

Domain General Change Detection Accounts for “Dishabituation” Effects in Temporal–Parietal Regions in Functional Magnetic Resonance Imaging Studies of Speech Perception

Jason D. Zevin,¹ Jianfeng Yang (杨剑峰),² Jeremy I. Skipper,¹ and Bruce D. McCandliss³

¹Sackler Institute for Developmental Psychobiology, Weill Cornell Medical College, New York, New York 10065, ²Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China, and ³Department of Psychology, Vanderbilt University Peabody College of Education and Human Development, Nashville, Tennessee 37240-7817

Functional magnetic resonance imaging (fMRI) studies of speech sound categorization often compare conditions in which a stimulus is presented repeatedly to conditions in which multiple stimuli are presented. This approach has established that a set of superior temporal and inferior parietal regions respond more strongly to conditions containing stimulus change. Here, we examine whether this contrast is driven by habituation to a repeating condition or by selective responding to change. Experiment 1 directly tests this by comparing the observed response to long trains of stimuli against a constructed hemodynamic response modeling the hypothesis that no habituation occurs. The results are consistent with the view that enhanced response to conditions involving phonemic variability reflect change detection. In a second experiment, the specificity of these responses to linguistically relevant stimulus variability was studied by including a condition in which the talker, rather than phonemic category, was variable from stimulus to stimulus. In this context, strong change detection responses were observed to changes in talker, but not to changes in phoneme category. The results prompt a reconsideration of two assumptions common to fMRI studies of speech sound categorization: they suggest that temporoparietal responses in passive paradigms such as those used here are better characterized as reflecting change detection than habituation, and that their apparent selectivity to speech sound categories may reflect a more general preference for variability in highly salient or behaviorally relevant stimulus dimensions.

Introduction

One popular approach to understanding the neural basis of perceptual categorization is to use passive presentation paradigms combined with the logic of habituation. In principle, measuring passive brain responses permits researchers to eschew difficulties with active categorization tasks, in which the degree to which performance is dependent on working memory and decision processes is difficult to ascertain. In habituation studies of speech perception, responses to repeated presentations of the same speech sound are compared with responses to trains of stimuli containing sounds from different categories. When stronger responses are observed in the condition that includes stimuli from multiple categories, it is interpreted as an index of sensitivity to the differences between categories (Celsis et al., 1999; Zevin and McCandliss, 2005). The logic of habituation suggests that this pattern indicates a role for the identified regions in representing

categories: if the response in a region diminishes with multiple presentations of stimuli from the same class, that region must be treating those stimuli as “the same” (Grill-Spector et al., 2006).

Specificity for speech has been explored in this paradigm by comparing responses to linguistically meaningful variability with responses elicited by acoustically matched nonlinguistic change along the same dimensions (Joanisse et al., 2007), and by comparing overlearned native-language contrasts to less familiar second-language contrasts (Jacquemot et al., 2003; Liu et al., 2009). The results of these studies have led to an emerging consensus that regions in and around the temporoparietal junction (TPJ, i.e., supramarginal and posterior superior temporal gyri) respond preferentially to conditions in which phonetically contrastive stimuli are presented.

Whether these responses actually reflect habituation to speech categories is ambiguous for two reasons. First, these studies have used physically identical stimuli in the “habituation” condition, so that any observed habituation could be stimulus specific rather than category specific. Second, direct contrasts cannot distinguish between diminished (habituated) responses when stimuli are repeated and enhanced responses to stimulus change. Further, although the results of these studies have typically been interpreted in terms of habituation (but see Raizada and

Received Aug. 20, 2009; revised Nov. 3, 2009; accepted Dec. 3, 2009.

This research was supported by National Institutes of Health Grants R01-DC007694 and R21-DC0008969 and the Brain CPR Network (James S. McDonnell Foundation).

Correspondence should be addressed to Jason D. Zevin, Sackler Institute for Developmental Psychobiology, Weill Cornell Medical College, 1300 York Avenue, Box 140, New York, NY 10065. E-mail: jdz2001@med.cornell.edu.

DOI:10.1523/JNEUROSCI.4599-09.2010

Copyright © 2010 the authors 0270-6474/10/301110-08\$15.00/0

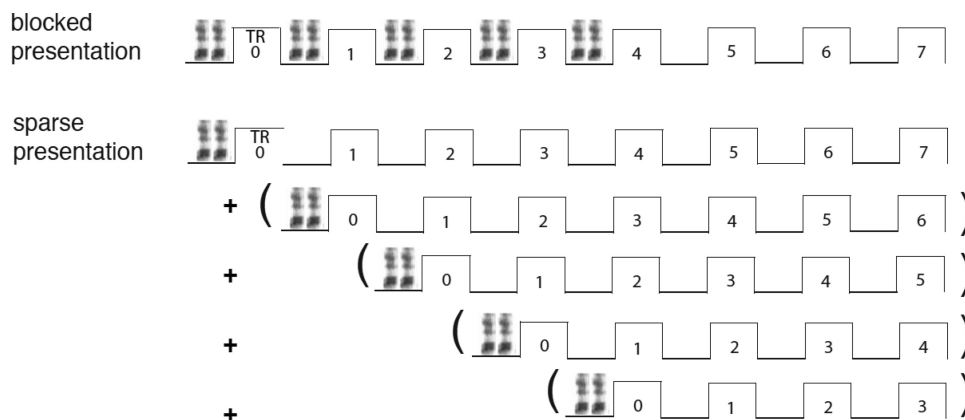


Figure 1. Stimulus presentation and construction of hypothetical HRFs for experiment 1. Stimuli were presented in pairs during the silent interval between acquisition periods (TR, or repeat time = 3 s). During blocked presentation, five pairs of stimuli were presented in one long (15 s) block. During sparse presentation, a single stimulus was presented at times corresponding to the time at which the first stimulus in each block was presented in blocked presentation. To construct a HRF representing the hypothesis that each subsequent response to pairs of stimuli in blocked presentation is identical to the first, we time-shifted the observed data from sparse presentation four times and added these together, as illustrated by the presentations shown in parentheses.

Poldrack, 2007), the anatomical locations typically overlap with regions identified as sensitive to change in behaviorally relevant stimulus features [although this response is often right lateralized (for review, see Corbetta et al., 2008)]. The role these regions play in speech perception is thus an open question, as is the nature and degree of their specialization for speech (for review, see Hickok and Poeppel, 2007; Obleser and Eisner, 2009).

