 11320440]

McMurray B, Aslin RN. Infants are sensitive to within-category variation in speech perception. Cognition. 2005; 95:B15–B26. [PubMed: 15694642]

McMurray B, Aslin RN, Tanenhaus MK, Spivey MJ, Subik D. Gradient sensitivity to within-category variation in words and syllables. Journal of Experimental Psychology: Human Perception and Performance. 2008; 34:1609–1631. [PubMed: 19045996]

McMurray B, Spivey M. The categorical perception of consonants: The interaction of learning and processing. Proceedings of the Chicago Linguistics Society. 2000; 34:205–220.

Meister IG, Wilson SM, Deblieck C, Wu AD, Iacoboni M. The essential role of premotor cortex in speech perception. Current Biology. 2007; 17:1692–1696. [PubMed: 17900904]

Molfese DL. Hemispheric specialization for temporal information: Implications for the perception of voicing cues during speech perception. Brain and Language. 1980; 11:285–299. [PubMed: 7470850]

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

Myers EB, Blumstein SE. The neural bases of the lexical effect: An fMRI investigation. Cerebral Cortex. 2008; 18:278–288. [PubMed: 17504782]

Myers EB, Blumstein SE, Walsh E, Eliassen J. Inferior frontal regions underlie the perception of phonetic category invariance. Psychological Science. 2009; 20:895–903. [PubMed: 19515116]

Newman SD, Twieg D. Differences in auditory processing of words and pseudowords: An fMRI study. Human Brain Mapping. 2001; 14:39–47. [PubMed: 11500989]

Norris D, McQueen JM, Cutler A. Merging information in speech recognition: Feedback is never necessary. Behavioral and Brain Sciences. 2000; 23:299–325. discussion 325–270. [PubMed: 11301575]

Obleser J, Eisner F. Pre-lexical abstraction of speech in the auditory cortex. Trends in Cognitive Sciences. 2009; 13:14–19. [PubMed: 19070534]

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

Pastore, RE. Categorical perception: Some psychophysical models. In: Harnad, S., editor. Categorical perception. Cambridge: Cambridge University Press; 1987. p. 29-52.

Poeppel D. Pure word deafness and the bilateral processing of the speech code. Cognitive Science. 2001; 25:679–693.

Poeppel D. The analysis of speech in different temporal integration windows: Cerebral lateralization as "asymmetric sampling in time.". Speech Communication. 2003; 41:245–255.

Pylkkanen L, Stringfellow A, Marantz A. Neuromagnetic evidence for the timing of lexical activation: An MEG component sensitive to phonotactic probability but not to neighborhood density. Brain and Language. 2002; 81:666–678. [PubMed: 12081430]

Raizada RD, Poldrack RA. Selective amplification of stimulus differences during categorical processing of speech. Neuron. 2007; 56:726–740. [PubMed: 18031688]

Raizada RD, Tsao FM, Liu HM, Kuhl PK. Quantifying the adequacy of neural representations for a cross-language phonetic discrimination task: Prediction of individual differences. Cerebral Cortex. 2009; 20:1–12. [PubMed: 19386636]

Rosen, S.; Howell, P. Auditory, articulatory, and learning explanations of categorical perception in speech. In: Harnad, S., editor. Categorical perception. Cambridge: Cambridge University Press; 1987. p. 113-160.

Salajegheh A, Link A, Elster C, Burghoff M, Sander T, Trahms L, et al. Systematic latency variation of the auditory evoked M100: From average to single-trial data. Neuroimage. 2004; 23:288–295. [PubMed: 15325376]

Salminen NH, Tiitinen H, May PJ. Modeling the categorical perception of speech sounds: A step toward biological plausibility. Cognitive, Affective & Behavioral Neuroscience. 2009; 9:304–313.

Schouten B. The end of categorical perception as we know it. Speech Communication. 2003; 41:71–80.

Simons J, Ralph ML. The auditory agnosias. Neurocase. 1999; 5:379–406.

Simos PG, Breier JI, Wheless JW, Maggio WW, Fletcher JM, Castillo EM, et al. Brain mechanisms for reading: The role of the superior temporal gyrus in word and pseudoword naming. NeuroReport. 2000; 11:2443–2447. [PubMed: 10943701]

Steinschneider M, Schroeder CE, Arezzo JC, Vaughan HG Jr. Physiologic correlates of the voice onset time boundary in primary auditory cortex (al) of the awake monkey: Temporal response patterns. Brain and Language. 1995; 48:326–340. [PubMed: 7757450]

Steinschneider M, Volkov IO, Noh MD, Garell PC, Howard MA III. Temporal encoding of the voice onset time phonetic parameter by field potentials recorded directly from human auditory cortex. Journal of Neurophysiology. 1999; 82:2346–2357. [PubMed: 10561410]

Wilson, C.; Wolmetz, M.; Smolensky, P. Replicator dynamics of speech perception and categorization; Paper presented at the Conference on Laboratory Phonology; July 2008; Victoria University Wellington, New Zealand. 2008.

Zaidel, E. Lexical organization in the right hemisphere. In: Buser, P.; Gougeul-Buser, A., editors. Cerebral correlates of conscious experience. Amsterdam: Elsevier; 1978. p. 177-197.

Zatorre, RJ. Cerebral correlates of human auditory processing: Perception of speech and musical sounds. In: Syka, J., editor. Acoustical signal processing in the central auditory system. New York: Plenum Press; 1997. p. 453-468.