Here we examine two aspects of the response to trains of stimuli similar to those used in previous habituation-based designs to better understand their role in the perception of behaviorally relevant variability in the speech signal. In experiment 1, we compare the response to long trains of stimuli to a model based on shorter trains of stimuli to directly test for habituation under conditions when stimuli are drawn from the same phonemic category. In experiment 2, we examine responses to variability in speech sound categories in the context of variability in talker identity, to test whether the observed specificity of these responses is related to a specialization for speech, or an orientation toward the most behaviorally relevant dimension of stimulus variability in a particular context.

Materials and Methods

Experiment 1

Overall design and predictions. To test whether responses to speech categorization in superior temporal and inferior parietal regions reflect habituation or change detection, we presented participants with long (10 stimulus) trains of acoustically varying stimuli drawn either from the same phonemic category (the “repeating” condition, e.g., all stimuli instances of the syllable /da/) or alternating between two categories (the “alternating” condition, e.g., between the syllables /da/ and /ga/). We also included separate runs in which shorter (two stimulus) trains were presented, to generate a hypothetical time series assuming an identical response to each successive stimulus pair. This was then used to test whether responses in conditions in which all stimuli were drawn from the same category could be interpreted as consistent with habituation. If the response to each successive presentation of stimuli from the same class were reduced (i.e., habituation), we predict that the response to the repeating condition will be smaller than the constructed response. Conversely, if the response is selective for conditions of stimulus change, the response in the alternating condition will be larger than the constructed response.

Participants. Sixteen right-handed monolingual native English speakers (eight female) participated in the study. All procedures, including forms for written consent and \$50 reimbursement, were approved by the Weill Cornell Medical College Institutional Review Board.

Stimuli. Stimuli were 60 unique recordings of each of the syllables /da/, /ga/, /la/, and /la/ made by a single native English speaker with variability in intonation, duration, and amplitude resulting from normal list order prosody. Recordings were made at 44,100 Hz (monaural) using Praat speech processing software, and were not processed further, except to divide them into individual files. This process retained variability in duration, amplitude, and various other linguistic and paralinguistic properties of the stimuli that was orthogonal to stimulus category, ensuring that responses to phonemic variability were not due to detection of variability in confounded acoustic properties.

Presentation parameters. Stimuli were presented over Resonance Technology Commander XG headphones using Paradigm experimental control software (Perception Research Systems). Two types of run (or “session”) were conducted: one with blocked presentation, and one with sparse presentation. The sparse presentation runs were used to generate a constructed hemodynamic response function (HRF), as illustrated in Figure 1.

In each of six blocked runs, stimuli were presented in trains of five pairs during 15 s blocks. Pairs were presented during the silent intervals of 3 s clustered acquisition periods, followed by 12 s of silence. In repeating blocks, stimuli were all drawn from the same phonemic category; in alternating blocks, stimuli were drawn from two categories (/da/ + /ga/ or /la/ + /la/). Alternation was pseudorandom: each pair of stimuli in this condition contained one stimulus from each category, but the order within a pair was randomly determined ahead of time. Each run comprised 18 blocks: three repeating blocks for each syllable by itself and three alternating blocks for each pair of stimuli, presented in a pseudorandomized order.

Subsequently, we collected sparse presentation runs to collect data that would permit us to estimate the response to a single pair of stimuli, and subsequently construct a hemodynamic response reflecting the hypothesis that each successive presentation of the same stimulus class elicits an identical response. In each of two single-pair runs (always collected after completing the blocked runs), pairs of stimuli were presented in the same temporal distribution as the first pair of each block in the blocked condition, i.e., a single pair of stimuli, followed by 24 s of silence.

Magnetic resonance imaging acquisition. Functional and anatomical images were collected using a 3T Signa scanner with an eight-channel phased-array headcoil (General Electric). Anatomical (MPRAGE) images were collected with 1.5-mm-thick sagittal slices and an in-plane resolution of 0.938 × 0.938 mm. Functional images were collected using the spiral in–out sequence (Preston et al., 2004). Fifteen slices were collected oblique to the axial plane, parallel to the Sylvian fissure, with 5 slices above and 10 below the anatomical landmark. This prescription was intended to permit optimal coverage of perisylvian regions of both superior temporal (STG) and supramarginal (SMG) gyri, but also resulted in incomplete coverage of superior aspects of SMG, and no cover-

age in middle frontal, superior frontal, or motor areas. Functional data were collected with a repetition time (TR) of 3000 ms, an echo time (TE) of 68 ms, and a flip angle of 90°. Acquisition was clustered, resulting in an ~1.5 s silent interval during each scan when stimuli were presented. Slice thickness was 3 mm, and in-plane resolution was 3.125×3.125 mm. In each of both the blocked and single-pair runs, 167 volumes were collected per run.

Magnetic resonance imaging data analysis. Functional data were analyzed using AFNI (Cox, 1996) (program names appearing in parentheses below are part of the AFNI suite). Cortical surface models were created with FreeSurfer (available at <http://surfer.nmr.mgh.harvard.edu/>), and functional data projected into anatomical space using SUMA (Saad et al., 2004; Argall et al., 2006) (AFNI and SUMA are available at <http://afni.nimh.nih.gov/afni>).

Preprocessing. Anatomical and functional datasets for each participant were coregistered using positioning information from the scanner and the `lpc_align` program to bring the oblique functional images into correct alignment. Functional datasets were preprocessed to correct slice timing (3dTshift) and head movements (3dvolreg), reduce extreme values (3dDespike), and detrend linear and quadratic drift from the time series of each run (3dDetrend). Percentage signal change also was calculated for each run.

General linear models and contrasts. Time series data for six blocked runs were analyzed with a general linear model (3dDeconvolve) comprising six regressors (one for each block type) and six regressors of no interest (head movement parameters output from 3dvolreg). Stimulus regressors were generated by convolving the timing of stimulus presentation with a canonical hemodynamic response function (waver, using the “WAV” model). Contrasts for main effects were then conducted by collapsing the two alternating conditions ($/da/ + /ga/$ and $/a/ + /la/$) into an alternating regressor and four repeating conditions ($/da/$, $/ga/$, $/a/$, and $/la/$) into a repeating regressor, then combining these to estimate a main effect of speech versus rest, and finally subtracting repeating from alternating to generate the alternating > repeating contrast.

Surface reconstruction and projection of functional data into surface space. The T1-weighted magnetic resonance imaging (MRI) structural images for each participant were converted into MGH-HMR format, cortical meshes were extracted from structural volumes, and the images were inflated to a sphere (Dale and Sereno, 1993; Dale et al., 1999; Fischl et al., 1999a, 2004) and anatomically registered to a standard sphere using FreeSurfer (Fischl et al., 1999b). Registered surfaces were imported into SUMA (Saad et al., 2004) so that they could be used in the AFNI software package (Cox, 1996). These surfaces were then converted to the mesh of an icosahedron, thus standardizing each of the participant’s registered surfaces to have the same number of corresponding nodes so that group-based analysis could be conducted. Next, volumetric functional data were mapped to these surface datasets, including both time series for calculation of constructed hemodynamic response functions and correlation analyses, as well as β -values from general linear tests to be used in group t tests. An average of all participants’ surface datasets was also created in FreeSurfer and imported into SUMA and given the standardized mesh, and this average brain was normalized to Talairach space. Group-based t tests and correlation maps are displayed on this averaged representation.

Group analyses. Group analyses were done on surface data (3dttest). The resulting surface was mapped to a volume based on a mesh of the averaged brain (3dSurf2Vol), resulting in a map with t statistics for each voxel for each condition and contrast. Activation maps were obtained by first thresholding individual voxels at $p < 0.005$ (uncorrected), and then applying a subsequent cluster-size threshold based on Monte Carlo simulations (AlphaSim), resulting in a corrected threshold of $p < 0.05$.

Meta-analysis. We identified six studies that have published stereotactic coordinates (five MRI, one PET) for regions that respond more strongly to trains of stimuli drawn from different speech categories than to stimuli drawn from the same category [with respect to the participants’ native language (Celsis et al., 1999; Tervaniemi et al., 2000; Jacquemot et al., 2003; Dehaene-Lambertz et al., 2005; Zevin and McCandliss, 2005; Joanisse et al., 2007)]. We generated activation likeli-

hood estimation (ALE) maps based on these studies using a sphere size with full-width at half-maximum diameter of 10 mm, using a voxelwise threshold of $p < 0.05$ (corrected) based on 10,000 permutations and false discovery rate correction, and a 100 mm^3 cluster size threshold [using GingerALE 1.2, available at <http://brainmap.org/ale/>, which implements methods from Turkeltaub et al. (2002) and Laird et al. (2005)].

Construction of hypothetical time series. To explore whether the alternating > repeating contrast was driven by habituation, we constructed hypothetical time series reflecting the assumption that the hemodynamic response did not change with serial presentation of the same stimulus class. To construct the hypothetical time series, data from the single-pair runs were first extracted from each subject, and a mean response was calculated. For comparison, a mean response to each condition in the blocked runs was also calculated. Hypothetical time series were then constructed from the single-pair data by iteratively shifting an empirically derived hemodynamic response to a single stimulus pair forward in time four times, and adding the results to the initial measurement. The result is a constructed response representing the predicted outcome if the response to each successive stimulus pair exactly replicated the response to the first. This hypothesis was tested with a 3 (condition: alternating, repeating, constructed HRF) \times 8 (time point) ANOVA with participants as the random factor.

Rather than repeat this analysis for all identified activations, we constructed hemodynamic response functions based on anatomical regions of interest (ROIs) selected on the basis of the meta-analysis. By this criterion, only supramarginal gyrus contained enough active voxels to justify including it in the analyses. For each participant, voxels were identified that were active at $p < 0.05$ uncorrected for alternating > repeating and fell within supramarginal gyrus, and these voxels used in the time series analysis. This resulted in the inclusion of a mean of 25.6 (SD = 14.7) voxels per participant.

Simple correlation analysis. To test for functional connectivity, a simple correlation was computed based on the mean time series from a seed volume selected by joint consideration of activation in this study and the ALE meta-analysis. Motion parameters and global signal change were first removed from the time series (3dSynthesize), and these data then used in all subsequent steps. The average time series was extracted from the seed volume, and correlated with activity in every voxel in the dataset. For the purpose of group analysis, the R^2 from the regression was converted to Fisher’s Z with normal distribution. A group t test was then conducted (3dttest), following general procedures for group analysis.

Experiment 2

Overall design and predictions. In this experiment, we explore the specificity of temporoparietal junction responses to speech sound categorization by including a condition with highly salient nonlinguistic change (voice alternating) in addition to phoneme alternating and repeating conditions that replicate experiment 1. If change-related responses in these regions are language specific, only the phoneme alternating condition will activate them; if instead they are tuned more generally to behavioral relevance, we predict equivalent or even greater responses in the voice alternating condition relative to the phoneme alternating condition.

Participants. Twelve right-handed monolingual native English speakers (seven female) participated in the study. All procedures, including forms for written consent and \$50 reimbursement, were approved by the Weill Cornell Medical College Institutional Review Board.

Stimuli. Stimuli were 60 unique recordings of each of the syllables $/da/$, $/ga/$, $/a/$, and $/la/$ made by two native English speakers (one male, one female) using techniques described for experiment 1. Half of the recordings were made in each voice. The male voice had a median F_0 (fundamental frequency) of 106 Hz, and the female voice had a median F_0 of 161 Hz.

Presentation parameters. Stimuli were presented using the same hardware and software as in experiment 1, in blocks of three types: repeating (repeating talker, repeating phoneme), voice alternating (same phoneme, changing talker), and phoneme alternating (changing phoneme, same talker). In each block, 10 stimuli were presented

Table 1. Activations for the alternating > repeating contrast in experiment 1

	BA	Voxels	Z value	Coordinates		
				x	y	z
Superior parietal lobule	7	18	3.72	−16	−70	42
Inferior parietal lobule	40	12	3.25	−34	−46	48
Supramarginal gyrus	40	61	4.29	−32	−50	32
Superior temporal gyrus		376	5.55	−56	−32	12
Middle temporal gyrus	21	14	3.55	−56	4	−28
Temporal pole	13	20	4.06	−38	10	−10
Lingual gyrus	30	14	3.48	−10	−38	2
Inferior parietal lobule	40	53	4.39	34	−46	42
	40	24	3.45	44	−56	42
Precuneus	31	21	3.29	8	−46	44
	27	19	3.73	16	−38	2
Supramarginal gyrus	40	72	4.59	58	−20	20
	40	15	3.48	58	−26	36
Superior temporal gyrus	22	33	3.89	62	−16	6
		26	4.11	44	−32	14
Insula	13	145	4.47	38	−20	6
	13	26	4.48	38	−2	−10
		14	3.67	38	16	2
Medial temporal pole	38	22	3.43	46	16	−24
Middle cingulate cortex	24	159	4.82	2	−22	36
Amygdala		19	3.38	22	4	−10
Parahippocampal gyrus		17	3.85	34	−26	−12

BA, Brodmann area. Volume is given in number of voxels ($3.125 \times 3.125 \times 3$). Z value, Peak value of inferential test for alternating responses being greater than repeating. Coordinates (x, y, and z) for the most active voxel in each cluster are given with reference to the Talairach atlas. All activations surpassing the critical value for extent determined in AlphaSim (12 voxels) are reported.