Zevin JD, McCandliss BD. Dishabituation of the bold response to speech sounds. Behavioral and Brain Functions. 2005; 1:4. [PubMed: 15953396]
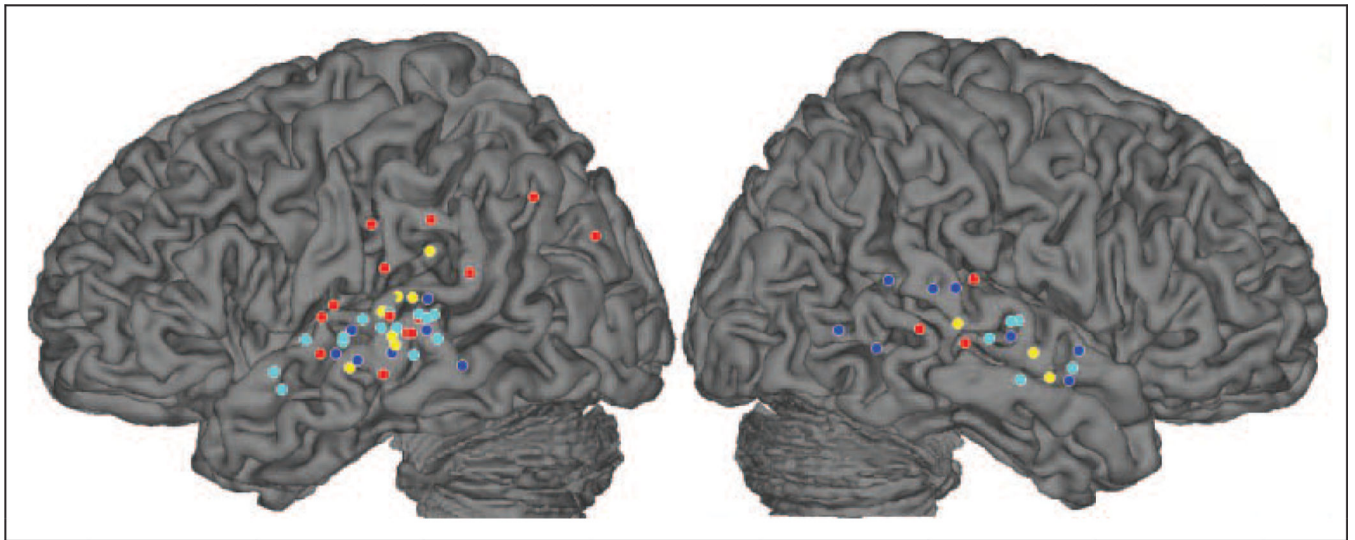
Figure 1.
Review of LH and RH temporal lobe and temporoparietal findings from 19 imaging studies. Colors indicate the level of processing targeted by the contrast plotted, with the cold colors representing the least abstract, sensitivity to spectral complexity and temporal cues (blue), and sensitivity to speech as compared with matched nonspeech (cyan) to the warmer colors representing sensitivity to syllabic and segmental dimensions (yellow) and phonemic category (red).