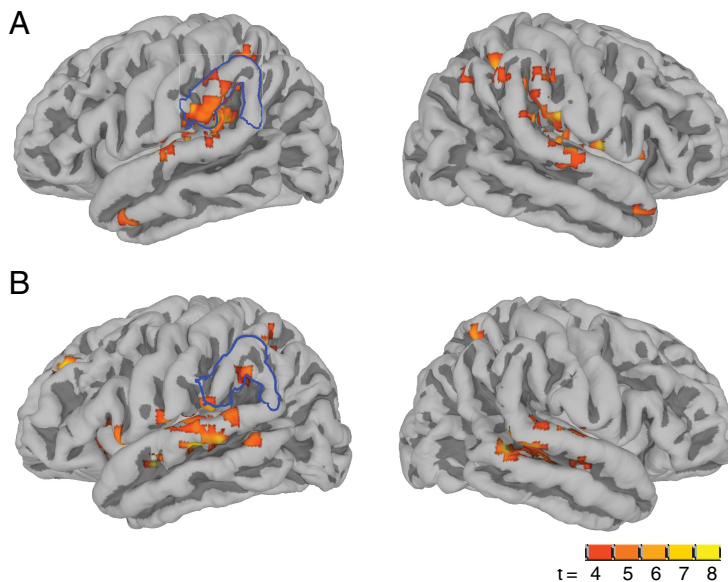


Figure 2. Whole-brain contrasts for experiments 1 and 2. **A, B,** Brain activation for alternating > repeating in experiment 1 (**A**) and voice alternating > repeating in experiment 2 (**B**). Blue outline indicates the anatomical location of left supramarginal gyrus according to Freesurfer parcellation of the average brain.

at the rate of 1 stimulus per second (continuous EPI acquisition was used, obviating the need to time stimulus presentation to silent periods in the TR). After each 10 s block of stimuli, a 12.5 s rest period was included. Stimuli were presented in six runs with four blocks of each condition (voice alternating, phoneme alternating, and repeating) presented in each run.

MRI acquisition. Functional and anatomical images were collected using a 3T Signa scanner with an eight-channel phased-array headcoil (General Electric). Anatomical (MPRAGE) images were collected using the same prescription as experiment 1. Functional images were collected using the standard GE echo-planar imaging (EPI) sequence, with TR = 1.5 s, flip angle = 90° , TE = 30 ms. Twenty-five axial slices were collected

at a resolution of $3.125 \times 3.125 \times 5$ mm, permitting coverage of the entire cerebral cortex and most of the cerebellum.

MRI data analysis. Preprocessing, regression analyses, and contrasts were performed as in experiment 1, with the following exceptions. First, because images were not acquired obliquely, *lpc_align* was not necessary to coregister the functionals and anatomicals. Further, the regression analyses were conducted for three conditions: repeating, phoneme alternating, and voice alternating. Separate contrasts for voice alternating > repeating and phoneme alternating > repeating are reported.

Simple correlation analysis. Correlation analyses based on a seed volume in SMG were conducted following the methods described for experiment 1.

Results

Experiment 1

Whole-brain analyses

Activations for the alternating > repeating contrast are shown in Table 1 and activation maps presented in Figure 2. Consistent with previous research, regions in the temporoparietal junction are more active for the condition involving phonemic variability (alternating) relative to the condition in which stimuli are drawn from the same phonemic category (repeating). Although stronger and more extensive activity is observed in the left hemisphere, substantial activation is also present in the right. Activations in previous studies have largely been restricted to the left hemisphere; this disparity is possibly related to the greater power of

the current study: By combining long trains of repeating stimuli with behaviorally interleaved gradients, we may have been more sensitive to right hemisphere activations than previous studies using shorter trains of stimuli (e.g., Jacquemot et al., 2003; Zevin and McCandliss, 2005; Joanisse et al., 2007) and/or sparser image acquisition schedules (Celsis et al., 1999).

Meta-analysis

To guide selection of ROIs for further analysis, we conducted a meta-analysis of six existing studies that report Talairach coordinates of peak activations for conditions of phonetic change compared to either repeating stimuli or nonphonetic, acoustic change (see Materials and Methods). Four consistently activated regions were identified, as shown in Table 2.

Time series analysis to test for adaptation

Time series constructed by concatenating the response to a single pair of stimuli with itself five times were compared to the responses in the alternating and repeating conditions. A 3 (condition) \times 8 (time points 1–8) ANOVA revealed a significant interaction between condition and time point, $F_{(14,210)} = 2.11$. As shown in Figure 3, activity in the alternating condition is greater than activity in both the repeating condition and the constructed response over TRs 3–6 (9–18 s after the onset of the block). *Post hoc t* tests revealed that the mean overall activity for TRs 1–8 was greater in the alternating condition than the repeating condition or the constructed responses ($Z_{(7)} = 2.97$ and $Z_{(7)} = 2.25$, respectively), whereas the repeating condition did not differ from the constructed response ($Z < 1$).

Table 2. Results of ALE meta-analysis including six studies of sensitivity to phonetic change

	BA	Voxels	Coordinates		
			x	y	z
Left angular gyrus	39	92	−28	−52	29
Left supramarginal gyrus	40	76	−40	−33	20
Left middle temporal gyrus	22	41	−62	−22	−10
Left middle frontal gyrus	8	30	−28	25	46

BA, Brodmann area. Volume is given in number of voxels (2 × 2 × 2). Coordinates (x, y, and z) for the most active voxel in each cluster are given with reference to the Talairach atlas.

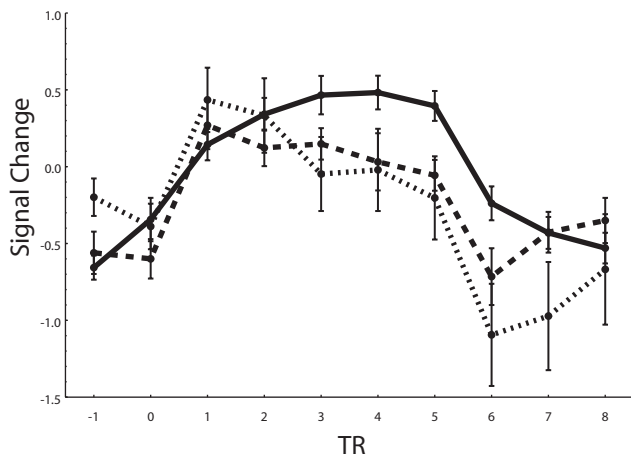


Figure 3. Time series analysis of supramarginal gyrus. Shown are empirically derived time series from the alternating (solid line) and repeating (dashed line) conditions and constructed time series (dotted line) based on the response to single pairs of stimuli, representing the hypothesis that responses do not change with successive presentations of the same stimulus class. The constructed hemodynamic response closely fits the response in the repeating condition, suggesting that no repetition suppression is observed in this condition.

Experiment 2

Whole-brain analyses

Activations for the voice alternating > repeating and phoneme alternating > repeating contrasts are shown in Tables 3 and 4 and activations for the voice alternating > repeating contrast shown in Figure 2. Results from the phoneme alternating > repeating contrast, which most closely replicate the procedures of experiment 1, do not show any activity in temporoparietal junction regions. In contrast, responses in the voice alternating > repeating contrast were broadly similar to both experiment 1 and the meta-analysis of studies of phonetic change.

ROI analyses

To determine whether there was any activity at all in the phoneme alternating condition in regions identified as preferring voice alternating to repeating, we extracted mean normalized parameter estimates from each participant for the ROIs in SMG, angular gyrus (AG), STG, middle temporal gyrus (MTG), and middle frontal gyrus (MFG). These ROIs were selected based on consideration of anatomical regions identified in the meta-analysis. As shown in Figure 4, activity is not significantly different from the repeating condition for the phoneme alternating condition, all Z values < 1.56, p values > 0.12. Activity differs marginally between the phoneme alternating and voice alternating conditions in the STG/MTG ($Z_{(11)} = 3.21$), MFG ($Z_{(11)} = 1.90$, $p = 0.06$), AG ($Z_{(11)} = 1.83$, $p = 0.07$), and SMG ($Z_{(11)} = 1.61$, $p = 0.11$). Thus, it appears that in the context of an experiment in which more salient behavioral variability is included, responses to phonemic

Table 3. Activations for voice alternating > repeating in experiment 2

	BA	Voxels	Z value	Coordinates		
				x	y	z
Inferior parietal lobule	40	32	4.31	−32	−52	44
Supramarginal gyrus	40	17	3.20	−52	−44	32
Superior temporal gyrus	40	37	4.10	−58	−28	18
Middle temporal gyrus	21	130	4.55	−58	−22	−0
	22	19	3.94	−56	−46	2
Insula	13	16	4.53	−28	16	−6
Temporal pole		16	3.69	−46	10	−0
Middle frontal gyrus	9	75	4.31	−26	40	36
Inferior frontal gyrus	13	26	3.89	−40	16	8
Precuneus	7	57	4.64	−4	−70	38
	7	25	4.39	−14	−64	50
Superior medial gyrus	9	20	3.73	−2	46	24
Angular gyrus	7	26	3.80	28	−58	48
Superior temporal gyrus	38	25	4.24	−50	−4	−6
	22	48	3.43	58	−14	6
Middle temporal gyrus	37	148	4.31	58	−44	−4
Cuneus	18	47	4.49	2	−76	14
	7	21	3.55	16	−76	38
Superior medial gyrus	8	71	4.90	2	20	50
Anterior cingulate cortex	22	44	3.69	10	38	14
	24	17	4.25	2	28	6
Middle cingulate cortex	24	177	5.31	2	2	36

BA, Brodmann area. Volume is given in number of voxels (3.125 × 3.125 × 5). Z value, Peak value of inferential test for voice alternating responses being greater than repeating. Coordinates (x, y, and z) for the most active voxel in each cluster are given with reference to the Talairach atlas. All activations surpassing the critical value for extent determined in AlphaSim (16 voxels) are reported.

Table 4. Activations for phoneme alternating > repeating in experiment 2

	BA	Voxels	Z value	Coordinates		
				x	y	z
Cuneus	7	20	4.50	−10	−74	30
Superior temporal gyrus	41	16	3.13	44	−28	14
Superior medial gyrus	6	33	4.08	2	20	54
	9	24	4.15	2	50	30
Middle cingulate cortex	24	17	4.03	2	−8	38

BA, Brodmann area. Volume is given in number of voxels (3.125 × 3.125 × 5); Z value, Peak value of inferential test for phoneme alternating responses being greater than repeating. Coordinates (x, y, and z) for the most active voxel in each cluster are given with reference to the Talairach atlas. All activations surpassing the critical value for extent determined in AlphaSim (16 voxels) are reported.

variability are greatly reduced in these regions, in that the response to phonemic alternation is intermediate between the voice alternating and repeating conditions.

Functional connectivity of the supramarginal gyrus

We measured functional connectivity for the supramarginal gyrus by computing correlations with the mean time series in the ROIs identified in each experiment. Heat maps of the correlations are shown in Figure 5. In both experiments, we observed no significant differences between conditions, indicating that SMG participated in a stable functional network despite its own activity being highly sensitive to experimental conditions.

Strong positive correlations were observed in both experiments throughout all of supramarginal gyrus, and in inferior frontal gyrus, both bilaterally. Negative correlations are observed in both studies in visual areas (precuneus, parahippocampal gyrus). Thus, the supramarginal gyrus forms part of a relatively stable functional network including canonical language regions and their right-hemisphere homologs. Negative correlations with visual areas in both experiments could be due to the fact that participants were watching movies during the passive presentation of auditory stimuli—so that behaviorally relevant change in

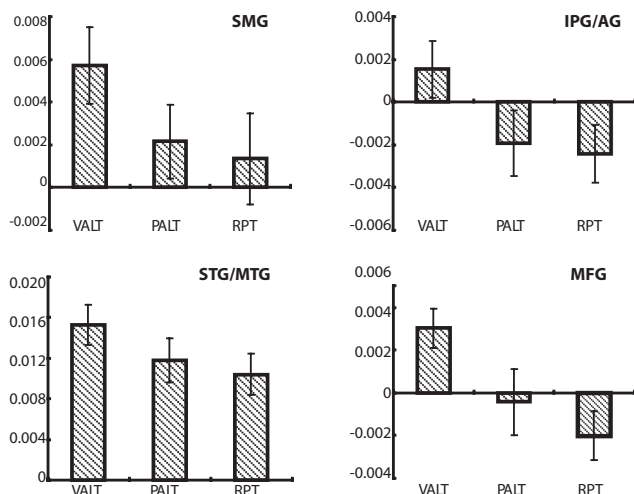


Figure 4. ROI analyses of the voice alternating and phoneme alternating conditions in experiment 2. Shown are β values for all three conditions in ROIs identified as active in voice alternating > repeating in experiment 2 and consistent with anatomical regions identified in the ALE meta-analysis.