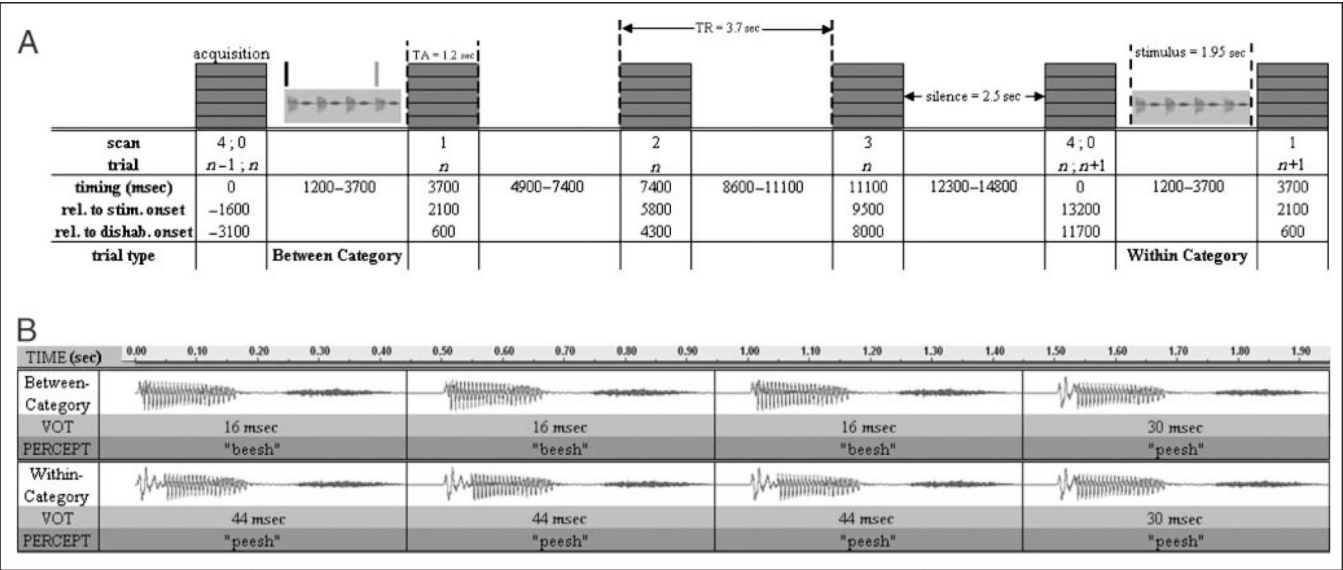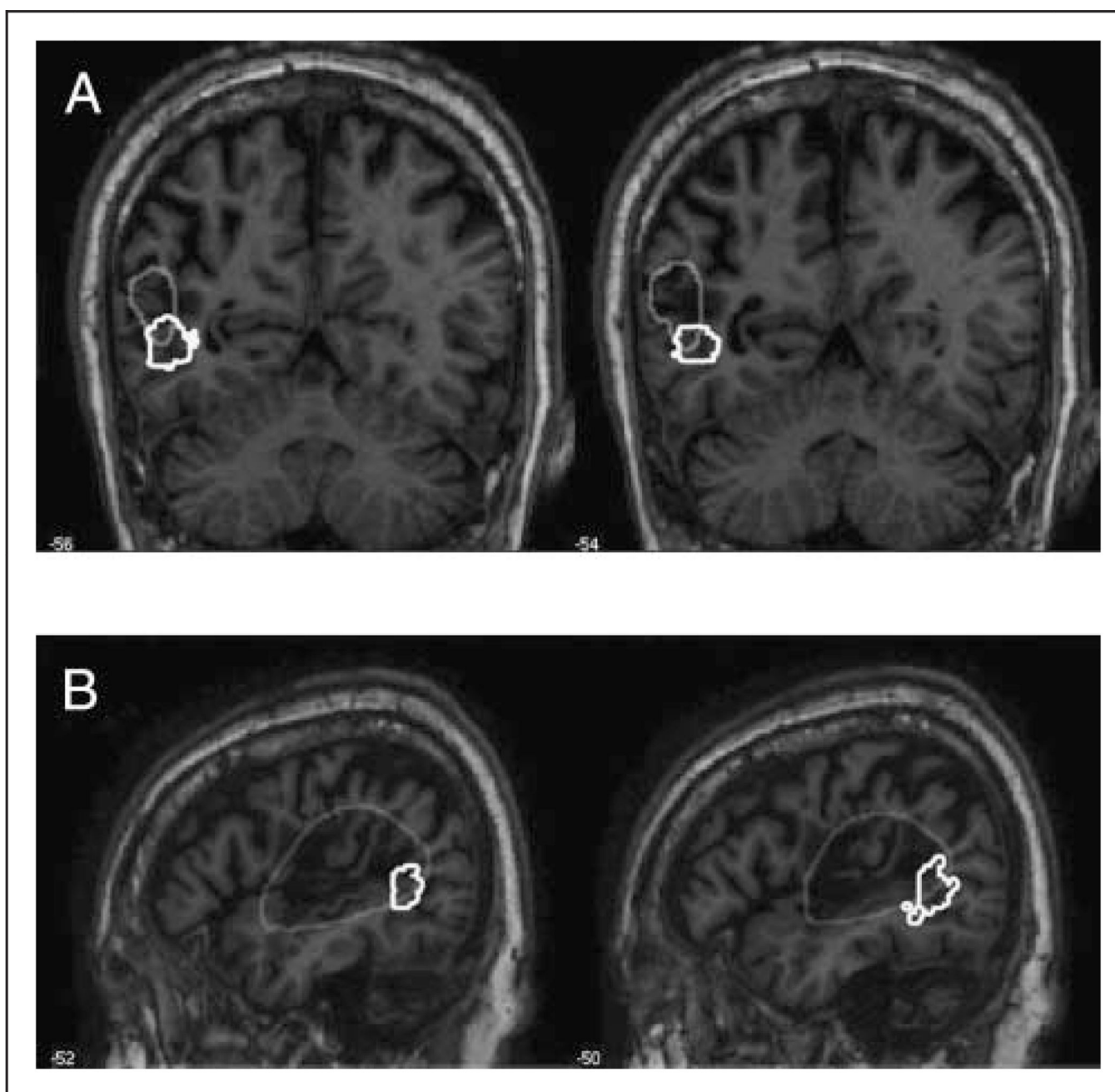
Figure 2.
Stimulus presentation and scan timing. Scanning schematic for one trial of the phonemic category adaptation paradigm (A), and examples of between-category and within-category trains (B, blown up and detailed from A). In A, the dark vertical bar shown above the first stimulus indicates the temporal placement of the speech onset regressor, and the light vertical bar indicates the temporal placement of the