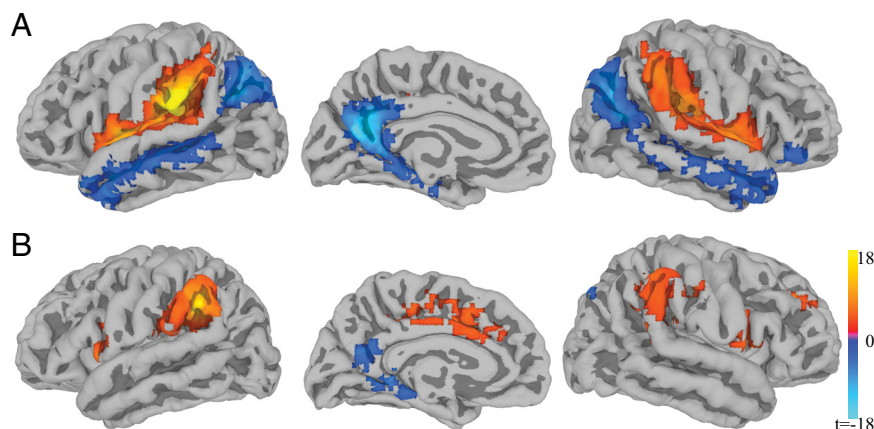


Figure 5. Correlations with left supramarginal gyrus in experiments 1 and 2. **A, B**, Correlation maps for ROIs in left supramarginal gyrus in experiments 1 (**A**) and 2 (**B**), at a voxelwise threshold of $p < 0.005$, corrected to $p < 0.05$ with AlphaSim (critical values = 38 voxels in experiment 1 and 32 voxels in experiment 2).

the auditory stream may have resulted in brief redirection of attention away from visual processing.

Some differences between experiments were also observed: correlations are stronger overall in experiment 1, and positive correlations are observed throughout the superior bank of the superior temporal gyrus as well as in precentral and postcentral gyri that are not observed in experiment 2. Negative correlations throughout anterior superior temporal sulcus, middle temporal sulcus, and the temporal pole, as well as posterior portions of the inferior parietal lobe. The only correlations unique to experiment 2 were throughout middle and anterior cingulate cortex. Negative correlations with the STS in experiment 1 may be related to focus on linguistic aspects of stimuli, consistent with the findings that this region prefers highly intelligible and predictable speech (e.g., Scott et al., 2000; Wise et al., 2001; Obleser et al., 2007). The cingulate activity in experiment 2 may be related to the fact that the relevant variability in that context was in the identity of the speaker; these regions have been shown to be involved in encoding self/other relationships (Chiu et al., 2008).

Discussion

We found that a set of regions in the posterior superior temporal gyrus and inferior parietal lobe respond selectively to behaviorally relevant variability in stimulus properties during passive presentation of speech sounds. The temporal properties of responses suggest that the observed activations reflect increased responding during conditions of unpredictable stimulus change, rather than habituation to repeated presentation of sounds from the same category. Further, selectivity to phonemic change is abolished when more salient nonlinguistic variability (i.e., talker change) is included in the experimental context. Together, these findings suggest that the change detection function of regions identified in previous studies of speech sound categorization is not the result of habituation, and is not in itself a specialization for speech categorization.

Habituation, change detection and representation

The basis of habituation, as observed in functional MRI (fMRI), is not well understood (see, e.g., Grill-Spector et al., 2006), although most descriptions of the phenomenon link habituation (albeit implicitly) to the notion of representation. That is, they share the premise that the particular pattern of activity that is reliably elicited by a stimulus (a neural representation), decreases in amplitude, spatial extent, or duration with successive presentations of the stimulus. In order for this to occur for multiple exemplars of the same category, as in the current study, the representation must be sufficiently general to count as a representation of the category itself.

This same logic is often applied to behavioral measures of habituation used to study categorization in infants (Werker and Tees, 1984; Kuhl et al., 1992). If experiment 1 had revealed a pattern consistent with habituation to multiple exemplars of the same speech category, this would have been strong evidence that the temporoparietal junction is involved in representing those categories. Instead, we found that the response to the repeating condition was well modeled by a hypothetical hemodynamic response function constructed to simulate a condition in which the response to each successive stimulus of the same type is identical. Although we found no evidence that supramarginal gyrus is specifically tuned to speech categories, this does not mean that it cannot serve as an index of language-specific categorization. To continue the analogy with behavioral habituation studies, an infant's increased rate of sucking on a pacifier or visual orientation toward a display of an interesting toy are not part of his or her representation of speech categories, but can nonetheless be used to indicate when a behaviorally relevant change was detected. Thus, the fact that, in most studies, linguistically relevant change elicits responses in SMG while other types of change do not (as in, e.g., Jacquemot et al., 2003; Joanisse et al., 2007) reveals something about what the listener treats as behaviorally relevant, even if it does not provide a localization of the mechanisms by which the brain categorizes speech sounds.

Change detection can also be related to representation, as in studies of an electrophysiological phenomenon, the mismatch negativity (MMN) elicited in experiments very similar to the current study. In MMN paradigms, the evoked response to a fre-

quently repeated “standard” stimulus is compared to the response to a relatively rare “deviant” stimulus. Between 100 and 250 ms, the response to the deviant stimulus is more negative, and the subtraction of the standard from the deviant yields a negative-going wave (for review, see Näätänen and Winkler, 1999), even when differences in the afferent patterns of stimuli are taken into account by examining standard and deviant responses to the same stimulus (the “identity MMN”) (Pulvermüller and Shtyrov, 2006). The MMN is most often interpreted in terms of change detection, specifically, the formation and updating of a model of the acoustic environment that predicts the occurrence of a particular sound at a particular time.

This notion of representation—in terms of a continuously updated predictive model of the auditory environment—is different from that intended in most studies of speech sound categorization. For example, many studies have shown an MMN to deviations from the standard tone that vary in size monotonically with the size of the acoustic difference between the standard and deviant (Sams et al., 1985). The MMN can be related to categorization depending on long-term memory, however [e.g., of phonetic categories (Winkler et al., 1999)]. Strong MMNs are elicited to unfamiliar speech contrasts, but the size of the response relative to the acoustic distance between contrasting stimuli is modulated by phonetic categorization such that large acoustic differences can elicit a relatively small MMN if the stimuli used are not categorized as members of native language categories (Näätänen et al., 1997). In contrast, change detection responses to speech in fMRI have not been observed to non-native (Jacquemot et al., 2003) or noncontrastive (Joanisse et al., 2007) change.

The responses observed in the current study are distinct from the MMN in other ways as well. Although the alternating conditions in the current study are in some ways comparable to a “deviant” condition in an MMN study—in that the stimulus class changes—the randomness of presentation order in the alternating conditions makes it impossible to construct a model of the environment that gives rise to deviance detection. In fact, designs with similar presentation parameters to the alternating condition do not give rise to an MMN response in electrophysiological studies. Finally, there is some ambiguity in the literature with respect to the anatomical location of the MMN generator(s), suggesting that this response is not necessarily localizable to the temporoparietal junction (Molholm et al., 2005; Baldeweg, 2007; Schönwiesner et al., 2007). The notion that the TPJ is involved in the detection of salient, behaviorally relevant events across modalities has a long history in cognitive neuroscience (e.g., Knight et al., 1989; Yamaguchi and Knight, 1991). In addition, change detection in TPJ is amodal (Downar et al., 2000) and task dependent (Downar et al., 2001) in a manner that is inconsistent with the MMN.

Thus, if we interpret them as TPJ responses to stimulus change, the responses observed in the current and related studies cannot be interpreted as neural correlates of either habituation to speech sound categories or the metabolic equivalent of the MMN. Instead, they are more consistent with change detection responses that are domain-general, amodal, and flexibly tuned to the relative behavioral relevance of different dimensions of stimulus contrast. This is perhaps not surprising, given that the regions involved are not particularly selective for speech (see Obleser and Eisner, 2009, for review), nor are they implicated in studies that use multivariate pattern analysis to identify patterns of activity that permit the discrimination of speech sounds from one another based on brain activity (Raizada and Poldrack, 2007; Formisano et al., 2008).

Linguistic contrast as a special case of behavioral relevance

Various designs have been used to explore the specificity of change detection responses to speech sound categorization. For example, Joanisse et al. (2007) presented participants with stimuli from a synthesized /da/–/ga/ continuum, and compared responses to within- and between-category stimulus change while controlling the magnitude of the acoustic change between conditions. The results indicated that supramarginal gyrus was selectively activated by acoustic changes that indicated a linguistic contrast (i.e., the difference between /da/ and /ga/) relative to changes that did not [i.e., two different exemplars of either /da/ or /ga/; see also Raizada and Poldrack (2007) for a more involved set of analyses indicating enhanced change detection for stimuli that cross a category boundary]. The pattern of results from previous studies thus suggests some specificity for speech sound categorization that requires explanation, even if we argue that the responses themselves are not a specialization for speech processing.

One critical observation is that change detection responses to speech sound categories have previously been elicited in studies that conflated linguistic contrast with behaviorally relevant stimulus variability. That is, an apparent specialization for speech categories may have been observed because change from one category to another was the most interesting thing that happened in the context of the experiment. In the current study, responses to phonemic contrast were examined both in a context in which this was the most salient source of variability (experiment 1) and in a context that contained more salient variability as well (experiment 2). When the stimulus set included talker variability, responses to phonemic variability were essentially abolished. Instead, regions previously identified as specialized for the perception of speech contrasts responded strongly to talker variability. Also consistent with the notion that engagement of the SMG with language-specific stimulus properties is flexible, functional connectivity of the SMG was observed to be task dependent with respect to regions that can be described as part of the “ventral stream” for speech perception (Rauschecker and Tian, 2000; Hickok and Poeppel, 2007; Rauschecker and Scott, 2009). The same analyses, however, indicate relatively stable connectivity of the SMG with posterior STG and inferior frontal gyrus, consistent with the notion of a “dorsal stream” for speech perception in this framework, and its likely role in relating acoustic and motor aspects of speech sounds. Specifically, we propose that change detection occurs via a predictive mechanism involving the motor system in which predicted behaviorally relevant stimulus features are compared with the current description of the sensory state of the listener in posterior superior temporal and inferior parietal cortices (Skipper et al., 2007).

Conclusions

We have replicated the response to phonemic change in temporoparietal regions observed in numerous studies of speech sound categorization using fMRI, and demonstrated that these responses are more consistent with change detection than with habituation to phonemic categories. These regions appear to be finely tuned to specific aspects of the speech signal when this is the most salient source of variability in the acoustic environment: In experiment 1, responses to phonemic change were measured against a background of random variability in intonation pattern present in both the alternating and repeating conditions. When other forms of behaviorally relevant stimulus variability are included, however, as in experiment 2, this apparent specificity disappears. This pattern has consequences for how we understand differences between conditions (or groups of participants)

in studies using “dishabituation” designs. While it is clear that these responses are influenced by language experience, they must be seen as reflecting acquired behavioral relevance rather than indicating a neural specialization for speech categorization in the location of the observed response.