dishabituation or rebound regressor, for this trial. Timing is listed relative to the onset of acquisition, relative to the stimulus onset, as well as relative to the dishabituation or rebound onset. Scan indicates acquisition within the trial: for example, the first acquisition shown is listed as scan 4 (of the previous trial, *n* − 1) and scan 0 (of the current trial, *n*).
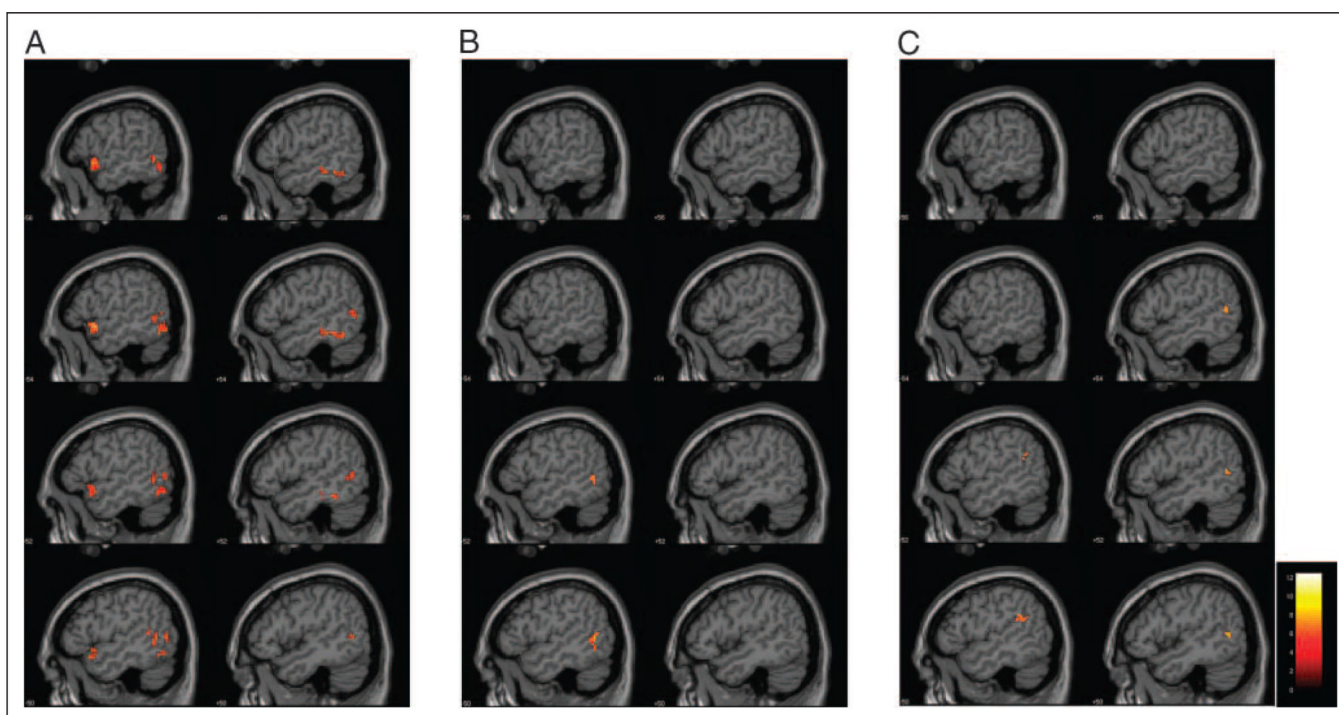
Figure 3.
Group *t* test results for the following contrasts: Between category > Within Category for Words (A), Between category > Within Category for Nonwords (B), and the interaction, Between category > Within Category for Words > Nonwords (C). LH on left, *x* ranging from − 56 to −50; RH on right, *x* ranging from 56 to 50.
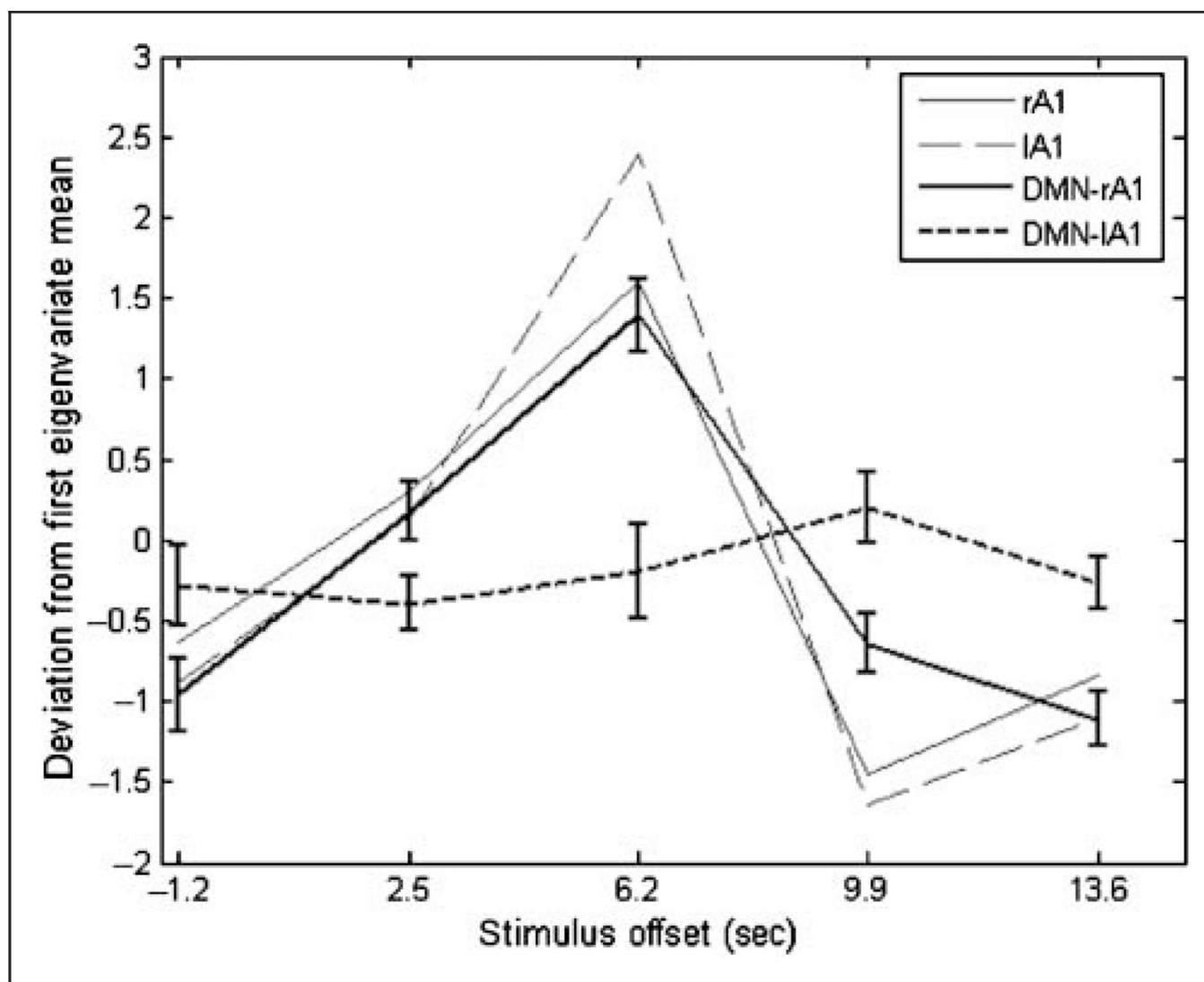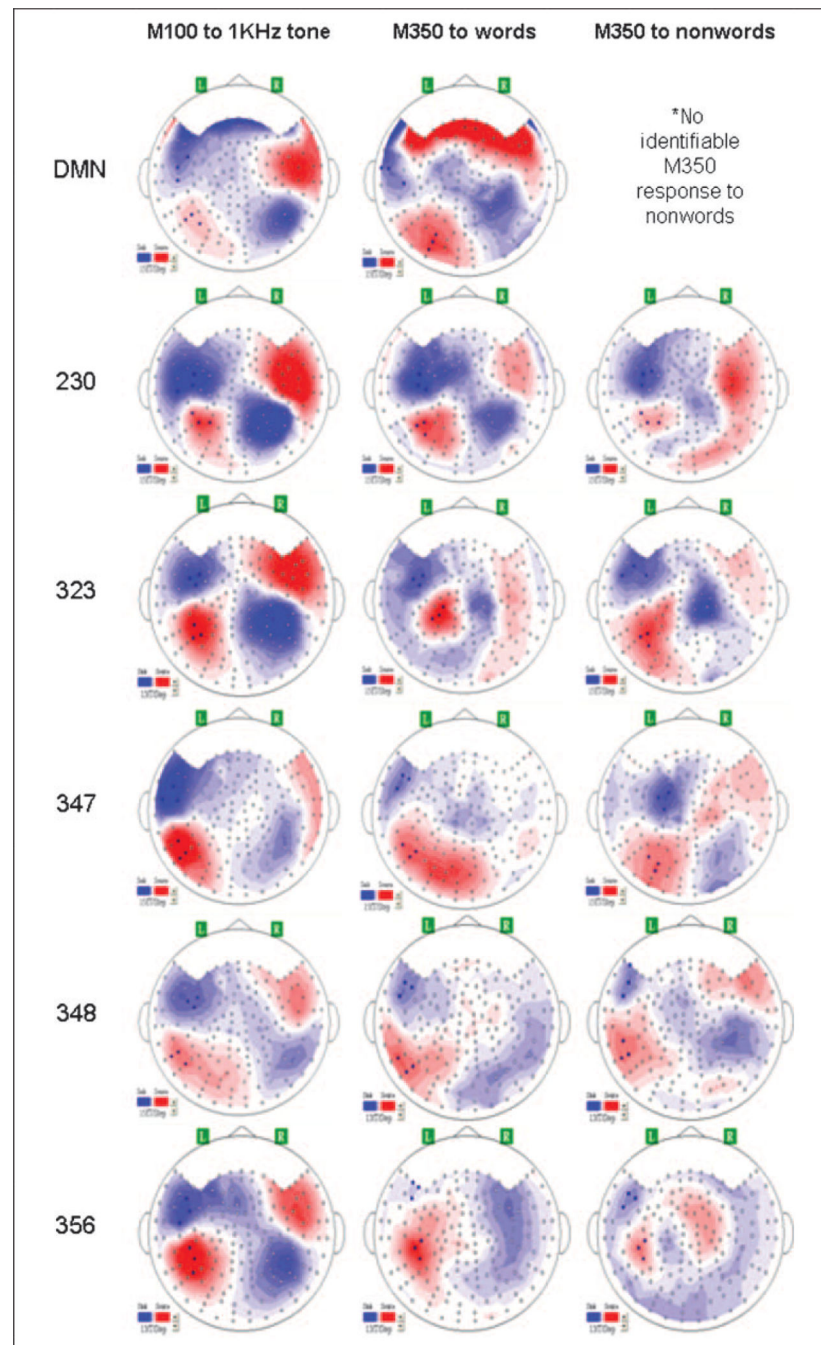
Figure 4.
Normalized MPRAGE of DMN's LH, as shown in coronal (A) and sagittal (B) slices. The gray border indicates outline of normalized lesion tracing, and the white border indicates outline of the main effect of phonemic dishabituation from Study 1. Coronal slices range from $y = -56$ to -54. Sagittal slices range from $x = -52$ to -50.
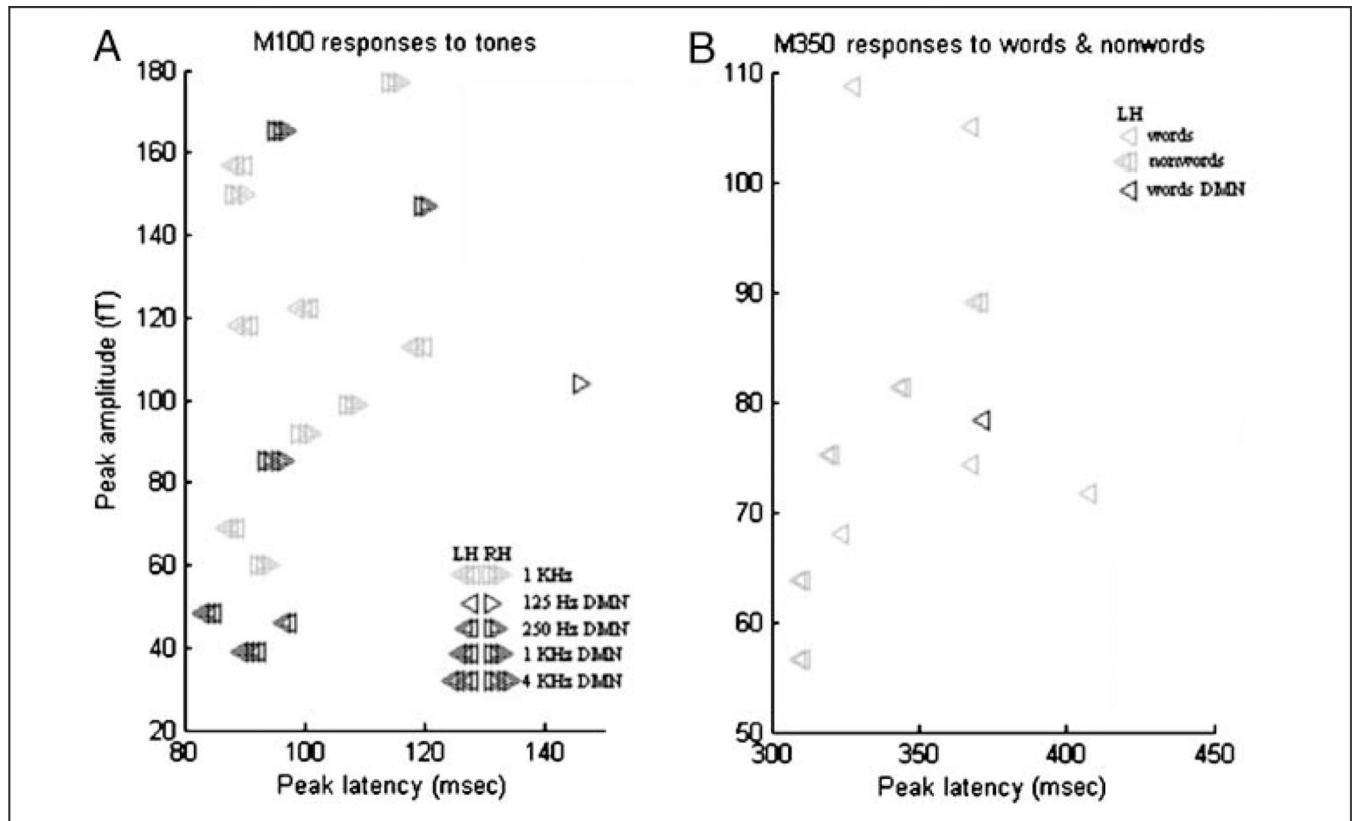
Figure 5.
Event-related response to speech in left and right HG ROIs for control participants in gray and DMN in black. Error bars denote one standard error of DMN's mean response for each time bin for each hemisphere.
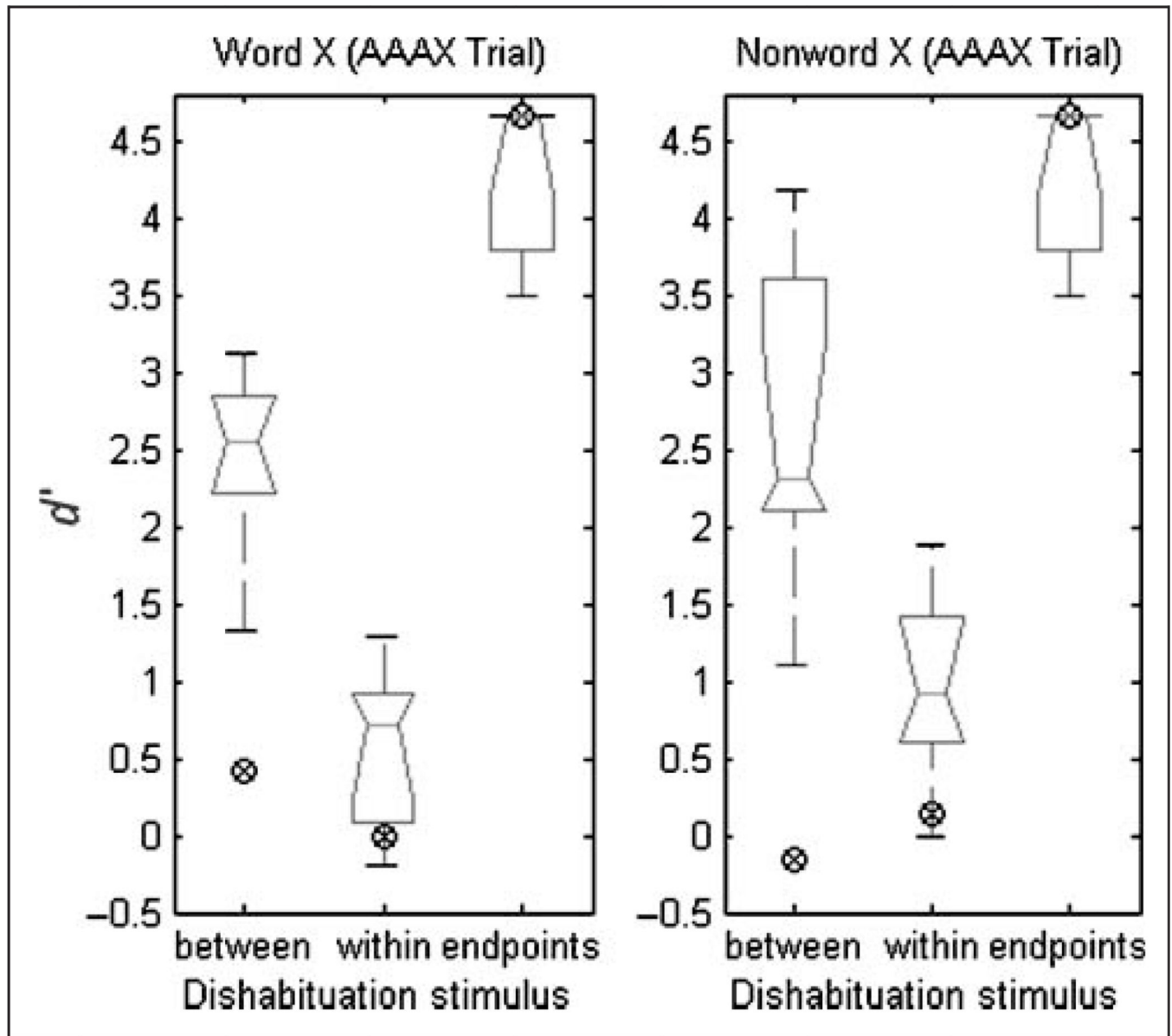
Figure 6.
Surface topographies for passive listening. Each topography is shown at the time point of peak amplitude for LH channels (scale = 15 fT/step). DMN's data are displayed in the first row, and each subsequent row corresponds to a different control participant.