References

- Argall BD, Saad ZS, Beauchamp MS (2006) Simplified intersubject averaging on the cortical surface using SUMA. *Hum Brain Mapp* 27:14–27.
- Baldeweg T (2007) ERP repetition effects and mismatch negativity generation: a predictive coding perspective. *J Psychophysiol* 21:204–213.
- Celsis P, Boulanouar K, Doyon B, Ranjeva JP, Berry I, Nespoulous JL, Chollet F (1999) Differential fMRI responses in the left posterior superior temporal gyrus and left supramarginal gyrus to habituation and change detection in syllables and tones. *Neuroimage* 9:135–144.
- Chiu PH, Kayali MA, Kishida KT, Tomlin D, Klinger LG, Klinger MR, Montague PR (2008) Self responses along cingulate cortex reveal quantitative neural phenotype for high-functioning autism. *Neuron* 57:463–473.
- Corbetta M, Patel G, Shulman GL (2008) The reorienting system of the human brain: from environment to theory of mind. *Neuron* 58:306–324.
- Cox RW (1996) AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res* 29:162–173.
- Dale AM, Sereno MI (1993) Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: a linear approach. *J Cogn Neurosci* 5:162–176.
- Dale AM, Fischl B, Sereno MI (1999) Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage* 9:179–194.
- Dehaene-Lambertz G, Pallier C, Serniclaes W, Sprenger-Charolles L, Jobert A, Dehaene S (2005) Neural correlates of switching from auditory to speech perception. *Neuroimage* 24:21–33.
- Downar J, Crawley AP, Mikulis DJ, Davis KD (2000) A multimodal cortical network for the detection of changes in the sensory environment. *Nat Neurosci* 3:277–283.
- Downar J, Crawley AP, Mikulis DJ, Davis KD (2001) The effect of task relevance on the cortical response to changes in visual and auditory stimuli: an event-related fMRI study. *Neuroimage* 14:1256–1267.
- Fischl B, Sereno MI, Dale AM (1999a) Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *Neuroimage* 9:195–207.
- Fischl B, Sereno MI, Tootell RB, Dale AM (1999b) High-resolution intersubject averaging and a coordinate system for the cortical surface. *Hum Brain Mapp* 8:272–284.
- Fischl B, van der Kouwe A, Destrieux C, Halgren E, Ségonne F, Salat DH, Busa E, Seidman LJ, Goldstein J, Kennedy D, Caviness V, Makris N, Rosen B, Dale AM (2004) Automatically parcellating the human cerebral cortex. *Cereb Cortex* 14:11–22.
- Formisano E, De Martino F, Bonte M, Goebel R (2008) “Who” is saying “what”? Brain-based decoding of human voice and speech. *Science* 322:970–973.
- Grill-Spector K, Henson R, Martin A (2006) Repetition and the brain: neural models of stimulus-specific effects. *Trends Cogn Sci* 10:14–23.
- Hickok G, Poeppel D (2007) The cortical organization of speech processing. *Nat Rev Neurosci* 8:393–402.
- Jacquemot C, Pallier C, LeBihan D, Dehaene S, Dupoux E (2003) Phonological grammar shapes the auditory cortex: a functional magnetic resonance imaging study. *J Neurosci* 23:9541–9546.
- Joanisse MF, Zevin JD, McCandliss BD (2007) Brain mechanisms implicated in the preattentive categorization of speech sounds revealed using fMRI and a short-interval habituation trial paradigm. *Cereb Cortex* 17:2084–2093.
- Knight RT, Scabini D, Woods DL, Clayworth CC (1989) Contributions of temporal-parietal junction to the human auditory P3. *Brain Res* 502:109–116.
- Kuhl PK, Williams KA, Lacerda F, Stevens KN, Lindblom B (1992) Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* 255:606–608.
- Laird AR, Fox PM, Price CJ, Glahn DC, Uecker AM, Lancaster JL, Turkeltaub PE, Kochunov P, Fox PT (2005) ALE meta-analysis: controlling the false discovery rate and performing statistical contrasts. *Hum Brain Mapp* 25:155–164.
- Liu R, Skipper JI, McCandliss BD, and Zevin JD (2009) Neural activity underlying passive perception of native vs. non-native phonetic contrasts. Poster presented at 15th Annual Meeting of the Cognitive Neuroscience Society, San Francisco, March.
- Molholm S, Martinez A, Ritter W, Javitt DC, Foxe JJ (2005) The neural circuitry of pre-attentive auditory change-detection: an fMRI study of pitch and duration mismatch negativity generators. *Cereb Cortex* 15:545–551.
- Näätänen R, Winkler I (1999) The concept of auditory stimulus representation in cognitive neuroscience. *Psychol Bull* 125:826–859.
- Näätänen R, Lehtokoski A, Lennes M, Cheour M, Huotilainen M, Iivonen A, Vainio M, Alku P, Ilmoniemi RJ, Luuk A, Allik J, Sinkkonen J, Alho K (1997) Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* 385:432–434.
- Obleser J, Eisner F (2009) Pre-lexical abstraction of speech in the auditory cortex. *Trends Cogn Sci* 13:14–19.
- Obleser J, Wise RJS, Dresner MA, Scott SK (2007) Functional integration across brain regions improves speech perception under adverse listening conditions. *J Neurosci* 27:2283–2289.
- Preston AR, Thomason ME, Ochsner KN, Cooper JC, Glover GH (2004) Comparison of spiral-in/out and spiral-out BOLD fMRI at 1.5T and 3T. *Neuroimage* 21:291–301.
- Pulvermüller F, Shtyrov Y (2006) Language outside the focus of attention: the mismatch negativity as a tool for studying higher cognitive processes. *Prog Neurobiol* 79:49–71.
- Raizada RD, Poldrack RA (2007) Selective amplification of stimulus differences during categorical processing of speech. *Neuron* 56:726–740.
- Rauschecker JP, Scott SK (2009) Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat Neurosci* 12:718–724.
- Rauschecker JP, Tian B (2000) Mechanisms and streams for processing of “what” and “where” in auditory cortex. *Proc Natl Acad Sci USA* 97:11800–11806.
- Saad Z, Reynolds R, Argall B, Japee S, Cox RW (2004) SUMA: an interface for surface-based intra- and inter-subject analysis with AFNI. *Proc IEEE Int Symp Biomed Imaging* 2004:1510–1513.
- Sams M, Paavilainen P, Alho K, Näätänen R (1985) Auditory frequency discrimination and event-related potentials. *EEG J* 62:437–448.
- Schönwiesner M, Novitski N, Pakarinen S, Carlson S, Tervaniemi M, Näätänen R (2007) Heschl’s gyrus, posterior superior temporal gyrus, and mid-ventrolateral prefrontal cortex have different roles in the detection of acoustic changes. *J Neurophysiol* 97:2075–2082.
- Scott SK, Blank CC, Rosen S, Wise RJS (2000) Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* 123:2400–2406.
- Skipper JI, van Wassenhove V, Nusbaum HC, Small SL (2007) Hearing lips and seeing voices: how cortical areas supporting speech production mediate audiovisual speech perception. *Cereb Cortex* 17:2387–2399.
- Tervaniemi M, Medvedev SV, Alho K, Pakhomov SV, Roudas MS, van Zuijen TL, Näätänen R (2000) Lateralized automatic auditory processing of phonetic versus musical information: A PET study. *Hum Brain Mapp* 10:74–79.
- Turkeltaub PE, Eden GF, Jones KM, Zeffiro TA (2002) Meta-analysis of the functional neuroanatomy of single-word reading: method and validation. *Neuroimage* 16:765–780.
- Werker JF, Tees RC (1984) Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behav Dev* 7:49–63.
- Winkler I, Kujala T, Tiihinen H, Sivonen P, Alku P, Lehtokoski A, Czigler I, Csépe V, Ilmoniemi RJ, Näätänen R (1999) Brain responses reveal the learning of foreign language phonemes. *Psychophysiology* 36:638–642.
- Wise RJ, Scott SK, Blank SC, Mummery CJ, Murphy K, Warburton EA (2001) Separate neural sub-systems within “Wernicke’s area”. *Brain* 124:83–95.
- Yamaguchi S, Knight RT (1991) Anterior and posterior association cortex contributions to the somatosensory P300. *J Neurosci* 11:2039–2054.
- Zevin JD, McCandliss BD (2005) Dishabituation of the BOLD response to speech sounds. *Behav Brain Funct* 1:4.