Figure 7.
Peak amplitudes and latencies at peak for the M100 response to passively heard tones (A) and the M350 response to passively heard words (B). Control participants' responses are limited to 1 kHz tones displayed in gray and DMN's responses in black. Leftward-facing arrows denote responses from the LH, whereas rightward-facing arrows denote responses from the RH. No markers are present for DMN's LH 125-Hz M100 or nonword LH M350 as no identifiable response could be found. Note that the scales are different in panels A and B.

Figure 8.
Box and whisker plots showing the median, upper and lower quartiles, and range of in-scanner discrimination sensitivities (quantified as d') from normal listeners for word stimuli on the left and nonword stimuli on the right. Indentations in the boxes denote the 95% confidence intervals of medians, such that boxes with indentations that do not overlap (e.g., nonwords, between and within) have significantly different medians ($p < .05$), whereas those that do overlap (e.g., word-between and nonword-between) do not. For comparison, sensitivity for discrimination of continuum end points is also shown as well as sensitivities for DMN represented by the x.
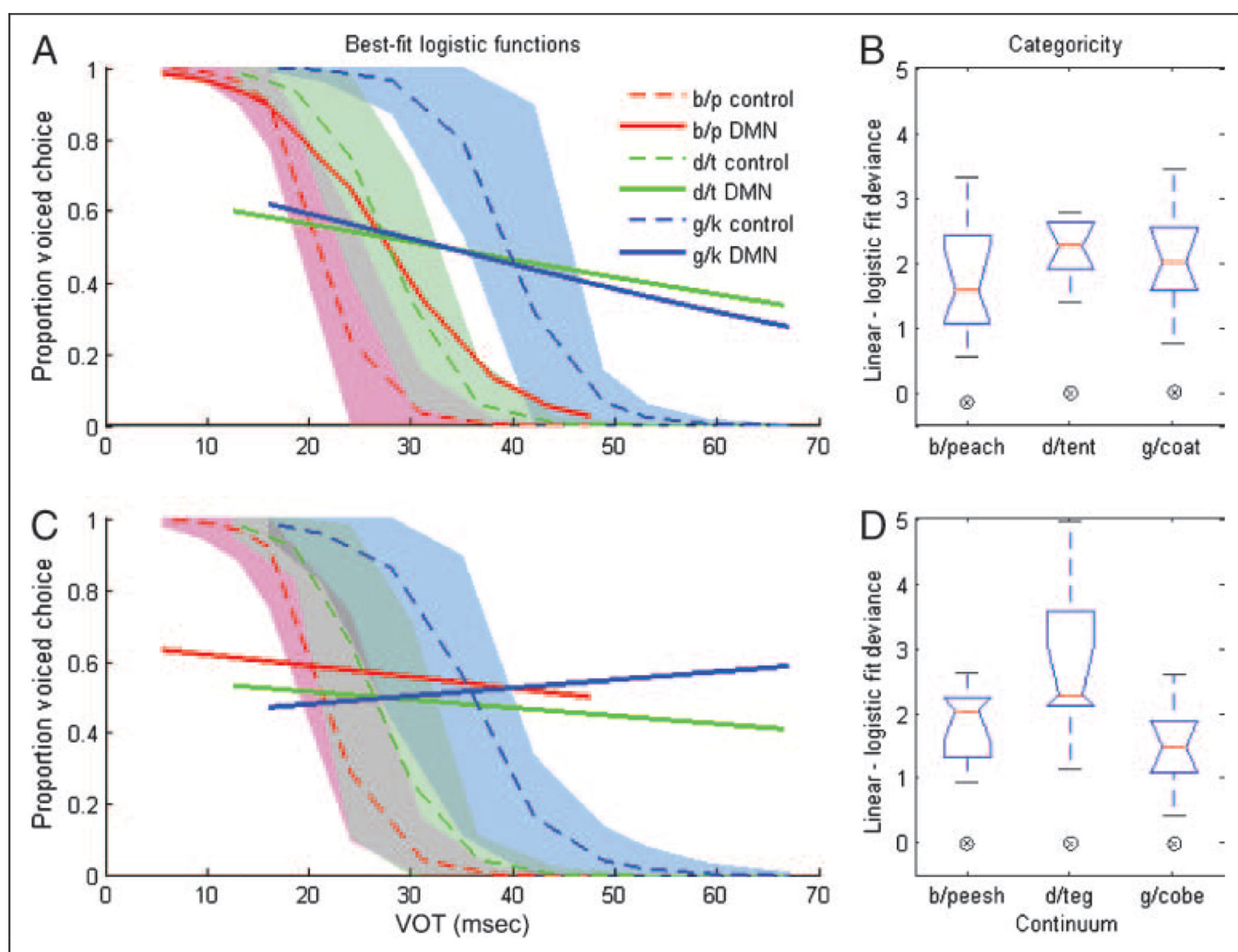
Figure 9.
Best-fit logistic functions for DMN (solid lines) and control identification (dashed) for word-embedded obstruent consonants (A) and matched nonword embedded obstruents (C) presented from continua. Color fields cover the range of control response. Box and whisker plots in panels B and D show the difference in deviances between linear function and logistic function fits for controls and DMN (marked as a ⊗) for words and nonwords, respectively (median, upper and lower quartiles, and range). Large deviance differences indicate a large categorical influence on identification, whereas small deviance differences indicate a small influence.
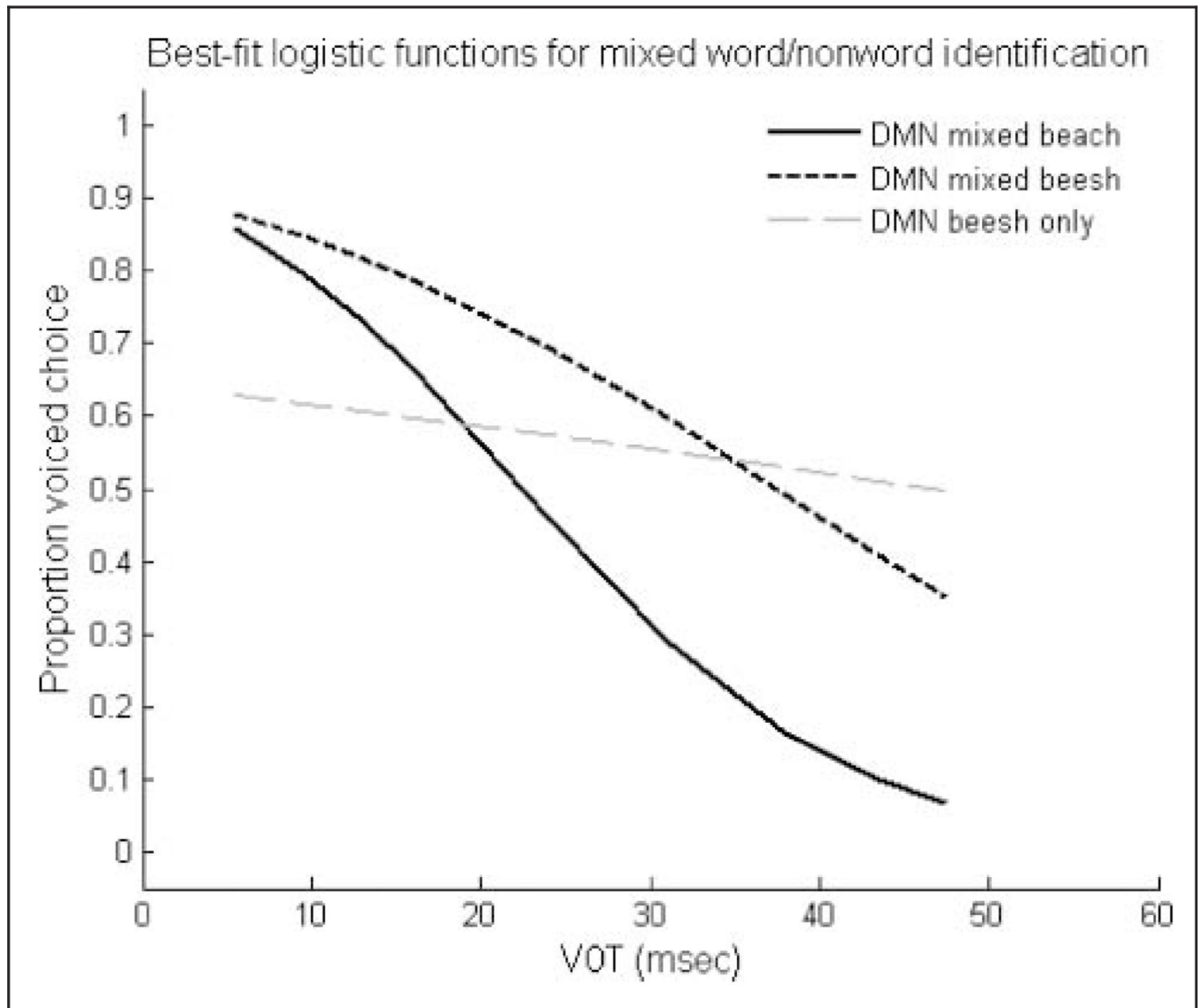
Figure 10.
Results from DMN's mixed word/nonword identification (dark lines). Although both words and nonwords were presented, DMN thought he was hearing and responding to words only. Nonword only results (gray) are shown for reference.

Table 1

Means and Standard Deviations for Sensitivities (d') and RTs (Post-offset) to the In-scanner Discrimination Task

	Words		Nonwords	
	Between Category	**Within Category**	**Between Category**	**Within Category**
d'	2.44 (0.59)	0.56 (0.53)	2.68 (1.07)	0.94 (0.63)
RT (msec)	352 (181)	495 (245)	401 (124)	395 (276)

Within-category sensitivity refers to the ability to hear an acoustic difference between members of the same phonemic category, whereas between-category sensitivity refers to the ability to hear an acoustic difference across a likely phonemic boundary.

Table 2

Results of the Single-Group *t* Tests Performed on Five Contrasts, Corrected by Alpha Probability Simulation to an α < .05

Hem	AAL Region	Local Max x	y	z	Voxel	Size (mm³)	Max t Value
Speech > Silence							
LH	MTG	−54	−36	−12	632	5056	12.93
LH	IPL, pre/postcentral gyrus	−44	−28	36	365	2920	12.81
LH	Cerebellum	−32	−48	−44	122	976	21.55
LH	Parahippocampal gyrus	−12	22	−20	502	4016	26.03
RH	Cerebellum	8	−52	−44	324	2592	13.3
RH	Calcarine	12	−92	−4	379	3032	12.6
RH	Postcentral gyrus	26	−34	32	195	1560	10.22
RH	MTG	46	−46	4	162	1296	32.1
RH	MTG	50	−6	−20	352	2816	18.59
RH	STG	52	−32	2	134	1072	11.6
Phoneme-specific Dishabituation (Between > Within)							
LH	MTG	−48	−56	4	87	696	12.47
Word-embedded Phoneme-specific Dishabituation (Between > Within Words)							
LH	Temporal Pole	−54	14	−8	256	2120	11.98
LH	MTG	−50	−56	4	355	2840	8.1
LH	Angular gyrus	−42	−56	34	181	1448	9.97
LH	Precentral gyrus	−40	8	46	50	400	5.45[a]
LH	Cerebellum	−36	−48	−40	106	848	8.31
LH	Insula	−34	−18	0	341	2728	11.04
LH	Hippocampus	−16	−20	−12	117	936	17.57
LH	Anterior cingulate gyrus	−16	42	8	84	672	7.86
RH	Cerebellum	16	−34	36	134	1072	12.04
RH	Parahippocampal gyrus	24	−10	−32	97	776	10.69
RH	Putamen	34	8	−10	463	3704	15.4
RH	Inferior parietal lobule	44	−46	42	77	616	10.21
RH	ITG	52	−40	−18	85	680	8.45

Hem	AAL Region	Local Max x	y	z	Voxel	Size (mm³)	Max t Value
RH	MTG	52	−60	6	99	792	8.19
Non-Word-embedded Phoneme-specific Dishabituation (Between > Within Nonwords)							
LH	MTG	−50	−58	6	94	752	6.78
LH	Anterior cingulate gyrus	−12	48	0	48	384	4.86[a]
RH	Hippocampus	24	−28	−6	81	648	7.46
RH	Temporal pole	24	12	−36	84	672	6.04
Interaction (Between > Within for Words > Nonwords)							
LH	Pre/postcentral gyrus	−36	−8	36	119	952	11.53
LH	Angular gyrus	−34	−58	22	804	6432	11.39
LH	Anterior cingulate gyrus	−8	48	6	86	688	9.4
RH	Cerebellum	44	−44	−32	92	736	6.35
RH	MTG	52	−58	10	190	1520	10.67
RH	ITG	52	−44	−14	72	576	6.56[a]

AAL regions reflect the locations of peak voxels as listed in the automated anatomical labeling atlas, and two regions are listed when peak voxels fall within 4 mm of another labeled region, but clusters often extend into neighboring regions regardless of peak voxel location.

AAL = automated anatomical labeling; IPL = inferior parietal lobule; MTG = middle temporal gyrus; STG = superior temporal gyrus.

[a] Refers to clusters showing a trend (α < .1).